Supplemental Text

Contents:

I. Supplemental Experimental Procedures

Luciferase reporter assays

RAR and ERα/RAR binding regions were amplified from human genomic DNA using primers tagged with KpnI/XhoI restriction site sequences (see Table S9) for subsequent cloning into pGL4.23 (Promega). The fragments were cloned in the same orientation to the transcription start site of the reporter gene they have to the transcription start site of their putative target genes in the human genome.

MCF-7 cells cultured in D-MEM with 10% charcoal-stripped FBS were transfected with 100 ng pGL4.23 containing the cloned fragments or the original vector, and 2 ng pGL4.73 using Lipofectamine LTX (Invitrogen) in white-bottom 96 well plates. Cells were treated with 50 nM of an AM580/CD437 mixture and the corresponding amount of DMSO as vehicle 24 hours after plasmid transfection, or with 10 nM E2, 50 nM of an AM580/CD437 mixture, or a mixture of 10 nM E2 and 50 nM of AM580/CD437, or a corresponding amount of DMSO and ethanol as vehicle (the same

concentrations of DMSO and ethanol were also used for all ligand treatments). Following incubation for 24 hours the Dual-Glo Luciferase assay (Promega) was performed. All experiments were performed with 4 replicates.

Reverse transfection with siRNAs

Reverse transfection was carried out at a concentration of 50 nM of control siRNA or 4 siRNAs directed against the same target gene. The siRNAs were then incubated with Dharmafect 1 transfection reagent (Dharmacon) for 20 minutes, which was pre-incubated with OptiMEM media (Invitrogen) for 5 minutes. MCF-7 cells were trypsinized and added onto the mixture of siRNA and transfection reagent. Gene specific siRNA oligonucleotides were ON-TARGET*plus* SMARTpool reagents (Dharmacon). The catalog numbers for siRNAs and controls used in this study are as follows: *RARA* (L-003437-00), *RARG* (L-003439-00), *ESR1* (L-003401-00), *FOXA1* (L-010319-00), *GATA3* (L-003781-00) and control siRNA (siNT1 or Non-targeting siRNA #1, D-001810-01).

Chromatin immunoprecipitation (ChIP) experiments

Cells at 80% confluency (~$5x10^6$ cells per ChIP) were subjected to chromatin immunoprecipitation as previously described with the following antibodies: goat anti-GFP (raised against His-tagged full-length eGFP and affinity-purified with GST-tagged full-length eGFP), goat anti-FoxA1 (ab5089) from Abcam, anti-panH3ac (06-599) from Millipore, anti-ERα (MC-20, sc-542x) and normal goat IgG (sc-2028) from Santa Cruz Biotechnologies.

For ChIP-qPCR assays, the fold enrichment of ChIPed DNA relative to input DNA at a given genomic site was determined by comparative C_T ($\Delta\Delta C_T$) method using StepOnePlus Real-Time PCR System and Power SYBR Green PCR Master Mix (Applied Biosystems) according to the manufacturer's protocol. An *ACTB* exonic region or *18S rRNA* genomic region was used for normalization. All primer sequences used for qPCR are described in Table S9.

For ChIP-chip, both ChIPed DNA and input DNA were subjected to linker-mediated PCR amplification, fragmentation and end-labeled with biotin using the GeneChip® WT Double-Stranded DNA Terminal Labeling Kit (Affymetrix) as previously described. The resulting labeled samples were hybridized to Affymetrix GeneChip® Human Tiling 2.0R Array Set following the Affymetrix® Chromatin Immunoprecipitation Assay Protocol. Independent biological triplicates were performed for each transcription factor, as well as the control (input DNA).

qRT-PCR and microarray gene expression profiling experiments

Total RNA from MCF-7 cells treated with different agonists and/or transfected with siRNAs was isolated and purified using RNeasy Mini Kits (Qiagen) according to the manufacturer's instructions. For qRT-PCR cDNA was synthesized with the SuperScriptIII kit (Invitrogen). Relative expression levels for specific genes was determined by comparative C_T ($\Delta\Delta C_T$) method using StepOnePlus Real-Time PCR System and Power SYBR Green PCR Master Mix (Applied Biosystems) according to the manufacturer's protocol. A *B2M* cDNA region was used for normalization. All primer sequences used for qRT-PCR are described in Table S9. Total RNA samples were labeled

by direct incorporation of cyanine 3-labeled CTP using the Agilent Low RNA Input Linear Amplification Kit PLUS (One-Color) (Agilent Technologies) according to the manufacturer's instructions. 1 μg of total RNA was the input for each labeling reaction. Labeled cRNAs were further purified with RNeasy Mini Kits (Qiagen). cRNA yield and the specific activity of Cy3 were determined using a NanoDrop ND-1000 UV-VIS Spectrophotometer (NanoDrop Technologies). Only samples with labeled cRNA yield > 1.65 μg and a specific activity > 9.0 pmol Cy3 per μg cRNA were used for further analysis. 1.65 μg of labeled cRNA was hybridized to Agilent Human Genome Oligo Microarrays (4 X 44K) containing 45,015 probes per array (Agilent Technologies). Hybridized microarrays were scanned using a GenePix 4000B scanner (Molecular Devices) at 5 μm resolution. All experiments were performed in triplicates.

Immunofluorescence and microscopy

RARα-LAP transgenic MCF-7 cells grown on coverslips were fixed in ice-cold methanol at -20°C for 8 minutes. After blocking in 0.2% gelatin from cold water fish (Sigma) in 1X PBS (PBS/FSG) for 20 minutes, coverslips were incubated with antibodies against eGFP (1:1000 dilution, mouse monoclonal, cat# A11120, Invitrogen) and RARα (1:1000 dilution, rabbit polyclonal, cat# 2554, Cell Signaling) and in PBS/FSG for 30 minutes. Following washes with 0.2% PBS/FSG, cells were incubated with 1:250 dilution of secondary antibody for 30min (donkey anti-rabbit conjugated to Alexa 488, anti-mouse conjugated to Alexa 555, Molecular Probes). Coverslips were counterstained with 1 μg/ml DAPI to visualize chromatin. After washing with 0.2% FSG/PBS, coverslips were mounted on glass slides by inverting them into mounting solution. Images were acquired on an imaging system (DM5000B, Leica) that was equipped with a microscope (DM5000B, Leica), a CCD camera (RETIGA 2000R, QImaging), with 63X objectives.

Western Blot

Whole cell lysates of wild-type and RARα-LAP transgenic MCF-7 cells in 1% SDS lysis buffer were quantified using Micro BCA Protein Assay Kit (Pierce). 25 μg of total protein per well were electrophoresed on a SDS-PAGE gel (NuPage 4-12% Bis-Tris, Invitrogen), blotted, and subsequently incubated with an antibody against RARα (1:1000 dilution, rabbit polyclonal, cat# SC551X, Santa Cruz Biotech) or β-Actin (1:10000 dilution, mouse monoclonal, cat# ab6276, Abcam) as loading control. After incubation with anti-rabbit or anti-mouse IgG-HRP Conjugate (BioRad), bands were detected using ECL^{TM} Western Blotting Detection Reagents (Amersham).

Chromatin immunoprecipitation (ChIP) experiments

Cells at 80% confluency (~5x10^6 cells per ChIP) were cross-linked with 1% formaldehyde for 10 minutes at 37^{o}C, and quenched with 125 mM glycine. The fixed cells were washed twice with cold PBS, scraped, and transferred into 1 ml PBS containing protease inhibitors (Roche). After centrifugation at 700 g for 4 minues at 4^{o}C, the cell pellets were resuspended in 100 μl ChIP lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl [pH 8.1] with protease inhibitors) and sonicated at 4^{o}C with Bioruptor (Diagenode) (30 seconds ON and 30 seconds OFF at highest power for 12 minutes). The sheared chromatin with an average fragment length of ~500 bp) was centrifuged at 10,000 g for 10 minutes at 4^{o}C). 100 μl of the supernatant was used for ChIP or as input. A 1:10 dilution of the solubilized chromatin in ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl 16.7 mM Tris-HCl [pH 8.1]) was incubated at 4^{o}C overnight with 4 μg/ml of specific antibodies. We used the following antibodies

for ChIP: goat anti-GFP (raised against His-tagged full-length eGFP and affinity-purified with GST-tagged full-length eGFP), goat anti-FoxA1 (ab5089) from Abcam, anti-panH3ac (06-599) from Millipore, anti-ERα (MC-20 or sc542x) and normal goat IgG (sc-2028) from Santa Cruz Biotechnologies (Santa Cruz). Immunoprecipitations were carried out by incubating with 40 µl pre-cleared Protein G Sepharose beads (Amersham Bioscience) for 1 hour at 4°C, followed by five washes for 10 minutes with 1ml of the following buffers: Buffer I: 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl [pH 8.1], 150 mM NaCl; Buffer II: 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl [pH 8.1], 500 mM NaCl; Buffer III: 0.25 M LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl [pH 8.1]; twice with TE buffer [pH 8.0]. Elution from the beads was performed twice with 100 µl ChIP elution buffer (1% SDS, 0.1 M $NaHCO_3$) at room temperature (RT) for 15 minutes. Protein-DNA complexes were de-crosslinked by heating at 65°C in 192 mM NaCl for 16 hours. DNA fragments were purified using QiaQuick PCR Purification kit (Qiagen) and eluted into 30 µl H_2O according to the manufacturer's protocol after treatment with RNase A and Proteinase K. For Re-ChIP experiments, immuno-complexes from the primary ChIP were eluted with 10 mM DTT for 30 min at 37°C and diluted 40 times with ChIP dilution buffer followed by the second ChIP.

For ChIP-qPCR assays, the fold enrichment of ChIPed DNA relative to input DNA at a given genomic site was determined by comparative C_T ($\Delta\Delta C_T$) method using StepOnePlus Real-Time PCR System and Power SYBR Green PCR Master Mix (Applied Biosystems) according to the manufacturer's protocol. An *ACTB* exonic region or *18S*

rRNA genomic region was used for normalization. All primer sequences used for real-time PCR are described in Table S9.

For ChIP-chip, both ChIPed DNA and input DNA were subjected to linker-mediated PCR amplification, i.e. treated with polynucleotide kinase (New England Biolabs), blunt-ended using Klenow polymerase (New England Biolabs), ligated to linkers (oligonucleotides 5'-AGAAGCTTGAATTCGAGCAGTCAG-3' annealed to 5'-CTGCTCGAATTCAAGCTTCT-3') using T4 DNA ligase (New England Biolabs) and amplified by PCR with the following conditions: initial denaturation at 94 ^{o}C for 5 minutes, 32 cycles of 94^{o}C for 40 seconds, 55^{o}C for 30 seconds, and 72 ^{o}C for 75 seconds, followed by a final elongation step of 72^{o}C for 10 minutes. A dNTP mixture containing dUTP (dTTP:dUTP = 4:1) was used in the above linker-mediated PCR amplification. ChIPed and Input DNA samples were further fragmented and end-labeled with biotin using the GeneChip® WT Double-Stranded DNA Terminal Labeling Kit (Affymetrix). The resulting labeled samples were hybridized to Affymetrix GeneChip® Human Tiling 2.0R Array Set following the Affymetrix® Chromatin Immunoprecipitation Assay Protocol. Independent biological triplicates were performed for each transcription factor, as well as the control (input DNA).

Analysis of ChIP-chip tiling array data

ChIP-chip tiling array data were normalized and analyzed with Affymetrix Tiling Analysis Software (TAS) as previously described (Bernstein et al., 2005; Cawley et al., 2004). Briefly, raw tiling array data were quantile-normalized within ChIP and control groups(Bolstad et al., 2003) and scaled to a median intensity of 500. The median intensity for the cloned BAC region was further normalized to the median intensity of the entire

chromosome on which the BAC region is located. For each genomic position of a 25-mer probe on the array, a local dataset composed of intensities for all adjacent probes within a window of ±250 bp was generated. A one-tailed Wilcoxon rank sum test was then applied to test the null hypothesis that intensities of ChIP and control experiments within the local dataset are from the same distribution, against the alternative hypothesis that the distribution of the ChIP data is positively shifted when compared to the distribution of the control data. This test procedure was performed in a sliding window across all tiled genomic regions. Significantly enriched probes, defined by applying a P-value cut-off of 1e-4, were locally extended by merging adjacent enriched probes within 100 bp and these merged regions (with a minimum length cutoff = 100 bp) were defined as transcription factor bound regions. Alternatively, we applied a P-value cut-off of 1e-3 to identify lower confidence binding regions of the mapped transcription factors, whose co-localization is displayed in the Venn diagrams of Figure S10C and S10D, S15A-S15C.

The preliminary binding sites were then subjected to three steps of filtering. First, the binding sites were checked for overlapping with repetitive elements identified by RepeatMasker (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/bigZips/chroOut.zip). ChIP regions with more than 70% overlapping with RepeatMasker repeats were filtered out. Secondly, ChIP regions were filtered out if more than 70% of the regions were annotated as tandem repeats identified by Tandem Repeat Finder (Benson, 1999). The ChIP regions were further controlled for non-specific chromatin binding of the eGFP antibody. The control dataset was generated by ChIP-chip experiments in wild-type or non-BAC transgenic MCF-7 cells using the eGFP antibody.

When comparing significant binding regions identified by ChIP-chip for different transcription factors, the regions were considered as common binding regions only if the centers of two regions were within 1 kb.

Analysis of microarray gene expression profiling experiments

Probe intensities were extracted using Agilent Feature Extraction software (version 9.5.3) (Agilent Technologies). To identify differentially expressed genes between the agonist treatment and control time course, the raw intensities were log2-transformed and quantile normalized. The software package LIMMA (Smyth, 2004) (Linear Models for Microarray Data) was applied to detect significantly differentially expressed probes using a Benjamini-Hochberg adjusted *P*-value cutoff 0.0005.

Conservation analysis of transcription factor binding regions.

Transcription factor binding regions and IR3 and DR5 regions in these regions were aligned at their centers and were extended 3 kb to each side. Base-by-base phastCons conservation scores (Siepel et al., 2005) based on the multiple alignment of 5 vertebrate species (human, rat, mouse, chicken, and *Fugu*) were retrieved (http://compgen.bscb.cornell.edu/~acs/conservation/) and the mean of phastCons scores at each position was calculated. 6 kb genomic regions with the distance of 10 kb to the center of RAR or ERα binding regions were used to estimate the conservation of the local genomic background (Figure 4C). For comparing site conservation in different categories (e.g. ERα and RAR common sites versus RAR unique sites), site conservation was represented by the mean of phastCons scores for the center 400 bases for each

binding site and Wilcoxon rank-sum test was then performed.

Enrichment analysis of transcription factor binding motifs in ChIP regions.

To test the enrichment of predicted transcription factor binding motifs in RAR or FoxA1 binding regions, we obtained position weight matrices for transcription factor binding motifs from the JASPAR database (Sandelin et al., 2004). PATSER (Hertz and Stormo, 1999) was applied to scan repeat masked human genomic sequence for matches to these weight matrices. Fold enrichment and significance were estimated by comparing the number of putative binding motifs within ChIP regions (unified length = 1 kb) with the number of predicted motifs within the same number of randomly selected 1 kb genomic regions (100,000 randomized sampling runs were performed).

Analysis of functional module enrichment in RA-regulated genes

A total of 5,463 functional modules from GSEA Molecular Signatures Database (MSigDB version 2.5) (http://www.broad.mit.edu/gsea/msigdb/) or gene signatures that have been previously associated with breast cancer (Table S7) were obtained for testing the enrichment of 1,413 RA regulated genes in MCF-7 cells. P values for the enrichment testing were calculated based on the hypergeometric distribution and further corrected for multiple testing as previously described (Rhodes et al., 2005). The module map was visualized using Cytoscape (Shannon et al., 2003).

Analysis of RAR binding and RA expression profiling on breast tumor samples

An RA signature score for each tumor sample was defined as Spearman's rank correlation

between the RA-dependent gene expression profile in MCF-7 cells (Log2 transformed fold changes in gene expression after RA agonist treatment [100 nM AM580 and 100 nM CD437 for 72 hours] relative to vehicle control/DMSO treatment were used here) and the gene expression profile in the given tumor sample for all putative RAR direct targets (defined as the genes containing at least one RARα or RARγ binding site within 50 kb to the TSSs). Associations between RA signature scores and other clinical and pathologic variables were examined using Chi-Square contingency test with the JMP7 software package (SAS Institute Inc.). The patient sample data were grouped into three categories based on their RA signature scores: P (positive RA score), N (negative RA score), and U (uncorrelated). The optimal two cutting-points of RA signature scores were determined by X-tile analysis (Camp et al., 2004). Kaplan-Meier survival curves for patient overall survival or relapse-free survival were plotted and the Log-Rank test was performed with WinStat software (Kalmia Inc.).

II. Supplemental Figure Legends

Figure S1. Localization and protein expression levels of endogenous and tagged RARA

(A-D) Localization of endogenous and LAP-tagged RARα. MCF7 cells expressing LAP-tagged RARα were fixed and immunostained with an antibody against eGFP (**A**), RARα (**B**) and DAPI (**D**). Note the marked co-localization of eGFP and RARα staining (**C**). Bar is 10 μm.

(**E**) Western Blot of wild-type and RARα-LAP transgenic MCF-7 cell lysates with an antibody against RARα shows the presence of endogenous RARα (predicted molecular weight of 51 kD) and LAP-tagged RARα (predicted molecular weight of 81 kD). Note the similar expression

levels of endogenous and LAP-tagged RARα in the transgenic MCF-7 cells. β-Actin was used as loading control.

Figure S2. Genomic distribution of RARγ, RARα, FoxA1 and GATA3 binding sites.

Distribution of 3,916 RARγ binding sites (**A**), 7,346 RARα sites (**B**), 8,061 FoxA1 sites (**C**), and 972 GATA3 sites (**D**) identified in MCF-7 cells relative to known genes. Within annotated genes, binding sites were classified as: within 5' untranslated regions (5' UTR), within coding sequences (CDS), within most TSS-proximal intron or the first intron, within other introns, and within 3' untranslated regions (3' UTR). Binding in intergenic regions was further classified based on distance to the nearest annotated gene (0-10 kb, 10-50 kb, and more than 50 kb).

Figure S3. Validation of binding regions detected by ChIP-chip with ChIP-qPCR

(**A**-**C**) The relative enrichment to genomic input DNA for 20 randomly selected RARγ binding regions within 50 kb to TSSs of known genes (**A**), or > 50 kb to TSSs of known genes (**B**), and 8 genomic regions that were not found to bind RARγ (**C**) was determined by ChIP-qPCR. Significant enrichment was observed for all 40 tested RARγ binding regions ($P < 0.01$) but not for any of 8 control regions. 18S rRNA genomic region was used as the endogenous control. Error bars represent s.d.

Figure S4. Pair-wise comparison of expression profiles obtained upon treatment with different RA agonists.

Pair-wise comparisons of gene expression profiles for all tested ligand combinations. MCF-7 cells were treated with 100 nM AM580 (RARα-specific), 100 nM CD437 (RARγ-specific), 100

nM all-*trans* retinoic acid (ATRA), or a combination of 100 nM AM580 and 100 nM CD437 (AM & CD) for 72 hours. X- and Y-axes show Log2 transformed fold changes in gene expression after RA agonist treatment relative to vehicle control (DMSO) treatment. r^2 represents the square of the linear correlation coefficient between two profiles.

Figure S5. RA agonist-mediated transcript regulation does not require *de novo* protein synthesis. MCF-7 cells grown in medium with charcoal-stripped FBS were pre-treated for one hour with 2.5 μg/ml cycloheximide were treated with vehicle (V) or 100 nM AM580/CD437 (RA) for 24 hours in the presence of 2.5 μg/ml cycloheximide. Relative expression levels of genes upregulated by RA agonists, *SOX9* (**A**), *FOXA1* (**B**), *DHRS3* (**C**), or downregulated by RA agonsists *CAV1* (**D**), *CAV2* (**E**), *FHL2* (**F**) were determined by qRT-PCR. Error bars represent s.d.

Figure S6. Effects of RA agonist, E2 and RA agonist/E2 co-treatment on the expression of putative direct target genes.

MCF-7 cells grown in medium with charcoal-stripped FBS were with vehicle (V), 10 nM estrogen (E2), 100 nM RA agonists AM580 and CD437 (RA), or a mixture of E2 and RA (E2+RA) for 24 hours. Relative expression levels of three putative RAR and ERα common target genes (**A**), unique ERα target genes (**B**), and unique RAR target genes (**C**) were determined by qRT-PCR. Error bars represent s.d.

Figure S7. Ratios of normalized ChIP versus input signal intensities for 18 ERα/RAR common binding regions.

Ratios were calculated from three replicates. Coordinates refer to UCSC hg16.

Figure S8. RA agonist/E2 responsiveness of putative regulatory elements in reporter gene assays (**A**-**D**) Binding regions for CYP26A1_1 (**A**), CYP26A1_2 (**B**), BTG2_2 (**C**) and GREB1 (**D**) were cloned into the firefly luciferase vector pGL4.23, which was co-transfected into MCF-7 cells with the Renilla luciferase vector pGL4.73 used to correct for transfection efficiency. Cells grown in medium with charcoal-stripped FBS were treated with vehicle, 10 nM estrogen (E2), 100 nM RA agonists AM580 and CD437 (RA), or a mixture of E2 and RA (E2+RA). Error bars represent s.d. In addition, we tested three additional ERα/RAR common regions for *BCL2*, *CCND1* and FOS_1 (Table S9), and found that these regions were not responsive to RA agonists (data not shown).

Figure S9. IR3, DR5, Forkhead, GATA and FOS motif enrichment in RAR and ERα binding regions

(**A**) Comparison of fold enrichment of the RARE DR5 in RAR unique regions with RAR and ERα common regions.

(**B**) Comparison of fold enrichment of the estrogen response element (IR3) in ERα unique regions with RAR and ERα common regions.

(**C**) Enrichment of Forkhead, GATA, and Fos binding motifs in RAR and ERα common regions and unique regions.

Figure S10. FoxA1 and GATA3 binding coincides with each other and with ERα and RAR binding.

(**A**,**B**) Venn diagram of FoxA1, GATA3, and RARs (union of RARγ and RARα) binding regions (**A**) or FoxA1, GATA3, and ERα, and RARs regions (**B**) using the high-stringency cutoff (P≤ 1e-4).

(**C**,**D**) Venn diagram of FoxA1, GATA3, and RARs (union of RARγ and RARα) binding regions (**C**) or FoxA1, GATA3, and ERα, and RARs regions (**D**) using the low-stringency cutoff (P≤ 1e-3).

Figure S11. Effects of FoxA1 and GATA3 knockdown on RA agonist regulated gene expression. (**A**,**B**) MCF-7 cells grown in medium with charcoal-stripped FBS were depleted from FoxA1 or GATA3 by RNAi (NT1 siRNA pool as control) for 48 hours, and were treated with vehicle (control) or 100 nM RA agonists AM580 and CD437 (RA) for 48 hours. X-axis shows Log2 transformed fold changes in gene expression after RA treatment relative to vehicle control (DMSO) treatment in RNAi control cells (siNT1). Y-axis denotes Log2 transformed fold changes in gene expression after RA agonist treatment relative to vehicle control treatment in FoxA1 (**A**) or GATA3 (**B**) knockdown cells. The red dashed lines were fit by linear least squares.

Figure S12. RARγ, RARα, ERα, FoxA1, and GATA3 binding regions close to breast cancer relevant genes.

This figure shows the binding sites of RARγ, RARα, ERα, FoxA1, and GATA3 for a selected list of breast cancer relevant genes. These genes include *RARA* (**A**), *ESR1* (**B**), *FOXA1* (**C**), *GATA3* (**D**), *FOS* (**E**), *CCND1* (**F**), *BCL2* (**G**), *NFIL3* (**H**), *LGALS3* (**I**), *CAPG* (**J**), *CTSL2* (**K**), *CTSD* (**L**), *CA12* (**M**), *BTG2* (**O**), *LASS2* (**P**), *RNF144B* (**Q**), *FGFR2* (**R**), and *NQO1* (**S**). The binding regions identified by ChIP-chip approach are denoted by black blocks.

Figure S13. Regulation of enzymes involved in the glycolytic pathway by RA.

(**A**) The glycolytic pathway and its interconnected pentose phosphate pathway. High rates of aerobic glycolysis, commonly known as Warburg effect (Warburg, 1956), has been frequently observed in cancer cells and recent studies showed that *PKM2* is essential for tumor growth *in vivo* and demonstrated the function of *PKM2* as a key regulator of the Warburg effect in tumor cells (Christofk et al., 2008a; Christofk et al., 2008b). We found the transcription levels of enzymes involved in the glycolytic pathway are typically down-regulated by RA in MCF-7 cells. The regulated genes include *PFKP, ALDOA, GAPDH, PGK1, PGAM1, ENO1, PKM2*, and *LDHA*. RAR binding sites were observed for *PFKP*, *ALDOA*, *ENO1*, *PKM2*, and *G6PD*, which suggests a novel mechanism by which RA may mediate its anti-growth effect, through which RA suppresses the glycolytic pathway by regulating the transcript levels of glycolytic enzymes. The highest expressed isozymes in MCF-7 cells are indicated by asterisks. RARγ, RARα and ERα binding are indicated by color-coded triangles. The nodes representing genes involved in this pathway are color-coded based on the expression values (Log2 fold changes of RA treatment for 72 hours relative to the vehicle treatment). (**B-F**) RAR, FoxA1, and ERα binding relative to genes involved in the glycolysis. The binding regions identified by ChIP-chip approach for *PFKP*, *ALDOA*, *ENO1*, *PKM2*, and *G6PD* are denoted by black blocks.

Figure S14. Kaplan-Meier plots for two independent breast cancer cohorts.

Kaplan-Meier curves of relapse-free survival for two breast cancer cohorts which contain 249 (Ivshina et al., 2006) (**A**) and 286 (Wang et al., 2005) (**B**) patient samples, respectively. The breast tumor samples were classified by the RA signature score as P (positive RA score), N

(negative RA score), or U (uncorrelated). P-values were obtained from log-rank tests. The datasets were obtained from Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) with accession number GSE4922 (**A**) and GSE2034 (**B**), respectively.

Figure S15. RARs, ERα, FoxA1 and GATA3 binding overlap for low-stringency binding sites (**A**-**C**) Venn diagram of RARα, RARγ and ERα binding regions (**A**), RARs (union of RARγ and RARα), ERα and FoxA1 regions (**B**), or RARs, ERα and GATA3 (**C**) using the low-stringency cutoff (P≤1e-3).

III. Supplemental Tables (Tables S1-S6 and S9 are provided as Excel worksheets)

Table S1. RARγ, RARα, FoxA1 and GATA3 binding sites in MCF-7 cells identified by ChIP-chip using the high-stringency cutoff (P≤1e-4).

.The coordinates for the binding sites are based on the human genome Mar. 2006 (hg18) assembly or NCBI Build 36.1. The binding sites can be visualized with UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) via the custom tracks. The instruction of generating BED format files for uploading binding sites to UCSC Browser is available at http://genome.ucsc.edu/goldenPath/help/customTrack.html#BED. Raw data will be deposited with NCBI Gene Expression Omnibus (GEO) upon acceptance of the manuscript for publication.

Table S2. RARγ, RARα, FoxA1 and GATA3 binding sites in MCF-7 cells identified by ChIP-chip using the low-stringency cutoff (P≤1e-3).

Coordinates refer to the human genome Mar. 2006 (hg18) assembly or NCBI Build 36.1.

Table S3. RA-regulated genes in MCF-7 cells.

This table displays RA and RA agonist regulated gene list in MCF-7 cells (Benjamini-Hochberg adjusted P <= 0.0005). Raw data will be deposited with NCBI Gene Expression Omnibus (GEO) upon acceptance of the manuscript for publication.

Table S4. Enrichment of transcription factor binding motifs in RARγ, RARα or FoxA1 binding regions.

Enriched motifs with P-value cut-off 1e-20 are shown in the table. The motifs were sorted by the transcription factor class.

Table S5. 15 selected RAR direct target genes that may be breast cancer-relevant.

The presence of binding sites for RARγ, RARα and ERα, RA or E2-dependent regulation of gene expression, functional annotation and references to literature linking these genes to cancer are provided. Detailed physical maps of RARγ, RARα and ERα sites associated with these genes are shown in Figure S12.

Table S6. BC1000 genes annotated with RAR binding and RA-regulated expression data.

The BC1000 gene set contains 1,347 manually curated genes implicated in breast cancer (Witt et al., 2006) (available at http://www.hip.harvard.edu/research/breast_cancer/index.htm). Genes containing binding sites for RARα, RARγ, ERα, FoxA1 or GATA3 with 50 kb to TSSs are labeled with

" √ " in the corresponding column. Genes significantly up- or down-regulated by RA in MCF-7 cells (Benjamini-Hochberg adjusted P <= 0.0005) are labeled with " √ " in the corresponding column.

Table S7. Annotations of nodes in the module map in Figure 6.

Node name in Figure 6	GSEA module ID(Subramanian et al., 2005)	Brief Description	References
BC1000	-	Gene set containing 1,347 manually curated genes implicated in breast cancer (http://www.hip.harvard.edu/research/breast_cancer/index.htm)	(Witt et al., 2006)
GATA3 regulated genes	-	Gene set containing 280 GATA3 regulated genes	(Oh et al., 2006; Usary et al., 2004)
positively correlated with ER status in breast cancer	BRCA_ER_POS	Genes, whose expression is positively correlated with estrogen receptor status in breast cancer - higher expression is associated with ER-positive tumors	(van 't Veer et al., 2002)
negatively correlated with ER status in breast cancer	BRCA_ER_NEG	Genes, whose expression is negatively correlated with estrogen receptor status in breast cancer - higher expression is associated with ER-negative tumors	(van 't Veer et al., 2002)
MYB regulated genes	LEI_MYB_REGULATED_GENES	Myb-regulated genes	(Lei et al., 2004)
overexpressed in CD31+ versus CD31- cells	BOQUEST_CD31PLUS_VS_CD31MINUS_UP	Genes over-expressed ≥3-fold in freshly isolated CD31+ versus freshly isolated CD31- cells	(Boquest et al., 2005)
overexpressed in leukemia cells (T-ALL)	CHIARETTI_T_ALL	Genes over-expressed in leukemia cells	(Chiaretti et al., 2004)
UV regulated genes	UVB_NHEK3_ALL	Genes regulated by UV-B light in normal human epidermal keratinocytes	(Sesto et al., 2002)
down-regulated following hGH expression in MCF-7	GH_AUTOCRINE_DN	Genes down-regulated following stable autocrine expression of human growth hormone in mammary carcinoma cells (MCF-7)	(Xu et al., 2005)
Promoters containing AP-1 motif	TGANTCA_V$AP1_C	Genes with promoter regions [±2kb to transcription start site] containing the motif TGANTCA which matches annotation for JUN: jun oncogene	(Xie et al., 2005)
upregulated in E2F- cells	IGLESIAS_E2FMINUS_UP	Genes that are upregulated in the absence of E2F1 and E2F2	(Iglesias et al., 2004)
Ras oncogenic signature	RAS_ONCOGENIC_SIGNATURE	Genes selected from supervised analyses to discriminate cells expressing activated H-Ras oncogene	(Bild et al., 2006)

		from control cells expressing GFP	
negatively correlated with brca1 germline status in breast cancer	BRCA_BRCA1_NEG	Genes, whose expression is negatively correlated with brca1 germline status in breast cancer - higher expression is associated with BRCA1 tumors	(van 't Veer et al., 2002)
Co-expressed with cadherin 3	GNF2_CDH3	Genes co-expressed cadherin 3, type 1, P-cadherin (placental) in the GNF2 expression compendium	(Su et al., 2001; Su et al., 2004)
Promoters containing TEF motif	WGGAATGY_V$TEF1_Q6	Genes with promoter regions [±2kb to transcription start site] containing the motif WGGAATGY which matches annotation for TEAD1: TEA domain family member 1 (SV40 transcriptional enhancer factor)	(Xie et al., 2005)
cancer module 1	module_1	Genes in cancer module 1 (http://robotics.stanford.edu/~erans/cancer/)	(Segal et al., 2004)
cancer module 2	module_2	Genes in cancer module 2	(Segal et al., 2004)
cancer module 3	module_3	Genes in cancer module 3	(Segal et al., 2004)
cancer module 5	module_5	Genes in cancer module 5	(Segal et al., 2004)
cancer module 6	module_6	Genes in cancer module 6	(Segal et al., 2004)
cancer module 12	module_12	Genes in cancer module 12	(Segal et al., 2004)
cancer module 52	module_52	Genes in cancer module 52	(Segal et al., 2004)

(Note: The detailed description for GSEA modules and the gene list for those modules are available at http://www.broad.mit.edu/gsea/msigdb/.)

Table S8. Associations between positive RA signature score and clinical/pathologic variables in 295 breast tumor samples.

Variable	***P-value****
Estrogen Receptor Status (Positive)	**<0.0001**
Tumor Grade (Grade 1)	**<0.0001**
Nodal Status (No Positive Nodes)	**0.0017**
Metastasis Event (No Metastasis)	**0.0058**
Tumor Size (<=2 cm)	**<0.0001**

* P-value is based on Chi-Square contingency test.

Table S9. Primers used in this study.

IV. Supplemental References

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res *27*, 573-580.

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R.*, et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Bild, A. H., Yao, G., Chang, J. T., Wang, Q., Potti, A., Chasse, D., Joshi, M. B., Harpole, D., Lancaster, J. M., Berchuck, A.*, et al.* (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature *439*, 353-357.

Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

Boquest, A. C., Shahdadfar, A., Fronsdal, K., Sigurjonsson, O., Tunheim, S. H., Collas, P., and Brinchmann, J. E. (2005). Isolation and transcription profiling of purified uncultured human stromal stem cells: alteration of gene expression after in vitro cell culture. Mol Biol Cell *16*, 1131-1141.

Camp, R. L., Dolled-Filhart, M., and Rimm, D. L. (2004). X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. Clin Cancer Res *10*, 7252-7259.

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

Chiaretti, S., Li, X., Gentleman, R., Vitale, A., Vignetti, M., Mandelli, F., Ritz, J., and Foa, R. (2004). Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. Blood *103*, 2771-2778.

Christofk, H. R., Vander Heiden, M. G., Harris, M. H., Ramanathan, A., Gerszten, R. E., Wei, R., Fleming, M. D., Schreiber, S. L., and Cantley, L. C. (2008a). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature *452*, 230-233.

Christofk, H. R., Vander Heiden, M. G., Wu, N., Asara, J. M., and Cantley, L. C. (2008b). Pyruvate kinase M2 is a phosphotyrosine-binding protein. Nature *452*, 181-186.

Hertz, G. Z., and Stormo, G. D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics *15*, 563-577.

Iglesias, A., Murga, M., Laresgoiti, U., Skoudy, A., Bernales, I., Fullaondo, A., Moreno, B., Lloreta, J., Field, S. J., Real, F. X., and Zubiaga, A. M. (2004). Diabetes and exocrine pancreatic insufficiency in E2F1/E2F2 double-mutant mice. J Clin Invest *113*, 1398-1407.

Ivshina, A. V., George, J., Senko, O., Mow, B., Putti, T. C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H.*, et al.* (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res *66*, 10292-10301.

Lei, W., Rushton, J. J., Davis, L. M., Liu, F., and Ness, S. A. (2004). Positive and negative determinants of target gene specificity in myb transcription factors. J Biol Chem *279*, 29519-29527.
Oh, D. S., Troester, M. A., Usary, J., Hu, Z., He, X., Fan, C., Wu, J., Carey, L. A., and Perou, C. M. (2006). Estrogen-regulated genes predict survival in hormone receptor-positive breast cancers. J Clin Oncol *24*, 1656-1664.
Rhodes, D. R., Kalyana-Sundaram, S., Mahavisno, V., Barrette, T. R., Ghosh, D., and Chinnaiyan, A. M. (2005). Mining for regulatory programs in the cancer transcriptome. Nat Genet *37*, 579-583.
Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res *32*, D91-94.
Segal, E., Friedman, N., Koller, D., and Regev, A. (2004). A module map showing conditional activity of expression modules in cancer. Nat Genet *36*, 1090-1098.
Sesto, A., Navarro, M., Burslem, F., and Jorcano, J. L. (2002). Analysis of the ultraviolet B response in primary human keratinocytes using oligonucleotide microarrays. Proc Natl Acad Sci U S A *99*, 2965-2970.
Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.
Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S.*, et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.
Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.
Su, A. I., Welsh, J. B., Sapinoso, L. M., Kern, S. G., Dimitrov, P., Lapp, H., Schultz, P. G., Powell, S. M., Moskaluk, C. A., Frierson, H. F., Jr., and Hampton, G. M. (2001). Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res *61*, 7388-7393.
Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G.*, et al.* (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A *101*, 6062-6067.
Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550.
Usary, J., Llaca, V., Karaca, G., Presswala, S., Karaca, M., He, X., Langerod, A., Karesen, R., Oh, D. S., Dressler, L. G.*, et al.* (2004). Mutation of GATA3 in human breast tumors. Oncogene *23*, 7669-7678.
van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T.*, et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530-536.
Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J.*, et al.* (2005). Gene-expression

profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet *365*, 671-679.
Warburg, O. (1956). On the origin of cancer cells. Science *123*, 309-314.
Witt, A. E., Hines, L. M., Collins, N. L., Hu, Y., Gunawardane, R. N., Moreira, D., Raphael, J., Jepson, D., Koundinya, M., Rolfs, A.*, et al.* (2006). Functional proteomics approach to investigate the biological activities of cDNAs implicated in breast cancer. J Proteome Res *5*, 599-610.
Xie, X., Lu, J., Kulbokas, E. J., Golub, T. R., Mootha, V., Lindblad-Toh, K., Lander, E. S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature *434*, 338-345.
Xu, X. Q., Emerald, B. S., Goh, E. L., Kannan, N., Miller, L. D., Gluckman, P. D., Liu, E. T., and Lobie, P. E. (2005). Gene expression profiling to identify oncogenic determinants of autocrine human growth hormone in human mammary carcinoma. J Biol Chem *280*, 23987-24003.