



Published in final edited form as:

Cancer Res. 2011 November 15; 71(22): 6940–6947. doi:10.1158/0008-5472.CAN-11-2091.

A Comprehensive View of Nuclear Receptor Cancer Cistromes

Qianzi Tang¹, Yiwen Chen², Clifford Meyer², Tim Geistlinger³, Mathieu Lupien³, Qian Wang¹, Tao Liu², Yong Zhang¹, Myles Brown³, and X. Shirley Liu²

¹Department of Bioinformatics, School of Life Science and Technology, Tongji University, Shanghai, 200092, China

²Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, 44 Binney St., Boston, MA 02115, USA

³Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney St., Boston, MA 02115, USA

Abstract

Nuclear receptors (NRs) comprise a superfamily of ligand-activated transcription factors that play important roles in both physiology and diseases including cancer. The technologies of Chromatin ImmunoPrecipitation followed by array hybridization (ChIP-chip) or massively parallel sequencing (ChIP-seq) has been used to map, at an unprecedented rate, the *in vivo* genome-wide binding (cistrome) of NRs in both normal and cancer cells. We developed a curated database of 88 NR cistrome datasets and other associated high-throughput datasets, including 121 collaborating factor cistromes, 94 epigenomes and 319 transcriptomes. Through integrative analysis of the curated NR ChIP-chip/seq datasets, we discovered novel factor-specific noncanonical motifs that may have important regulatory roles. We also revealed a common feature of NR pioneering factors to recognize relatively short and AT-rich motifs. Most NRs bind predominantly to introns and distal intergenetic regions, and binding sites closer to transcription start sites (TSSs) were found to be neither stronger nor more evolutionarily conserved. Interestingly, while most NRs appear to be predominantly transcriptional activators, our analysis suggests that the binding of ESR1, RARA and RARG has both activating and repressive effects. Through meta-analysis of different omic data of the same cancer cell line model from multiple studies, we generated consensus cistrome and expression profiles. We further made probabilistic predictions of the NR target genes by integrating cistrome and transcriptome data, and validated the predictions using expression data from tumor samples. The final database, with comprehensive cistrome, epigenome, transcriptome datasets, and downstream analysis results, constitutes a valuable resource for the nuclear receptor and cancer community.

Introduction

Nuclear receptors (NRs) form a large class of transcription factors that can bind directly to DNA to regulate gene expression upon ligand activation. The ligands can be steroid, hormones, or other molecules, although some NRs, called orphan receptors, have no known ligands. The human and mouse genomes encode 48 and 49 NRs, respectively. These nuclear receptors play important roles in the development, homeostasis and metabolism of higher organisms.

Corresponding Author: X. Shirley Liu, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, 44 Binney St., Boston, MA 02115, USA. Phone: 617-632-2472; Fax: 617-632-2444; xshliu@jimmy.harvard.edu. Myles Brown, Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney St., Boston, MA 02115, USA. Phone: 617-632-3948; Fax: 617-632-5417; myles_brown@dfci.harvard.edu.

Nuclear receptors play key roles not only in normal physiology but also in many pathological processes, most notably cancer. Estrogen receptor (ESR) is over expressed in over 70% of breast cancers and is the archetypal molecular therapeutic target (1). Progesterone receptor has been shown to enhance breast cancer motility and invasiveness (2). Androgen receptor (AR) overactivation by androgens is essential for the initiation and progression of prostate cancers (3, 4). Retinoic acid receptor (RAR), upon activation by retinoic acid (RA), has antiproliferative effects in tumor cells (5). The translocation and subsequent oncofusion of PML with retinoic acid receptor α (RAR α) in hematopoietic myeloid cells causes acute promyelocytic leukemia (6). Recent studies have linked cancer to lipid metabolism and cell inflammations (7, 8), and the major NRs regulating these processes include glucocorticoid receptor (GR), peroxisome proliferator activated receptor (PPAR), and liver X-receptor (LXR) (9).

NRs often bind to DNA as homo- or heterodimers, each recognizing a half-site of six nucleotides. Thus their DNA-binding sequences, called hormone response elements (HREs), often consist of two half-sites in directed, everted or inverted configurations, separated by a variable gap (10). Much effort has been devoted to de novo prediction of NR binding sites based solely on genomic DNA sequence, without much success. Recently, the application of ChIP-chip/seq techniques has enabled the accurate and effective detection of the genome-wide in vivo binding sites, or cistromes of NRs (Supplementary Table S1). Herein, we define the cistrome as the set of cis-acting elements bound by a trans-factor at the genomic scale, i.e., binding sites identified by ChIP-chip/seq experiments. Publicly available cistrome data has been growing rapidly, and sometimes multiple cistrome profiles of the same trans-factor in the same biological system are available. Meta-analysis of related cistrome profiles can often yield much more biologically relevant insights than the examination of single profiles.

Previous efforts to identify NR target genes have mostly relied on differential expression profiles before and after NR activation. However, the differential expression cutoff selected may not be ideal, and for many genes, differential expression may be due to secondary or tertiary effects of NR activity. With the availability of cistrome data, target gene prediction based on the presence of a binding site within a certain distance from the transcription start site (TSS) of the gene has also been used, although the distance cutoff is often arbitrary. In addition, many genes with nearby binding sites show no differential expression upon binding, due to the gene's promoter chromatin status, missing essential cofactors, and other confounding effects. Intuitively, the combination of cistrome and differential expression profiles should allow for a much more robust prediction of the direct target genes of nuclear receptors than either data alone.

In this study, we systematically collected and preprocessed all of the publicly available genome-wide ChIP-chip/seq data for NRs, their collaborating factors and histone modifications in humans and mice using a standardized computational pipeline. We compared the hormone response element (HRE) patterns, distance (to TSSs of genes) distributions, evolutionary conservation, and collaborating partners of different NRs. We also conducted meta-analyses to generate consensus cistrome and expression profiles. Finally, we integrated cistrome and transcriptome data to make probabilistic predictions of NR target genes, including 10 NRs in various cancer cell line models. The resultant cistromes, epigenomes, transcriptomes, motif analyses, and target gene lists are publicly available at (11).

Materials and Methods

Target gene prediction

In some systems such as ESR1 activation in the breast cancer cell line MCF-7, multiple cistrome and transcriptome data are available from different studies using the same or different platforms. We first used Stouffer's p-value combination method (12) to combine different transcriptome datasets, giving each gene a consensus differential expression z-score. We also utilized MM-chip (13) to combine redundant cistrome datasets to create a consensus peak list. Based on the characterization of higher order chromatin interactions and our preliminary analysis, we calculated the regulatory potential for a given gene, S_g , as the sum of the nearby binding sites weighted by the distance from each site to the TSS of the

gene: $S_g = \sum_{i=1}^k e^{-(0.5+4\Delta_i)}$, where k is the number of binding sites within 100 kb of gene g and Δ_i is the distance between site i and the TSS of gene g normalized to 100 kb (e.g., 0.5 for a 50 kb distance). This equation models the influence of each binding site on gene regulation as a function that decreases monotonically with increasing distance from the TSS. The shape of this function approximates empirical observations of the distance between binding sites and differentially expressed genes in multiple ChIP-seq experiments. The constant in the equation enables the exponential function to adopt more flexible shapes, and 0.5 was derived to better fit ChIA-PET and Hi-C data. As rank product was finally used to predict targets, the exact value of this constant would not change the regulatory potential ranking of genes. Incorporating binding affinity into the model does not significantly improve the prediction power, therefore were excluded from the model. We represented each gene using two parameters: the differential gene expression z-score (if multiple transcriptome data are available) or t-value (if single transcriptome data is available) and the regulatory potential. For target prediction, we only considered genes with at least one binding site within 100 kb from its TSS and a differential expression z-score or t-value above the 75th percentile. We applied the Breitlings rank product method (14,15) to combine transcription factor binding potentials with differential expression values (shown in Fig. 3a is an example of the rank product result from integration of one ER ChIP-chip dataset and one differential expression dataset of estrogen 12hr treatment). The FDR of each predicted target is estimated by a permutation method proposed in (14).

Results

Dataset summary

We collected a total of 88 cistrome datasets for 13 NRs, 121 cistrome datasets for 21 collaborating factors and 94 genome-wide analyses of 12 histone modifications, which were profiled in the same cell systems as the NRs. These datasets encompass all of the published genome-wide ChIP-chip/seq studies on NRs and their related factors in humans and mice before 2011, as far as we are aware. We included the ESRRB ChIP-seq conducted in mouse embryonic stem cells but did not include other stem cell ChIP-chip/seq data because of the large number of such datasets that are not necessarily related to the cancer focus of this study. For ChIP-chip data consistency, we did not include any chromosome-wide, custom tiling, or spotted cDNA arrays but did include ChIP-chip on Affymetrix whole genome or promoter tiling arrays because of their stable designs. MAT (16) and MACS (17) were used for ChIP-chip and ChIP-seq peak calling, respectively. In addition, we analyzed 40 gene expression datasets for 11 activation and/or deactivation experiments on NRs, totaling 319 microarray profiles. A summary of the data analyzed is shown in Table 1.

Motif analyses

Previous protein structure analysis has suggested that in NR dimers, one monomer often binds to DNA much more strongly than the other (10). However, when we applied MDscan for de novo motif discovery in the NR cistrome sites, the NR full-site motifs identified were surprisingly symmetric between the two half-sites. In addition, when we collected full-site motif hits in the cistrome sites having sufficiently good overall matches (the summation of the two half-site matching scores) to the consensus sequence, the two half-sites were also symmetric (Fig. 1a). This suggests that the two monomers contribute similarly to the *in vivo* binding, which may differ from *in vitro* binding.

We then examined how the two monomers were arranged in directed (DR), everted (ER) or inverted (IR) configurations with variable gaps for different NRs (see Fig. 1a and Supplementary Fig. S1–7). Most of the NRs investigated had only one strong full-site motif, corresponding to their previously known canonical motif. Many other noncanonical motifs, while significantly enriched compared with the genome background, were much weaker than the canonical ones (Fig. 1a), suggesting that the binding sites with non-canonical motifs may be functional in a more context-dependent manner. One interesting exception was ESR1, which had strong enrichment of DR, ER, and IR motifs.

Some NRs that form heterodimers with other NRs can recognize different full-site motifs. For example, RXR recognizes DR5 when dimerizing with RARA in human NB4 cells and DR1 when dimerizing with PPARG in mouse adipocyte cells. Note that RXR and its dimerization partner in adipocytes, PPARG, show noncanonical ER14 and IR3 motifs (Supplementary Fig. S1–S2) in pre-adipocytes. For PPARG, the enrichment level of ER14 and IR3 motifs became weaker during adipogenesis and disappeared in mature adipocytes, whereas that of DR1 became stronger. For PPARG's dimerization partner RXR, ER14 and IR3 enrichment was also observed in early adipogenesis, and DR1 enrichment was observed in mature adipocyte. This suggests that PPARG may have different interaction partners and recognition patterns in early adipogenesis. Further studies are needed to identify these factors and their transcriptional consequences.

Previous studies using *in vitro* gel-shift and protein structure analysis implied that some NRs could bind half-site motifs and function as monomers *in vivo*. We took all of the NR cistrome sites that do not contain a full-site (with DR0-20, ER0-20, and IR0-20 patterns), and searched for half-site occurrences. Using regions 2 kb away from cistrome sites as a random control, we found that RARG had the strongest pattern of half-site enrichment (Supplementary Fig. S8). Other factors showing half-site enrichment include AR, ESR2, NR3C1, PPARG, and RARA, suggesting that they indeed bind to DNA *in vivo* as a monomer in addition to a dimer (Supplementary Fig. S8). Interestingly, the ESRRB ChIP-seq data from mouse embryonic stem cells showed three equally enriched full-site motifs, DR0, DR5, and DR8 (Fig. 1a), but no half-site enrichment (Supplementary Fig. S8). Although ESRRB was previously suggested to bind as a monomer (18), its ChIP-seq data indicates that ESRRB functions as a dimer *in vivo*. We then examined whether the NR peaks with only half-sites were associated (within 50 bp between peak and gene TSS) with significantly more differentially expressed genes than random genes. We indeed observed significant association, especially between the half-sites and up-regulated genes (Supplementary Fig. S9), which indicates that NR binding to half-sites is very likely to be functional.

ChIP-chip/seq can pull down targets of transcription factors that interact with the ChIP-ed factor of interest. We therefore conducted a motif analysis to find the most significant collaborating motifs for each NR (Supplementary Fig. 1b; Supplementary Table S2). Among these motifs, there are previously reported and experimentally validated collaborating

motifs, including FoxA1 for AR, ESR1, RARA and RARG, C/EBP for HNF4A, NR3C1, PPARG and RXR, PU.1 for NR1H2, and CDX2 for HNF4A; there are also newly discovered collaborating motifs, including FoxA1 for PGR, AP-1 for ESR2, NR3C1 and VDR, and PU.1 for RARA and RXR. One interesting phenomenon we noticed is that many of the transcription factors for NRs such as FoxA1, C/EBP, PU.1 and CDX2 have relatively short and AT-rich motifs. These motifs are likely the cell-type-specific chromatin remodelers that can more easily bind to nucleosome-free regions. Once these pioneering factors pry open the chromatin, NRs can bind to the DNA and, with their relatively longer motif patterns, convey specific transcriptional effects.

Binding site distributions

When transcription factor cistromes were first published (19), a surprising finding was that, despite the significant promoter enrichment observed, most binding sites were located in introns and distal intergenic regions far away from transcription start sites (TSSs). With the cistromes of many NRs at hand, we investigated these findings in a more comprehensive way. For most of the NRs, the median distance between a binding site and its nearest gene was over 10 kb. Stronger binding sites were no closer to the genes, although binding sites were much closer to genes than random genomic regions (Fig. 2a). However, binding sites were often significantly closer to genes that are differentially expressed upon NR activation (Fig. 2b). Interestingly, although factors AR, ESR2, HNF4A, NR3C1, PGR, VDR, and PPARG were only closer to up-regulated genes, ESR1, RARA and RARG in MCF7 were closer to both up- and down-regulated genes (Fig. 2b). This finding suggests that ESR1, RARA and RARG have dual functions as transcriptional activators and repressors, while other NRs mainly function as activators.

We investigated the evolutionary conservation of NR cistromes over 46 vertebrate species. Previous studies from our group and others have reported that the majority of binding sites are not conserved at the sequence level (1). Indeed, the average phastCons conservation score at the 200 bp binding summits was only approximately 0.15. We then divided binding sites into two categories on the basis of their distance from a TSS. Interestingly, binding sites closer to genes were not more conserved than those farther away (Fig. 2c; Supplementary Fig. S10). In addition, binding sites near upregulated genes were not more conserved than those near random genes, and binding sites closer to upregulated genes were not always more conserved than those farther away from upregulated genes (Supplementary Fig. S11).

Target gene prediction

One of the most important goals of transcription factor ChIP-chip/seq studies is the identification of the factor's direct target genes. However, as most binding sites land in distal intergenic regions or introns, target gene prediction is not straightforward. Prior studies have often used cutoffs such as differential expression FDR (false discovery rate) < 0.05 and at least one binding within 10 kb from the gene TSS to identify targets. However, such cutoffs are arbitrary and ignore the fact that some target genes are more strongly regulated by one factor than others. Techniques such as Hi-C and ChIA-PET have been developed to study the genome-wide chromatin interactions but do not have the sensitivity or resolution to link each binding site to its regulated genes. However, these studies found that the general trend of chromatin interactions diminishes in a predictable way with increasing genomic distance. In addition, our preliminary analysis found that enhancer regulation potential is proportional to the number of binding sites near a gene, and this finding suggests that transcription factor binding and regulatory gene target follows a many-to-many relationship. Therefore, by combining differential gene expression profiles with

transcription factor cistromes, we should be able to make improved probabilistic prediction of a factor's direct target genes.

Through meta-analysis of different omic data of the same cancer cell line model from multiple studies, we generated consensus cistrome and expression profiles: we combined multiple ChIP-chip/seq datasets for the same NR in the same cell line model to create a consensus peak list, and we combined multiple expression datasets in the same cell line model and condition to give each gene a consensus differential expression z-score (see Materials and Methods). We further made probabilistic predictions of the NR target genes by integrating cistrome and transcriptome data (see Materials and Methods).

As a validation of our integrated target prediction method that was applied to identify ESR1 gene targets in MCF7, we calculated the correlation of all the other genes with ESR1 using van de Vijver's breast tumor expression data (20). By defining genes with expression correlations larger than 0.3 as true positives and those with correlations between -0.2 and 0.2 as true negatives, we generated a ROC-like curve of our predictions. Combining multiple expression or ChIP data gave better results than using single expression or ChIP data, and integrating expression with ChIP gave better results than each data type alone and also better results than the simple cutoff method (Fig. 3b).

Discussion

ChIP-chip/seq methods have been increasingly adopted as a powerful approach to study transcription factor regulation in normal physiology and disease. Nuclear receptors are important gene regulators in many cancer systems. We systematically collected publicly available cistrome data for nuclear receptors in cancer cells, for their collaborating transcription factors, and for histone modifications. We also integrated the cistrome data with related differential gene expression data to identify the direct targets of different nuclear receptors in these cancers. Together, these integrated data not only create a useful resource for the nuclear receptor and cancer community but also provide a more comprehensive view of the genome-wide binding characteristics and regulatory mechanisms of nuclear receptors involved in cancer.

As more related cistrome and transcriptome data become available, we will add them to the current database, such as the NR1D1 ChIP-seq dataset published in 2011 (21). We will refine the regulatory modules, including the collaborating transcription factors and gene targets, of different nuclear receptors in cancers. We are also working on a comprehensive data analysis pipeline (22), so researchers can reuse the public data in combination with their own genomic and epigenomic data to better understand gene regulation in cancers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Mitch Lazar, Chris Glass and Xiaopeng Cai for their thoughtful discussions, and Len Taing and Scott Taing for their assistance with the Amazon Compute Cloud. The project was supported by the National Basic Research (973) Program of China No. 2010CB944904 (to QT, QW, and YZ), NIH grants DK062434 (to YC, CAM, TL, and XSL) and DK074967 (to TG, ML, and MB).

References

1. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoutte J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, Wang Q, Bekiranov S, Sementchenko V, Fox EA, Silver PA, Gingeras TR, Liu XS,

- Brown M. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet.* 2006; 38:1289–1297. [PubMed: 17013392]
2. Fu XD, Goglia L, Sanchez AM, Flamini M, Giretti MS, Tosi V, Genazzani AR, Simoncini T. Progesterone receptor enhances breast cancer cell motility and invasion via extranuclear activation of focal adhesion kinase. *Endocr Relat Cancer.* 2010; 17:431–443. [PubMed: 20233709]
 3. Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J, Chen Z, Beroukhir R, Wang H, Lupien M, Wu T, Regan MM, Meyer CA, Carroll JS, Manrai AK, Janne OA, Balk SP, Mehra R, Han B, Chinnaiyan AM, Rubin MA, True L, Fiorentino M, Fiore C, Loda M, Kantoff PW, Liu XS, Brown M. Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell.* 2009; 138:245–256. [PubMed: 19632176]
 4. Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, Wang X, Wu L, Li J, Hu M, Gong Y, Cheng H, Laxman B, Vellaichamy A, Shankar S, Li Y, Dhanasekaran SM, Morey R, Barrette T, Lonigro RJ, Tomlins SA, Varambally S, Qin ZS, Chinnaiyan AM. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell.* 2010; 17:443–454. [PubMed: 20478527]
 5. Hua S, Kittler R, White KP. Genomic antagonism between retinoic acid and estrogen signaling in breast cancer. *Cell.* 2009; 137:1259–1271. [PubMed: 19563758]
 6. Martens JH, Brinkman AB, Simmer F, Francoijs KJ, Nebbioso A, Ferrara F, Altucci L, Stunnenberg HG. PML-RARalpha/RXR Alters the Epigenetic Landscape in Acute Promyelocytic Leukemia. *Cancer Cell.* 2010; 17:173–185. [PubMed: 20159609]
 7. Iliopoulos D, Jaeger SA, Hirsch HA, Bulyk ML, Struhl K. STAT3 activation of miR-21 and miR-181b-1 via PTEN and CYLD are part of the epigenetic switch linking inflammation to cancer. *Mol Cell.* 2010; 39:493–506. [PubMed: 20797623]
 8. Hirsch HA, Iliopoulos D, Joshi A, Zhang Y, Jaeger SA, Bulyk M, Tschlis PN, Shirley Liu X, Struhl K. A transcriptional signature and common gene networks link cancer with lipid metabolism and diverse human diseases. *Cancer Cell.* 2010; 17:348–361. [PubMed: 20385360]
 9. Glass CK, Saijo K. Nuclear receptor transrepression pathways that regulate inflammation in macrophages and T cells. *Nat Rev Immunol.* 2010; 10:365–376. [PubMed: 20414208]
 10. Kumar R, Thompson EB. The structure of the nuclear hormone receptors. *Steroids.* 1999; 64:310–319. [PubMed: 10406480]
 11. Nuclear Receptor Cancer Cistromes. [Accessed 2011 Jun 10] http://cistrome.dfci.harvard.edu/NR_Cistrome
 12. Ochsner SA, Steffen DL, Hilsenbeck SG, Chen ES, Watkins C, McKenna NJ. GEMS (Gene Expression MetaSignatures), a Web resource for querying meta-analysis of expression microarray datasets: 17beta-estradiol in MCF-7 cells. *Cancer Res.* 2009; 69:23–26. [PubMed: 19117983]
 13. Chen Y, Meyer CA, Liu T, Li W, Liu JS, Liu XS. MM-ChIP enables integrative analysis of cross-platform and between-laboratory ChIP-chip or ChIP-seq data. *Genome Biol.* 2011; 12:R11. [PubMed: 21284836]
 14. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* 2004; 573:83–92. [PubMed: 15327980]
 15. Klisch TJ, Xi Y, Flora A, Wang L, Li W, Zoghbi HY. In vivo Atoh1 targetome reveals how a proneural transcription factor regulates cerebellar development. *Proc Natl Acad Sci U S A.* 2011; 108:3288–3293. [PubMed: 21300888]
 16. Johnson WE, Li W, Meyer CA, Gottardo R, Carroll JS, Brown M, Liu XS. Model-based analysis of tiling-arrays for ChIP-chip. *Proc Natl Acad Sci U S A.* 2006; 103:12457–12462. [PubMed: 16895995]
 17. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008; 9:R137. [PubMed: 18798982]
 18. Gearhart MD, Holmbeck SM, Evans RM, Dyson HJ, Wright PE. Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition. *J Mol Biol.* 2003; 327:819–832. [PubMed: 12654265]

19. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoutte J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*. 2005; 122:33–43. [PubMed: 16009131]
20. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002; 347:1999–2009. [PubMed: 12490681]
21. Feng D, Liu T, Sun Z, Bugge A, Mullican SE, Alenghat T, Liu XS, Lazar MA. A circadian rhythm orchestrated by histone deacetylase 3 controls hepatic lipid metabolism. *Science*. 2011; 331:1315–1319. [PubMed: 21393543]
22. Cistrome. [Accessed 2010 Oct 20] <http://cistrome.org/ap/>

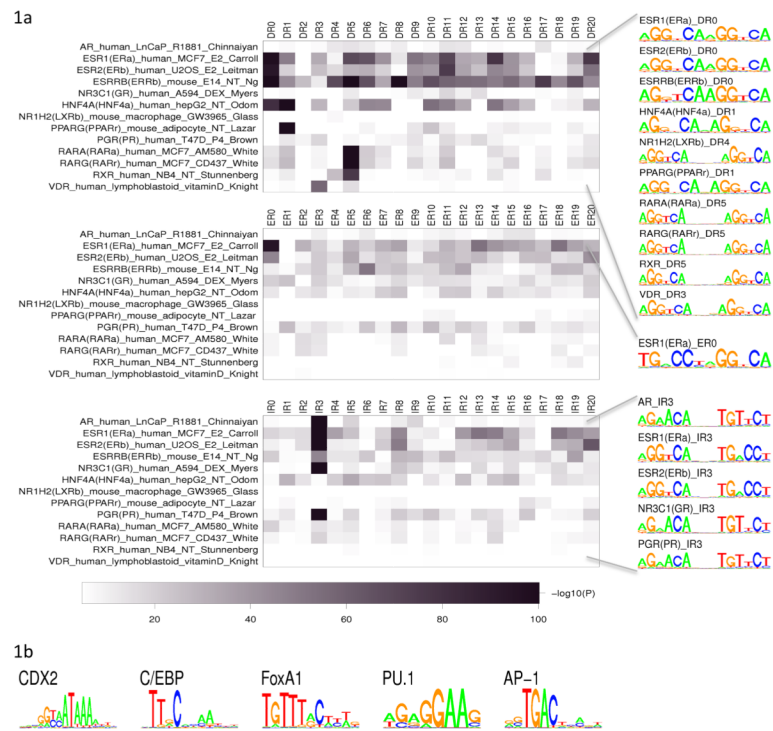


Figure 1.
 (a) Enrichment heat map of NR full-site motifs arranged in directed (DR), everted (ER) or inverted (IR) patterns spaced by 0–20 random nucleotides. One representative dataset for each NR was selected and is shown in the figure. To the right of the heat map are sequence logos of NR full-site motifs retrieved from ChIP-chip/seq peak regions. (b) Sequence logos of identified NR collaborating factors.

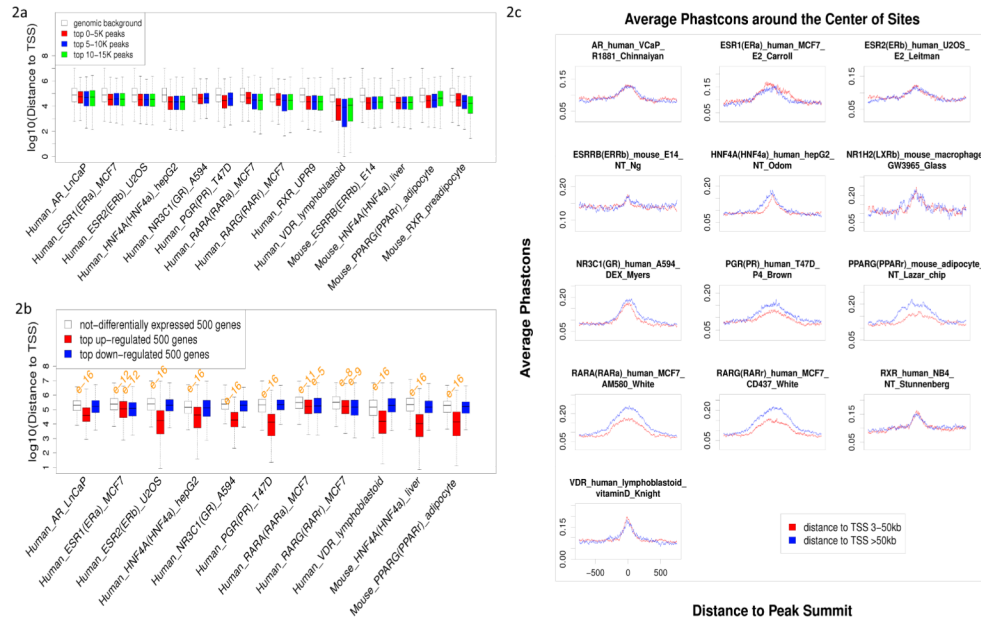


Figure 2. (a) Top 0–5 K/5–10 K/10–15 K peaks (ranked by p-values from small to large) were compared by their distances to the nearest genes, and all the NRs with at least one dataset of over 10 K peaks are included. No significant differences in peak distance were observed among the three groups. (b) Two groups of genes (upregulated and downregulated; 500 genes) were compared to one group (nondifferentially expressed; 500 genes) by their distances to the nearest peaks, and all the NRs with at least one ChIP-chip/seq dataset coupled with de/activation experimental dataset were included. Wilcoxon rank sum test was utilized to test significance of distance disparity; significantly small p-values are marked above the corresponding boxplot. For all the NRs, upregulated genes were closer to binding sites than nondifferentially expressed genes; however, similar phenomena were observed in downregulated genes only for ESR1, RARA and RARG. (c) Average phasCons conservation score profiles around the 1,200 bp summits of NR binding sites. Two groups of binding sites, with different distances to their nearest genes, were compared by their phasCons conservation score profiles, and no significant variations were observed.

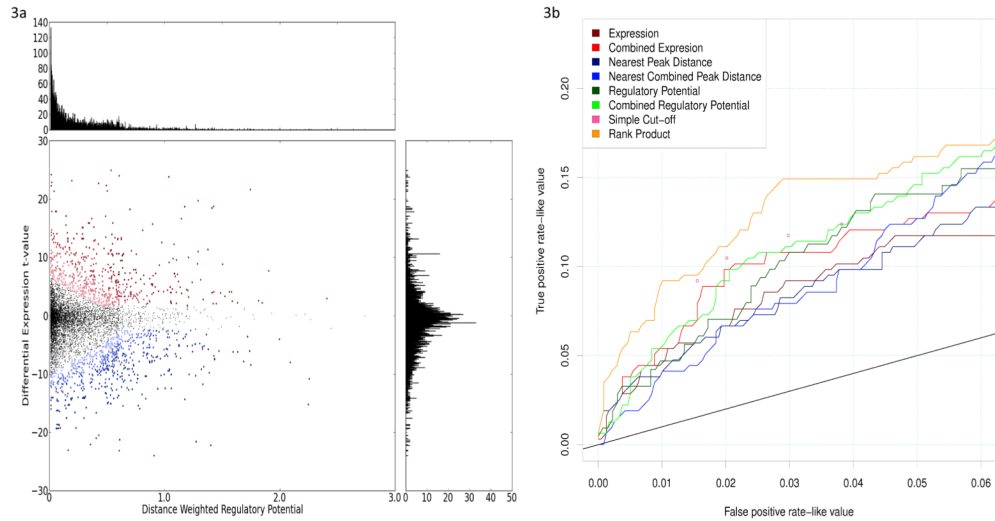


Figure 3. (a) Scatter plot of genes represented by two parameters: the regulatory potential calculated from Brown labs's ER ChIP-chip dataset and the differential expression t-value calculated from Brown lab's expression dataset of 12hr estrogen treatment. Rank product method was utilized to integrate the two parameters and render a rank order list of genes according to their likelihoods of being ER targets. Red dots represent the top 800 genes that are most likely to be upregulated ER targets; red dots with darker colors are more likely to be targets than those with lighter colors. Similarly, blue dots represent the top 800 genes that are most likely to be downregulated ER targets; blue dots with darker colors are more likely to be targets than those with lighter colors. The horizontal histogram represents the distribution of regulatory potential, and the vertical histogram represents the distribution differential expression values. (b) ROC-like curve for ESR1 as a validation of our integrated target prediction method. We calculated the correlation values of all the other genes with ESR1 using van de Vijver's breast tumor expression data, and defined genes with expression correlations larger than 0.3 as true positives and those with correlations between -0.2 and 0.2 as true negatives. ESR1 target genes predicted by different approaches were compared.

Table 1

Summary of cistrome, epigenome and transcriptome datasets included in the NRCistrome web interface.

	Species	Factor number	Factor	Dataset number
Nuclear receptor	Human	10	AR, ESR1, ESR2, HNF4A, NR3C1, PGR, RARA, RARG, RXR, VDR	66
Collaborating factor	Mouse	7	ESRRB, HNF4A, NR1H2, NR3C1, PPARC, RAR, RXR	22
	Human	18	CDX2, CEBP, CTCF, ERG, FoxA1, FoxA2, GABP, GATA3, GATA6, PML, PolII, RAD21, SRC-3, STAG1, TRIM24, c-Fos, c-Jun, c_MYC	82
Epigenome	Mouse	6	CEBP, FoxA2, Oct2, PDX1, PU1, PolII	39
	Human	12	AcC_H3, H3K14ac, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me3, H3K17me2, Pan-H3	75
Transcriptome	Mouse	3	H3K4me1, H3K4me3, H3K9ac	19
	Human	9	AR, ESR1, ESR2, HNF4A, NR3C1, PGR, RARA, RARG, VDR	35
	Mouse	3	HNF4A, NR1H2, PPARC	5