

RESEARCH ARTICLE

Open Access

# Inference of hierarchical regulatory network of estrogen-dependent breast cancer through ChIP-based data

Fei Gu<sup>1</sup>, Hang-Kai Hsu<sup>2</sup>, Pei-Yin Hsu<sup>2</sup>, Jiejun Wu<sup>2</sup>, Yilin Ma<sup>1</sup>, Jeffrey Parvin<sup>1</sup>, Tim H-M Huang<sup>2</sup>, Victor X Jin<sup>1\*</sup>

## Abstract

**Background:** Global profiling of in vivo protein-DNA interactions using ChIP-based technologies has evolved rapidly in recent years. Although many genome-wide studies have identified thousands of ER $\alpha$  binding sites and have revealed the associated transcription factor (TF) partners, such as AP1, FOXA1 and CEBP, little is known about ER $\alpha$  associated hierarchical transcriptional regulatory networks.

**Results:** In this study, we applied computational approaches to analyze three public available ChIP-based datasets: ChIP-seq, ChIP-PET and ChIP-chip, and to investigate the hierarchical regulatory network for ER $\alpha$  and ER $\alpha$  partner TFs regulation in estrogen-dependent breast cancer MCF7 cells. 16 common TFs and two common new TF partners (RORA and PITX2) were found among ChIP-seq, ChIP-chip and ChIP-PET datasets. The regulatory networks were constructed by scanning the ChIP-peak region with TF specific position weight matrix (PWM). A permutation test was performed to test the reliability of each connection of the network. We then used DREM software to perform gene ontology function analysis on the common genes. We found that FOS, PITX2, RORA and FOXA1 were involved in the up-regulated genes.

We also conducted the ER $\alpha$  and Pol-II ChIP-seq experiments in tamoxifen resistance MCF7 cells (denoted as MCF7-T in this study) and compared the difference between MCF7 and MCF7-T cells. The result showed very little overlap between these two cells in terms of targeted genes (21.2% of common genes) and targeted TFs (25% of common TFs). The significant dissimilarity may indicate totally different transcriptional regulatory mechanisms between these two cancer cells.

**Conclusions:** Our study uncovers new estrogen-mediated regulatory networks by mining three ChIP-based data in MCF7 cells and ChIP-seq data in MCF7-T cells. We compared the different ChIP-based technologies as well as different breast cancer cells. Our computational analytical approach may guide biologists to further study the underlying mechanisms in breast cancer cells or other human diseases.

## Background

Global level profiling of in vivo protein-DNA interactions using ChIP-based technologies has evolved rapidly in recent years, from hybridization with spotted or tiling microarray (ChIP-chip) [1-4], to SAGE-like tags (ChIP-SAGE) [5] or pair-end tag sequencing (ChIP-PET) [6], to current massively parallel sequencing (ChIP-seq) [7-10].

Estrogen-mediated gene regulation is such a challenging question that it may require powerful genome-wide profiling tools like ChIP-based technologies. In breast cancer cells, ER $\alpha$  can mediate genomic transcription regulation with nuclear initiated steroid signalling and non-genomic activation of various protein kinase cascades [11]. In the classical genomic pathway, estrogen receptor binds to estrogen response elements (ERE) at the regulatory region of the target genes and recruits co-activators or co-repressors to modulate gene transcription [12]. The non-classical genomic pathway does not require ERE but mediates transcription by the interactions of ER $\alpha$  with other proteins such as AP1 [13],

\* Correspondence: victor.jin@osumc.edu

<sup>1</sup>Department of Biomedical Informatics, The Ohio State University, Columbus, USA

Full list of author information is available at the end of the article

NF- $\kappa$ B [14], SP1 [15,16] and others. At molecular level, we need to identify those genes targeted and regulated by estrogen receptors, and the more challenging task is to delineate the architectures and the underlying mechanisms of such regulation. Estrogen receptors, once activated, may induce increased or decreased transcription of its numerous targets, which have been investigated by expression arrays [17,18]. In some recent publications, profiling the distribution of ER $\alpha$  by ChIP-seq ChIP-PET and ChIP-chip indicated a highly complicated regulation network involved with both ER $\alpha$  and other relative transcription factors [17-19]. For instance, only a small portion of ER $\alpha$  binding sites were located in the promoter regions of known genes and many unforeseen binding sites could be far away from the TSS, up to 50-100 kb. It was also found that a large number of transcription factors had binding sites co-enriched with ER $\alpha$  binding sites, which indicated a close collaboration between ER $\alpha$  and other factors. All these findings support the evolving concept of estrogen receptor regulation from the conventional interaction between ERE and ER $\alpha$  to the long-range chromatin loop [20].

Tamoxifen is one of selective ER modulators (SERMs) and is widely used to block ER $\alpha$  function for breast cancer treatment [11]; however, this endocrine therapy is limited by the onset of drug resistance. Tamoxifen resistance could be induced through both genomic and non-genomic estrogen pathways mentioned above and understanding estrogen regulation network will offer therapeutic advantages. Our recent expression array study showed that, in breast cancer cells with acquired tamoxifen resistance, different groups of genes were targeted by estrogen treatment compared with the parental cells [20]. Delineating the changed architectures of the ER $\alpha$  regulation network in tamoxifen resistance cells by ChIP-based assays may provide direct and useful information on tamoxifen resistance.

In this study, we collected three public available ChIP-based datasets—ChIP-chip, ChIP-PET and ChIP-seq for ER $\alpha$  binding sites in breast cancer MCF7 cells upon estrogen exposure, which also include RNA polymerase II (Pol-II) binding sites in these cells since the binding of Pol-II could provide direct information of potential transcription activation. We then applied computational approaches to investigate the hierarchical regulatory information for ER $\alpha$  regulation in MCF7 cells. We have been able to construct hierarchical regulatory networks with target hubs and have established the regulatory pathways between TFs and genes. We also have applied ChIP-seq technology to systematically compare the estrogen-mediated regulatory information between MCF7 and MCF7-T cells.

## Results

### Whole-Genome-Wide localization of *in vivo* ER $\alpha$ /Pol-II binding peaks to genes in breast cancer cells

It is well recognized that the transcription is driven by Pol-II and the general transcriptional machinery; therefore, it is likely that the genes (or promoters) bound by ER $\alpha$  may not be transcribed without Pol-II binding. As such, we need to use the whole-genome-wide binding peaks information combining ER $\alpha$  with Pol-II antibodies.

The peak-calling procedure was described in the **Methods** section. The number of ChIP peaks from ChIP-seq is 12,516 for ER $\alpha$  and 13,261 for Pol-II upon E2-treated compared to the vehicle control, the number from ChIP-PET is 14,703 for ER $\alpha$  and 13,133 for Pol-II, the number from ChIP-chip is 10,409 for ER $\alpha$  and 11,455 for Pol-II (Additional file 1, Table S1).

Our previous studies [20-23] and others [24-28] have tried to establish and characterize the molecular mechanisms of estrogen-dependent breast cancer cells. In order to further elucidate the more detailed underlying mechanisms, we then created a whole-genome-wide localization of *in vivo* ER $\alpha$  binding peaks in the MCF7. We located these identified ER $\alpha$  and Pol-II binding peaks relative to a known annotated gene from the RefSeq database (UCSC HG18 Assembly). Our results showed that 18% for ChIP-seq, 44% for ChIP-PET, 45% for ChIP-chip of Pol-II binding peaks are located within Promoter regions (defined as 2 kb upstream to 2 kb downstream, including 5' Core, 5'TSS and WithinGene Core regions) of a known transcription start site (5'TSS) (Additional file 1, Table S1 and Additional file 2, Figure S1). We also found that a relatively small number of ER $\alpha$  binding peaks are located in a known 5'TSS region (9% for ChIP-seq, 6% for ChIP-PET, 6% for ChIP-chip). A big portion of ER $\alpha$  binding peaks are located within intra-genic regions (38% for ChIP-seq, 37% for ChIP-PET, 58% for ChIP-chip) as well as gene desert regions (14% for ChIP-seq, 20% for ChIP-PET, 12% for ChIP-chip), 100 kb far away from known 5'TSSs and 3'TSSs (Additional file 2, Figure S1B). While our location analysis confirms the results from previous studies [17-19] where the data were collected, our comparative analysis further showed ER $\alpha$  binding patterns are essentially similar in which the majority of ER $\alpha$  binding peaks are outside of proximal promoter regions regardless of different ChIP technologies. Our analysis also suggests that most ER $\alpha$  associated genes may be regulated by the long-distance interaction between ER-bound distal enhancer and proximal promoter regions. We also found that some portions (18% for ChIP-seq, 17% for ChIP-PET, 11% for ChIP-chip) of Pol-II binding peaks are located in the 3' ends (including 3' Core, 3' Proximal

and 3' Distal regions). This finding might imply some alternative transcripts are transcribed in the 3' end and is consistent to the previous report of alternative promoters identified in the 3' end [29]. This could also be simply due to either stalled RNA Pol II that has finished transcription or loops formed with the promoters as has been proposed [30].

#### Regulation of ER $\alpha$ target gene expression in MCF7 cells

We then systematically compared three different ChIP based technologies. We aim to correlate 1,513 E2-induced genes identified by Carroll et al. [17] to ER $\alpha$  and Pol-II binding peaks identified from different ChIP-based datasets. For Pol-II binding peaks, we only considered those peaks within proximal upstream regions (less than 10 kb) relative of TSS, within intra-genic regions, and within proximal downstream regions (less than 10 kb) relative of TTS. For ER $\alpha$  binding peaks, we considered all binding peaks except those in the gene desert regions (larger than 100 kb away from a TSS). For ChIP-seq data, it turned out to have a total of 12,516 ER $\alpha$  binding peaks corresponded to 5,693 annotated genes and 13,261 Pol-II binding peaks corresponded to 5,186 genes. And 2,661 genes were identified to have both Pol-II and ER $\alpha$  binding peaks in ChIP-seq dataset (Figure 1A, top panel). Among these 2,661 genes with enriched double (ER $\alpha$  and Pol-II) binding peaks, only 273 of them overlapped with 1,513 E2-induced genes in MCF7 show differential expression (Figure 1A, top panel). Surprisingly, a majority (1,240 of 1,513) of genes in ChIP-seq dataset with differential expression lack ER $\alpha$  and Pol-II binding sites. After examining the correlation of ER $\alpha$  and Pol-II identified by ChIP-chip dataset in MCF7, we obtained very similar results, where 307 genes have both Pol-II and ER enriched binding peaks. (Figure 1A, bottom-right panel). In order to exclude different experiments' discrepancy, we also compared the ER $\alpha$  and Pol-II binding data identified by ChIP-PET, and E2-induced expression data in MCF7. Similarly, only 320 common genes were obtained (Figure 1A, bottom-left panel). The overlap of gene number of three different technologies was shown in Figure 1B. Overall, the similar findings in all three ChIP-based dataset seem to indicate that ER $\alpha$  binding and E2-mediated Pol-II binding in most of ER $\alpha$  direct targets may not necessarily lead to detectable activated or repressed expression changes. Instead, they may be in paused states as the previous reports suggested [18,31,32]. Meanwhile, it also seems that a majority of E2-induced expression changes could be caused by downstream estrogen effect and may have not direct and close correlation with ER $\alpha$ -recruited Pol-II complex for transcription initiation [33].

We also found that 164 genes (~60%) are common between ChIP-seq and ChIP-chip (Figure 1A) while 183

(~59.6%) common genes between ChIP-chip and ChIP-PET (Figure 1A). A comparison of ChIP-seq and ChIP-PET has shown that they have 172 (~63%) common genes (Figure 1A).

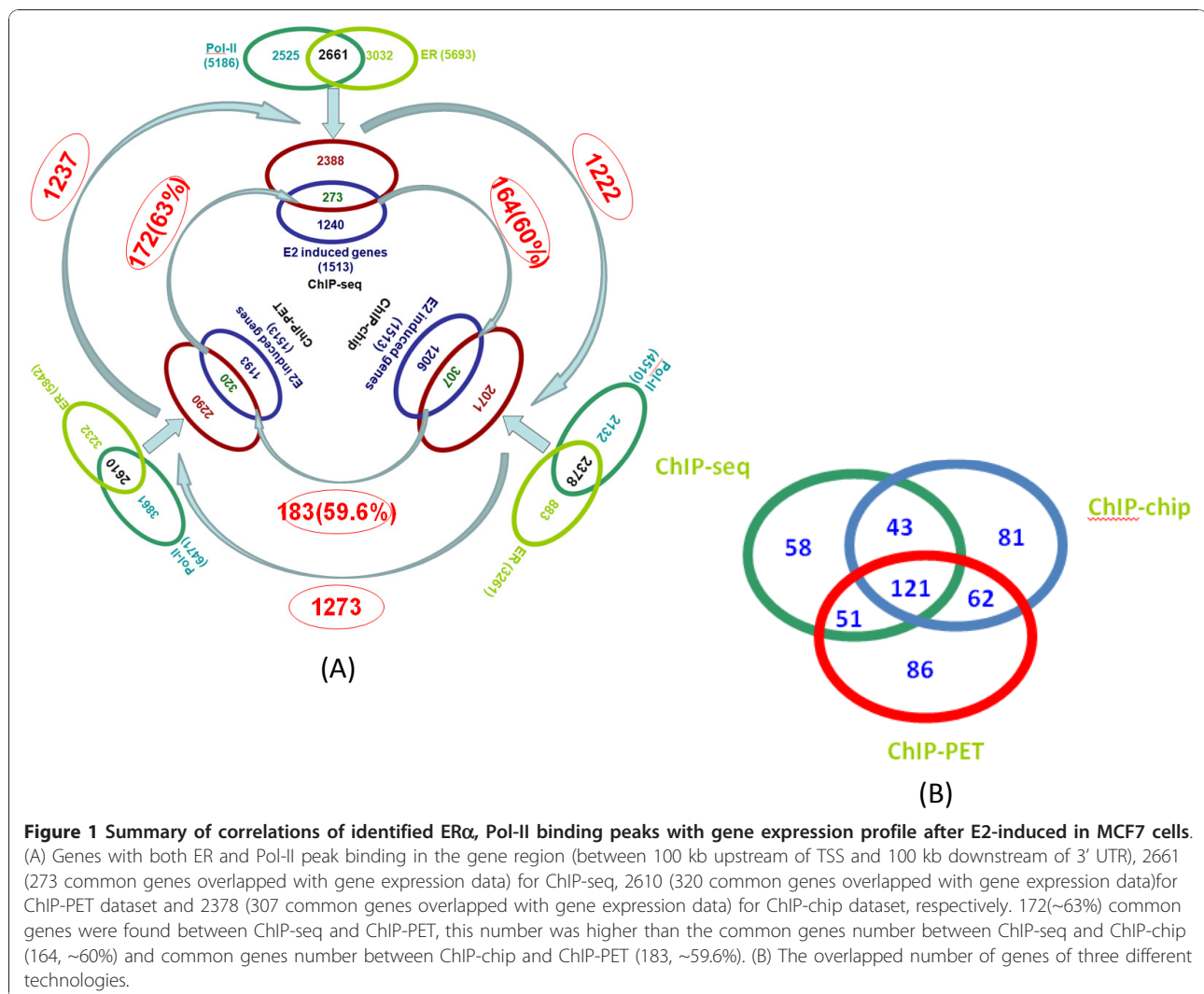
#### De Novo identification for ER $\alpha$ binding sites and its binding partners in MCF7 cells

One possible scenario to explain why the majority of ER $\alpha$  direct targets are not activated or repressed is that they may lack certain binding TF partners serving as helpers for ER $\alpha$  to co-regulate these genes. Thus we applied the *de novo* motif discovery approach (ChIP-Motifs) [34,35] developed in our previous study to identify the known or novel ER $\alpha$  TF partners. ChIP-Motifs is online software designed for searching the most significant motifs in given peaks based on the known factor motifs in TRANSFAC [36] and JASPAR [37] databases.

In this study, for each ChIP-based dataset, we selected the top 2000 peaks with high scores (enrichments) as the input data for ChIPMotifs. After running ChIPMotifs, ERE, PAX6, PITX2 and RORA were identified in ChIP-seq data; ERE, PITX2, RORA and GATA2 were identified in ChIP-chip data; RORA, PAX6, PITX2 and ERE were identified in ChIP-PET data (Figure 2). We failed to identify several TFs (CEBP [38], FOS (AP1) [39], FOXA1 [17]) reported to be associated with ER $\alpha$  in previous studies. One possible reason is that our *de novo* ChIPMotifs first *ab initio* identify motifs at a set of relatively short sequences (~300 - 500 bp), then find possible matched TFs from the TRANSFAC database after obtaining significant motifs, therefore, this might miss some co-TFs if they locate outside 500 bp distance from ER $\alpha$ . But if we used longer sequences, that might lose the specificity for the identified motifs such as missing identifying ERE. Regardless, we chose to include these three co-TFs (CEBP, FOS and FOXA1) in our further analysis. These TFs' Seq-LOGOs in ChIP-seq data, ChIP-chip data and ChIP-PET data were shown in Figure 2. Further analysis of these data indicates that in addition to ERE, there are two TF motifs common among ChIP-seq data, ChIP-chip data and ChIP-PET data—PITX2 and RORA (Figure 2).

#### ER $\alpha$ regulatory network in estrogen-dependent MCF7 cells

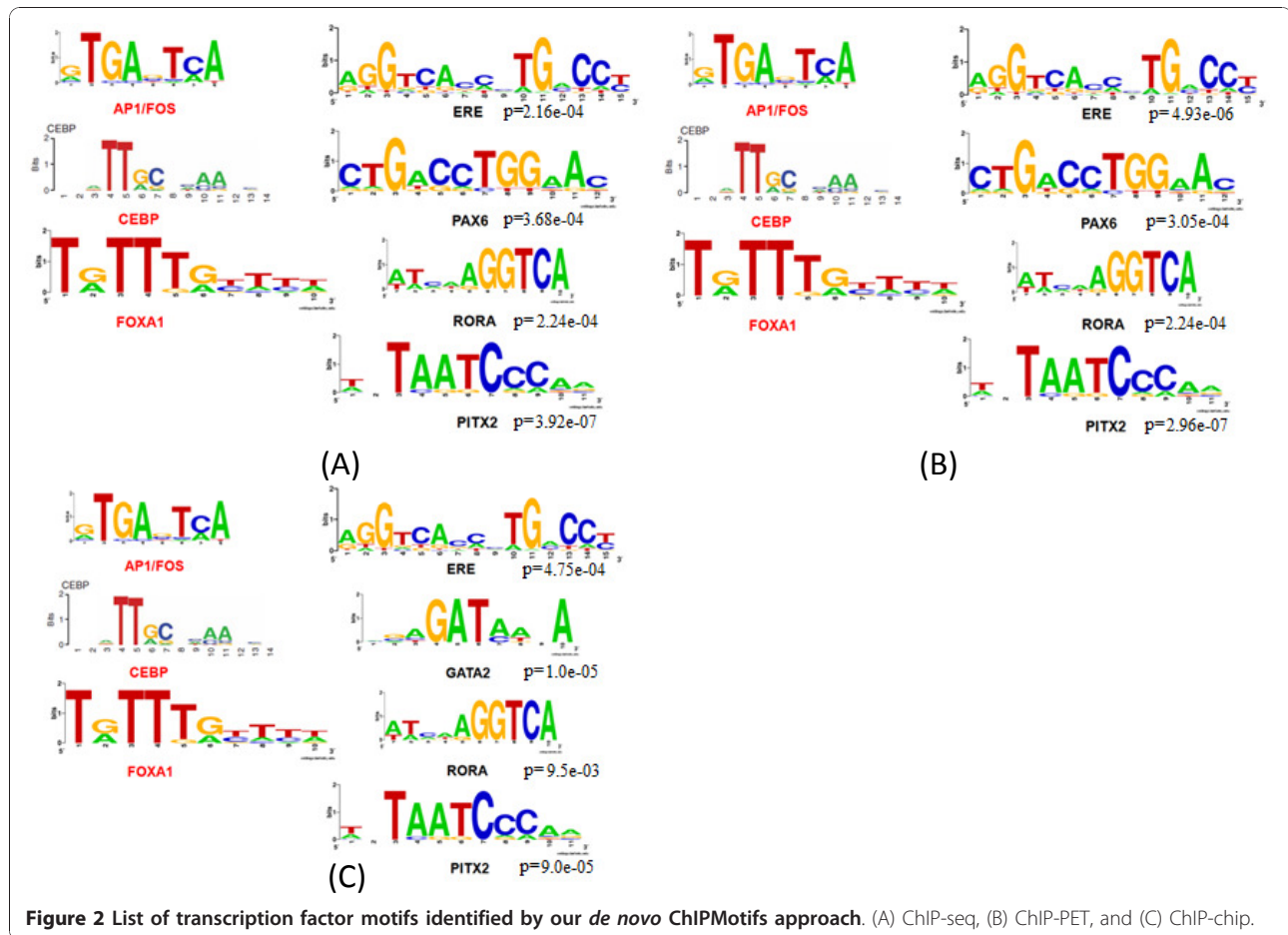
Previous studies have shown ER $\alpha$  regulates its target genes through three major binding models a) direct binding to ERE (estrogen response element); b) indirect binding, through which it binds to other TF partners which bind to DNA; and c) co-occurrent binding, where both ER $\alpha$  and other TF partners bind to their own specific DNA motifs (Additional file 3, Figure S2) [20-22,40,41].



**Figure 1 Summary of correlations of identified ER $\alpha$ , Pol-II binding peaks with gene expression profile after E2-induced in MCF7 cells.** (A) Genes with both ER and Pol-II peak binding in the gene region (between 100 kb upstream of TSS and 100 kb downstream of 3' UTR), 2661 (273 common genes overlapped with gene expression data) for ChIP-seq, 2610 (320 common genes overlapped with gene expression data) for ChIP-PET dataset and 2378 (307 common genes overlapped with gene expression data) for ChIP-chip dataset, respectively. 172 (~63%) common genes were found between ChIP-seq and ChIP-PET, this number was higher than the common genes number between ChIP-seq and ChIP-chip (164, ~60%) and common genes number between ChIP-chip and ChIP-PET (183, ~59.6%). (B) The overlapped number of genes of three different technologies.

In order to understand how the regulatory network responds to E2 treatment in three ChIP-based technologies, we here integrated the ChIP data with the time series of E2 induced gene expression data. In order to determine if a gene is differentially expressed, we used the difference of expression levels between time point 12 hr and 0 hr, positive value as up-regulated and negative value as down-regulated. Thus all genes in the networks were differentially expressed and with ER $\alpha$  and PolII binding ChIP-peaks (as mentioned in “**Regulation of ER $\alpha$  target gene expression in MCF7 cells**” section, the ER $\alpha$  peak location is between 100 kb upstream of 5'TSS and 100 kb downstream of 3'TSS, while the Pol-II peak location is between 10 kb upstream of 5'TSS and 10 kb downstream of 3'TSS. The ER $\alpha$  peak and Pol-II peak do not necessarily be overlapped, but they must be located in the same gene). Furthermore, for the transcriptional regulatory network (as a part of gene regulatory network), only transcription factors were used for

the network construction (called normal TFs). ER $\alpha$  binding peaks associated with those normal TFs were further scanned by Hub TF (the TF which motif was enriched in top 2000 ER $\alpha$  binding peaks of the genes with both ER $\alpha$  and Pol-II binding sites, and was identified from the previous section “*De Novo* identification for ER $\alpha$  binding sites and its binding partners”) PWMs to determine if there is any connection between a Hub TF and a normal TF. A shuffling test was performed to test the reliability of each connection of the network. The resulted regulatory networks were thus constructed and topologically visualized using Cytoscape [42] software platform (see **Methods** section and Figure 3). In the network, all the normal/Hub TFs were represented as a node (red nodes represented for up-regulated genes, green nodes represented for down-regulated genes, and blue nodes represented for Hub TFs), and all the connections were represented as edge between 2 nodes (Figure 3). An edge has a direction, where it starts



**Figure 2** List of transcription factor motifs identified by our *de novo* ChIPMotifs approach. (A) ChIP-seq, (B) ChIP-PET, and (C) ChIP-chip.

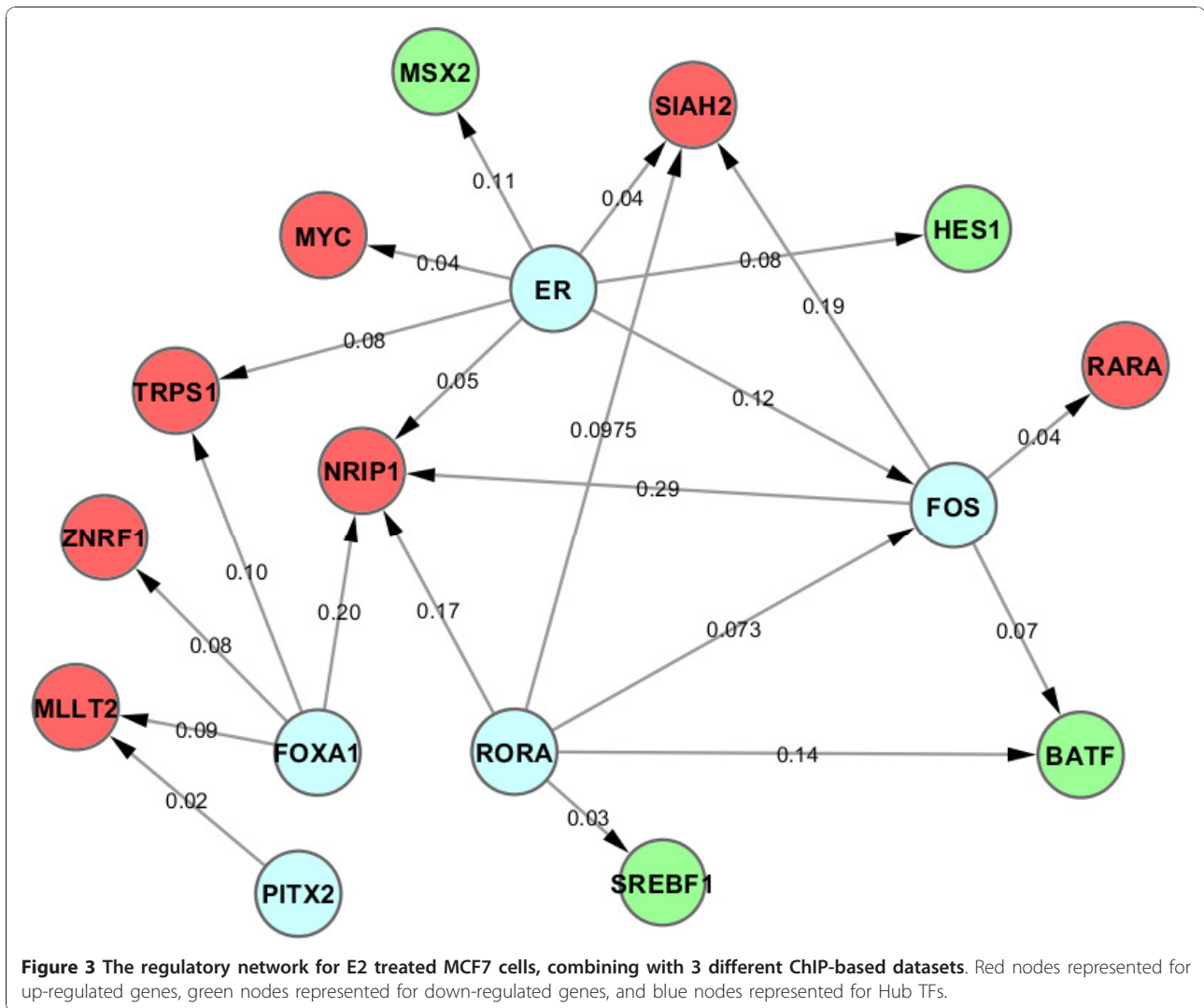
from Hub TFs to normal TFs. The edge between a Hub TF PWM (ex: FOXA1) and a normal TF differentially expressed following estrogen stimulation (ex: MYC) means that a motif of Hub TF (i.e.: FOXA1) is found within ER $\alpha$  peak region(s) associated with another TF gene (i.e: MYC). Since every normal TF was with both ER $\alpha$  and Pol-II peaks, the edge represented for the possible direct/indirect binding of Hub TF (ex: FOXA1) together with ER $\alpha$  to regulate normal TF (i.e: MYC).

In summary, 40 nodes (Additional file 4, Figure S3A) were identified for ChIP-seq data, 50 for ChIP-PET and 37 for ChIP-chip (Additional file 4, Figure S3B, C). A comparison of the three regulatory networks shows that there are 6 common hubs (ER, FOS, RORA, FOXA1, CEBP and PITX2) and 16 (~40%) common targeted TFs (Additional file 5, Table S2) in MCF7. The final combined ER $\alpha$  regulatory network for MCF7 cell from all three ChIP-based datasets was shown in Figure 3.

#### ER $\alpha$ regulatory pathway analysis in estrogen-dependent MCF7 cells

We next applied DREM program to model, analyze, and visualize dynamic gene regulatory functions [43]. The

maps would potentially infer major bifurcation events in the time series expression data and transcription factors responsible for them. Take the ChIP-seq data as example, we used 6 ER $\alpha$  associated TF partners—RORA, PITX2 and PAX6 were identified by *de novo* ChIPMotifs from our data, and CEBP, FOS (AP1) and FOXA1 were known TF partners of ER $\alpha$ . After integrating the TF-DNA interactions data with 1,513 genes (see **Methods** section), a total of 5 paths were clearly diverging for these genes (Additional file 6, Figure S4 (I)). The same TF partners were used in ChIP-PET dataset (Additional file 6, Figure S4 (II)). The TF PAX6 was replaced with GATA2 in ChIP-chip dataset (Additional file 6, Figure S4 (III)). All the TF partners were selected based on the ChIPMotifs results (see **Results** section “*De Novo* identification for ER $\alpha$  binding sites and its binding partners”). A final combined regulatory pathway map obtained from three ChIP data showed that genes were traversing the 3 splits are shown with (a) corresponding to the split at 0 hr, (b) corresponding to the split at 3 hr and (c) corresponding to the split at 6 hr (Figure 4). The paths out of the splits were annotated with TFs determined by DREM to be associated with the genes

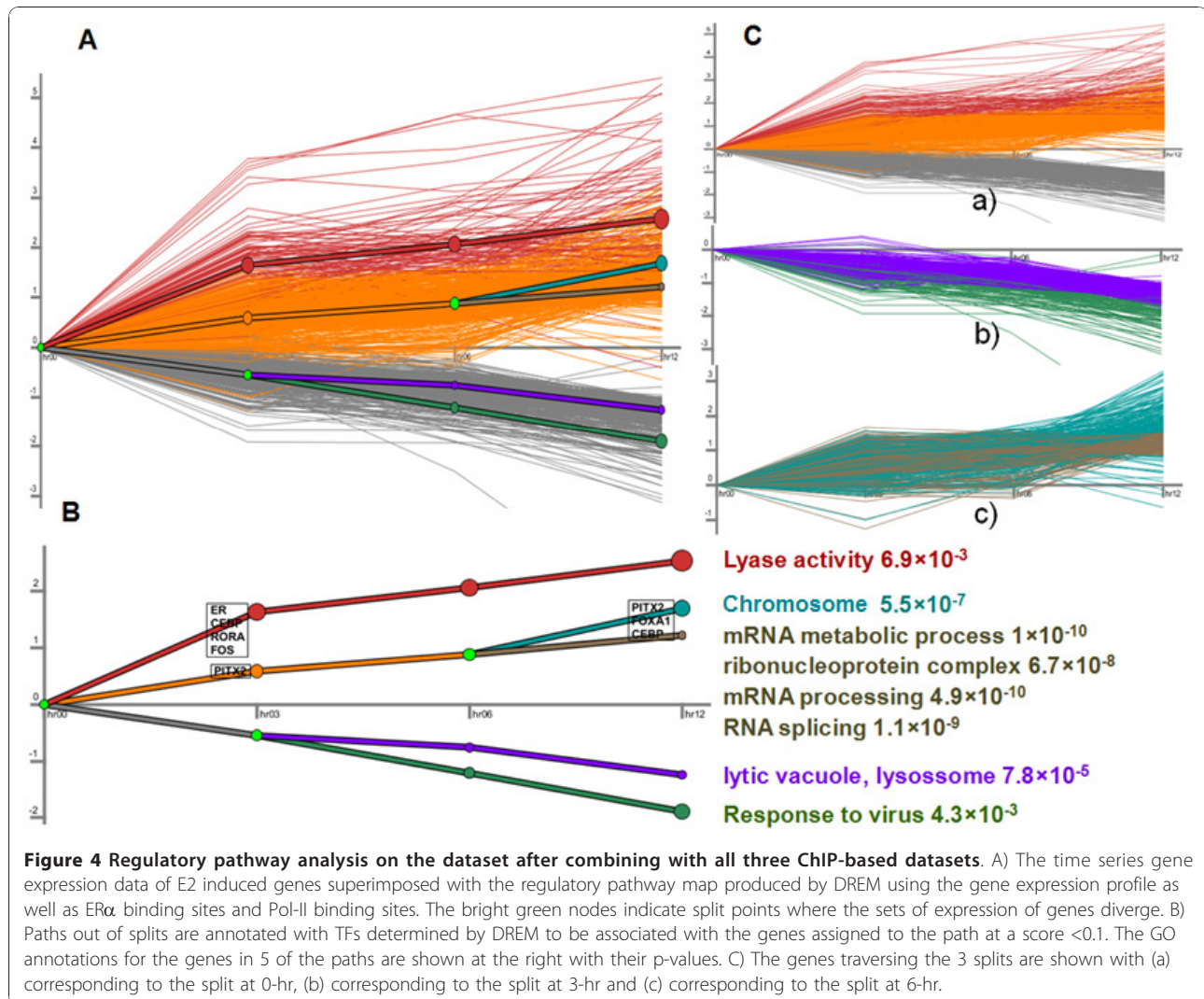


assigned to the path at a score < 0.1. The GO annotations for genes in 5 of the paths were shown at the right with their p-values (Figure 4). Our finding showed E2 mediated a set of repressed genes with direct ER $\alpha$  binding sites where they are overrepresented with lytic vacuole/lysosome ( $p = 7.8 \times 10^{-5}$ ). We further revealed a new estrogen-mediated up-regulated path requiring a TF partner PITX2, and the function of this set of genes is associated with ribonucleoprotein complex ( $p = 6.7 \times 10^{-8}$ ), mRNA metabolic process ( $p = 1 \times 10^{-10}$ ), mRNA processing ( $p = 4.9 \times 10^{-10}$ ) and chromosome ( $p = 5.5 \times 10^{-7}$ ). Our analysis also showed other paths with less significant genes: Lyase activity with a p-value of  $6.9 \times 10^{-3}$  (RORA, FOS and CEBP), response to virus with a p-value of  $4.3 \times 10^{-3}$ . We also found that the TF FOS, CEBP, FOXA1 and PITX2 are related to the up-regulated genes, while FOXA1 was shown up in both up and down-regulated genes. The gene lists of five paths were shown in

Additional file 7, Table S3. The relationship of the 6 TFs (ER $\alpha$ , RORA, FOS, FOXA1 PITX2 and CEBP) and their related genes were shown in Additional file 8, Table S4.

#### A comparison of MCF7-T and MCF7 cells

To further understand the difference of the ER $\alpha$  associated gene regulatory information between MCF7 and MCF7-T, a tamoxifen-resistant breast cancer cell line, we conducted ER $\alpha$  and Pol-II ChIP-seq experiments in MCF7-T cells (see **Methods** section). Around 22 million reads were obtained for Pol-II control data, and 19 million reads in Pol-II E2-treated data. For ER $\alpha$  antibody, the number is 14 million for control data and 16 million for E2-treated data (Table 1). After uniquely mapping to the human genome, the reads number reduced to 7 million for Pol-II control data, 5 million for Pol-II E2 treated data, 5 million for ER $\alpha$  control data and 4 million for E2 treated data, respectively (Table 1). In general,



25-30% reads were uniquely mapped to the human genome.

In MCF7-T cells, 3,596 ER $\alpha$  binding peaks corresponded to 1,743 annotated genes and 4,644 Pol-II binding peaks corresponded to 3,207 genes were found. 530 genes were identified to have both Pol-II and ER $\alpha$  binding peaks (Figure 5A). 438 of them show differential expression overlapped with 1,667 E2-induced genes identified in MCF7-T expression profile

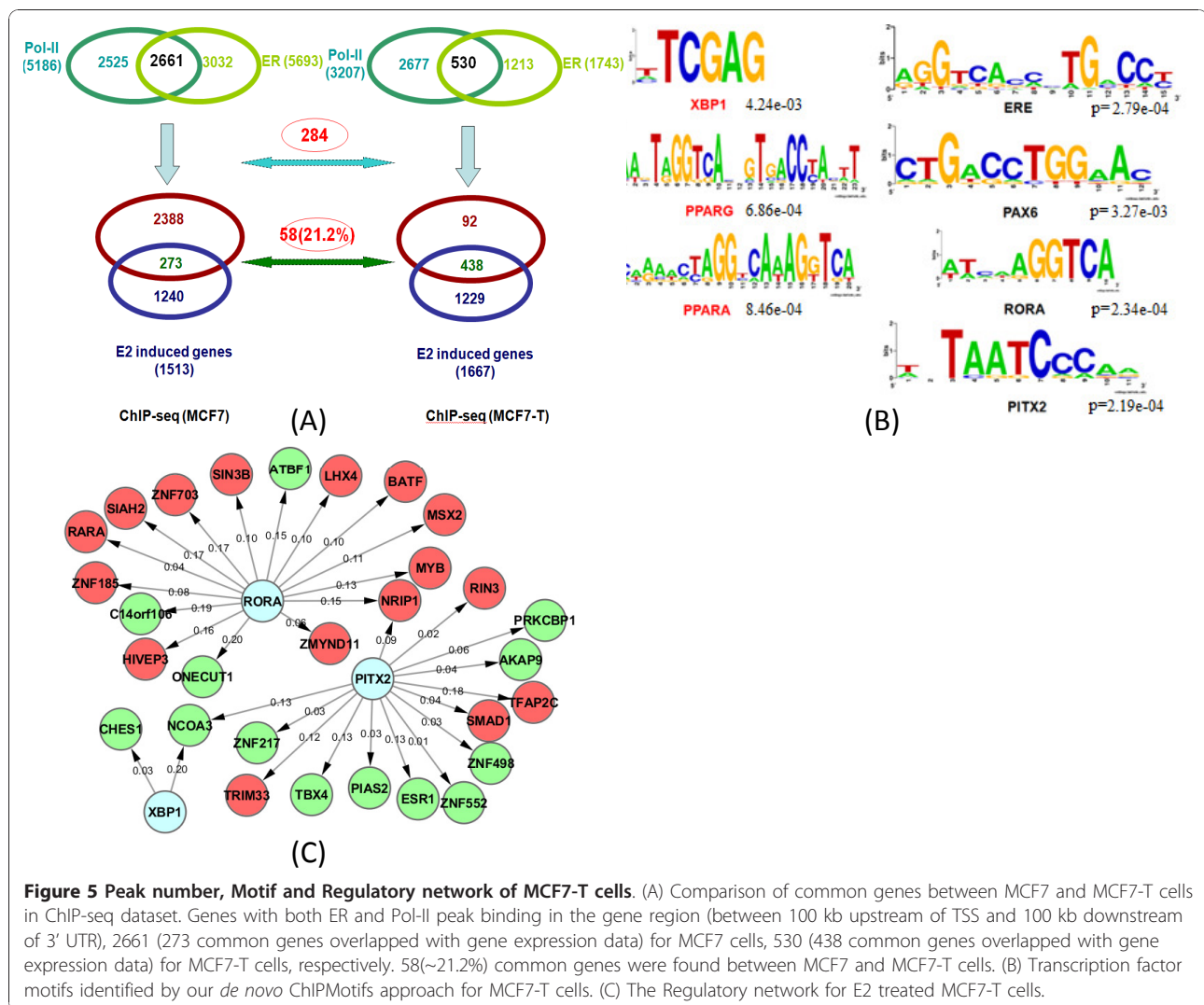
**Table 1 A summary of binding peaks of Pol-II and ER $\alpha$  in control and E2-treated MCF7-T cells identified by CHIP-seq**

TFs	Cell conditions	Reads	Unique Mapped Reads
Pol-II	Control	22,168,614	7,434,175
	E2-treated MCF7-T	19,573,995	5,715,418
ER $\alpha$	Control	14,075,733	5,161,146
	E2-treated MCF7-T	16,401,648	4,445,403

analysis (Figure 5A). Similar to the MCF7 cells, a majority (1,229 of 1,667) of genes with differential expression lack ER $\alpha$  and Pol-II binding peaks.

Since we used CHIP-seq technology for MCF7-T cells, we compared the number of ER $\alpha$  regulated genes between MCF7-T and MCF7 cells using CHIP-seq technology. To illustrate the difference of these two cell lines, we first compared the overlapped number of genes with ER $\alpha$  binding peak and Pol-II binding peak, respectively. We found that 1075 genes were overlapped with ER $\alpha$  binding peak for both MCF7 and MCF7-T cell lines, and this number is 1508 for Pol-II binding peak. We also found that only 58 (~21.2%) common genes (Figure 5A) between these two breast cancer cells. This low overlapped number demonstrated that ER $\alpha$  targeted different sets of genes in these two cancer cells.

We next applied the same computational analytical approach to examine the ER $\alpha$  regulated network in MCF7-T cells. We started with training the top 2000



ER $\alpha$  binding peaks with high scores (enrichments) by the ChIPMotifs, a total of 7 TFs were identified, including ERE, PAX6, PITX2, RORA, XBP1, PPARG and PPARA (Figure 5B). Of them, ERE, PAX6, PITX2 and RORA are the same as the ones we identified in ChIP-seq data of MCF7 cells.

The ER $\alpha$  associated transcriptional regulatory network in MCF7-T cells was then constructed and visualized as shown in Figure 5C, including 32 nodes. A comparison of the networks between MCF7-T and MCF7 cells (using ChIP-seq dataset) showed that there are 2 common Hub TFs (RORA and PITX2) and 8 (about 25%) common targeted TFs. However, in MCF7 cells we found that there are 6 common Hub TFs and 16 (about 40%) common targeted TFs among different ChIP technologies. Taken together, our results strongly suggested that E2 induces a different ER $\alpha$  associated regulatory mechanism in MCF7-T cells compared to MCF7 cells,

in other words, a rewired ER $\alpha$  regulation network in tamoxifen resistance cells.

### Discussion

In this study, we applied computational approaches to analyze and integrate three ChIP-based datasets and one time-series gene expression data to investigate the dynamic regulatory information for ER $\alpha$  in estrogen-dependent breast cancer MCF7 cells. Our studies not only compared the results of three ChIP-based protocols, but also have inferred the regulatory network and pathway for E2 induced MCF7 cells. Moreover, we used the same approach to compare the difference between MCF7 and MCF7-T cells.

Our comprehensive analysis of the estrogen-mediated regulatory network in MCF7 indicated that the three ChIP-based technologies have similar peak distribution patterns for the same antibody. However, for different



antibody (ER $\alpha$  and Pol-II), the binding preference is totally different. Pol-II tends to binding to the promoter region, while ER $\alpha$  has no specific preference.

Our analysis (Figure 1A) also showed that there were more common regulated genes (in percentage) between ChIP-seq data and ChIP-PET data than ChIP-chip data. This may be due to a similar sequencing-based technology used for ChIP-seq and ChIP-PET, but an array-based technology used for ChIP-chip data. The regulatory network analysis indicates some common regulatory Hub TFs were formed in response to estrogen signaling and may lead to the same regulatory paths in MCF7. We also found that there are ~40% common TFs regulated by ER $\alpha$  among three different ChIP-based technologies.

We also compared the regulatory network in MCF7 constructed by our approach with another method—ReMoDiscovery [44]. We used 0.9 (for Motif threshold), and 0.5 (for expression correlation threshold) as the input threshold. Total 26 (of 42) nodes were overlapped between our method (Additional file 4, Figure S3A) and ReMoDiscovery (Additional file 9, Table S5).

In Figure 4, all the genes were with both ER $\alpha$  and Pol-II peak binding, and differentially expression between 12 hours of E2 treat and control. In Figure 4A, the genes were clustered based on the expression values. Each thin line meant the gene expression data (relative value compared with the first time point) at different time point. The genes were classified into 3 groups (the second time point), 4 groups (the third time point) and 5 groups (the fourth time point), respectively. The thick lines corresponded to the gene expression trend of each group. In Figure 4B, the important Hub TFs (Hub TFs with statistically significant number of motifs in the ER $\alpha$  binding peaks of groups of genes) were identified for each group at every time point (4 Hub TFs for time point 2, group 1; 1 for time point 2, group 2 and 3; 3 for time point 4, group 2). The 5 groups of genes were also showed significantly different gene functions. We listed the functions and *p*-values. Figure 4C (a) showed both up and down regulated genes, (b) corresponded to down regulated genes, and (c) corresponded to up regulated genes. A TF-gene relationship table is prepared as the input of DREM (Additional file 8, Table S4). In this table, 1 represented motif of Hub TF existed in the peak region of a gene.

Since a totally different technology is used for ChIP-chip data, in order to reduce the inconsistency caused by the technology itself, we further examined only ChIP-seq and ChIP-PET data to reproduce the network (Figure 6). We found a total of 23 TFs in the network in which there is about 50% increase in terms of the number of nodes. After comparing the combined network (ChIP-seq and ChIP-PET) with ChIP-seq network

(Additional file 4, Figure S3A) or ChIP-PET network (Additional file 4, Figure S3B), about 60% of the TFs were overlapped. However, this number is only 40% for the combined network of three technologies together.

The hierarchical regulatory network analysis with DREM and Cytoscape revealed that RORA could be a potential ER $\alpha$  partner, which is consistent to other reports [45] showing that RORA interacts with ER $\alpha$  and enhances ER transcriptional activity in breast cancer. This finding also indicates that RORA was required for a subset of E2-mediated up-regulated genes associated with functions of Lyase activity. Our results may provide the guidance for further investigation on the role of this co-regulation pathway played in breast cancer.

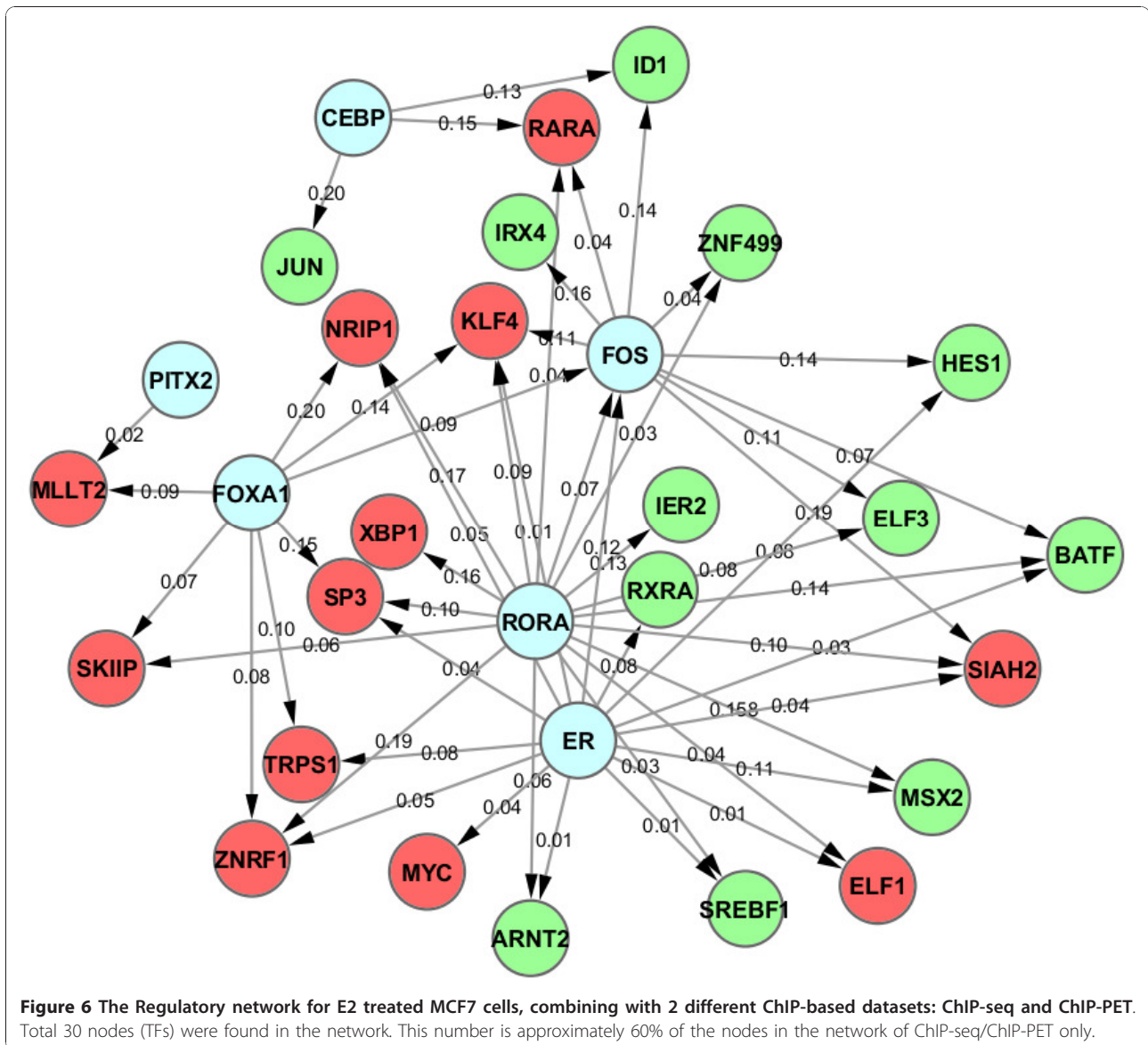
In MCF7-T cells, we found very small proportion (21.2%) of differential expressed genes overlapped with MCF7 cells (ChIP-seq dataset) with both ER $\alpha$  and Pol-II binding peaks. In addition, the number of common TFs of the network between MCF7-T and MCF7 cells is only 25%. Moreover, we also found two TFs with opposite expression trend between MCF7 and MCF7-T cells, in which MSX2 is up-regulated in MCF7-T cells and down-regulated in MCF7 cells, whereas ESR1 (ER $\alpha$ ) is down-regulated in MCF7-T cells. We also identified 3 new Hub TFs in MCF7-T cells which were not found in MCF7 cells. This indicates that ER $\alpha$  may be no longer the most important transcription factor, other transcriptional regulators and signalling pathways may play important role in tamoxifen resistant MCF7 cells. Our comparative analysis also suggests that the ER $\alpha$  associated regulatory network in MCF7-T cells is rewired upon E2 induced.

We have performed additional experiments using both RT-qPCR and ChIP-qPCR to validate randomly select eight ER regulated binding loci in MCF7-T cell, and the result demonstrated that the predicted ER regulated binding loci with differential gene expression after E2 treatment were validated as shown in Additional file 10, Figure S5.

We further compared the E2 induced gene functions to specify the difference between MCF7 and MCF7-T cell line. We selected the genes that were differentially expressed and with ER $\alpha$  and PolIII binding site. These genes were performed by GO function by the DAVID program [46] and the top 10 (the close functions were removed) categories were selected based on the *p*-value. The result was shown in Additional file 11, Figure S6 and Additional file 12, Table S6. 4 of the 10 functions were the same between those two groups of genes.

## Conclusions

In summary, we analyzed hierarchical regulatory networks for estrogen-dependent regulation in MCF7 and MCF7-T cells. We systematically compared different



ChIP-based technologies as well as different breast cancer cells. Our results revealed extended hierarchal regulatory networks with new target hubs in breast cancer cells. Our computational analytical approach may also provide a framework for dissecting transcriptional regulatory networks in response to breast cancer and other human diseases.

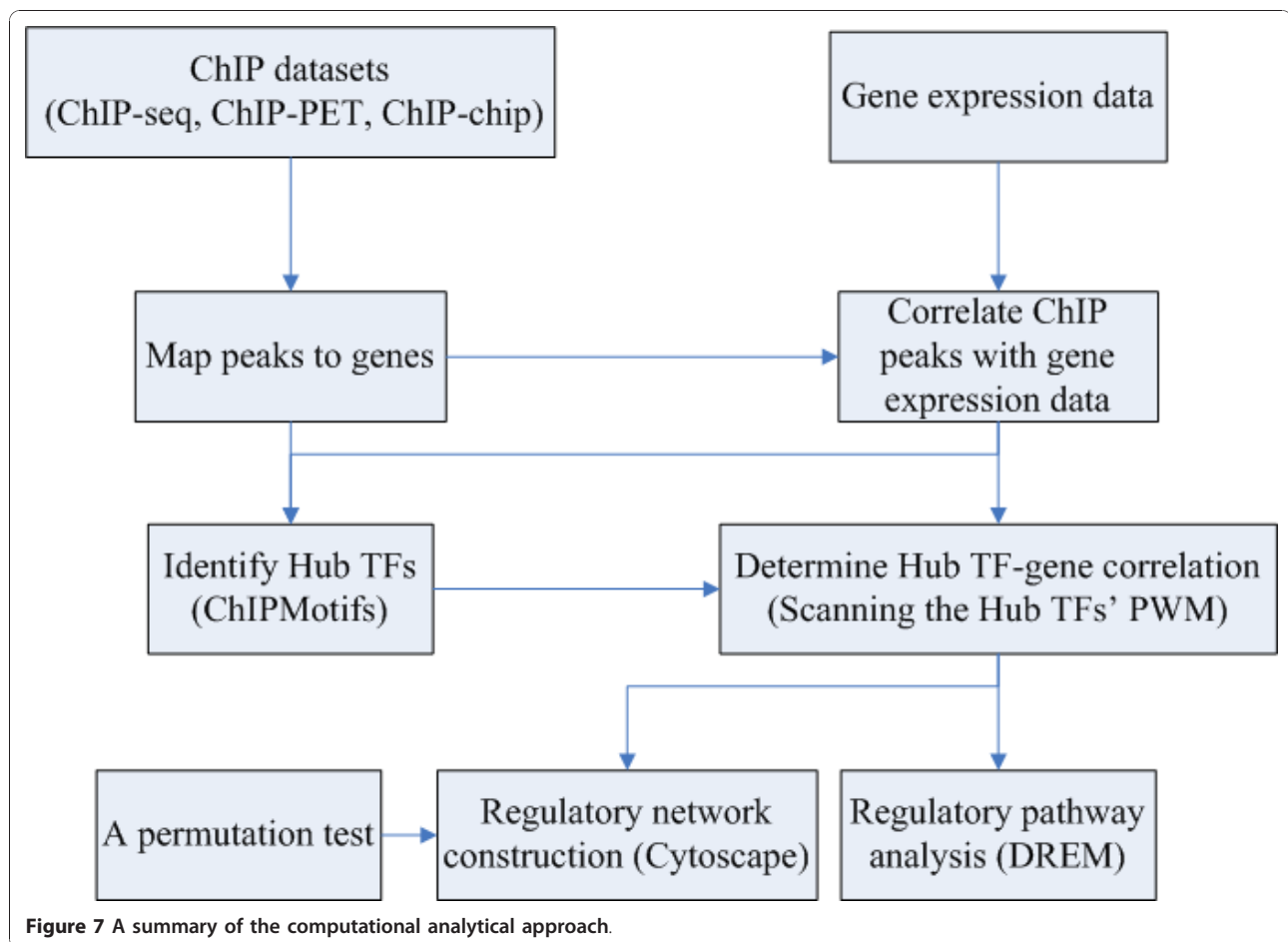
### Methods

Our computational analytical approach (Figure 7) started with ChIP based datasets and gene expression data. The identified binding peaks of a given TF and Pol-II are then located to known genes, and genes having both the given TF and Pol-II binding peaks are further correlated with gene expression data based on

RefSeq Gene ID. The given TF binding peaks are further used for finding the most significant motifs by the ChIP-Motifs, in which they are used as Hub TFs. The Hub TF-gene connection is determined by scanning the Hub TFs' PWMs in all binding peaks and a permutation test (described in detail in **Methods** section "ER $\alpha$  regulatory network analysis") is used to test the reliability of each connection of the network. The resulted regulatory network is visualized by Cytoscape and the pathway is analyzed by DREM. In this study, a given TF is ER $\alpha$ .

### ChIP and microarray data of MCF7 cells

For MCF7 cells, the ChIP-chip data is from Carroll's paper [17], ChIP-seq data is from Welboren's paper [18], and ChIP-PET data is from Melissa's paper [19].



**Figure 7** A summary of the computational analytical approach.

The time-series E2-induced gene expression data in MCF7 cells for all three ChIP-based datasets was obtained from Carroll et al. 2006 [17]. The data was with E2 treatment for 4 time points: 0 hr, 3 hr, 6 hr and 12 hr. We considered those genes were up-regulated if the gene expression values of time point 12 hr are bigger than time point 0, and vice versa.

#### ChIP and microarray data of MCF7-T cells

ChIP-seq of E2-induced MCF7-T cells were maintained in a hormone-free medium (phenol red-free MEM with 2 mmol/L L-glutamine, 0.1 mmol/L nonessential amino acids, 50 units/mL penicillin, 50 Ag/mL streptomycin, 6 ng/mL insulin, and 10% charcoal-stripped FBS) supplemented with  $10^{-7}$  mol/L 4-hydroxytamoxifen. Prior to all experiments, MCF7-T cells were cultured in hormone-free medium for 1 week to deplete any residual OHT, and then MCF7-T cells were cultured in hormone-free basal medium (phenol red free MEM with 2 mM L-glutamine, 0.1 mM nonessential amino acids, 50 units/ml penicillin, 50 µg/ml streptomycin, and 3% charcoal-dextran stripped FBS)

for three days. MCF7-T cells were treated with E2 (108 mol/L) for 3 hours.

$5 \times 10^7$  cells were crosslinked with 1% formaldehyde for 10 min, at which point 0.125 M glycine was used to stop the crosslinking. In brief, after crosslinking, cells were treated by lysis buffers and sonicated to fragment the chromatin to a size range of 200 bp-1 kb. Chromatin fragments were then immunoprecipitated with 10 µg of antibody/magnetic beads. The antibodies against Pol-II and ER $\alpha$  were purchased from Santa Cruz Biotechnology (Santa Cruz, sc-899  $\times$  and sc-8005 X). After immunoprecipitation, washing, and elution, ChIP DNA was purified by phenol:chloroform:isoamyl alcohol and solubilized in 70 µl of water. Then Illumina library was constructed and sequenced with Illumina/Solexa Genome Analyzer (Canada's Michael Smith Genome Sciences Centre, Vancouver, CA). Briefly, the ChIP DNA sample was run in 12% PAGE and the 100-300 bp DNA fraction was excised and eluted from the gel slice overnight at 4°C in 300 µl of elution buffer (5:1, LoTE buffer (3 mM Tris-HCl, pH 7.5, 0.2 mM EDTA)-7.5 M

ammonium acetate) and was purified using a QIA-quick purification kit (Qiagen, Cat#28104). The library was constructed using Illumina genomic DNA prep kit by following its protocol (Illumina, cat# FC-102-1002), clusters were generated on the Illumina cluster station (Illumina, cat# FC-103-1002), and the sequence was run on Illumina 1G Analyzer following the manufacturer's instructions (Illumina, cat# FC-104-1003). The unique mapped ChIP-seq datasets (bed format, total four files: control using ER antibody, E2 treated using ER antibody, control using Pol-II antibody and E2 treated using Pol-II antibody) for MCF7-T cells are available from GEO database (accession number: GSE26083) or from our webpage <http://motif.bmi.ohio-state.edu/ERNetwork/>.

For E2-induced gene expression data in MCF7-T cells, Affymetrix U133 Plus2.0 array platform consisting of ~55,000 transcripts was used for measuring E2-induced gene expression in MCF7-T cells. Three replicates from different biological samples were performed and normalized using MASS5 in the statistic package R <http://www.R-project.org> (Additional file 13, Table S7). A p-value less than 0.05 using a Welch's t-test and 2 fold changes for each gene were used as cutoff thresholds to determine a set of differential expressed genes. Finally, we mapped the genes with Human TFs and get 1,667 genes remains.

#### RT-qPCR and ChIP-qPCR

After 3 hr of E2 treatment, control and treated MCF7-T cells were subjected to total RNA extraction by Trizol reagent (Invitrogen). Total RNA (2 µg) was reverse transcribed to cDNA with oligo-dT (SuperScript III; Invitrogen). Quantitative RT-PCR was performed by using SYBR Green dye chemistry (Applied Biosystems) on a 7500 Real-time PCR system apparatus (Applied Biosystems). Gene Expression was measured by the  $\Delta\Delta C_t$  method using *GAPDH* as the internal control. Statistical analyses were carried out by using a two-tailed t test. Specific primers for amplification are available on request.

To confirm candidate ER $\alpha$  target genes determined by ChIP-seq, PCR primers targeting a region within 200 bp of the predicted ERE were used to measure the enrichment of this sequence in anti-ER $\alpha$ -immunoprecipitated samples by quantitative PCR with SYBR Green-based detection method (Applied Biosystems). Quantitative values measured by a standard curve (50 to 0.08 ng, 5-fold dilution, R<sup>2</sup> > 0.99) of input DNA amplified with the same primer set. Results are presented as the mean of triplicates with standard derivation.

#### Calling peaks of ChIP-based data

A standard procedure for extracting image files, mapping the reads onto human genome, and filtering the mapped reads to unique reads was followed with the Solexa 1.6 pipeline [47]. Only uniquely mapped reads with a length of 36 bp were then used further for determining the binding regions by our BELT program [48]. For all datasets, we narrowed the peaks with 500 bp in length. All the common peak number in Figures 1 and 5A were overlapped according to the gene ID of each file. We used the fisher exact test [49] to calculate the p-value of each overlapped number, and assumed that the human genome number is 30,000 (Additional file 14, Table S8).

#### ER $\alpha$ regulatory network analysis

The following steps were used to calculate the position weight matrix (PWM).

Suppose there are m sequences, and the sequence length is n. For each column j(1 ≤ j ≤ n), the occurrence frequency of every nucleotide (A or T or C or G) was calculated by counting the number in m sequences (equation 1).

$$f_{i,j} = \frac{\sum_{i=1}^m (S_{i,j} = c)}{m} \quad (1)$$

Where c is one of base type (A/T/C/G), S<sub>i,j</sub> is the nucleotide at row i, column j.

Hence, the PWM is represented as a 4\*n matrix. For row i(1 ≤ i ≤ 4) and column j(1 ≤ j ≤ n), the PWM can be calculated:

$$w(i, j) = \log_2 \frac{f_{i,j}}{p(i)} \quad (2)$$

Where p(i) is the background frequency of nucleotide i. For Human genome, p(i = A) = p(T) = 30% and p(i = C) = p(G) = 20%, approximately.

For the case that the number sample is small, the equation 1 needed to be replaced as

$$f_{i,j} = \frac{\sum_{i=1}^m (S_{i,j} = c) + \frac{\sqrt{m}}{4}}{m + \sqrt{m}} \quad (3)$$

Then, the significant PWMs were mapped to TransFac [36] database and JASPAR [37] database to find the most similar TFs. And ER $\alpha$  and its partner TFs were found.

To construct the regulatory network, the PWM was used to scan the peak region of each gene. To make the result more reliable, we use stringent threshold (1 for core score, 0.95 for PWM score) to determine the underlying TF binding site. The PWM score and core score is calculated as follows:

Given a sequence with the same length (column number  $n$ ) of PWM, we can first calculate the sequence score:

$$S_{seq} = \sum_{j=1}^n w(S_i, j) \quad (4)$$

Where  $w(S_i, j)$  is the score at row  $i$  which the nucleotide is the same as the given sequence at column  $j$ .

Then, the minimum and maximum score of PWM can be calculated as:

$$S_{min} = \sum_{j=1}^n \min\{w(i, j)\} \quad (5)$$

$$S_{max} = \sum_{j=1}^n \max\{w(i, j)\} \quad (6)$$

The PWM score for the given sequence is shown in equation 7

$$S_{PWM} = \frac{S_{seq} - S_{min}}{S_{max} - S_{min}} \quad (7)$$

Our program ChIPMotifs will give the length ( $k$ ) and start ( $k_1$ ) and end ( $k_2$ ) position of Core region in a PWM. The given sequence will be scanned ( $n-k+1$ ) times (move forward one position each time) to find a maximum Core score. For the  $d$ th time, the sequence score can be represented as:

$$S_{seq,d} = \sum_{j=d}^{d+k-1} w(S_i, j) \quad (8)$$

For equation 5 and 6, only the boundary of  $j$  was changed from  $1-n$  to  $k_1-k_2$ . The Core score for the  $d$ th time is

$$S_{Core,d} = \frac{S_{seq,d} - S_{min}}{S_{max} - S_{min}} \quad (9)$$

And the Core score for the given sequence is

$$S_{Core} = \max_d(S_{Core,d}) \quad (10)$$

If there is Motif for ER $\alpha$  and Hub TFs for certain gene, a connection between the Hub TF (include ER $\alpha$ ) and the gene was made. We further facilitated these relationships by visualizing an estrogen regulatory network with Cytoscape [42] software platform (see Results section).

To test the significance of the network proposed in this study, a statistical strategy (permutation) was used to determine the probability of each edge of the network under random circumstances. Since the TF binding site region is composed of specific sequences, and only by scanning the sequence region using PWM, we get the network edges. Hence, we shuffled the sequence of each peak region for 1000 times, and to see how many times a specific TF binding site is hit by the scanning process. The ratio (times hit by scanning divide by 1000) of each edge was calculated. The number with low value was considered high statistical significance (we used 0.2 as a cutoff to include as more connection as possible while keeping the relationship reliable).

#### ER $\alpha$ regulatory pathway analysis

The up/down-regulated genes were grouped into several classes (sub up/down-regulated classes). For each group, the common gene functions were identified from GO database with p-value represented for the significance. Then, we used the PWMs of the 6 TFs (ER $\alpha$ , RORA, FOS, FOXA1 PITX2 and CEBP) to identify all possible binding sites in all genes with peaks. Thus we established the relationship between Hub TFs and grouped genes. A DREM diverging score less than 0.1 was used as a significant score threshold for TF-grouped gene relationship.

#### Additional material

**Additional file 1: Table S1.** A distribution of locations of Pol-II and ER $\alpha$  binding sites relative to Human HG18 RefSeq Genes three ChIP-based datasets in MCF7 cell line.

**Additional file 2: Figure S1. A plot of the distribution of identified ER $\alpha$  and Pol-II binding loci relative to a known gene's TSS.** A) Definition of different regions of a gene. B) The histogram of the distribution of peak location. A big portion of Pol-II bind in promoter regions (2 kb around TSS); A small portion of ER $\alpha$  binding loci are located in promoter region. These observations confirm that the majority of ER $\alpha$  binding loci are outside of proximal promoter regions in which it is consistent with the results from other studies.

**Additional file 3: Figure S2. Three major binding models for ER $\alpha$  regulated gene expression.** A) direct binding to ERE (estrogen response element); B) indirect binding, through which it binds to other TF partners which bind to DNA; C) co-occurrent binding, where both ER $\alpha$  and other TF partners bind to their own specific DNA motifs.

**Additional file 4: Figure S3. The Regulatory network for E2 treated MCF7 cell line of three different ChIP-based datasets.** A) The Regulatory network for E2 treated MCF7 cell line from ChIP-seq dataset. B) The Regulatory network for E2 treated MCF7 cell line from ChIP-PET dataset. C) The Regulatory network for E2 treated MCF7 cell line from ChIP-chip dataset.

**Additional file 5: Table S2.** The list of common TFs among all three regulatory networks.

**Additional file 6: Figure S4. Regulatory pathway analysis from three ChIP-based dataset. Figure S4 (I). Regulatory pathway analysis from ChIP-seq dataset.**

A) The time series gene expression data of E2 induced genes superimposed with the regulatory pathway map produced by DREM using the gene expression profile as well as ER $\alpha$  binding sites and PolII binding sites. The bright green nodes indicate split points where the sets of expression of genes diverge. B) Paths out of splits are annotated with TFs determined by DREM to be associated with the genes assigned to the path at a score <0.1. The GO annotations for the genes in 5 of the paths are shown at the right with their p-values. C) The genes traversing the 3 splits are shown with (a) corresponding to the split at 0-hr, (b) corresponding to the split at 3-hr and (c) corresponding to the split at 6-hr. Figure S4 (II). Regulatory pathway analysis from ChIP-PET dataset. A) The time series gene expression data of E2 induced genes superimposed with the regulatory pathway map produced by DREM using the gene expression profile as well as ER $\alpha$  binding sites and PolII binding sites. The bright green nodes indicate split points where the sets of expression of genes diverge. B) Paths out of splits are annotated with TFs determined by DREM to be associated with the genes assigned to the path at a score <0.1. The GO annotations for the genes in 5 of the paths are shown at the right with their p-values. C) The genes traversing the 3 splits are shown with (a) corresponding to the split at 0-hr, (b) corresponding to the split at 3-hr and (c) corresponding to the split at 6-hr. Figure S4 (III). Regulatory pathway analysis from ChIP-chip dataset. A) The time series gene expression data of E2 induced genes superimposed with the regulatory pathway map produced by DREM using the gene expression profile as well as ER $\alpha$  binding sites and PolII binding sites. The bright green nodes indicate split points where the sets of expression of genes diverge. B) Paths out of splits are annotated with TFs determined by DREM to be associated with the genes assigned to the path at a score <0.1. The GO annotations for the genes in 5 of the paths are shown at the right with their p-values. C) The genes traversing the 3 splits are shown with (a) corresponding to the split at 0-hr, (b) corresponding to the split at 3-hr and (c) corresponding to the split at 6-hr.

**Additional file 7: Table S3.** A. The five paths for ER $\alpha$  regulated genes identified by DREM analysis in ChIP-seq data. B. The five paths for ER $\alpha$  regulated genes identified by DREM analysis in ChIP-PET data. C. The five paths for ER $\alpha$  regulated genes identified by DREM analysis in ChIP-chip data.

**Additional file 8: Table S4.** A. The transcription factors identified in ChIP-seq with predicted binding motifs. B. The transcription factors identified in ChIP-PET with predicted binding motifs. C. The transcription factors identified in MCF7 with predicted binding motifs.

**Additional file 9: Table S5.** Transcription factor network identified by ReMoDiscovery.

**Additional file 10: Figure S5. Validation of the regulatory network of MCF7-T cell line.** 8 TFs were selected from MCF7-T network. All the TFs were found with ER $\alpha$  binding peaks. (A) mRNA levels derived from RT-qPCR of 8 ER $\alpha$  regulated target genes were shown under E2 (10 nM, 3hr) stimulation in MCF7-T cells. GAPDH was as internal control. Mean  $\pm$  SD (n = 3). (B) Validations of predicted ER $\alpha$ -binding regions for 8 binding loci by ChIP-qPCR. Control and E2-treated MCF7-T cells were subjected to ChIP-qPCR with ER $\alpha$  antibody. Mean  $\pm$  SD (n = 3).

**Additional file 11: Figure S6. A comparison of GO functions between MCF7 and MCF7-T cells using genes from ChIP-seq technology.** (A) GO functions of MCF7 cells. (B) GO functions of MCF7-T cells. (C) the GO function of two cells in the same figure.

**Additional file 12: Table S6.** A list of genes used for GO function analysis for MCF7 and MCF7-T ChIP-seq data.

**Additional file 13: Table S7.** Gene expression data of MCF7-T cells.

**Additional file 14: Table S8.** Fisher exact test of the significance of overlapped genes in Figures 1A and 5A.

## Acknowledgements

This study was supported by grants from NIH U54CA113001 and R01CA069065 and by funds from the Ohio State University Comprehensive Cancer Center and The Ohio State University Biomedical Informatics.

## Author details

<sup>1</sup>Department of Biomedical Informatics, The Ohio State University, Columbus, USA. <sup>2</sup>Human Cancer Genetics Program, The Ohio State University, Columbus, USA.

## Authors' contributions

FG, HKH, PYH, JW, YM, JP, THMH and VXJ conceived and designed the study. VXJ and FG performed the computational analysis and HKH, PYH and JW performed MCF7-T experiments and RT-, ChIP-qPCR validations. VXJ, FG, JP and THMH drafted the manuscript. All authors provided comments on the paper and read and approved the final manuscript.

Received: 30 June 2010 Accepted: 17 December 2010

Published: 17 December 2010

## References

1. Kim TH, Barrera LO, Zheng M, Qu CX, Singer MA, Richmond TA, Wu YN, Green RD, Ren B: **A high-resolution map of active promoters in the human genome.** *Nature* 2005, **436**:876-880.
2. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA: **Genome-wide location and function of DNA binding proteins.** *Science* 2000, **290**:2306-2309.
3. Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ: **Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis.** *Genes Dev* 2002, **16**:235-244.
4. Wu J, Smith LT, Plass C, Huang TH: **ChIP-on-chip comes of age for genome-wide functional analysis.** *Cancer Res* 2006, **66**:6899-6902.
5. Roh TY, Cuddapah S, Cui K, Zhao K: **The genomic landscape of histone modifications in human T cells.** *Proc Natl Acad Sci USA* 2006, **103**:15782-15787.
6. Loh Y-H, Wu Q, Chew J-J, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong K-Y, Sung KW, Lee CWH, Zhao X-D, Chiu K-P, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei C-L, Ruan Y, Lim B, Ng H-H: **The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells.** *Nature Genetics* 2006, **38**:431-440.
7. Robertson G, Hirst M, Bainbridge M, Bilenyk M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S: **Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing.** *Nat Methods* 2007, **4**:651-657.
8. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome.** *Cell* 2007, **129**:823-837.
9. Johnson DS, Mortazavi A, Myers RM, Wold B: **Genome-wide mapping of in vivo protein-DNA interactions.** *Science* 2007, **316**:1497-1502.
10. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448**:553-560.
11. Osborne CK, Schiff R: **Estrogen-receptor biology: continuing progress and therapeutic implications.** *J Clin Oncol* 2005, **23**:1616-1622.
12. Klinge CM: **Estrogen receptor interaction with co-activators and co-repressors.** *Steroids* 2000, **65**:227-251.
13. Jakacka M, Ito M, Weiss J, Chien PY, Gehm BD, Jameson JL: **Estrogen Receptor Binding to DNA Is Not Required for Its Activity through the Nonclassical AP1 Pathway.** *J Biol Chem* 2001, **276**:13615-13621.
14. Speir E, Yu ZX, Takeda K, Ferrans VJ, Cannon RO: **Competition for p300 regulates transcription by estrogen receptors and nuclear factor-kappaB in human coronary smooth muscle cells.** *Circ Res* 2000, **87**:1006-1011.

15. Abdelrahim A, Samudio I, Smith R III, Burghardt R, Safe S: **Small inhibitory RNA duplexes for Sp1 mRNA block basal and estrogen-induced gene expression and cell cycle progression in MCF-7 breast cancer cells.** *J Biol Chem* 2002, **277**:28815-28822.
16. Gaub MP, Bellard M, Scheuer I, Chambon P, Sassone-Corsi P: **Activation of the ovalbumin gene by the estrogen receptor involves the fos-jun complex.** *Cell* 1990, **63**:1267-1276.
17. 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.** *Nature Genetics* 2006, **38**:1289-1297.
18. Welboren WJ, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FCGJ, Span PN, Stunnenberg HG: **ChIP-Seq of ER $\alpha$  and RNA polymerase II defines genes differentially responding to ligands.** *EMBO J* 2009, **28**:1418-1428.
19. Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EGY, Huang PYH, Welboren WJ, Han YY, Ooi HS, Ariyaratne PN, Vega VB, Luo YQ, Tan PY, Choy PY, Wansa KDSA, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li HX, Desai KV, Thomsen JS, Lee YK: **An oestrogen-receptor- $\alpha$ -bound human chromatin interactome.** *Nature* 2009, **462**:58-64.
20. Fan M, Yan PS, Hartman-Frey C, Chen L, Paik H, Oyer SL, Salisbury JD, Cheng AS, Li L, Abbosh PH, Huang TH, Nephew KP: **Diverse gene expression and DNA methylation profiles correlate with differential adaptation of breast cancer cells to the antiestrogens tamoxifen and fulvestrant.** *Cancer Res* 2006, **66**:11954-11966.
21. Cheng AS, Jin VX, Fan M, Smith LT, Liyanarachchi S, Yan PS, Leu YW, Chan MW, Plass C, Nephew KP, Davuluri RV, Huang TH: **Combinatorial analysis of transcription factor partners reveals recruitment of c-MYC to estrogen receptor- $\alpha$  responsive promoters.** *Mol Cell* 2006, **21**:393-404.
22. Jin VX, Leu YW, Liyanarachchi S, Sun H, Fan M, Nephew KP, Huang TH, Davuluri RV: **Identifying estrogen receptor  $\alpha$  target genes using integrated computational genomics and chromatin immunoprecipitation microarray.** *Nucleic Acids Res* 2004, **32**:6627-6635.
23. Leu YW, Yan PS, Fan M, Jin VX, Liu JC, Curran EM, Welshons WW, Wei SH, Plass C, Nephew KP, Huang TH: **Loss of estrogen receptor signaling triggers epigenetic silencing of downstream targets in breast cancer.** *Cancer Res* 2004, **64**:8184-8192.
24. 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.
25. Hurtado A, Holmes KA, Geistlinger TR, Hutcheson IR, Nicholson RI, Brown M, Jiang J, Howat WJ, Ali S, Carroll JS: **Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen.** *Nature* 2008, **456**:663-666.
26. Laganière J, Deblois G, Giguère V: **Functional genomics identifies a mechanism for estrogen activation of the retinoic acid receptor  $\alpha$ 1 gene in breast cancer cells.** *Mol Endocrinol* 2005, **19**:1584-1592.
27. Massarweh S, Osborne CK, Creighton CJ, Qin L, Tsimelzon A, Huang S, Weiss H, Rimawi M, Schiff R: **Tamoxifen resistance in breast tumors is driven by growth factor receptor signaling with repression of classic estrogen receptor genomic function.** *Cancer Res* 2008, **68**:826-833.
28. Wu M, Soler DR, Abba MC, Nunez MI, Baer R, Hatzis C, Llombart-Cussac A, Llombart-Bosch A, Aldaz CM: **CtIP silencing as a novel mechanism of tamoxifen resistance in breast cancer.** *Mol Cancer Res* 2007, **5**:1285-1295.
29. Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH-M: **The functional consequences of alternative promoter use in mammalian genomes.** *Trends in Genetics* 2008, **24**:168-177.
30. Carroll JS, Brown M: **Estrogen Receptor Target Gene: An Evolving Concept.** *Molecular Endocrinology* 2006, **20**:1707-1714.
31. Komashko VM, Acevedo LG, Squazzo SL, Iyengar S, Rabinovich A, O'Geen H, Farnham PJ: **Using ChIP-chip technology to reveal common principles of transcriptional repression in normal and cancer cells.** *Genome Res* 2008, **18**:521-532.
32. Kinnis M, Chen BS, Diehl AG, Isaacs GD, Zhang T, Siepel AC, Clark AG, Kraus WL: **Genomic analysis of transcription factor binding, histone acetylation, and gene expression reveal mechanistically distinct classes of estrogen-regulated promoters.** *Molecular and Cellular Biology* 2007, **27**:5090-5104.
33. Lin CY, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F, Yeo A, George J, Kuznetsov VA, Lee YK, Charn TH, Palanisamy N, Miller LD, Cheung E, Katzenellenbogen BS, Ruan Y, Bourque G, Wei CL, Liu ET: **Whole-genome cartography of estrogen receptor  $\alpha$  binding sites.** *PLoS Genet* 2007, **3**:e87.
34. Jin VX, O'Geen H, Iyengar S, Green R, Farnham PJ: **Identification of an OCT4 and SRY regulatory module using integrated computational and experimental genomics approaches.** *Genome Res* 2007, **17**:807-817.
35. Jin VX, Apostolos J, Nagisetty NSVR, Farnham PJ: **W-ChIPMotifs: a web application tool for *de novo* motif discovery from ChIP-based high-throughput data.** *Bioinformatics* 2009, **25**:3191-3193.
36. Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Prüß M, Reuter I, Schacherer F: **TRANSFAC $^{\circ}$ : an integrated system for gene expression regulation.** *Nucleic Acids Res* 2000, **28**:316-319.
37. Wasserman WW, Sandelin A: **Applied bioinformatics for the identification of regulatory elements.** *Nat Rev Genet* 2004, **5**:276-287.
38. Levy N, Zhao X, Tang H, Jaffe RB, Speed TP, Leitman DC: **Multiple transcription factor elements collaborate with estrogen receptor  $\alpha$  to activate an inducible estrogen response element in the NKG2E gene.** *Endocrinology* 2007, **148**:3449-3458.
39. Jakacka M, Ito M, Weiss J, Chien PY, Gehm BD, Jameson JL: **Estrogen Receptor Binding to DNA Is Not Required for Its Activity through the Nonclassical AP1 Pathway.** *J Biol Chem* 2001, **276**:13615-13621.
40. Carroll JS, Brown M: **Estrogen receptor target gene: an evolving concept.** *Mol Endocrinol* 2006, **20**:1707-1714.
41. Laganière J, Deblois G, Lefebvre C, Bataille AR, Robert F, Giguère V: **Location analysis of estrogen receptor  $\alpha$  target promoters reveals that FOXA1 defines a domain of the estrogen response.** *Proc Natl Acad Sci USA* 2005, **102**:11651-11656.
42. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Research* 2003, **13**:2498-504.
43. Ernst J, Vainas O, Harbison CT, Simon I, Bar-Joseph Z: **Reconstructing Dynamic Regulatory Maps.** *Molecular Systems Biology* 2007, **3**:74.
44. Lemmens K, Dhollander T, Bie TD, Monsieurs P, Engelen K, Smets B, Winderickx J, Moor BD, Marchal K: **Inferring transcriptional module networks from ChIP-chip-, motif- and microarray data.** *Genome Biology* 2006, **7**:R37.
45. Ross-Innes CS, Stark R, Holmes KA, Schmidt D, Spyrou C, Russell R, Massie CE, Vowler SL, Eldridge M, Carroll JS: **Cooperative interaction between retinoic acid receptor- $\alpha$  and estrogen receptor in breast cancer.** *Genes Dev* 2010, **24**:171-182.
46. Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources.** *Nature Protoc* 2009, **4**:44-57.
47. Robertson G, Hirst M, Bainbridge M, Bilenyi M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S: **Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing.** *Nat Methods* 2007, **4**:651-657.
48. Frieze S, Lan X, Jin VX, Farnham PJ: **Genomic targets of the KRAB and SCAN domain-containing zinc finger protein 263.** *J Biol Chem* 2010, **285**:1393-1403.
49. Fisher RA: **On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P.** *Journal of the Royal Statistical Society* 1922, **85**:87-94.

doi:10.1186/1752-0509-4-170

Cite this article as: Gu et al: Inference of hierarchical regulatory network of estrogen-dependent breast cancer through ChIP-based data. *BMC Systems Biology* 2010 **4**:170.