

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

***In Silico* Prediction for Regulation of Transcription Factors on Their Shared Target Genes Indicates Relevant Clinical Implications in a Breast Cancer Population**

Li-Yu D Liu^{2,*}, Li-Yun Chang^{1,*}, Wen-Hung Kuo³, Hsiao-Lin Hwa¹, Ming-Kwang Shyu¹, King-Jen Chang³ and Fon-Jou Hsieh¹

¹Department of Obstetrics and Gynecology, College of Medicine, National Taiwan University, Taipei, Taiwan. ²Department of Agronomy, Biometry Division, National Taiwan University, Taipei, Taiwan. ³Department of Surgery, College of Medicine, National Taiwan University, Taipei, Taiwan. *These authors contributed equally to this work.
Corresponding author email: fjhsieh@ntu.edu.tw

Abstract: Aberrant transcriptional activities have been documented in breast cancers. Studies often find some transcription factors to be inappropriately regulated and enriched in certain pathological states. The promoter regions of most target genes have binding sites for their transcription factors. An ample of evidence supports their combinatorial effect on their shared target gene expressions. Here, we used a new statistic method, bivariate CID, to predict combinatorial interaction activity between ER α and a transcription factor (E2F1 or GATA3 or ERR α) in regulating target gene expression via four regulatory mechanisms. We identified gene sets in three signal transduction pathways perturbed in breast tumors: cell cycle, VEGF, and PDGFRB. Bivariate network analysis revealed several target genes previously implicated in tumor angiogenesis are among the predicted shared targets, including *VEGFA*, *PDGFRB*. In summary, our analysis suggests the importance for the multivariate space of an inferred ER α transcriptional regulatory network in breast cancer diagnostic and therapeutic development.

Keywords: bivariate CID, network, transcription factor, shared target gene expression, angiogenesis

Cancer Informatics 2012:11 113–137

doi: [10.4137/CIN.S8470](https://doi.org/10.4137/CIN.S8470)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

To better characterize estrogen receptor positive (ER(+)) breast cancer has its importance when traditional pathological information can not fully evaluate the clinical outcomes of patients, especially for those who eventually have worse prognosis. Oh et al¹ identified two subtypes of ER(+) breast cancers, group IE and group IIE, to be clinically distinct in their estrogen induced functional estrogen receptor α (ER α) activities in addition to their intrinsic phenotypes at molecular level (e.g. Normal like, luminal A and B). Our research findings² also suggested that aberrant ER α phenotype in breast cancer may be demonstrated by ER α directed transcriptional regulatory network, at least in part. We thus hypothesize that the ongoing promoter use pathways within the functional ER α transcriptional regulatory network may provide insights into a better understanding the transcriptional regulatory activities of both ER α alone and ER α with other transcription factor(s) (TFs) on ER α target genes in breast cancer.

Both regular and alternative promoter uses of ER α to its target genes have been studied intensively *in vitro*.³⁻⁷ Davuluri et al⁷ pointed out the proposed mechanisms of promoter use that may be different between tumor and normal cells. The above research evidence, in part, supported our hypothesis that some of the clinical pathophenotypic differences between groups IE and IIE breast cancers may be derived from the alternative promoter use pathways of ER α and/or other TFs on their shared target genes. In theory, the clue of subtype specific promoter use on subsets of genes in the ER α transcriptional regulatory network would assist in functional subtyping of ER(+) breast cancers.

In this study, we aimed at building up a network with clusters of gene sets based on their relevant nature of promoter use, and filling the multivariate space of the previously established network.² This effort is proposed to be achieved by applying univariate and multivariate versions of coefficient of intrinsic dependence (CID) as well as the Galton Pearson's correlation coefficient (GPCC) to statistically identify the promoter use mechanisms for shared target genes of functional transcription factors in a given population. CID in combination with GPCC for measuring univariate association has been applied to identify both direct and remote associations of

TF-target (Liu et al. 2009).² Here, we would propose that CID itself should be able to evaluate the association between two transcription factors (e.g. ER α and other TF) and one of their shared target genes when ER α may interact with a TF, which is one of ER α primary target genes. The simultaneous or sequential interaction between ER α and other TF results in regulating the expression of their shared target gene that is predicted to occur. Such regulatory association is expected to be recognized by one version of multivariate CID called bivariate CID (see Methods). Three pairs of transcription factors (ER α , E2F1), (ER α , GATA3) and (ER α , ERR α) were used to test this hypothesis. Finally, we briefly predicted how interactions of ER α with three transcription partners to achieve a visible pathological phenotype, which is a new clinical parameter, based on results of bivariate CID. The transcriptional regulatory mechanisms of those interactions may interpret a transcriptional switch made for altering pre-programmed physiological state to a pathophysiological state.

Methods

Clinical breast cancer expression arrays

The one hundred and ninety nine clinical arrays (abbreviated as "199A") were from a patient cohort (started from 1998 to 2007) collected at National Taiwan University Hospital (NTUH). The tumor samples, defined by average greater than 50% tumor cells per high-power field examined in a section adjacent to the tissue used,⁸ were included in this study. These clinical arrays were generated using the Human 1A (version 2) oligonucleotide microarray (Agilent technologies, USA) according to the methods provided by the manufacture. The pre-processing, normalization of microarray data were performed following methods described.⁸ The total array content is half a genome not a whole genome. This will influence the size of gene pools to be extracted *in silico* for building a network described in this study. All patients had given informed consent according to guidelines approved by the Institutional Review Board (IRB) at NTUH. The dataset can be retrieved at NCBI-GEO (accession number GSE24124). Three subsets of the clinical arrays were used in the analyses. The first data set (abbreviated as "152A") included group IE breast cancer (61A) and ER(-) breast cancer (91A). Both ER α status



and progesterone receptor A (PR) status are positive for group IE breast cancer.¹ The second data set (abbreviated as “120A”) included group IIE breast cancer (29A) and ER(-) breast cancer (91A). Group IIE has positive ER α status and negative PR status¹ (Table S10). As a control of this study, we selected the third data set including eighteen non-tumor samples (18A) that were surgically taken from breast tissue adjacent to some of 90 IDC breast tumors with ER(+) described below.

Immunohistochemical staining of ER α and progesterone receptor A (PR)

All the paraffin sections of breast cancer specimens (3–5 μ m in thickness) on slides were processed in Ventana’s automated staining system (BenchMark[®] LT) (Ventana Medical Systems, Inc) for the immunohistochemical stain (IHC). All the procedure for IHC stain has been documented.² To detect the IHC of progesterone receptor A, mouse anti-human PGR monoclonal antibody, unconjugated clone 5D10 (Catalog # H00005241-M07, Abnova Corporation, Taiwan) with dilution ratio at 1:50 was used as the specific antibody to bind PR protein on the tumor section of 181 samples. And, the definition of positive IHC stain for ER α protein (ER(+)) or PR protein (PR(+)) in this study is for tumor slide that has shown greater than or equal to 10% tumor cells with moderate to high amount of immunoreactive nuclear ER α protein or PR protein. To prevent extracting less meaningful data in this study, we used both IHC and real-time quantitative polymerase chain reaction (QPCR) data of both ER α and PR (data not shown) to be the supporting information for this study.

Statistical analysis on univariate association between a TF and a target gene using part of gene expression dataset from 181A

The statistical methods applied for identifying the gene lists of estrogen regulated transcriptional activities were the univariate association measured by the coefficient of intrinsic dependence (CID)^{2,9} and that by Galton-Pearson’s correlation coefficient (GPCC).² The univariate CID result for a given TF was designated as CID-TF. Instead of all subgroups having an equal size ($N \approx 10$),² we divided

the cohort by hierarchical clustering (described in method below) to mimic biological systems in which similar expression pattern in a subgroup may reflect the similar biological event shared by the members within a subgroup. As a result, the subgroup was designated as j . For instance, each subCID value of the assigned subgroup (j) is determined in part by the sample size of the subgroup j , a constant value and the two times square of difference between cumulative distribution function (CDF) of gene Y in the subgroup j and the average CDF of gene Y in a given population. Total CID value demonstrates the degree of dependence between TF and its target gene.² We have optimized number of subgroups chosen for CID measurements. That is, we set rounding number for one tenth of total array numbers as the final number of subgroups (15 subgroups for 152A) after expression profile of a variable was hierarchically pre-clustered in a given population. The GPCC measures the linear expression relationship between the TF and its target gene. The univariate association results were derived by combining CID-TF and GPCC (designated as CID-TFUGPCC).²

ER α _E2F1, ER α _GATA3 and ER α _ERR α pathways predicted *in silico* by bivariate CID

There are four steps included in this analysis. The analysis of ER α _E2F1 pathway is demonstrated here as an example (Fig. 2A). First, a scatter plot for mRNA levels of two TFs was produced. The left panel in Figure 2A provided an example demonstrating the scatter plot of *E2F1* mRNA versus *ESR1* mRNA in 152A. A scatter plot represents a two-dimensional scattering pattern of two co-expressed TFs of interest at mRNA level. Second, a hierarchical clustering¹⁰ based on the distances between spots and between groups of spots was performed. In this study, the distance between two spots was described by Euclidean Distance (i.e. $D = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, where (x_i, y_i) are mRNA levels of (TF1, TF2) for the i th spot). The distance between two groups of spots was defined by complete linkage, which is the Euclidean Distance between the farthest pair of spots in two groups. The shorter distance between two spots/groups indicates more similarity in their expression levels. A tree-shape diagram, or dendrogram, is typically used to illustrate the result of hierarchical clustering (lower panel in Fig. 2A). Third,



the dendrogram was used to further cluster scattered spots into fifteen subgroups (containing $N = 9, 8, 5, 4, 11, 9, 33, 4, 3, 11, 10, 6, 17, 7$ and 15 spots in fifteen subgroups, respectively). The lower part in Figure 2A shows the array ID numbers labeled by rainbow colors (a color per subgroup). We denoted the labels of subgroups for spots by variable Z ($Z = 1-15$). CID only used Z to represent the assigned subgroup of two co-expressed TFs. The same equation for calculating both univariate CID and bivariate CID was performed except we replaced j by Z . The right panel in Figure 2A describes the characteristics of *ESR1* and *E2F1* mRNA levels in each clustered subgroup, where the height of the bar indicates a net value derived from the median of mRNA expression levels in each subgroup subtracted by global median of mRNA expression levels of *ESR1* and *E2F1*, respectively. Thus, fifteen subgroups are represented by the paired color bars, respectively. In majority of ER (+) subgroups, low *E2F1* mRNA levels were co-expressed with relatively high levels of *ESR1* mRNA. We observed both *ESR1* and *E2F1* mRNAs to be co-expressed at high level in three subgroups of ER(+) IDCs (see a heatmap in Fig. 9A and/or a bar plot in Fig. 2A). In ER(-) group, however, all subgroups were having inverted expression pattern on *E2F1* mRNA while comparing to *ESR1* mRNA except one subgroup showing low *E2F1* mRNA level. Final step is the CID evaluation on two TFs and their shared target genes that was taken place by the procedure previously described² using the subgroup label, Z , as the predictor variable. In brief, CID-*ESR1*n*E2F1* was resulted from the above analysis. Both Figures 2B and C show (1) a scatter plot of *ESR1* mRNA level vs. a TF mRNA level; (2) the mRNA expression patterns of both *ESR1* and a TF for 15 subgroups; (3) a dendrogram based on co-expressed *ESR1* and a TF at mRNA level in 152A. A CID-*ESR1*nTF was also evaluated based on the same procedure described for *E2F1* except that TF stands for *GATA3* or *ESRRA* in this study.

Computing P -values for results from univariate CID, bivariate CID and GPCC analyses

We compute both univariate CID and bivariate CID using the same equation² except the label of subgroups after hierarchical clustering were designated as j , z , respectively (see the method described above).

To access the significance of univariate or bivariate CID values generated via *in silico* analyses and to facilitate comparison among data derived from applying different types of methods, the univariate or bivariate CID value, S , was compared with the values generated by random sampling mimicking the mRNA expression data distribution of gene Y that is independent on the data partitioning of the assigned variable (a TF or TFs). In 152A study, the independent data distribution was derived from randomly drawing 152 simulated values for an artificial gene and put appropriate number of data for the gene in each subgroup that is the same as the sample size in a pre-clustered subgroup. We re-computed the subCID value of each subgroup and added them together to be a new CID value (K). This was repeated 1,000 times and yielded 1,000 of CID values (K_i , $i = 1-1,000$). The P value was determined by an equation (i.e. $P = (1 + N_{(k_i \geq S)})/1,001$, where $N_{(k_i \geq S)}$ is the number of K_i values greater than S). The P values for GPCC measurements were computed using asymptotic normal theory.¹¹ We set the cut off point for P value of both methods to be significant when $P \leq 0.05$.

Finding biological implications of statistically identified gene pools

We used web-based software that helps in identifying the Gene Ontology and Pathway information for a gene list of interest. The web site is <http://vortex.cs.wayne.edu/projects.htm>. The Gene Ontology tool is called “Onto-Express” and the gene annotation database is ontotools database. For pathway analysis, a method to gather all possible pathway information involving the gene list of interest is called “Pathway-Express”.¹² In addition, we also used commercial software called Gene Spring GX 7.3.1 (Agilent Technologies, USA) in this study.

Results and Discussion

A flow chart of statistical analysis and the confounders controlled during multivariate analysis

Figure 1 briefly describes a pictorial sketch of the analysis steps used to carry out the statistical methods described above and indicates their predicted feature results to be demonstrated in figures “(see results and discussion below). The implementation of both univariate

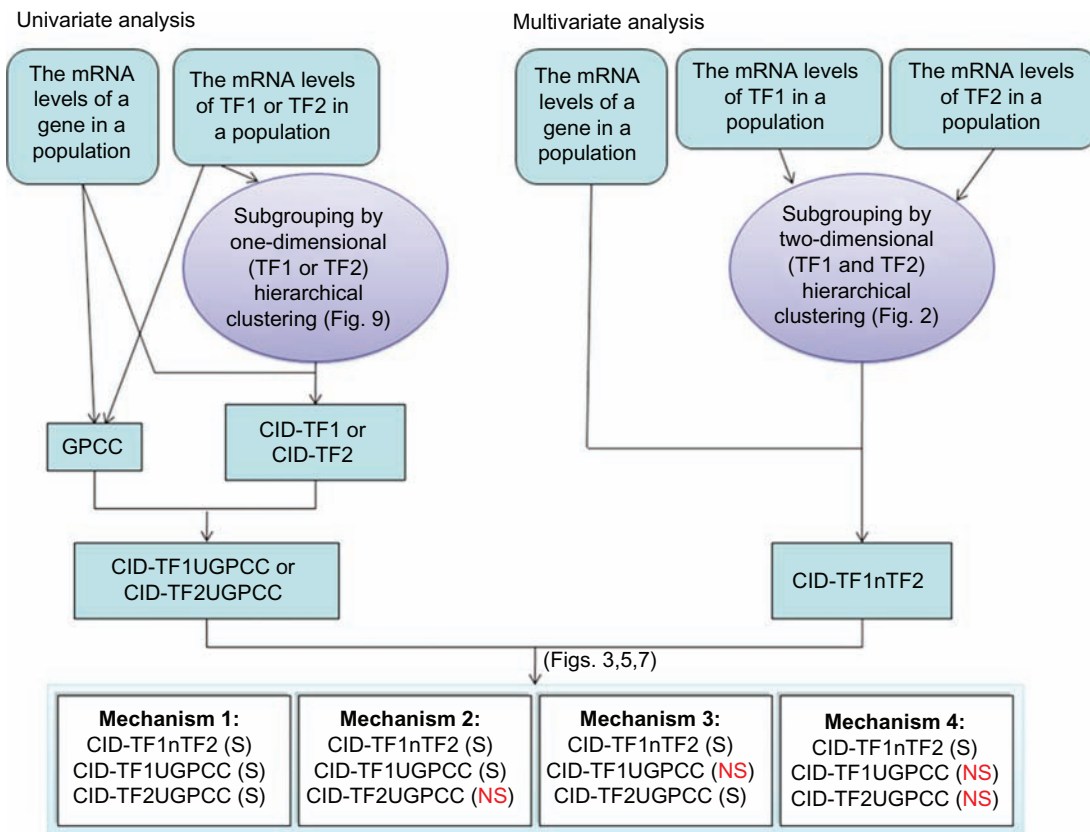


Figure 1. Flow chart for genome-wide univariate and multivariate analyses. The procedure for identification of a statistically relevant promoter use pathway and its regulatory mechanisms at molecular level includes (1) univariate analysis (left panel) of both linear association (e.g. GPCC) and non-linear association (e.g. univariate CID) of TF-target in a population; (2) multivariate analysis (right panel) of non-linear association (e.g. bivariate CID) of TF1nTF2-target in a population; (3) combined results from univariate and bivariate analyses to be divided into four categories. They are designated as mechanism 1, 2, 3 and 4, respectively.

and bivariate CID in R codes can be downloaded from <http://homepage.ntu.edu.tw/~lyliu/multCID/>.

Statistically, the definition for confounders is for all variables that are unknown, not available or not of interest at this moment but affecting the systems. The same definition is also applied to biological systems.

In this study, we designed a few key steps to control confounders during multivariate analyses. They are as follows.

(1) We first selected two patient cohorts for multivariate analysis based on the statuses of immunohistochemical stains for ER and PR on their breast tumor sections (see Table S10); (2) We selected the active regulator—estrogen receptor alpha as the TF1 because its aberrant transcriptional activities have been observed in a large subset of breast cancer; (3) We chose hierarchical clustering to facilitate in locating the potentially pre-programmed interaction pathway between two functionally co-expressed transcription factors on their promoter use in a given cohort.

Importantly, we picked up a target gene of TF1 to be TF2 in this study.^{13–16} As a result, such expression pattern clustering strategy enhanced the ability of CID to identify a network of bivariate associations for both TF1 and TF2 to their shared target genes that in part validate a relevant event *in vivo* (see Fig. 8).

The predicted molecular features of a new multivariate analysis reveals the subtypic differences in their pre-programmed transcriptional regulation on three gene sets in a breast cancer population

The promoter content of statistically identified genes that are shared targets of two functional transcription factors

Firstly, we reasoned the structure/function relationship between transcription factors (e.g. TF1, TF2) and their binding sites on promoter regions of their shared target

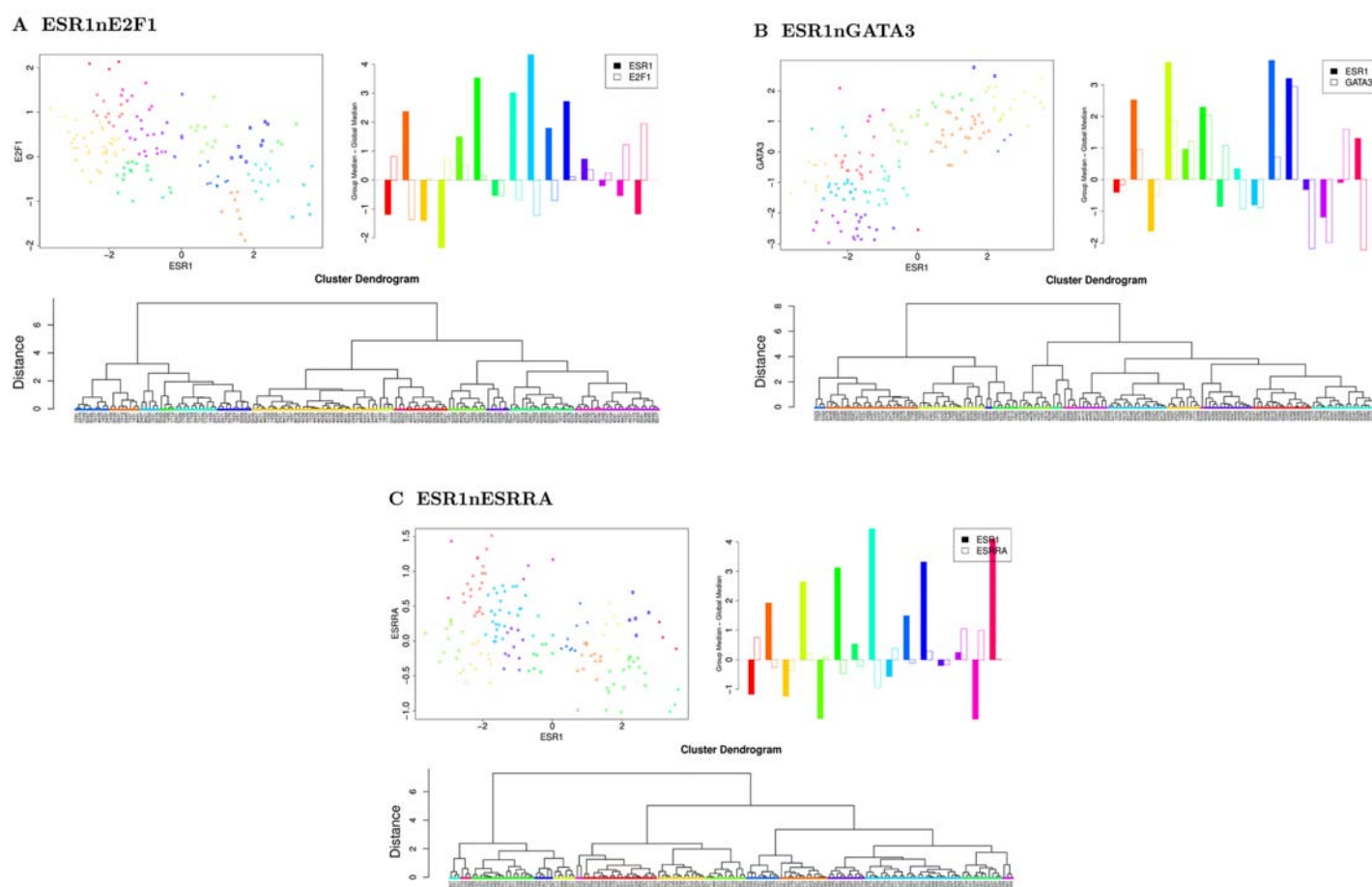


Figure 2. The illustration for hierarchical clustering on two variables of interest.

Panel A. Upper left panel, a scatter plot for co-expression of *ESR1* and *E2F1* mRNAs in 152A. Fifteen clusters are labeled by a series of rainbow colors, respectively. The number of patients or N for 15 subgroups are presented in different color spots to be 9, 8, 5, 4, 11, 9, 33, 4, 3, 11, 10, 6, 17, 7 and 15, respectively. Upper right panel, paired bar plots showing the mRNA expression patterns of both *ESR1* and *E2F1* in each subgroup, respectively. Lower panel, a dendrogram indicating the fifteen subgroups labeled with their corresponding array ID number that was determined by hierarchical clustering.

Panel B. Upper left panel, a scatter plot for co-expression of *ESR1* and *GATA3* mRNAs in 152A. Fifteen clusters are labeled by a series of rainbow colors, respectively. The numbers of patients or N for 15 subgroups are presented in different color spots to be 3, 25, 16, 2, 7, 9, 3, 4, 8, 16, 9, 1, 13, 16 and 18, respectively. Upper right panel, paired bar plots showing the mRNA expression patterns of both *ESR1* and *GATA3* in each subgroup, respectively. Lower panel, a dendrogram indicating the 15 subgroups labeled with their corresponding array ID number that was determined by hierarchical clustering.

Panel C. Upper left panel, a scatter plot for co-expression of *ESR1* and *ESRRA* mRNAs in 152A. Fifteen clusters are labeled by a series of rainbow colors, respectively. The numbers of patients or N for 15 subgroups are presented in different color spots to be 3, 3, 17, 5, 6, 2, 20, 13, 11, 9, 13, 10, 26, 11 and 3, respectively. Upper right panel, paired bar plots showing the mRNA expression patterns of both *ESR1* and *ESRRA* in each subgroup, respectively. Lower panel, a dendrogram indicating the 15 subgroups labeled with their corresponding array ID number that was determined by hierarchical clustering.

genes (Fig. 3) to be predicted via bivariate CID. Thus, the gene expression relationship of co-expressed TF1 (e.g. *ESR1*) and TF2 (e.g. *E2F1* or *GATA3* or *ESRRA*) to their common target gene expression was measured by bivariate CID (designated as CID-TF1nTF2). Secondly, we screened gene pools statistically identified by CID-TF1nTF2 via their promoter contents published by others.

We ran the bivariate CID on a dataset of 152A to measure gene expression relationship between (*ESR1*, *E2F1*) and their common target gene. As a result, 8,616 probes were identified to be significant ($P \leq 0.05$)

for CID-ESR1nE2F1. Only 194 probes among these 8,616 probes are known to have E2F binding site(s)¹⁷ (Table S3, Fig. 4). In addition, we identified 10,294 probes via analysis of CID-ESR1nGATA3 in 152A. Only 102 probes (102/10,294) are known to have both ER α and GATA binding site(s) at their promoter regions⁴ (Fig. 4). Finally, we identified 4,169 probes via analysis of CID-ESR1nESRRA in 152A. Only 56 probes (56/4,169) are known to have both ER α and ERR α binding site(s) at their promoter regions.¹⁸ Those gene sets were used to further predict for the combinatorial actions of their transcription factors on

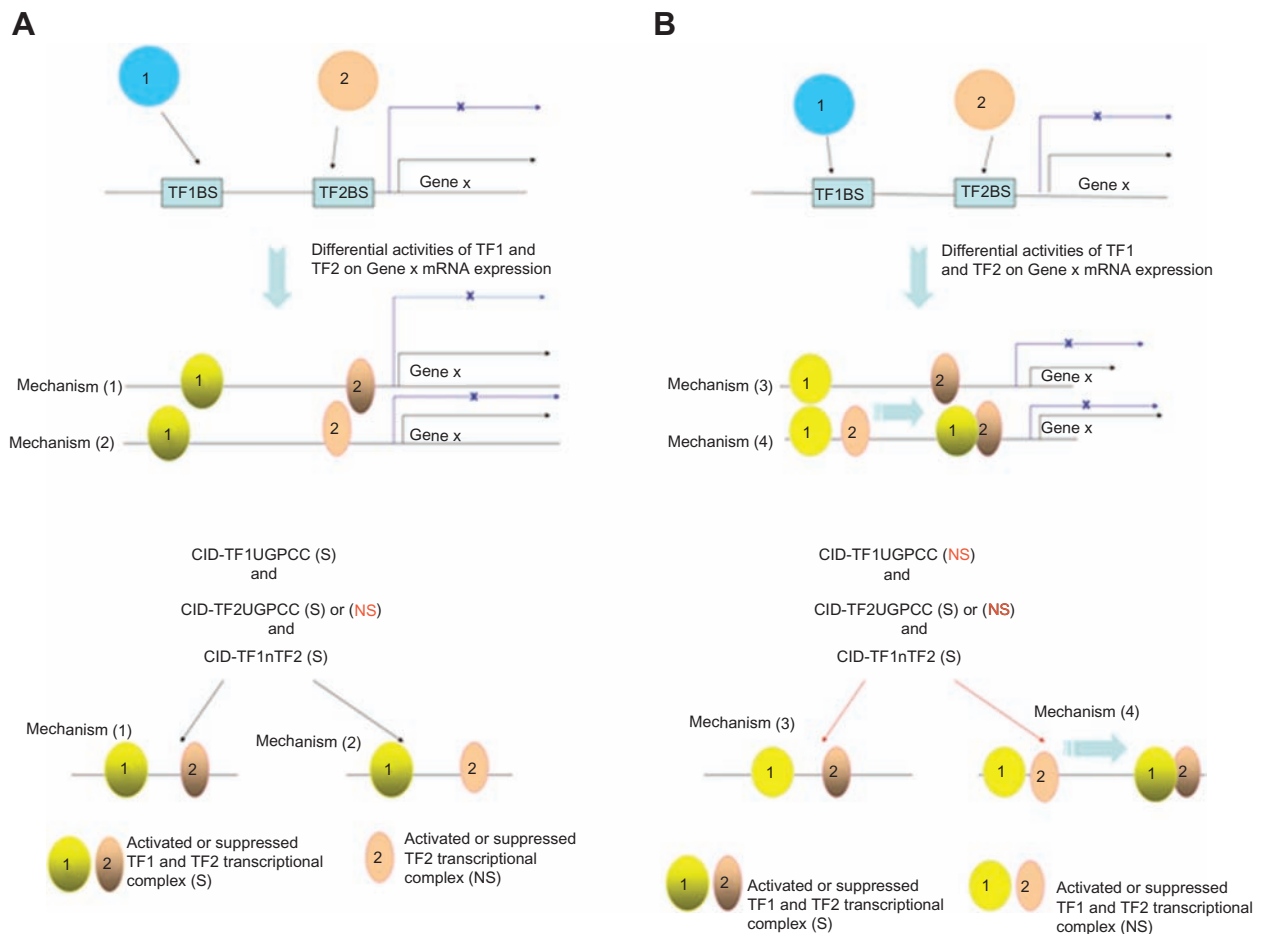


Figure 3. Molecular model mechanisms of TF1 and TF2 on their shared target gene. The hypothesized transcriptional mechanisms for TF1 (also labeled as “1”), TF2 (also labeled as “2”) and target gene (labeled as “Gene x”) are shown in panels A and B. They were predicted *in silico* by CID-TF1nTF2 along with both CID-TF1UGPCC and CID-TF2UGPCC ($P \leq 0.05$) (see Methods). The suppressive regulation on gene x mRNA expression is indicated by the blue solid arrow with a stop sign (x), which is located above the chromosomal DNA (black line). The increase mode on expression of gene x mRNA is indicated by the black solid arrow without a stop sign. TF1BS, TF2BS stand for the binding site of TF1, TF2, respectively. Current model for transcription factor families in regulating their target genes has been simplified in this figure based on the most updated literatures.^{5–10,33}

Notes: “S” stands for “significant”. “NS” stands for “not significant”.

regulating their gene expressions. We describe below how both univariate and multivariate associations (i.e. CID-TF1UGPCC, CID-TF2UGPCC and CID-TF1nTF2) predict the features of functional promoter use pathway by two TFs of interest in a clinical breast cancer setting.

Statistical prediction on the mechanisms of promoter use for shared target genes by two TFs in group IE breast cancer population

We defined statistically relevant mechanisms of the promoter use by both transcription factors of interest based on the hypothesis of ours. We hypothesized that the mRNA expression level of a gene, which contains both estrogen receptor binding site (ERBS) and E2F binding site (E2FBS) at the promoter region, would be

dependent on either one or both functional transcription factors (e.g. ER α and E2F1) interacting with those corresponding binding sites. As a result, the differential activities of two transcription factors determine the final mRNA expression levels of their shared target genes, in part (Fig. 3). The same scheme was also applied in analyzing the mechanisms for ER α and GATA3, ER α and ERR α involved promoter use programs on their target genes, respectively. Statistically, we further dissected the promoter use features of identified genes by comparing results among the multivariate association measured by bivariate CID (i.e. CID-TF1nTF2) and two univariate associations measured by coupling univariate CID (i.e. CID-TF1 or CID-TF2) and GPCC. The statistically predicted results were partially validated via literature search.

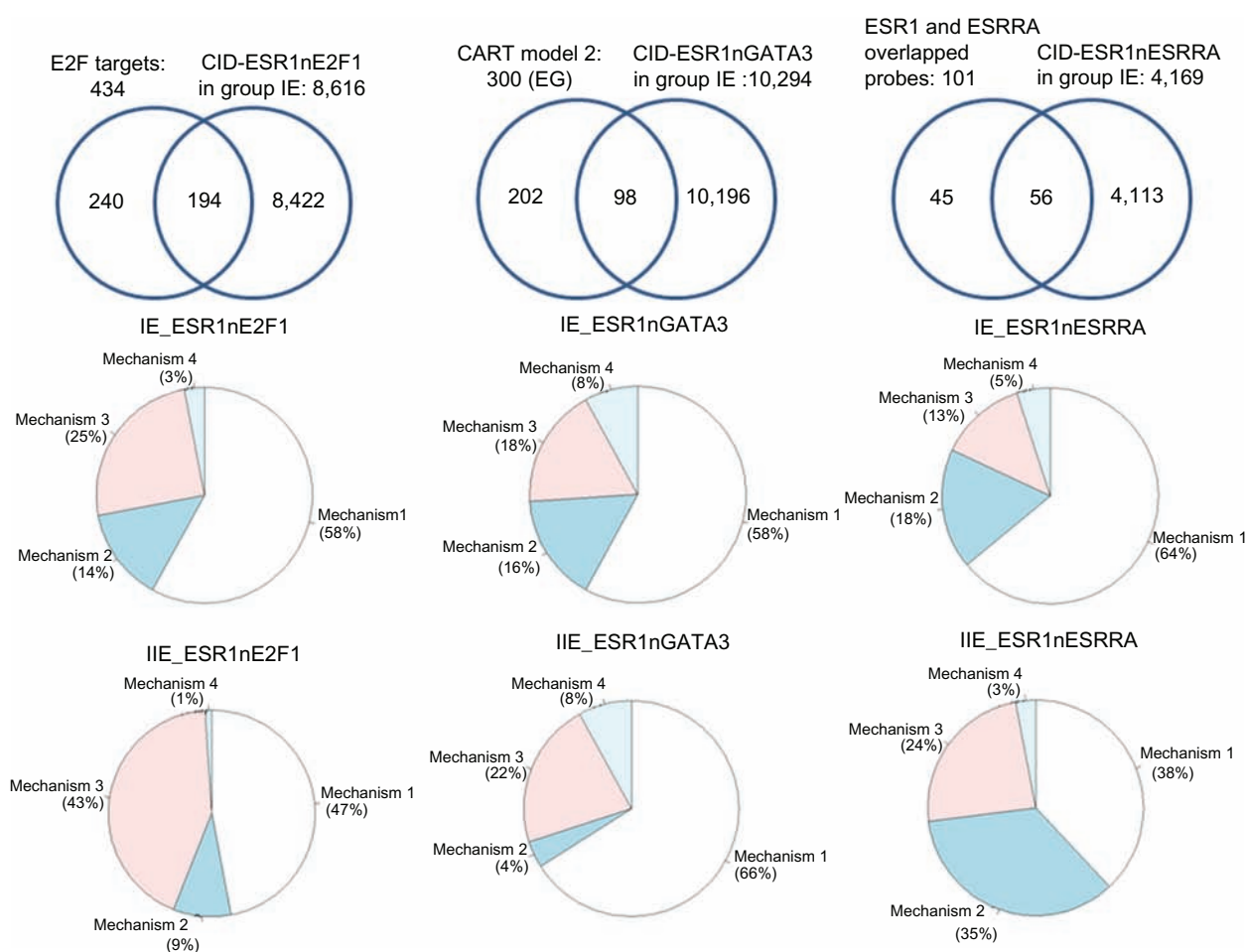


Figure 4. Summary of genes identified by bivariate CID and their pie distribution among four regulatory mechanisms in groups IE and IIE breast cancers. We performed dissecting transcriptional regulatory mechanisms in two TFs directed common target gene expressions by coupling the results derived from univariate associations (CID U GPC) with those derived from bivariate CID in groups IE and IIE breast cancers. Upper panel, first, it represents one hundred ninety four probes (194 probes) identified significantly by CID-ESR1nE2F1 in group IE (8,616 probes). They belong to E2F targets (434 probes) identified by others.¹⁷ Second, it represents one hundred and two probes (102 probes) identified significantly by both CID-ESR1nGATA3 in group IE (10,294 probes) and CART model 2 identified shared target genes of ER α and GATA3 (300 probes).⁴ Third, it represents fifty six probes identified significantly by both CID-ESR1nESRRA in group IE (4,169 probes) and the overlapped target genes of ER α and ESRRA (101 probes).¹⁸ Lower panel shows subtype specific pie distribution of four regulatory mechanisms for 194, 102 and 56 probes, respectively.

Subgrouping of one hundred ninety four probes (194 probes) based on their gene regulatory mechanisms via differential interactions between ER α and E2F1 on their promoters

We summarized in Table S3 that 194 probes were selected by both significance in multivariate association (CID-ESR1nE2F1) and their promoter content¹⁷ (i.e. E2FBS) in group IE breast cancer. They were further analyzed using two univariate measures of association (i.e. CID-ESR1UGPCC and CID-E2F1UGPCC). We claimed the target genes to be regulated via mechanism 1 through mechanism 4 if CID-ESR1UGPCC and CID-E2F1UGPCC yielded results of (S, S), (S, NS), (NS, S) and (NS, NS), respectively, where S denotes significance and NS

denotes insignificance at significant level $\alpha = 0.05$. They are listed in Table S1.

One hundred thirteen probes (113/194) were predicted to be regulated via mechanism 1 (Fig. 3 and Table S1) in group IE breast cancer. They were predicted *in silico* to be regulated by ER α , E2F1 alone (univariate association, $P \leq 0.05$). But, most importantly both transcription factors may cross talk to regulate this set of genes (multivariate association, $P \leq 0.05$). Nine of them (9/113) have also been identified *in vitro* by Bourdeau et al. (2008).¹⁹ They concluded *ASF1B*, *BRIP1*, *CDCA5*, *RAD51API*, *SPBC25*, *TOP2A*, *UBE2T*, *CCNG2* and *UBE2C* to be estrogen responsive target genes. Table S3 shows twenty of them (20/113) having ERE site(s) at their



promoter regions.²⁰ Fourteen of them (14/113) follow indirect tethering mechanism.⁴ The detailed promoter uses by ER α and E2F1 either simultaneously or sequentially via mechanism 1 still need to be validated *in vitro*.

Twenty seven probes (27/194) were predicted to be regulated via mechanism 2 (Fig. 3 and Table S1). Those probes were predicted statistically relevant to be preferentially regulated by ER α *in vivo* but not to be significantly regulated by E2F1. Only CD44 (1/27) was documented to have ERE site at its promoter region²⁰ and RNF130 follows indirect tethering mechanism of classical ER pathway.⁴ The remaining twenty five probes (25/27) may follow ERE independent ER pathways to be preferentially regulated by ER α , while their E2FBS were not statistically relevant in functional interacting with E2F1 that was predicted by univariate CID and/or GPCC ($P > 0.05$) (Fig. 3, Tables S1 and S3). But, when both TFs are co-expressed in tumor section, either sequentially or simultaneously cross talking between ER α and E2F1 may take place (CID-ESR1nE2F1, $P \leq 0.05$) (Table S1). These potential mechanisms are awaited to be proved *in vitro*.

Forty eight probes (48/194) have been identified in group IE breast cancer to be significantly regulated by mechanism 3 (Fig. 3 and Table S1). It indicates the promoter use by both ER α and E2F1 to the promoter regions of those probes via mechanism 3 to be a statistically relevant event in group IE breast cancer (CID-ESR1nE2F1, $P \leq 0.05$). In addition, a gene being expressed via this mechanism is predicted to partially contributed by the functional interaction between E2F1 and E2FBS alone (CID-E2F1UGPCC). However, ER α alone may have a limited contribution for such regulatory mechanism. No statistically significant regulatory event for these genes was predicted to be regulated by ER α (CID-ESR1UGPCC, $P > 0.05$). For example, only ATAD2 was found to be an estrogen responsive target gene.¹⁹ It has both ERBS²⁰ and E2FBS¹⁷ at its promoter region. However, its promoter use by ER α and E2F1 follows mechanism 3 in group IE breast cancer. Interestingly, seven of them (7/48) have known ERE sites²⁰ (Table S3). Three of them (3/48) are known to be regulated via indirect tethering mechanism of classical ER pathway.⁴ The predicted regulatory mechanism of those probes still need to be validated *in vitro*.

Six (6/194) probes were predicted to be regulated by both ER α and E2F1 via mechanism 4 (Fig. 3 and Table S1), which suggests solely an ER α and E2F1 cooperative regulation on those gene expressions. Only SFRS1 (1/6) has both ERBS²⁰ and E2FBS¹⁷ sites. The remaining probes (5/6) have E2FBS¹⁷ sites but no validated ERBS. Thus, it indicates two types of ER α pathways, classical and non-classical ones, together with E2F1 to allow transcriptional regulation on their target gene expression via mechanism 4 in group IE breast cancer population. However, further evidence *in vitro* would be needed to support this prediction.

Sixteen (16/194) probes have more than one mRNA products, which have been indicated by their unique feature numbers. There are mRNA expressions of seven pair variants following the same mechanism (i.e. AGGF1, DZIP3, H2AFX, KHSRP, PLDN, RAD54B and TRAIIP). But for SP1 transcript variants, their expressions follow different mechanisms of promoter use, which are via differential interactions of ER α and E2F1 on promoter region of SP1 (Table S1). Those mechanisms, which indicate the possible switching in promoter use for two TFs in regulating the expression of transcript variants, are awaited to be validated by well defined experiments *in vitro*.

Subgrouping one hundred and two probes (102 probes) based on their gene regulatory mechanisms via differential interactions between ER α and GATA3 on their promoters 102 probes, which are known to have both ERBS and GATA3 binding site (GATABS) in their promoter regions,⁴ were detected to be significant by CID-ESR1nGATA3 in 152A (Fig. 4, Tables S4 and S6). Further dissecting the relevant transcriptional activities involving both TFs, we concluded four mechanisms either simultaneously or sequentially regulating those gene expressions at mRNA level. Those target genes were claimed to be regulated by ER α and GATA3 via mechanism 1 through mechanism 4 based on the same definition described for the combinatorial interaction between ESR1 and E2F1 to the promoter region of their common target gene.

We summarized the predicted regulatory mechanisms among those probes following ESR1_GATA3 pathway as follows. There were fifty nine probes (59/102) to be regulated via mechanism 1 (Table S4). Sixteen of them (16/102) were identified to be regulated by mechanism 2. Nineteen (19/102) were



regulated via mechanism 3. Eight (8/102) were regulated via mechanism 4. Two genes (4/102)—*EVXI* and *SLC9A3R1* were identified to share the same mechanism in their transcript variant expressions (i.e. both transcriptional and post-transcriptional regulation) involving actions of ER α and GATA3 on their promoter regions. Moreover, two transcript variants of a gene (2/102)—*IGFBP5* were regulated by ER α and GATA3 via different promoter use mechanisms. Only 12 probes (12/102) were documented to have ERE sites²⁰ (Table S6). Therefore, it would be highly desirable for one in finding more supporting evidence *in vitro* for those statistically relevant regulatory mechanisms.

Subgrouping fifty six probes (56 probes) based on their gene regulatory mechanisms via differential interactions between ER α and ESRR α on their promoters
 ERR α (or *ESRR α*) does not respond to estrogen stimuli in general. However, its ligands are recently listed to be isoflavones, chlordane and diethylstilbestrol (DES).²¹ The activity of ERR α is strongly stimulated by the co-expression of coactivators without adding a ligand.²² Moreover, ERR α is evolutionarily related to estrogen receptors and can efficiently bind to EREs.²³ We selected a gene list (101 probes) as estrogen induced and shared target genes of both ER α and ERR α from the gene expression data of an estrogen responsive breast cancer cell line (MCF7). Their transcription factor binding sites for ER α and ERR α have been validated *in vitro*.¹⁸ Only 56 probes (56/101) were overlapped with the shared target gene pool of both *ESR1* and *ESRR α* in 152A that were predicted by bivariate CID (Fig. 4).

Thirty six of them (36/56) indicate the regulatory mechanism as mechanism 1 (Table 1). Six of them (6/36) have ERE sites²⁰ (Table S9). One among those 6 probes—*DDIT4* not only has an ERE site²⁰ at promoter region but has been documented to be possibly regulated following indirect tethering mechanism.⁴ Six of them (6/36) could be regulated following tethering mechanism⁴ (Table S9).

Ten probes (10/56) were predicted to be regulated by *ESR1* and *ESRR α* via mechanism 2. Only *LMO4*(17053) (1/10) may be regulated not only by mechanism 2 of *ESR1*_ESRR α promoter use pathway but by tethering mechanism.⁴ Seven probes (7/56) were predicted to be significant for

mechanism 3. However, only *EFNA1* (1/7) has an ERE site²⁰ at its promoter region. In addition, *SYP* (1/7) follows tethering mechanism.⁴ We observed three probes (3/56) to be possibly regulated via mechanism 4 in group IE breast cancer. *PTMA* (1/3) has been found to have an ERE site.¹⁹

In summary, there is a trend for three sets of gene expressions at mRNA level to be potentially regulated by their transcription factors shown in Figure 4. Majority of probes (58% for *ESR1*_E2F1 pathway; 58% for *ESR1*_GATA3 pathway; 64% for *ESR1*_ESRR α pathway) are predicted to follow mechanism 1 (Figs. 4–6). The number of probes being regulated via mechanisms 2, 3 and 4, are less than those via mechanism 1. The mixed mechanism (i.e. Two TFs regulate the expression of a target gene which has more than one mRNA products.) is the rare case that may involve both transcriptional and post-transcriptional regulatory event. In addition, both promoter use pathways and regulatory mechanisms in each pathway in regulating three gene sets are quite different between groups IE and IIE (Tables S2, S5 and S8).

The role of bivariate CID in multivariate space of a transcriptional regulatory network and its predicted subtypic difference in response to pre-programmed promoter use pathways upon estrogen exposure

We hypothesized that the mRNA expression level of a gene, which contains binding sites for both functional TFs at its promoter region, would depend on either one or both TFs interacting with those corresponding binding sites to initiate its gene expression. As a result, we briefly illustrated the statistically relevant mechanisms of the promoter use by both transcription factors in determining the final mRNA expression levels of their shared target genes (Fig. 3). This scheme was applied in analyzing the mechanisms for ER α _E2F1, ER α _GATA3 and ER α _ERR α involved promoter uses on their target genes in a given breast cancer population (Tables S1, S4 and S7).

Four statistical conclusions derived from three statistical measures (i.e. CID-TF1nTF2, CID-TF1UGPCC and CID-TF2UGPCC) were grouped into four proposed biological regulatory pathways, respectively (Figs. 3–6). It indicates a preferential

**Table 1.** Summarized statistical analyses of 56 probes on their gene expression regulatory mechanisms by both ER α and GATA3.

Probe no	Featurenum	Genename	ESR1	ESRRA	ESR1ESRRA	Mechanism
1	16378	ANKMY2	S	S	S	1
2	8540	ATP1B1	S	S	S	1
3	1487	ATP6AP2	S	S	S	1
4	10261	ATP6V0A1	S	S	S	1
5	13294	C6orf211	S	S	S	1
6	15121	CA12	S	S	S	1
7	18269	CALU	S	S	S	1
8	12896	CASD1	S	S	S	1
9	3580	CSAD	S	S	S	1
10	2085	CYCS	S	S	S	1
11	11758	DDIT4	S	S	S	1
12	18926	DHRS3	S	S	S	1
13	12004	ENO1	S	S	S	1
14	5480	ESRRA	S	S	S	1
15	7339	FAM100A	S	S	S	1
16	1034	FAM102A	S	S	S	1
17	13981	FRAT2	S	S	S	1
18	14967	GATA3	S	S	S	1
19	16477	KCNN4	S	S	S	1
20	13712	KIAA0664	S	S	S	1
21	19715	KRT18	S	S	S	1
22	20526	LETMD1	S	S	S	1
23	8255	LMO4	S	S	S	1
24	15258	LSMD1	S	S	S	1
25	11537	MRPS30	S	S	S	1
26	14171	MYBPC3	S	S	S	1
27	2111	NDUFA11	S	S	S	1
28	11771	PCYT1A	S	S	S	1
29	1098	PFN1	S	S	S	1
30	19493	PLCD3	S	S	S	1
31	19203	PREX1	S	S	S	1
32	6791	PTP4A2	S	S	S	1
33	6795	STEAP3	S	S	S	1
34	9377	TBKBP1	S	S	S	1
35	11917	WWP1	S	S	S	1
36	12247	ZBTB20	S	S	S	1
37	12023	C11orf21	S	NS	S	2
38	383	COQ7	S	NS	S	2
39	17802	FDXR	S	NS	S	2
40	21403	GPD1 L	S	NS	S	2
41	17053	LMO4	S	NS	S	2
42	2749	MCM6	S	NS	S	2
43	17911	NYX	S	NS	S	2
44	3707	PGPEP1	S	NS	S	2
45	10567	PIK3AP1	S	NS	S	2
46	8719	TARS	S	NS	S	2
47	1409	C6orf62	NS	S	S	3
48	12111	EFNA1	NS	S	S	3
49	12417	MRPL34	NS	S	S	3
50	1636	NDRG4	NS	S	S	3
51	6418	SULT2B1	NS	S	S	3
52	9539	SYP	NS	S	S	3
52	13484	TYSND1	NS	S	S	3
54	6397	PHF15	NS	NS	S	4
55	11333	PHF8	NS	NS	S	4
56	4335	PTMA	NS	NS	S	4

Note: "S" means "significant" and "NS" means "not significant".

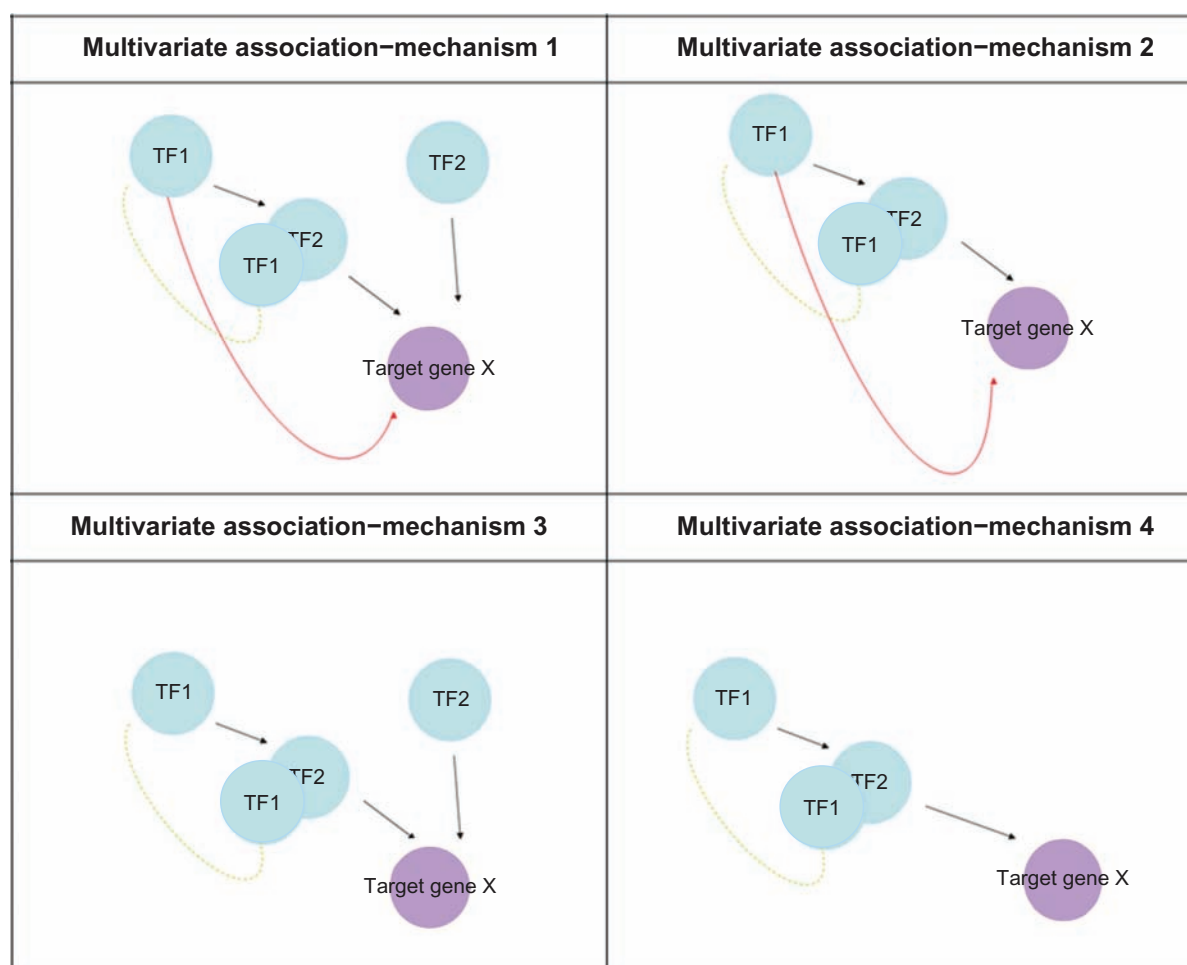


Figure 5. The distinctive combinations of regulatory relationships for four functional gene expression relationships between (TF1, TF2) and target gene X. Univariate association measured by both CID and GPCC is to predict direct (a) and/or remote (b) regulatory relationships between a TF and a gene of interest (i.e. $TF1 \rightarrow TF2$, $TF2 \rightarrow$ target gene X and $TF1 \rightarrow$ target gene X). For case (a), TF2 is known a primary target gene of TF1. In the meantime, target gene X is known as a primary target of TF2. Moreover, TF1 is a transcription factor of target gene X. For case (b), TF2 is a remote target of TF1. Target gene X is a remote target of TF2. In addition, target gene X is a remote target of TF1. On the other hand, multivariate association includes (1) multivariate association partially dependent of univariate association (i.e. mechanisms 1, 2 and 3); (2) multivariate association independent of univariate association (i.e. mechanism 4). The individual mechanism is demonstrated by the regulatory association derived from functional interactions of TF1, TF2 at the promoter region of target gene X. As a result, it is clear that our findings in Tables 1 and 2 have included estrogen responsive genes which have been discussed in main text. This indicates that bivariate CID in general measures direct and/or remote associations between co-expressed TF1, TF2 and their shared target genes. However, these regulatory mechanisms occurring either simultaneously or sequentially can not be distinguished through this approach. Red solid curve with arrow stands for a regulatory relationship between TF1 and target gene X. Black solid arrows stand for the basic regulatory relationship of a TF and its target gene or of TF1_TF2 and their shared target genes. Dashed yellow-green curve stands for position switching of TF1 to join with TF2. Five basic regulatory interactions are predicted via this approach. They are (1) TF1 directly regulates TF2; (2) TF1 indirectly regulates target gene X via TF2; (3) TF1 may also simultaneously join TF2 to co-regulate target gene X; (4) TF1 regulates target gene X; (5) TF2 regulates target gene X. As such, they become four small networks further demonstrating four types of interactions between TF1 and TF2 (or four transcriptional regulatory mechanisms) in controlling the expression of gene X.

promoter use via mechanisms 1, 2, 3 and 4 by two functional transcription factors on their shared target genes in a breast cancer population that can be statistically identified. Whether the transcriptional program is simultaneously or sequentially involving two TFs, its mechanism remains unclear based on this prediction. For instance, 302 probes were expressed due to the pre-programmed estrogen effects on estrogen responsive gene expression in MCF-7 as Increase, Increase, Increase in a time dependent

manner (III mode) (Table S11). Only six probes (6/302) with such expression mode were predicted to stimulate cell cycle signal transduction pathway (Table S12). Eight (8/302) probes were also predicted to be regulated by $ER\alpha_E2F1$ pathway in group IE subtype via mechanisms 1, 2 and 3 (Table S13). Both *ATAD2* and *DTL* have been identified in estrogen responsive event¹⁹ *in vitro*. Our data suggests that group IE breast cancer showed the suppressive effect on *E2F1* gene expression at mRNA level by *ESR1*

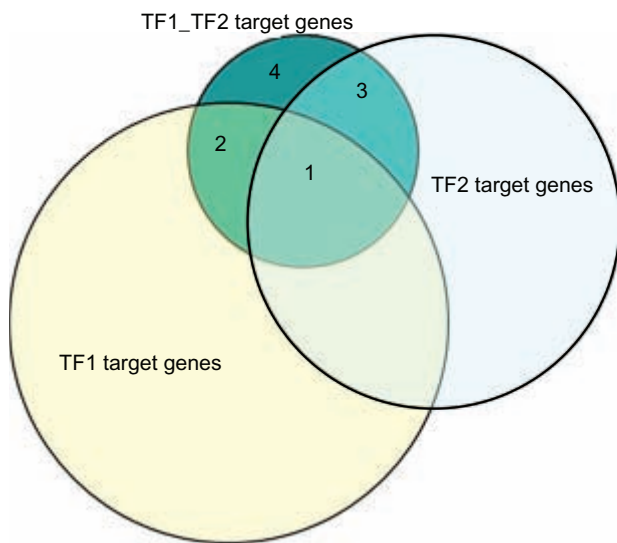


Figure 6. Distinctive features of a shared target gene pool of both TF1 and TF2 based on their overlapped and non-overlapped genes with target genes of TF1 and/or of TF2. They were divided into four categories. We designated them to be mechanism 1, 2, 3 and 4, respectively.

Mechanism 1: the feature definition for mechanism 1 is a target gene expression via cross talking between TF1 and TF2 that is significantly identified by the bivariate CID. In addition, the identified gene expression relationship of TF1_TF2 and their shared target gene is also partially involved the regulatory action of TF1 on their shared target gene. However, this event occurring simultaneously or sequentially can not be distinguished. **Mechanism 2:** the feature definition for mechanism 2 is a target gene expression via cross talking between TF1 and TF2 that is significantly identified by the bivariate CID. In addition, the identified gene expression relationship of TF1_TF2 and their shared target gene is also partially involved the regulatory action of TF1 on their shared target gene. However, this event occurring simultaneously or sequentially can not be distinguished. **Mechanism 3:** the feature definition for mechanism 3 is a target gene expression via cross talking between TF1 and TF2 that is significantly identified by the bivariate CID. In addition, the identified gene expression relationship of TF1_TF2 and their shared target gene partially is dependent on the regulatory action of TF2 on their shared target gene. However, this event occurring simultaneously or sequentially can not be distinguished. **Mechanism 4:** the feature definition for mechanism 4 is a target gene expression via cross talking between TF1 and TF2 that is significantly identified by the bivariate CID. In addition, the identified gene expression relationship of TF1_TF2 and their shared target genes is independent on the individual regulatory actions of both TF1 and TF2 on their target gene. However, this event occurring simultaneously or sequentially can not be distinguished. Three circles represent three statistical measures. The open circles filled with light yellow and light blue colors stand for significance when the regulatory association is measured by univariate association (i.e. CID-TF1UGPCC; CID-TF2UGPCC, $P \leq 0.05$), respectively. The open circle filled with dark green color stands for significance when the regulatory association is measured by multivariate association (i.e. CID-TF1nTF2, $P \leq 0.05$). The overlapped area between multivariate association and univariate association(s) represents functional mechanisms 1, 2 and 3. Functional mechanism 4 shows the area only for multivariate association but no overlapping area between multivariate association and two univariate associations.

in general (Fig. 9A). As a result, both expression of *CD44* and *IVNSIABP* were suppressed at mRNA level (see GPCC-ESR1 results for *ESR1* vs. *CD44* and *ESR1* vs. *IVNSIABP* in Table S13) via mechanism 2 in group IE breast cancer. *RACGAP1*, *RFC3* and *YBX1* were predicted to be down regulated by ER α

and E2F1 via mechanism 1. However, *ALG8* was found to be up-regulated in group IE breast cancer (see GPCC-ESR1 result for *ESR1* vs. *ALG8* in Table S13) via mechanism 2 and in estrogen treated MCF-7. In the meantime, bivariate CID predicted five genes in III mode to be differentially co-regulated by *ESR1* and *GATA3*. *CCT5* and *GART* were possibly down regulated via mechanism 1. Both *DHFR* and *KPNB1* were predicted to be up regulated via mechanisms 1 and 3, respectively. *CPSF2* was predicted to be down regulated via mechanism 3. Furthermore, *KIF2C* and *CDCA8* are two shared target genes of ER α , E2F1 and *GATA3*. They have ERE,²⁰ E2FBS¹⁷ and GATA3BS⁴ at their promoter regions. They were predicted to be significantly regulated via both ESR1_E2F1 and ESR1_GATA3 promoter use pathways in group IE subtype (Tables S1 and S4). Therefore, we suspected that their mRNA expression mode may be operated by switching promoter use pathways and/or switching regulatory mechanisms under a pre-programmed condition. Thus, by gathering these predicted mechanisms, it would be worth to further prove the switching mechanism of III mode to a suppressive mode, in part, due to alternative promoter use by different interactions among TFs in a time dependent manner to achieve a particular phenotype in the model systems.

An inferred transcriptional regulatory network and future medicine

Our experience in a microarray approach

In silico established ER α transcriptional regulatory network predicts a global effect of functional ER α alone (univariate portion of network) and of ER α and TF(s) (multivariate portion of network). Here, we use breast cancer functional transcriptome to briefly test how ER α regulates its target genes in an estrogen-dependent manner. In particular, we propose the network to predict that ER α interacts with different transcription factors to achieve gene expression responses in the subgroup of patients with breast cancer who were diagnosed with a pathological feature of interest. Multiple signaling pathways are known to be activated by functional ER α . We observed the altered mRNA expression patterns for three signal transduction pathways in groups IE and IIE breast tumors as compared to those in non-tumor part in the heatmap generated by unsupervised hierarchical clustering (Figs. S1, S2 and S3). They are cell cycle, VEGF

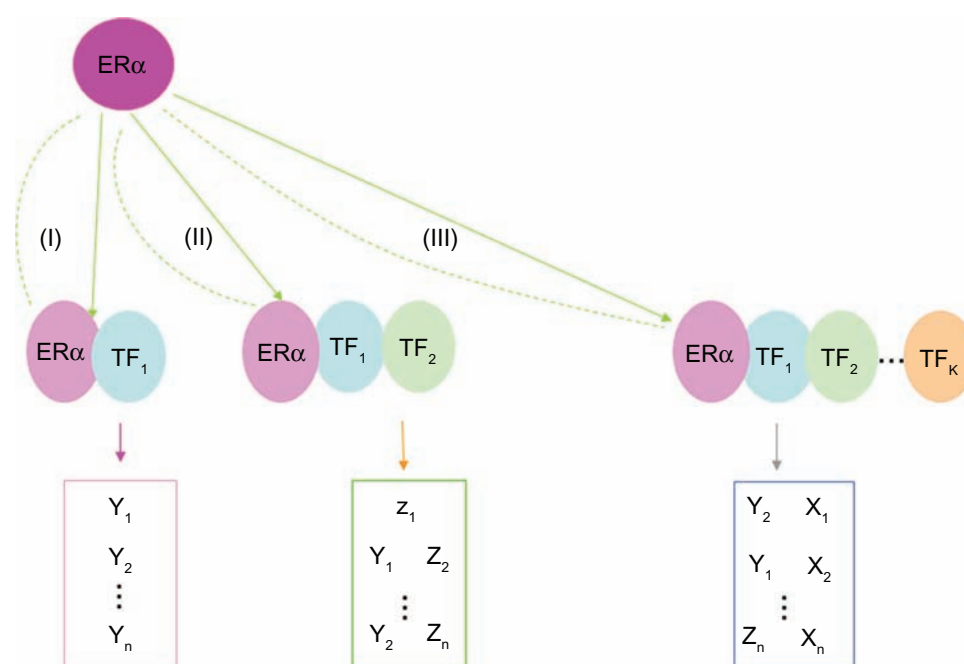


Figure 7. The simplified model of co-regulatory networks predicted by multivariate CID.

An example of ER α initiated transcriptional co-regulatory network is illustrated. When ER α interacts with other TF(s) to co-regulate a gene pool, some of those target genes (e.g. Y_1, Y_2) may have more than one type of binding site (TFBS) at their promoter regions to permit cross-talk among ER α , TF $_1$, TF $_2$ and TF $_k$. As a result, different clusters of gene pools indicated in the figure show their gene expression relationships with the assigned TFs due to a preference in promoter use pathways (i.e. I, II, III). While comparing 194 probes with 102 probes in this study, only *KIF2C*, *CDC48* were found to be the shared targets of ER α , GATA3 and E2F1 both *in vitro* and *in silico* (see Tables S1 and S4). Green solid arrows stand for the causal regulatory relationship of a TF and its target gene or of more than one TF and their shared target genes. Dashed green curve stands for position switching of ER α . The combinatorial regulatory interactions between or among ER α and TF $_k$ ($K = 1, 2, 3, k > 3$) are indicated in the figure that ER α either directly regulates TF $_k$ or indirectly regulates target gene (X, Y, Z) via TF $_k$ or ER α may also simultaneously join TF $_k$ to regulate their target genes (e.g. $X_{1\dots n}, Y_{1\dots n}, Z_{1\dots n}$).

and PDGFRB signal transduction pathways. Then, we suspected coregulatory networks of ESR1_E2F1, ESR1_ESRRA and ESR1_GATA3 in group IE to have their overlapped target gene pools that differentially contribute to those signal transduction pathways, respectively. The combined interactions among *ESR1*, *E2F1*, *ESRRA* and *GATA3* to regulate their shared target gene pools were demonstrated by Venn Diagrams (Tables S18, S20 and S22). We have summarized in Table 2 for ER α mediated transcription activities via combinatorial interactions between

or among transcription factors in groups IE and IIE to potentially contribute to three signal transduction pathways. Interestingly, we validated this multivariate portion of network to be functionally linked to a clinical pathological phenotype (i.e. Vascularity index) in a subset of patients. This suggests its predicting power in dissecting the regulator(s) of disease phenotypic features through functional interactions among transcription factors (Fig. 8). In addition, the inferred network (Fig. 8B) shows that groups IE and IIE shared most of promoter use pathways except for an

Table 2. The numbers of probes in gene pools of cell cycle, VEGF and PDGFRB signal transduction pathways to be regulated by ESR1_E2F1 (EE1), ESR1_GATA3 (EG), ESR1_ESRRA (EESRRA) as well as by their combined promoter use pathways in groups IE and IIE breast cancers.

Promoter use pathway ER(+) subtypes	EE1		EG		EESRRA		EE1_EG_EESRRA	
	IE	IIE	IE	IIE	IE	IIE	IE	IIE
Cell cycle (128 probes)	71	62	76	46	62	43	44	17
VEGF (110 probes)	27	30	63	34	52	25	17	7
PDGFRB (65 probes)	26	18	41	16	26	17	17	6

Note: For biochemical pathway profiling of a transcriptional regulatory network, we included the signal transduction pathways from both KEGG database and NCBI pathway interaction database.

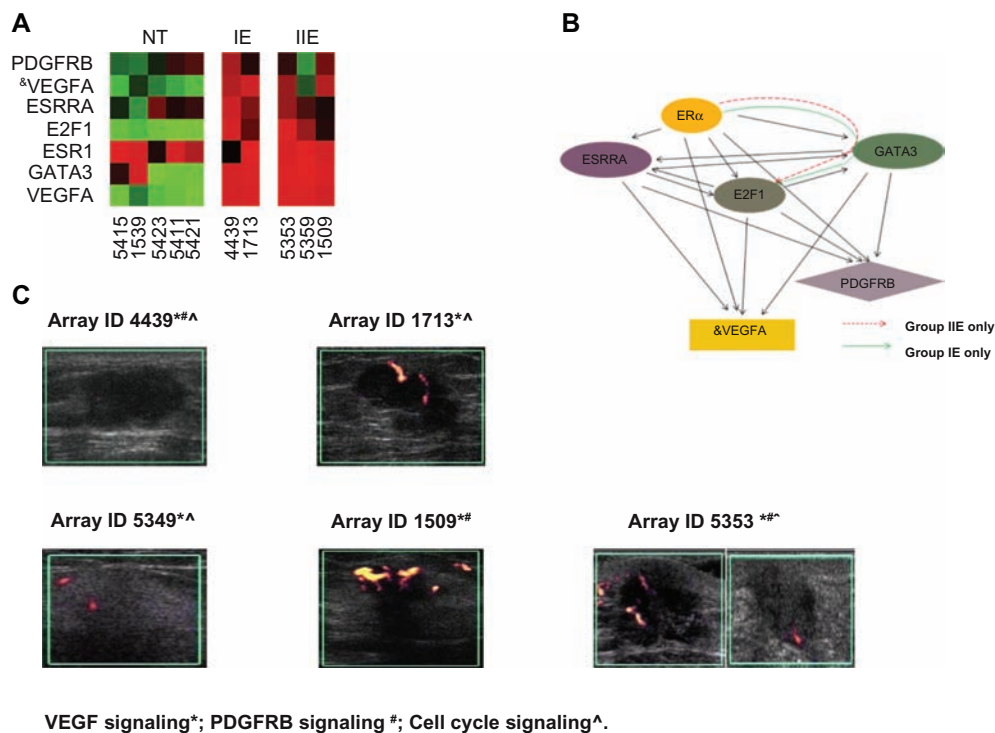


Figure 8. An example of tumor angiogenesis activities in part co-driven by ESR1_E2F1, ESR1_GATA3 and ESR1_ESRRA promoter use pathways in a few ER(+) breast cancer patients.

The vascularity index is a clinical parameter—a new measure for angiogenesis activities using ultrasound that now is routinely used clinically. Panel A shows the heatmaps of a gene profile in a network, which is shown in Panel B, for both groups IE and IIE breast cancer patients. Non-tumor part (NT) is a control. Panel B shows shared multivariate space of a network in two ER(+) breast cancer subtypes except *E2F1* is only significantly regulated by ESR1_GATA3 pathway in group IE breast cancer. Panel C shows the corresponding sonograms for each breast tumor piece per patient array ID. A labeled *VEGFA* (&) has the Agilent feature number as 15367.

ESR1_GATA3 pathway to indirectly regulate *VEGFA* and *PDGFRB* via *E2F1*. Both *VEGFA* and *PDGFRB* are functionally related to tumor vascularity.^{24,25}

Tumor angiogenesis is required for tumor out-growth and metastasis. It is a complex and highly regulated process involving many different cell types and extracellular factors. PDGF receptor beta staining was particularly localized in the cell surface membrane portion of periepithelial stroma in breast carcinoma, suggesting a paracrine stimulation of adjacent stromal tissue by breast tumor cells.²⁶ Interestingly, the statistical correlation between high level of estrogen receptor protein expression and the presence of PDGFR beta protein in cytoplasmic compartment was found in 24% (6/25) breast cancer specimens.²⁶ This may support our network prediction. In addition, the promoter context of both VEGF (ERE,^{27,28} E2F²⁹ and Sp1²⁹) and PDGFRB³⁰ (NF-Y, AP-1, Sp1) also supports network prediction (i.e. direct and/or indirect regulations on gene expressions of *VEGF* and *PDGFRB* by differential interactions among *ESR1*, *E2F1*, *GATA3* and

ESRRA). *VEGFA* is known to be a druggable target for anticancer treatment.

For a population with similar expression ratios between two co-expressed TFs to be grouped together (roughly $N = 10$), we assumed that those TFs may operate their interactions to a similar extent. In this case, a hierarchical clustering was applied to perform subgrouping. As such, if the association for TF1nTF2-target gene measured by bivariate CID is significant, we could easily locate a subgroup with high subCID value in a given population to show more contribution for such relationship. Theoretically, a network was built by linking together multiple significant TF1nTF2-target relationships. To overall evaluate a functional network in a given population, a supervised heatmap (A dendrogram at X axis) will show the gene expression pattern of a network component (i.e. a probe) to be clustered following a preclustered order of TF(s) (that is exactly the same order shown in a dendrogram) (Fig. 9). For instance, eight genes in EE1 pathway following mechanisms 1, 2, 3 and 4 are shown in a heatmap for 152A (Fig. 9B).

The heatmap with unsupervised clustering on those genes is as the control.

In silico established transcriptional regulatory network approach demonstrates a qualitative measure for predicting cause/result relationships between TFn ($n \geq 1$) and target. Therefore, a range of expression levels from low to high for gene components within this network is expected. Some of them are expressed in a pattern as a cluster shown in a heatmap based on their similarity in expression levels in several subgroups (Fig. 9A). Many of them are expressed in a scatter pattern shown in a heatmap (Fig. 9B). When the suppressive mode of ER α on E2F1 expression becomes weak, *E2F1* mRNA level is increased (Fig. 9). Some of those genes in cell cycle signal transduction pathway are switched to an increased mode same as *E2F1*. We picked four significant probes as the example and displayed them in a heatmap (Fig. 9A). They were predicted to have more direct control in TF-target relationship (Significance in linear relationship evaluated by GPCC analysis, see Table S14) and less remote control in TF-target relationship (Significance in nonlinear relationship evaluated by CID analysis,

see Table S14). But all of them were significant based on bivariate CID analysis (Table S14). We found *in vitro* evidence showing three of them (3/8) to have both E2F1 and ER α binding sites (Table S3). They are *ADAMTS5*, *CD44* and *SFRS1*. We also observed only two subgroups in 152A to have high mRNA levels of both *ESR1* and *E2F1*. They showed 8 tested probes to form the clustered gene expression pattern in a heatmap (Fig. 9B). In Figure 9, we demonstrated the feature of CID in picking up the most relevant probes ($P = 0-0.05$). The gene expression patterns of those probes show either clustered or scattered patterns in a heatmap. The display of a heatmap follows two dimensionally supervised (X axis represents an order of a Dendrogram; Y axis represents an order of assigned gene list) but unsupervised clustering made within each subgroup. Such heatmap display indicates that the enriched target gene expression pattern shown in the heatmap may allow a clear vision on linking them with the transcription regulator(s) to form a functional subnetwork. Further supporting information will prove this approach to be powerful. For instance, *in vitro* and/or *in vivo* validations on the novel relationship(s)

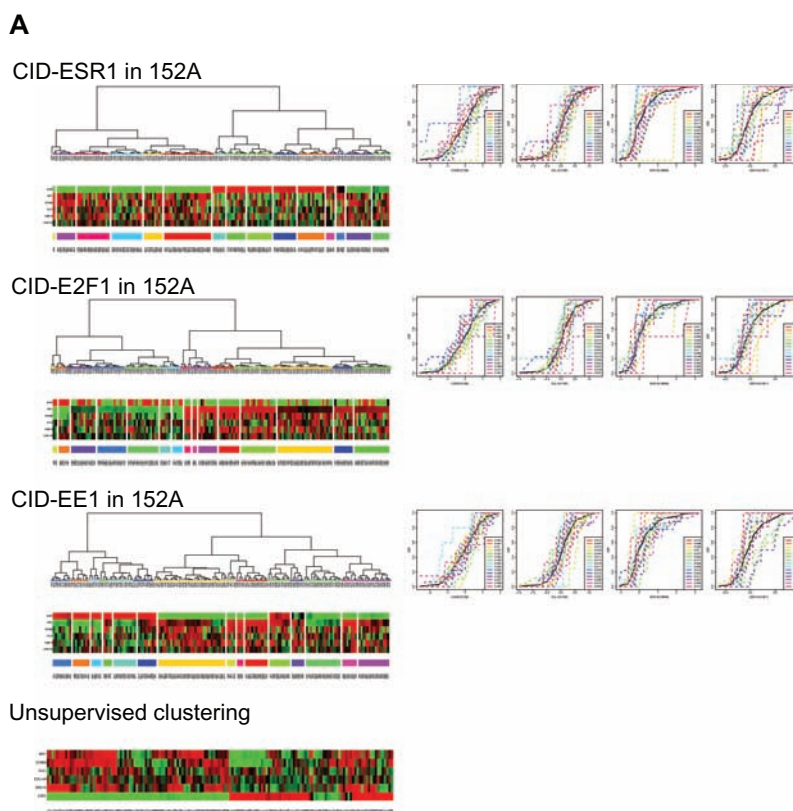


Figure 9. (Continued)

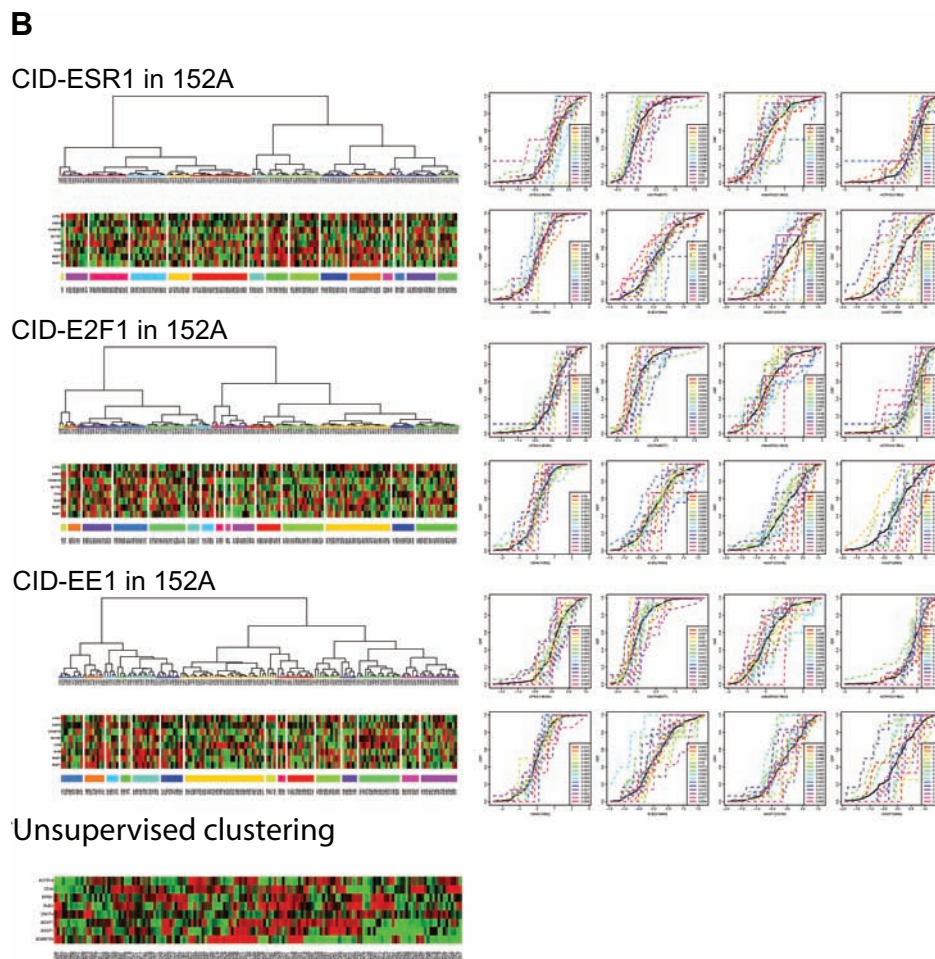


Figure 9. Two examples of cumulative distribution function (CDF) estimation plots for target genes co-driven by ESR1_E2F1 promoter use pathway in 152A and their corresponding heatmaps.

Panel A. Three sets of CDF plots and heatmaps for four shared target probes of both *ESR1* and *E2F1* in cell cycle signal transduction pathway. Upper panel shows the display of a heatmap supervised with a dendrogram for *ESR1* (left) and the CDF plots of CID-*ESR1* measurements on probes of interest (right). Middle panel shows the display of a heatmap supervised with a dendrogram for *E2F1* (left) and the CDF plots of CID-*E2F1* measurements on probes of interest (right). Lower panel shows the display of a heatmap supervised with a dendrogram for co-expressed *ESR1* and *E2F1* (left) and the CDF plots of CID-*ESR1nE2F1* measurements on probes of interest (right). The bivariate CID predicts 4 probes (*CCNB2*, *CDC14A*, *CUL1* and *SMC1B*) to be regulated by *ESR1* and *E2F1*. Fifteen subCID values are listed in a CDF plot per tested probe. This includes two univariate associations (CID-*ESR1*, CID-*E2F1*) and one multivariate association (CID-*ESR1nE2F1*). A heatmap of six probes based on unsupervised clustering is as a control. Fifteen subheatmaps are displayed side by side to form a heatmap that was supervised with its corresponding dendrogram showing a simple, powerful visualization for the regulatory effect of *ESR1* and/or *E2F1* on four target gene expression patterns when each heatmap is compared to the control. **Panel B.** Three sets of CDF plots and heatmaps for eight shared target probes of both *ESR1* and *E2F1*. Upper panel shows the display of a heatmap supervised with a dendrogram for *ESR1* (left) and the CDF plots of CID-*ESR1* measurements on probes of interest (right). Middle panel shows the display of a heatmap supervised with a dendrogram for *E2F1* (left) and the CDF plots of CID-*E2F1* measurements on probes of interest (right). Lower panel shows the display of a heatmap supervised with a dendrogram for co-expressed *ESR1* and *E2F1* (left) and the CDF plots of CID-*ESR1nE2F1* measurements (right). 8 probes were predicted to be regulated by *ESR1* and *E2F1* following mechanisms 1 (two variants *AGGF1*), 2 (*BUB3*, *CD44*), 3 (*ACTR10*, *ADAMTS5*) and 4 (*CNOT4*, *SFRS1*) by bivariate CID. Fifteen subCID values are listed in a CDF plot per tested probe. This includes two univariate associations (CID-*ESR1*, CID-*E2F1*) and one multivariate association (CID-*ESR1nE2F1*). A heatmap of eight probes generated based on unsupervised clustering is as a control. Fifteen subheatmaps are displayed side by side to form a heatmap that was supervised with its corresponding dendrogram show a simple, powerful visualization for the regulatory effect of *ESR1* and/or *E2F1* on eight target gene expression patterns when each heatmap is compared to the control. The Agilent feature number for each probe is within a parenthesis right behind its corresponding gene name shown in CDF plots of panel A and panel B. Those color bars underneath the heatmap are used to label each subgroup which was clustered by hierarchical clustering.

in network(s) will be needed. To link network with phenotypic changes, to further validate a subnetwork to be essential to a particular pathological state etc. are a few experimental directions in adding our knowledge for how transcriptional dynamics between or among transcriptional regulatory subnetworks in determining

tumor fate. Our ongoing effort in discovering druggable targets as well as prognostic factors with network approach has been encouraging. It is becoming clear that we may unravel the action of a transcription factor in coupling with its major partner transcription factor(s) at a system-wide scale with network approach.



The potential technical concerns in interpreting results from network approach

In silico established transcriptional regulatory network, which can be built via a series of unbiased analyses using appropriate algorithms, enables us in dissecting complex disease biology. Here, a few technical concerns are concluded based on our experience in using *in silico* built transcriptional regulatory network constructed mainly by CID in a clinical breast cancer model system. Most of them may be from the shortcomings of high throughput biology and the heterogeneity of disease. But, only a few are from network approach.

- (1) When one only measures the association between or among continuous variables using hierarchical clustering for subgrouping, a few results are not similar to subgrouping via expression levels in dichotomous or multichotomous categories (e.g.s immunohistochemical staining), which is more relevant to clinical status. For instance, the bivariate CID measurement for $ESR1_TF_x$ -target may provide the result to be less relevant than that for $ER\alpha_TF_x$ -target. $ESR1/TF_x$ mRNA ratio may not truly reflect protein ratio for $ER\alpha/TF_x$ because Jarzabek et al.³¹ concluded that the gene expression of *ESR1* at mRNA and protein levels are not in a linear relationship in breast cancer due to post-transcriptional or post-translational mechanisms. In some cases, $ER\alpha$ is more dysregulated in breast cancer. But, the TF of interest may be more tightly regulated in breast cancer. As such, more supporting experiments will be performed to validate prediction from network approach.
- (2) This network architecture is highly dependent on the content of a given population. When dealing with heterogeneous disease like cancer, this network approach may conclude the functional relevant phenotype(s) to be significant and to be reproducible only if the population has been well selected.
- (3) A small bias in *P* value determination between two different runs for the same experiment can not be avoided. This is due to random sampling during 1,000 times simulations per experiment may alter the chance of a few probes to be recruited as the significant ones.
- (4) The limitation in sampling options for clinical samples (a sample per patient and this sample is a portion of tumor tissue not a whole one) may reduce

the accuracy of prediction. For instance, we have considered the sample source as well as the tumor proportion among samples to be different when our samples for groups IE and IIE breast cancers are compared to the samples described by Oh et al.¹

- (5) More than one gene products—transcription variants are differentially co-regulated by two transcription factors of interest. For example, the bivariate network approach predicts tumor angiogenesis activities for five patients shown in Figure 8. One patient (array ID 5349) has low mRNA levels for both *PDGFRB*(4503) and *VEGFA*(15367) but showing visible vascularity on the sonogram. For this patient, we found an elevated mRNA expression of another variant—*VEGFA*(1135) but is only predicted to be regulated by promoter use pathways—EE1 and EG not EESRRA in group IIE that may explain the positive result shown on angiosonogram of array 5349, in part. However, we still need to confirm expression status of *PDGFRB* for this patient in our future experiment.
- (6) CID has the advantage of measuring association (designated as subCID) in a small subgroup ($N \approx 10$). All subCID values were added together as the final CID value and it was evaluated as significance by its corresponding *p* value. The subCID values in some subgroups show relatively high as compared to those in other subgroups within a given population. This only indicates more contribution for those subgroups in determining a particular TF1nTF2-target relationship in a given population. To relate CID results with the heatmap display and to claim a few subgroups to be statistically significant for a given TF1nTF2-target relationship, we may need increase *N* number in those subgroups to confirm results derived from studies using small *N*.
- (7) It is not unusual for uneven data partitioning. After hierarchical clustering, some subgroups in this study have *N* number ranging from 1 to 33. Statistically, it ($N < 10$) reaches a critical point for CID prediction not to be reliable. Excluding those small *N*s before CID measurement could be an option. However, such strategy has the disadvantage for each bivariate CID measurement not based on the same population while biologically reproducible investigations are made in the same population for



establishing the common relationship(s) among many transcription factors ($N > 2$) to control their shared target gene expressions.

- (8) To effectively increase the power of network approach, we will consider including other clustering methods³² for identifying a variety of combinations on transcription factors that selectively carry out their functional interactions within the same expression profiling data.

Conclusions

Together, we have demonstrated a supervised statistical approach using bivariate CID to predict the most relevant network of interactions between transcription factors in regulating their shared target gene expressions in a population. Typically, our data mainly indicates bivariate CID to be successful in finding the most relevant, alternative pathways for promoter use by two given transcription factors in regulating their shared target gene expressions in a breast cancer population.

First, we have predicted three relevant promoter use pathways that were operated by combinatorial interactions between ER α and a TF (E2F1 or GATA3 or ER α) on the promoter region of their shared target gene in a breast cancer population, respectively. Four unique regulatory mechanisms predicted in each promoter use pathway can be statistically identified by combined methods of bivariate CID, univariate CID and Galton-Pearson's correlation coefficient. Biologically, those regulatory events involving cross-talk between two TFs are divided into four distinct mechanisms due to the co-regulatory event partially dependent on (a) individual transcriptional activities from those two TFs (Mechanism 1); (b) a TF dominant regulation (Mechanism 2 and Mechanism 3); and (c) no individual transcriptional activity from those two TFs (Mechanism 4). The expression of a gene may be co-regulated by many TFs ($N \geq 2$) simultaneously and/or sequentially. Thus, a simplified model of biologically pre-programmed gene expression relevant in a given population for the ER α involved co-regulatory networks predicted by multivariate CID was proposed (Fig. 7). Those predictions have been partially validated *in vitro*.

Second, this approach can be very sensitive to describe the subtle difference between two subtypes of breast cancers or between tumor and non-tumor

tissues. For instance, Figure 8 indicates a promise for the multivariate space of an inferred ER α transcriptional regulatory subnetwork, which differentially controls a subset of angiogenesis related gene expressions, to be relevant in leading the phenotypic difference (i.e. vascularity) between tumor and non-tumor parts (Figs. 8A and C). We observed the subtle difference in combinatorial actions of *ESR1*, *E2F1*, *GATA3* and *ESRRA* on up regulating genes that promote tumor vascularity in groups IE and IIE predicted by this subnetwork (Fig. 8B). This suggests a possible niche for an *in silico* established multivariate space of ER α transcriptional regulatory network in breast cancer diagnostic and therapeutic development.

Abbreviations

A few key abbreviations in this study are listed below.

PDGFRB, platelet-derived growth factor receptor, beta polypeptide; VEGF, vascular endothelial growth factor; VEGFA, vascular endothelial growth factor A; ER(-), negative status for immunochemical stain of estrogen receptor; ER(+), positive status for immunochemical stain of estrogen receptor; PR(-), negative status for immunochemical stain of progesterone receptor; PR(+), positive status for immunochemical stain of progesterone receptor; CID, coefficient of intrinsic dependence; GPCC, Galton Pearson's Correlation Coefficient; KEGG, Kyoto Encyclopedia of Genes and Genomes; NCBI, National Center for Biotechnology Information; CDF, cumulative distribution function; E2F1, E2F transcription factor 1; ESR1, estrogen receptor 1; ESRRA, estrogen-related receptor alpha; Group IE, ER(+)/PR(+); Group IIE, ER(+)/PR(-).

in vitro: Literally, "in glass". A term describing biological reactions that occur in a laboratory apparatus or test tube.

in vivo: Literally, "in life". A term describing processes that occur within a living cell or organism.

in silico: Literally, "in computer". A term describing analyses that are performed in a computer using the tools derived from specific algorithms.

Acknowledgements

We wish to thank for the valuable dataset from Dr. Myles Brown's group as part of data analyzed in this study and Dr. Tzu L. Phang provided part of results for trajectory clustering analysis on that



dataset. In addition, we wish to express our sincere thanks to Dr. Sorin Draghici for the on-line free package of Onto-tools. We also feel grateful with the kindness from both Dr. Huang-Chun Lien and Dr. Shih-Ming Jung for examining IHC stain of this project. Dr. Yi-Shing Lin at Welgene Biotechnology Company also kindly provided some of the technical support for this work. Thanks to Cancer Registry, Medical Information Management Office and IRB at NTUH in providing the great assistance for accessing breast cancer patient clinicopathological information. Thanks to Miss Fu-Chin Chen for gathering sonograms for this study. This work was supported mainly by the financial support from NSC95-2314-B-002-255-MY3(FJH), NSC 98-2314-B-002-093-MY2(FJH) and NSC 99-2118-M-002-004(LDL).

Disclosures

Author(s) have provided signed confirmations to the publisher of their compliance with all applicable legal and ethical obligations in respect to declaration of conflicts of interest, funding, authorship and contributorship, and compliance with ethical requirements in respect to treatment of human and animal test subjects. If this article contains identifiable human subject(s) author(s) were required to supply signed patient consent prior to publication. Author(s) have confirmed that the published article is unique and not under consideration nor published by any other publication and that they have consent to reproduce any copyrighted material. The peer reviewers declared no conflicts of interest.

References

- Oh DS, Troester MA, Usary J, et al. Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. *J Clin Oncol*. 2006;24:1656–64.
- Liu LYD, Chen CY, Chen MJM, et al. Identification of gene expression relationships by CID to build a transcriptional regulatory network. *BMC Bioinformatics*. 2009;10:85–97.
- DeNardo DG, Cuba VL, Kim HT, Wu K, Lee AV, Brown PH. Estrogen receptor DNA binding is not required for estrogen-induced breast cell growth. *Mol Cell Endocrinol*. 2007;277:13–25.
- Jin VX, Leu YW, Liyanarachchi S, et al. Identifying estrogen receptor alpha target genes using integrated computational genomics and chromatin immunoprecipitation microarray. *Nucleic Acids Res*. 2004;32:6627–35.
- Manavathi B, Kumar R. Steering estrogen signals from the plasma membrane to the nucleus: two sides of the coin. *J Cell Physiol*. 2006;207:594–604.
- Shen Q, Uray IP, Li Y, et al. The AP-1 transcription factor regulates breast cancer cell growth via cyclins and E2F factors. *Oncogene*. 2008;27:366–77.
- Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang THM. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet*. 2008;24(4):167–77.
- Lien HC, Hsiao YH, Lin YS, et al. Molecular signatures of metaplastic carcinoma of the breast by large-scale transcriptional profiling: identification of genes potentially related to epithelial-mesenchymal transition. *Oncogene*. 2007;26(57):7859–71.
- Liu LYD. Coefficient of intrinsic dependence: a new measure of association. *Ph.D Thesis*. Texas A&M University, Department of Statistics; 2005.
- Goldstein DR, Ghosh D, Conlon EM. Statistical issues in the clustering of gene expression data. *Stat Sin*. 2002;12:219–40.
- Draper NR, Smith H. *Applied Regression Analysis*. Wiley Series in Probability and Statistics; 1998.
- Draghici S, Khatri P, Tarca AL, et al. A systems biology approach for pathway level analysis. *Genome Res*. 2007;17:1537–45.
- Eeckhoutte J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M. Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Res*. 2007;67(13):6477–83.
- Jin VX, Leu YW, Liyanarachchi S, et al. Identifying estrogen receptor alpha target genes using integrated computational genomics and chromatin immunoprecipitation microarray. *Nucleic Acids Res*. 2004;32:6627–35.
- Louie MC, McClellan A, Siewit C, Kawabata L. Estrogen receptor regulates E2F1 expression to mediate tamoxifen resistance. *Mol Cancer Res*. 2010;8(3):343–52.
- Liu D, Zhang Z, Gladwell W, Teng CT. Estrogen stimulates estrogen-related receptor α gene expression through conserved hormone response elements. *Endocrinology*. 2003;144(11):4894–904.
- Xu X, Bieda M, Jin VX, et al. A comprehensive ChIP-chip analysis of E2F1, E2F4, and E2F6 in normal and tumor cells reveals interchangeable roles of E2F family members. *Genome Res*. 2007;17:1550–61.
- Chisamore MJ, Cunningham ME, Flores O, Wilkinson HA, Chen JD. Characterization of a novel small molecule subtype specific estrogen-related receptor α antagonist in MCF-7 breast cancer cells. *PLoS ONE*. 2009;4(5):e5624–37.
- Bourdeau V, Deschênes J, Laperrière D, Aid M, White JH, Mader S. Mechanisms of primary and secondary estrogen target gene regulation in breast cancer cells. *Nucleic Acids Res*. 2008;36:76–93.
- Hua S, Kallen CB, Dhar R, et al. Genomic analysis of estrogen cascade reveals histone variant H2A.Z associated with breast cancer progression. *Mol Syst Biol*. 2008;4:188–201.
- Riggins RB, Mazzotta MM, Maniya OZ, Clarke R. Orphan nuclear receptors in breast cancer pathogenesis and therapeutic response. *Endocrine-Related Cancer*. 2010;17:R213–31.
- Gaillard S, Dwyer MA, McDonnell DP. Definition of the molecular basis for estrogen receptor-related receptor- α -cofactor interactions. *Mol Endocrinol*. 2007;21(1):62–76.
- Kallen J, Schlaeppli JM, Bitsch F, et al. Evidence for ligand-independent transcriptional activation of the human estrogen-related receptor alpha (ERR alpha): crystal structure of ERR alpha ligand binding domain in complex with peroxisome proliferator-activated receptor coactivator-1alpha. *J Biol Chem*. 2004;279:49330–7.
- Cheng WF, Lee CN, Chen CA, et al. Comparison between “in vivo” and “in vitro” methods for evaluating tumor angiogenesis using cervical carcinoma as a model. *Angiogenesis*. 1999;3:295–304.
- Suzuki S, Dobashi Y, Hatakeyama Y, et al. Clinicopathological significance of platelet-derived growth factor (PDGF)-B and vascular endothelial growth factor-A expression, PDGF receptor- β phosphorylation, and microvessel density in gastric cancer. *BMC Cancer*. 2010;10:659–69.
- Coltrera MD, Wang J, Porter PL, Gown AM. Expression of platelet-derived growth factor B-chain and the platelet-derived growth factor receptor beta subunit in human breast tissue and breast carcinoma. *Cancer Res*. 1995;55:2703–8.
- Mueller MD, Vigne JL, Minchenko A, Lebovic DI, Leitman DC, Taylor RN. Regulation of vascular endothelial growth factor (VEGF) gene transcription by estrogen receptors alpha and beta. *Proc Natl Acad Sci U S A*. 2000;7(20):10972–7.



28. Luo H, Rankin GO, Liu L, Daddysman MK, Jiang BH, Chen YC. Kaempferol inhibits angiogenesis and VEGF expression through both HIF dependent and independent pathways in human ovarian cancer cells. *Nutr Cancer*. 2009;61(4):554–63.
29. Merdzhanova G, Gout S, Keramidas M, et al. The transcription factor E2F1 and the SR protein SC35 control the ratio of pro-angiogenic versus antiangiogenic isoforms of vascular endothelial growth factor-A to inhibit neovascularization in vivo. *Oncogene*. 2010;29:5392–403.
30. Qin Y, Fortin JS, Tye D, Gleason-Guzman M, Brooks TA, Hurley LH. Molecular cloning of the human PDGFR- β promoter and drug targeting of the G-quadruplex-forming region to repress PDGFR- β expression. *Biochemistry*. 2010;49(19):4208–19.
31. Jarzabek K, Koda M, Kozlowski L, et al. Distinct mRNA, protein expression patterns and distribution of oestrogen receptors α and β in human primary breast cancer: Correlation with proliferation marker Ki-67 and clinicopathological factors. *Eur J Cancer*. 2005;41:2924–34.
32. Katagiri F, Glazebrook J. Pattern discovery in expression profiling data. *Curr Protoc Mol Biol*. 2009;Chapter 22:Unit 22.5.1–22.5.15.
33. Tsai SY, Opavsky R, Sharma N, et al. Mouse development with a single E2F activator. *Nature*. 2008;454(7208):1137–41.



Supplementary Data

Additional file—a PDF file contains Tables S1–S23 and Figures S1–S3. However, Table of content and Figures S1-S3 are shown on pages 21-23 of this article.

Supplementary Information

Table of content

Table S1. The subtypic differences in promoter use by ER α and E2F1 on their shared target genes (194 probes) in estrogen responsive breast cancers.

Table S2. Differences between two subtypes of estrogen responsive breast cancers in the combinatorial interactions between *ESR1* and *E2F1* for regulation of their shared target gene expressions.

Table S3. Literature cited additional feature(s) of shared target genes of *ESR1* and *E2F1* in an estrogen responsive breast cancer subtype—group IE.

Table S4. The subtypic differences in promoter use by ER α and GATA3 on their shared target genes (102 probes) in estrogen responsive breast cancers.

Table S5. Differences between two subtypes of estrogen responsive breast cancers in the combinatorial interactions between *ESR1* and *GATA3* for regulation of their shared target gene expressions.

Table S6. Literature cited additional feature(s) of shared target genes of *ESR1* and *GATA3* in an estrogen responsive breast cancer subtype—group IE.

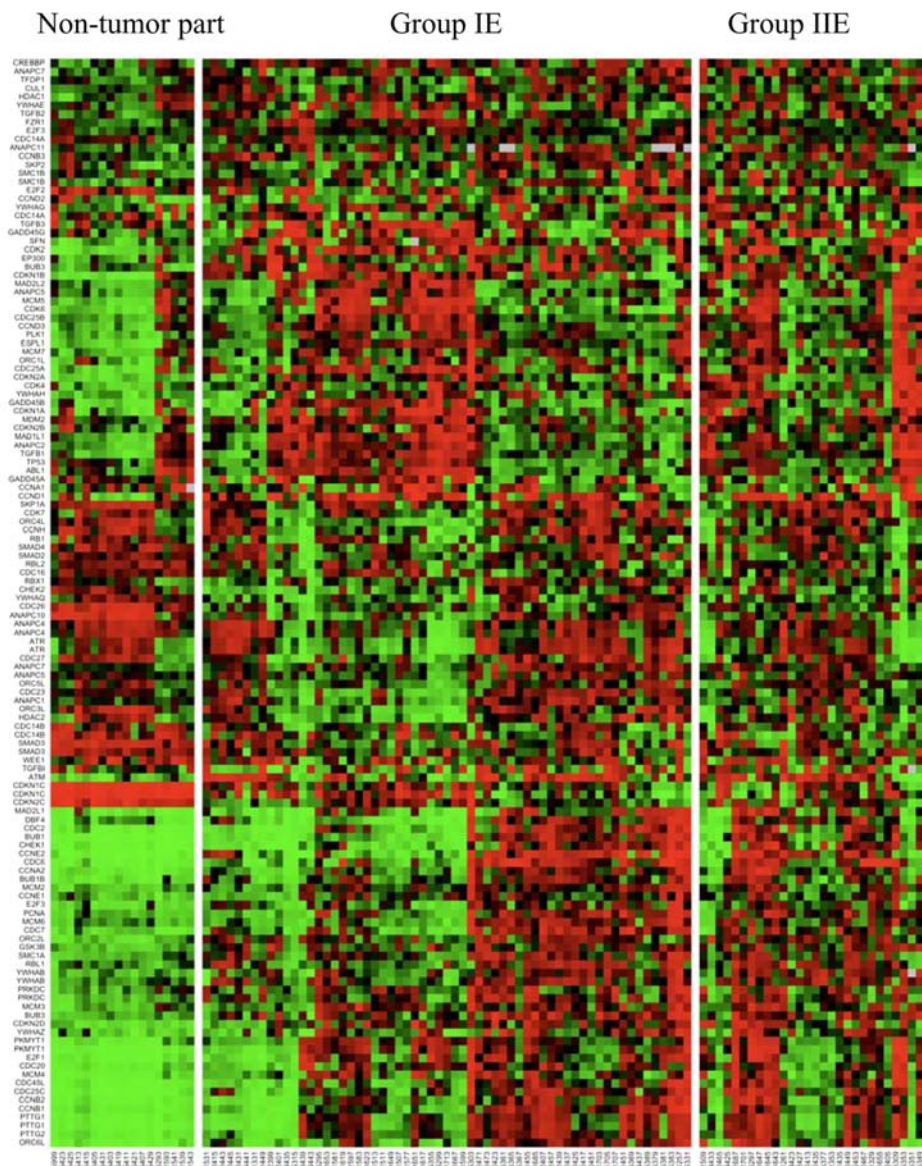


Figure S1. A heatmap of gene expression pattern for cell cycle signal transduction pathway in groups IE and IIE as compared to non-tumor part.

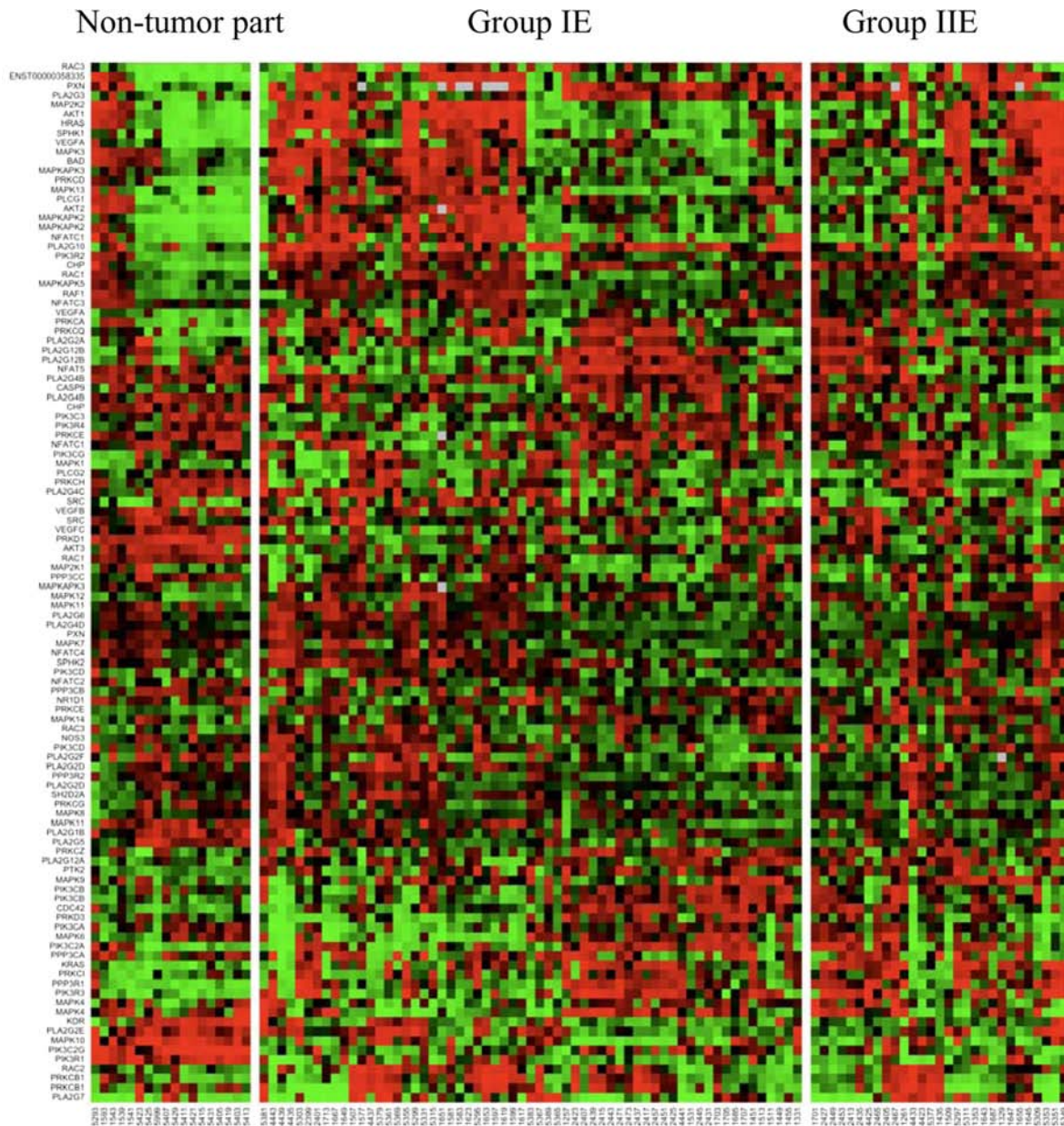


Figure S2. A heatmap of gene expression pattern for vascular endothelial growth factor (VEGF) signal transduction pathway in groups IE and IIE as compared to non-tumor part.

Table S7. The subtypic differences in promoter use by ER α and ERR α on their shared target genes (56 probes) in estrogen responsive breast cancers.

Table S8. Differences between two subtypes of estrogen responsive breast cancers in the combinatorial interactions between *ESR1* and *ESRRA* for regulation of their shared target gene expressions.

Table S9. Literature cited additional feature(s) of shared target genes of *ESR1* and *ESRRA* in an estrogen responsive breast cancer subtype—group IE.

Table S10. The clinical pathological data for 90 clinical breast cancer arrays (90A).

Table S11. The gene list of III mode in estrogen treated MCF-7 (Total 302 probes) extracted by trajectory clustering.

Table S12. Six probes in III mode of estrogen treated MCF-7 predicted to be in cell cycle signal transduction pathway (112 genes).

Table S13. Probes in III expression mode (302 probes) of estrogen treated MCF-7 partially predicted to be shared target genes of two promoter use pathways in group IE breast cancer.

Table S14. The results for both univariate and bivariate associations in determining regulatory mechanisms of

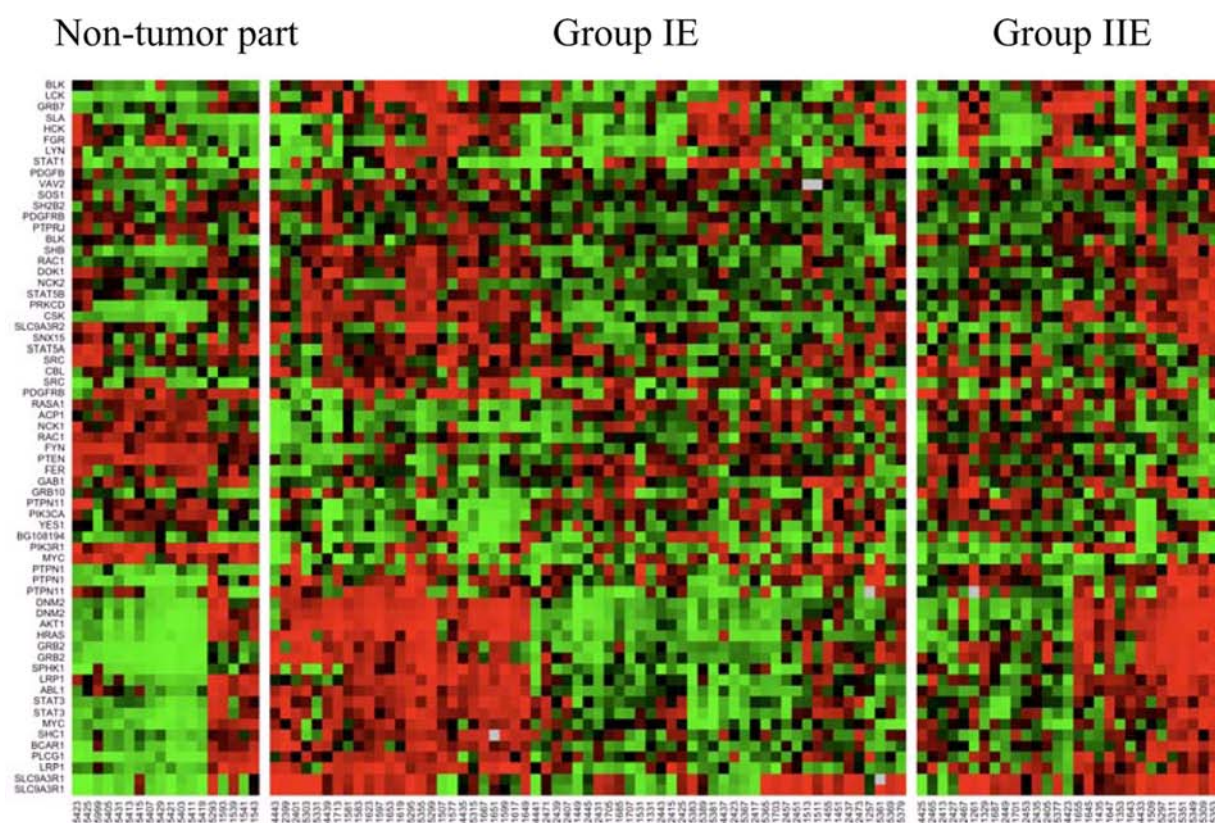


Figure S3. A heatmap of gene expression pattern for platelet-derived growth factor receptor, beta polypeptide (PDGFRB) signal transduction pathway in groups IE and IIE as compared to non-tumor part.

both four probes and eight probes in ESR1_E2F1 promoter use pathway in group IE breast cancer.

Table S15. The overlapped gene pool between VEGF signal transduction pathway and ESR1_E2F1 (EE1) promoter use pathway in group IIE breast cancer.

Table S16. The overlapped gene pool between VEGF signal transduction pathway and ESR1_GATA3 (EG) promoter use pathway in group IIE breast cancer.

Table S17. The overlapped gene pool between VEGF signal transduction pathway and ESR1_ESRRA (ESRRA) promoter use pathway in group IIE breast cancer.

Table S18. The shared gene list for cell cycle signal transduction pathway co-regulated by ESR1_E2F1, ESR1_GATA3, and ESR1_ESRRA promoter use pathways in group IE breast cancer.

Table S19. The shared gene list for cell cycle signal transduction pathway co-regulated by ESR1_E2F1, ESR1_GATA3, and ESR1_ESRRA promoter use pathways in group IIE breast cancer.

Table S20. The shared gene list for VEGF signal transduction pathway co-regulated by ESR1_E2F1, ESR1_GATA3, and ESR1_ESRRA promoter use pathways in group IE breast cancer.

Table S21. The shared gene list for VEGF signal transduction pathway co-regulated by ESR1_E2F1, ESR1_GATA3, and ESR1_ESRRA promoter use pathways in group IIE breast cancer.

Table S22. The shared gene list for PDGFRB signal transduction pathway co-regulated by ESR1_E2F1, ESR1_GATA3, and ESR1_ESRRA promoter use pathways in group IE breast cancer.

Table S23. The shared gene list for PDGFRB signal transduction pathway co-regulated by ESR1_E2F1, ESR1_GATA3, and ESR1_ESRRA promoter use pathways in group IIE breast cancer.



Publish with Libertas Academica and every scientist working in your field can read your article

"I would like to say that this is the most author-friendly editing process I have experienced in over 150 publications. Thank you most sincerely."

"The communication between your staff and me has been terrific. Whenever progress is made with the manuscript, I receive notice. Quite honestly, I've never had such complete communication with a journal."

"LA is different, and hopefully represents a kind of scientific publication machinery that removes the hurdles from free flow of scientific thought."

Your paper will be:

- Available to your entire community free of charge
- Fairly and quickly peer reviewed
- Yours! You retain copyright

<http://www.la-press.com>