Supplemental information for

Cellular reprogramming by the conjoint action of estrogen receptor α, FOXA1 and GATA3 to a ligand inducible growth state

Say Li Kong[#], Guoliang Li[#], Siang Lin Loh, Wing-Kin Sung, Edison T. Liu[*]

[#] These authors contributed equally to this work.

[*] Corresponding author:

Edison T. Liu

Genome Institute of Singapore

60 Biopolis, Singapore 138672

T. 65 6808 8038; F. 65 6808 9051

Email: liue@gis.a-star.edu.sg

Submitted: December 30th, 2010

Revised: June 15th, 2011

This supplemental information includes:

Supplemental Methods

Supplemental Tables I-XIII

Supplemental Figure S1-S16

SUPPLEMENTAL METHODS

Short reads mapping

The reads of 25 bp from each library were mapped independently to the reference genome hg18 using the BatMan package (Tennakoon et al, manuscript in preparation). BatMan is a Burrows-Wheeler-Transform-based (BWT) method which maps short sequences to a reference genome at a very high speed. Up to 2 bp mismatches were allowed. The non-mappable reads and reads with more than 2 "best mapping" genomic locations were removed. The best mapping means, if a read has mapping locations without mis-match (as 0 mis-match), the best mapping will be the locations without mis-match; if a read has no mapping location with 0 mis-match, the mapping locations with 1 mis-match will be the best mapping locations; and the same for 2 mis-matches, etc. Only the best unique mapped location for each mappable read was kept for further processing.

Fold change of the numbers of binding sites before and after E_2 treatment

From Supplemental Table S1, there may be concerns that the sequencing depth of the library has some effects on the number of binding sites. To minimize the effect of the sequencing depth on the fold change of the numbers of binding sites before and after E_2 treatment, we normalized the fold change of the numbers of binding sites with the sequencing depth. Using ERα library as an example, the sequencing depths are 13.6 million reads before E_2 treatment and 8 million reads after E_2 treatment. There are 1990 binding sites before E_2 treatment and 19412 binding sites after E_2 treatment. We computed the fold change of the numbers of binding sites as follows:

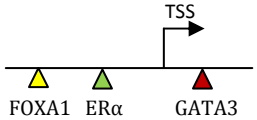
$$\text{normalized fold change of the number of binding sites} = \frac{\frac{19412}{1990}}{\frac{8}{13.6}} = 16.58$$

Assigning peaks to genes

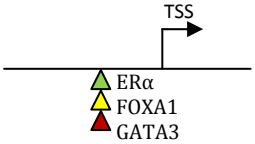
In order to assign peaks to genes, we look for the nearest TSS from the UCSC RefSeq genes for each peak. If the distance from the peak to the nearest TSS is less than 20Kb, the peak will be assigned to the TSS. In this way, each peak is assigned to at most one TSS. However, each TSS may have a few peaks assigned from the same TF. If a TSS is the nearest TSS to at least one ERα peak, we will associate this TSS to the ERα peak, regardless of the number of possible ERα peaks within the distance of 20kb. Similarly, the same TSS can be associated to FOXA1 and GATA3 peaks. Based on the condition whether the TSS is associated with ERα, FOXA1 and GATA3 peaks, the TSS will be grouped into 8 different categories: 1) with ERα+FOXA1+ GATA3 peaks; 2) with ERα+FOXA1 peaks; 3) with ERα+GATA3 peaks; 4) with FOXA1+GATA3 peaks; 5) with ERα peaks only; 6) with FOXA1 peaks only; 7) with GATA3 peaks only; and 8) without any peaks.

For the category with ERα+FOXA1+GATA3 peaks, the situation can be sub-divided into two sub-categories with/without ERα+FOXA1+GATA3 conjoint peaks as shown in the following illustrations. In the left figure, we associated the TSS with ERα+FOXA1+GATA3 non-overlapped peaks. In the right figure, we associated the TSS with overlapped (or conjoint) ERα+FOXA1+ GATA3peaks.

Non-overlapped ERα+FOXA1+ GATA3 peaks



Overlapped ERα+FOXA1+ GATA3 peaks



Analysis on the Expression Profiles of MDA-MB-231 Transfectant Cells

We have three different MDA-MB-231 transfectant cells: transfected with vector control, ERα-only, and ERα+FOXA1+GATA3. These different MDA-MB-231 transfectants were subjected to estrogen and vehicle treatment, the cells were collected at day 2 and day 10 before proceeding to microarray experiment performed on 3 biological replicates.

The gene expression was calculated in the following procedure. The expression of the probes is measured by the fold change of the raw intensities from E_2 treatment over the raw intensities from vehicle treatment, and followed by median-normalization in each replicate. The normalized expression from the probes was averaged from three replicates. The expression of the probes from the same genes was averaged to generate the final expression level of the genes. The genes with fold change over 1.2 are called as up-regulated genes, and the genes with fold change lower than 0.83 are called as down-regulated genes. The E_2-regulated genes from the vector-control MDA-MB-231 cells, ERα-expressing MDA-MB-231 cells and ERα+FOXA1+GATA3-expressing MDA-MB-231 cells are overlapped with the E_2-regulated genes from MCF-7, and this list of genes are used to compare the expression from different cells.

For the comparison of MDA-MB-231 transfectants and MCF-7, we used a growth normalized strategy. This means that the time points that are compared are selected by the time of equivalent phenotype. This strategy was used in defining *MYC* effects on apoptosis where different conditions gave the same gene responses but with different phasing (Yu Q. *et al*, 2002). In our experimental case, we found that MDA-MB-231-ERα+FOXA1+GATA3 triple transfectant exhibited E_2 stimulated growth only after day 7 and maximally by day 10. By contrast, MCF-7 showed near optimal growth induction after 24 hours of E_2 exposure.

Therefore, our array comparison was between the 24 hours time-point for MCF-7 and 10 days for MDA-MB-231- ERα+FOXA1+GATA3 (see Supplemental Table X).

We assessed the expression level of E_2-responsive genes in the MDA-MB-231-ERα+FOXA1+GATA3 normalized to the MDA-MB-231-vector control cells at 10 days and observed that specific genes previously known to be E_2-regulated in the ERα responsive cell lines (Frasor J. *et al*, 2003) such as *CCND1, STC2, ADCY9 and BTG1* were also regulated in the same direction by ligand in the triple factor transfected MDA-MB-231 cells (Supplemental Table XI).

Analysis on the Expression Levels of Luminal and Basal Marker Genes

Based on the recently published luminal and basal marker gene list from Kao *et al* (Kao *et al*, 2009), we included 45 luminal marker genes and 49 basal marker genes (referred to as the luminal/basal cassette) that present in both our MCF-7 and transfected MDA-MB-231 expression data. We generated the box plots for the expression levels of these luminal and basal marker genes in the MCF-7 and transfected MDA-MB-231 cells 2 days after estrogen stimulation (see Supplemental Figure S14A-B). Additionally, we assessed the expression changes of the luminal/basal cassette from MDA-MB-231 cells transfected with vector control to the MDA-MB-231 cells transfected with ERα+FOXA1+GATA3. The expression level of a gene from the MDA-MB-231 cells transfected with ERα+FOXA1+GATA3 was deducted with the expression level of the same gene from the MDA-MB-231 cells transfected with vector control. The average expression differences from luminal and basal marker genes were shown in the Supplemental Figure S14C.

Co-motif Analysis Using Pomoda

Peak Oriented Motif Discovery Algorithm (Pomoda) (Zhang *et al*, in preparation) is a position-weight-matrix-optimization method for de novo motif finding from ChIP-seq peaks. Pomoda takes into consideration the peak locations and intensities to give more weights to the motifs near the center of the peaks and from the peaks with high peak intensities. It utilizes all the peaks for motif finding, which is possible to identify all the potential co-factors for a transcription factor of interest.

SUPPLEMENTAL TABLES

Table I. The tabulation for the number of reads and binding sites from individual ChIP-seq libraries used in this study.

ChIP-seq libraries	Number of binding sites	Number of uniquely mapped reads	Reference
ERα-E_2	19412	8.0 millions	(Joseph *et al*, 2010)
ERα-vehicle	1990	13.6 millions	(Joseph *et al*, 2010)
FOXA1-E_2	15852	13.5 millions	(Joseph *et al*, 2010)
FOXA1-vehicle	9337	19.6 millions	(Joseph *et al*, 2010)
GATA3-E_2	38530	23.6 millions	This study
GATA3-vehicle	20707	16.8 millions	This study
RNA pol II-E_2	Not applicable	7.6 millions	(Joseph *et al*, 2010)
RNA pol II-vehicle	Not applicable	9.6 millions	(Joseph *et al*, 2010)
P300-E_2	Not applicable	12.8 millions	This study
P300-vehicle	Not applicable	16.9 millions	This study
FAIRE-E_2	Not applicable	12.6 millions	(Joseph *et al*, 2010)
FAIRE-vehicle	Not applicable	12.3 millions	(Joseph *et al*, 2010)
MCF-7 Input	Not applicable	9.3 millions	(Joseph *et al*, 2010)

Table II. The Gene Ontology analysis (Pathway) for genes with ERα+FOXA1+GATA3 bindings as compared to genes with ERα unique, FOXA1 unique or GATA3 unique binding.

Pathway	Genes with ERα+FOXA1+GATA3 overlapped peaks (p-value)	Genes with ERα unique peaks (p-value)	Genes with FOXA1 unique peaks (p-value)	Genes with GATA3 unique peaks (p-value)
PDGF signaling	1.41E-07	N.S.	N.S.	N.S.
Heterotrimeric G-protein signaling	3.92E-05	N.S.	N.S.	N.S.
Inflammation mediated by chemokine and cytokine signaling	2.81E-05	N.S.	N.S.	N.S.
Histamine H1 receptor mediated signaling	1.86E-05	N.S.	N.S.	N.S.

Note: N.S. means "not significant with p-value larger than 1E-04"

Table III. The Gene Ontology analysis (Molecular Function) for genes with ERα+FOXA1+GATA3 bindings as compared to genes with ERα unique, FOXA1 unique or GATA3 unique binding.

Molecular Function	Genes with ERα+FOXA1+GATA3 overlapped peaks (p-value)	Genes with ERα unique peaks (p-value)	Genes with FOXA1 unique peaks (p-value)	Genes with GATA3 unique peaks (p-value)
Protein binding	1.59E-15	3.90E-07	N.S	7.29E-06
Small GTPase regular activity	6.29E-15	N.S	N.S	N.S
Enzyme regulator activity	2.51E-12	N.S	N.S	1.55E-05
Catalytic activity	7.68E-09	5.30E-08	N.S	3.13E-11
Guanyl-nucleotide exchange factor activity	9.82E-09	N.S	N.S	N.S
Binding	1.09E-08	N.S	N.S	1.76E-06
Transporter activity	6.08E-08	N.S	N.S	N.S
Transmembrane transporter activity	1.23E-07	N.S	N.S	N.S
Transferase activity	7.96E-07	N.S	N.S	3.37E-07
Kinase activity	2.91E-06	N.S	N.S	8.11E-05
Structural constituent of cytoskeleton	8.41E-06	N.S	N.S	N.S
Calcium ion binding	1.71E-05	N.S	N.S	N.S

Note: N.S. means “not significant with p-value larger than 1E-04”

Table IV. The list of antibodies used in this study.

Antibody	Catalog number	Supplier
ERα	Sc-543	Santa-Cruz
FOXA1	AB-4124	Chemicon
GATA3	Sc-22205X	Santa-Cruz
RNA pol II	Ab-5408	Abcam
p300	Sc-584X	Santa-Cruz

Table V. List of primers used in the Re-ChIP experiments to investigate the co-occupancy of ERα+FOXA1 and ERα+GATA3 to the target sites.

No	Name	Sequence
1	GPR37L1	TAACCCTCCTCTTTGGCTCTG CTAGCCATGTCCTTTCTGCC
2	GREB1	GGCTCCAGTCCAAGTACACAAACTTC TTTTGCTGGGTCACAGTGCTCTCC
3	ITPK1	CTGCCTGCAATCTGTTCCATAC TCAGGTGACGCTGACTGTTG
4	LRRN6A	GTTTGCTGACCAAACACTAGGAAGT CCCACGGAAGCTTAGCTTTA
5	SLC25A25	GCTTTTCCTGTGGAGGCTTC TGGTAGGTACTCGGCAAACC
6	RAD51L1	GTAACAGAACAGGCTGTGCC CAAACAGATGCAAGACAAGG
7	PRKCBP1	GAAGGAACCAAGAGGAAGGAG CCCTGTTTACCTTTGTTTCC

Table VI. List of primers used to study the progressive recruitment of ERα and FOXA1 in synchronized MCF-7 cells.

No	Name	Sequence
1	GPR37L1	TCTCTGGCAGCTTGTGTCAG GGCCACGACAAACAGATTAT
2	KIAA0649	GTGCTTCCTAGCTTTTCTAATGCAGC ATGAGAGCAGAGCAGTTGGGAGCTTC
3	SLC25A25	TTCTGAGGTTGCTCCAGCCT TGAACCTGGTAGGTACTCGG
4	DNPEPP	ACAGGGCTGTTTACTTTCAG CCCTGGAACTTCATAGACAT
5	RPLP1	AGCTTCCTGTTCTGCTGCCGGTTATC TTTGCAGCCTAACACTGGTG

Table VII. List of primers used for the validation of FOXA1 and GATA3 binding sites by ChIP-qPCR. The coordinates (in hg18) for the binding sites represent the genomic site at the center of the peak (submitted as separate Excel file).

Table VIII. The list of E_2 up-regulated genes at 3, 6, 9, 12, 24 and 48 hours identified in MCF-7 cells with the ERα, FOXA1, GATA3, RNA Pol II bindings support information (submitted as separate Excel file).

Table IX. The list of E_2 down-regulated genes at 3, 6, 9, 12, 24 and 48 hours identified in MCF-7 cells with the ERα, FOXA1, GATA3, RNA Pol II bindings support information (submitted as separate Excel file).

Table X. The list of all the E_2-regulated genes as well as the cell cycle, cellular proliferation and DNA replication genes in the transfected MDA-MB-231 cells compare to MCF-7 cells (submitted as separate file).

Table XI. The list of previously known E_2-responsive genes identified by Frasor *et al* that regulate in the same direction in the transfected MDA-MB-231-ER+FOXA1+GATA3 cells (submitted as separate Excel file).

Table XII. The expression of luminal and basal cassette in the MCF-7 and MDA-MB-231 transfectants (submitted as separate Excel file).

Table XIII. The presence of ERα, FOXA1 and GATA3 binding support within 20kb of TSS of the luminal and basal marker genes defined by Kao *et al*. (submitted as separate Excel file).

REFERENCES

Frasor J., Danes J. M., Komm B., Chang K. C. N., Lyttle R., Katzenellenbogen BS (2003) Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. *Endocrinology* **144:** 4562-4574.

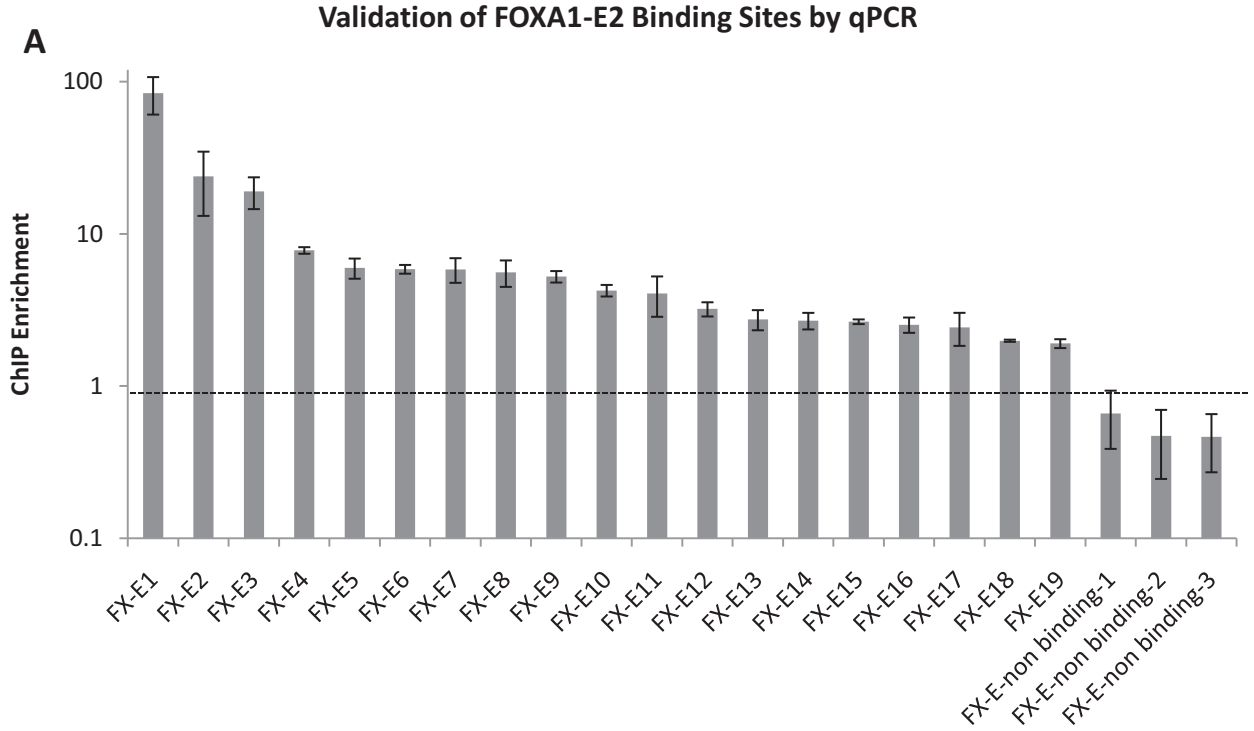
Joseph R, Orlov YL, Huss M, Sun W, Kong SL, Ukil L, Pan YF, Li G, Lim M, Thomsen JS, Ruan Y, Clarke ND, Prabhakar S, Cheung E, Liu ET (2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor α. . *Mol System Biol* **6:456**.

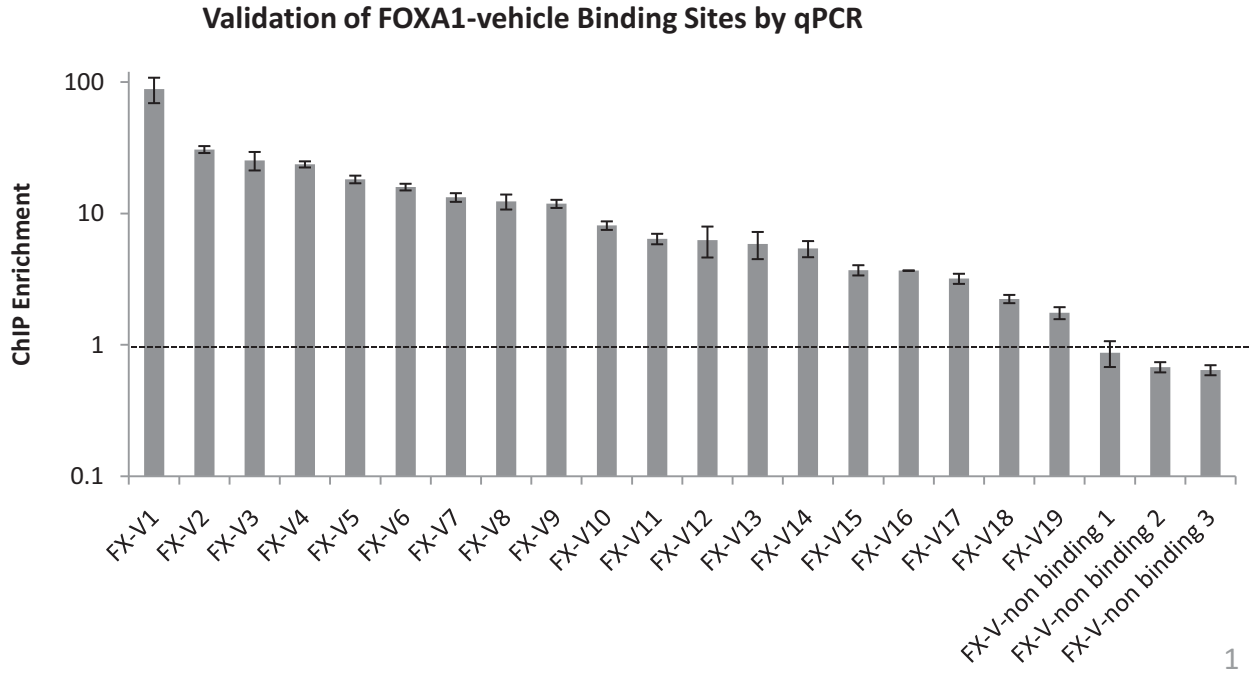
Kao J, Salari K, Bocanegra M, Choi YL, Girard L, Gandhi J, Kwei KA, Hernandez-Boussard T, Wang P, Gazdar AF, Minna JD, Pollack JR (2009) Molecular profiling of breast cancer cell lines

defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* **4:** e6146.

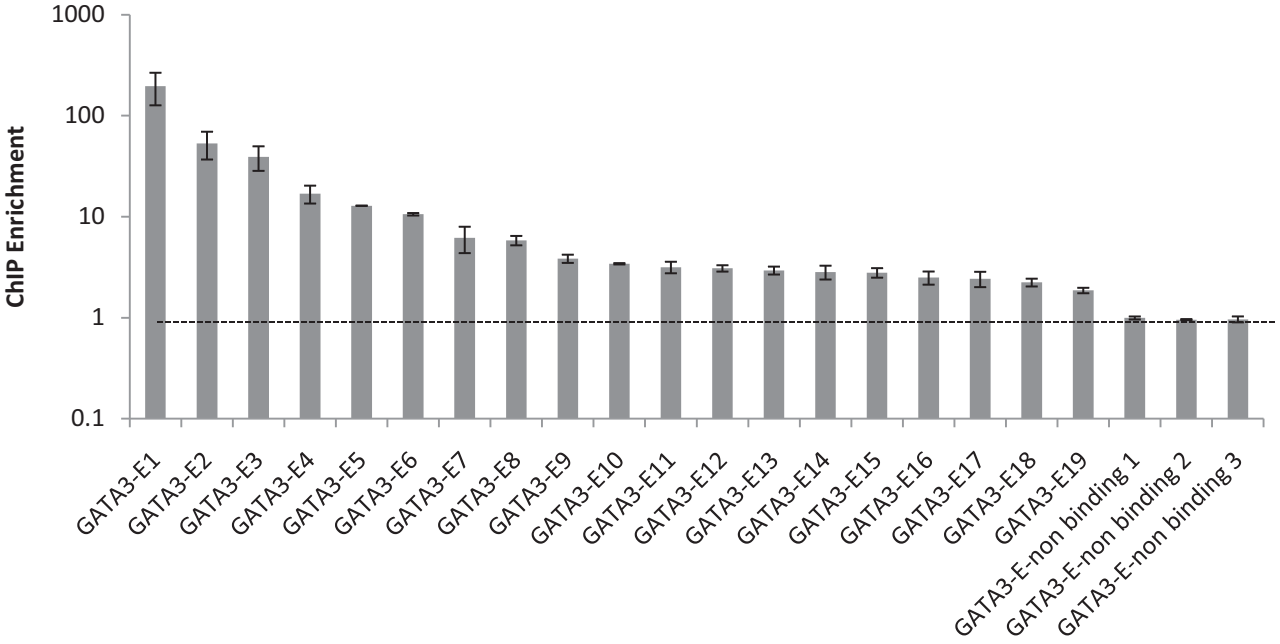
Yu Q., He M., Lee N.M., E.T. L (2002) identification of myc-mediated death response pathways by microarray analysis. *journal of biological chemistry* **277:** 13059-13066.

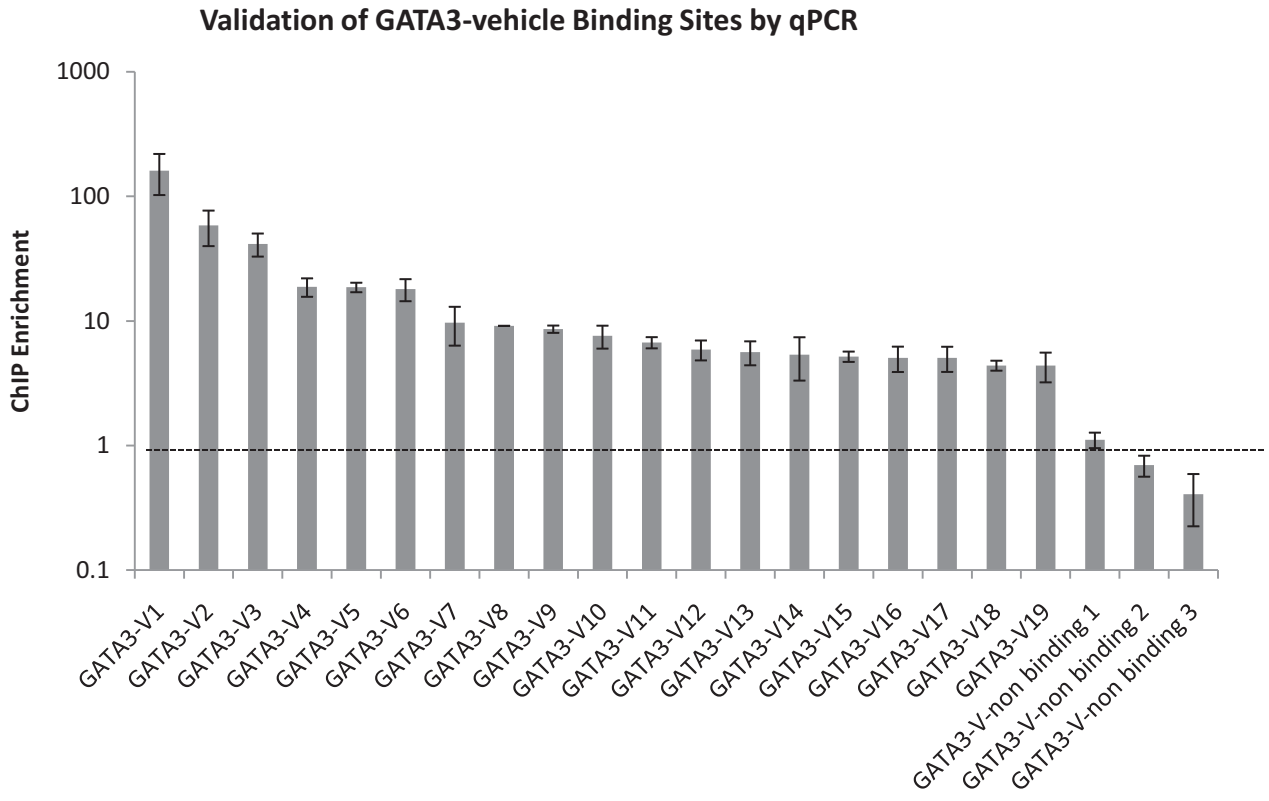
Figure S1. The validation of (A) FOXA1 and (B) GATA3 bindings on randomly selected sites with different binding intensity by ChIP-qPCR. A few non-binding sites are included as negative control. The error bars show the standard errors of the means of binding enrichments from three replicates. (C) The scatter plots on the correlation between FOXA1 and GATA3 peak intensity measured by ChIP-seq after normalization to input control and binding enrichment assayed by qPCR.



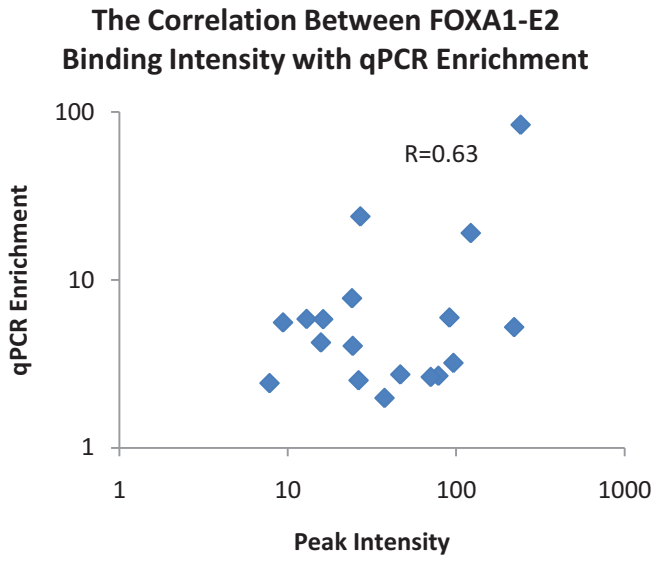


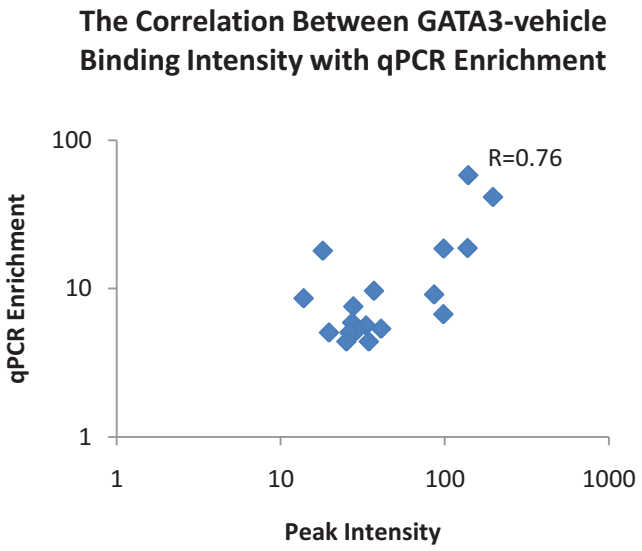
B

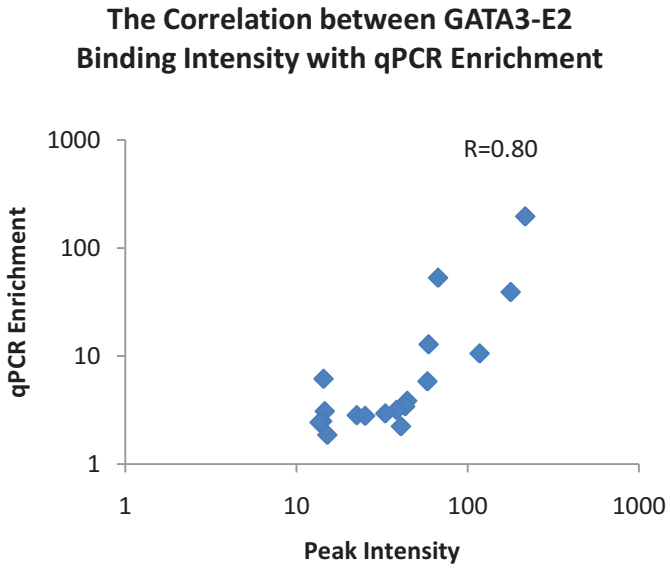

Validation of GATA3-E2 Binding Sites by qPCR
ChIP Enrichment
1000
100
10
1
0.1
GATA3-E1
GATA3-E2
GATA3-E3
GATA3-E4
GATA3-E5
GATA3-E6
GATA3-E7
GATA3-E8
GATA3-E9
GATA3-E10
GATA3-E11
GATA3-E12
GATA3-E13
GATA3-E14
GATA3-E15
GATA3-E16
GATA3-E17
GATA3-E18
GATA3-E19
GATA3-E-non binding 1
GATA3-E-non binding 2
GATA3-E-non binding 3


Validation of GATA3-vehicle Binding Sites by qPCR
ChIP Enrichment
1000
100
10
1
0.1
GATA3-V1
GATA3-V2
GATA3-V3
GATA3-V4
GATA3-V5
GATA3-V6
GATA3-V7
GATA3-V8
GATA3-V9
GATA3-V10
GATA3-V11
GATA3-V12
GATA3-V13
GATA3-V14
GATA3-V15
GATA3-V16
GATA3-V17
GATA3-V18
GATA3-V19
GATA3-V-non binding 1
GATA3-V-non binding 2
GATA3-V-non binding 3

C


The Correlation Between FOXA1-E2 Binding Intensity with qPCR Enrichment
R=0.63
qPCR Enrichment
Peak Intensity
100
10
1
1
10
100
1000


The Correlation Between GATA3-vehicle Binding Intensity with qPCR Enrichment
R=0.76
qPCR Enrichment
Peak Intensity
100
10
1
1
10
100
1000


The Correlation between GATA3-E2 Binding Intensity with qPCR Enrichment
R=0.80
qPCR Enrichment
Peak Intensity
1000
100
10
1
1
10
100
1000

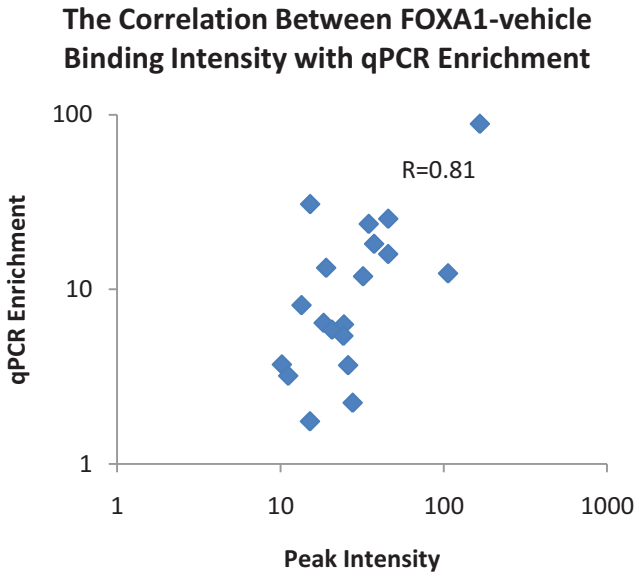

The Correlation Between FOXA1-vehicle Binding Intensity with qPCR Enrichment
R=0.81
qPCR Enrichment
Peak Intensity
100
10
1
1
10
100
1000

Figure S2. The correlation of ER peaks intensity with (A) FOXA1 and (B) GATA3 peaks intensity. The peak intensities are in log10 scale.

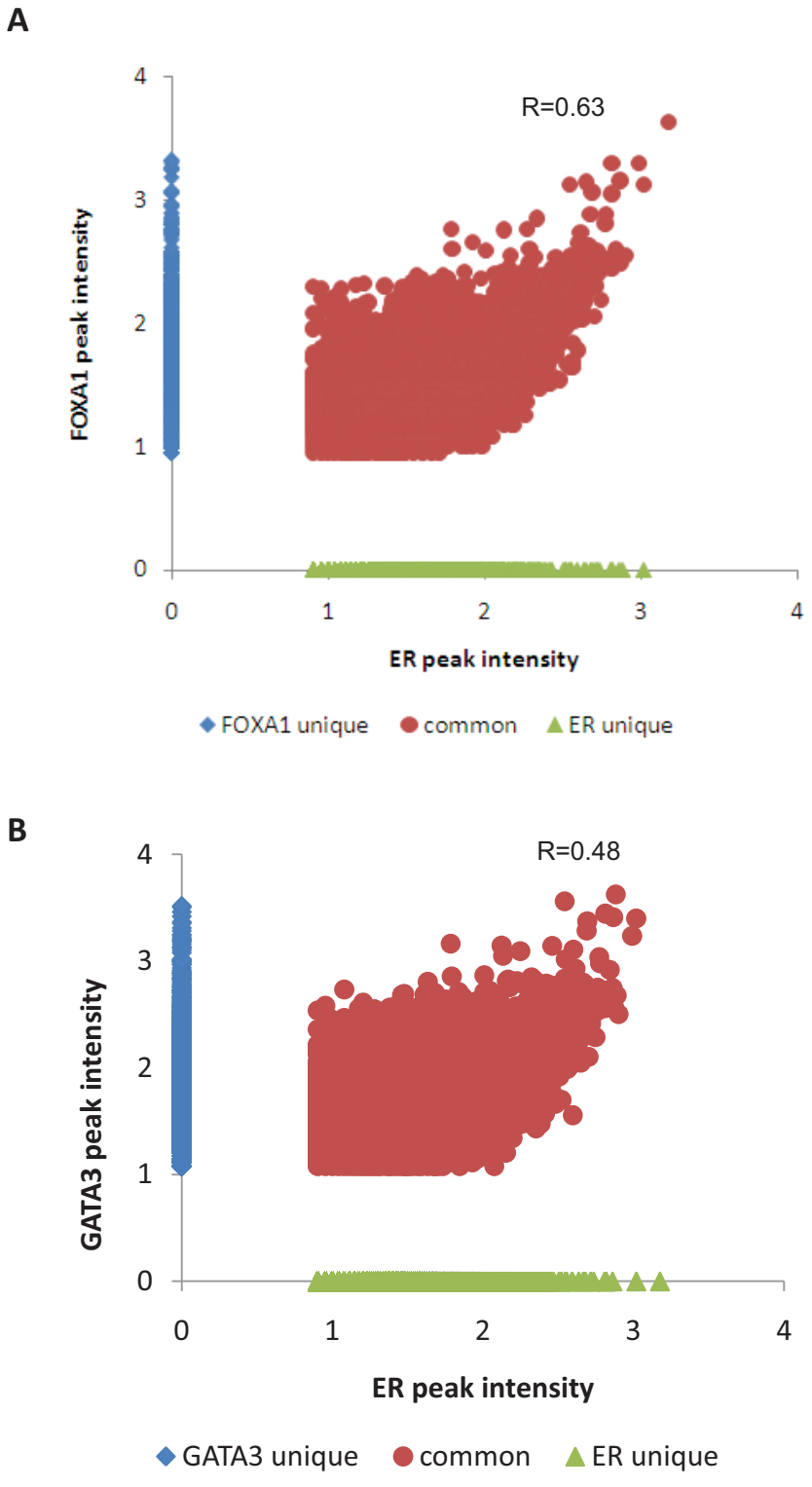
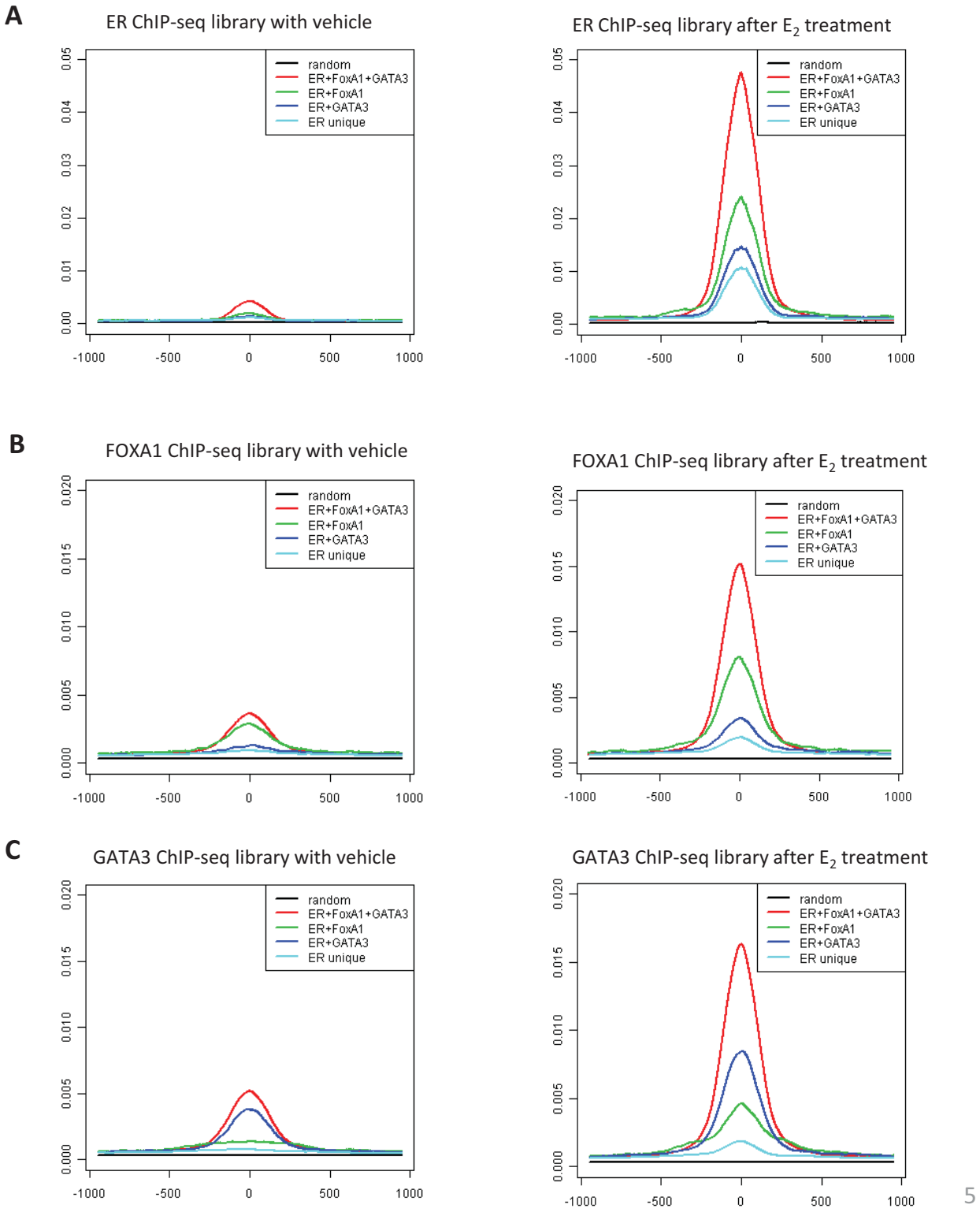
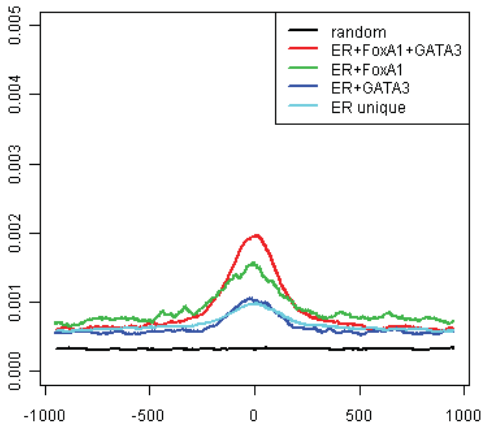


Figure S3. The tabulation of ERα unique, ERα+GATA3, ERα+FOXA1, ERα+FOXA1+GATA3 sites with the (A) ERα occupancy (B) FOXA1 occupancy, (C) GATA3 occupancy, (D) p300 recruitment, (E) RNA Pol II recruitment and (F) chromatin opening measured by FAIRE in before and after E_2 stimulation.

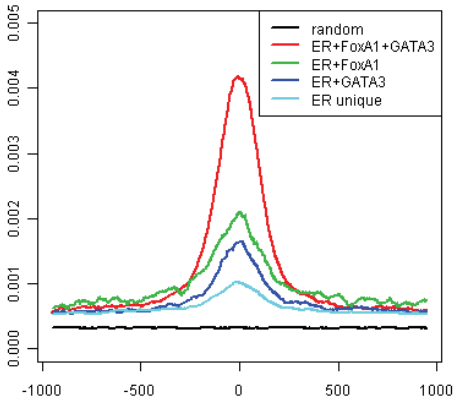


D

p300 ChIP-seq library with vehicle

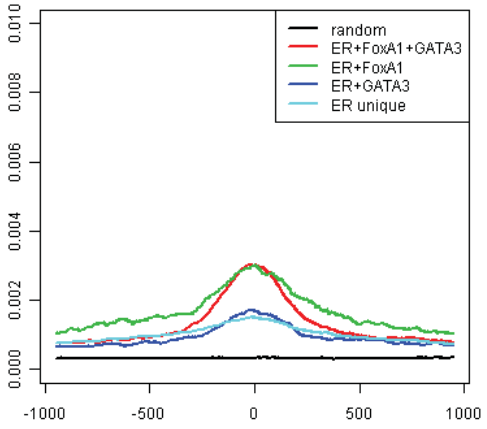


p300 ChIP-seq library after E_2 treatment

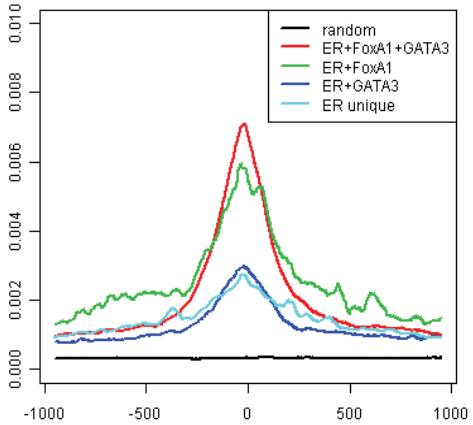


E

RNA Pol II ChIP-seq library with vehicle

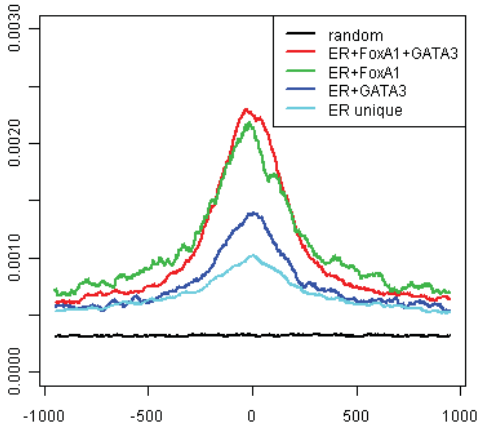


RNA Pol II ChIP-seq library after E2 treatment



F

FAIRE ChIP-seq library with vehicle



FAIRE ChIP-seq library after E_2 treatment

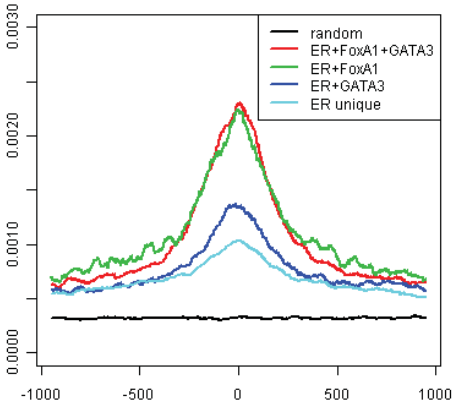


Figure S4. (A) The distribution of enhanceosome consists of ERα+FOXA1+GATA3 in the ~19k ERα binding sites in MCF-7 cells. Majority of the enhanceosomes were found in the top quartile ERα binding sites with the highest binding intensity. (B) The comparison between enhanceosome (red) vs non-enhanceosome (green) from the top quartile ERα sites with the occupancy of ERα, FOXA1 and GATA3; (C) p300 recruitment and (D) RNA pol II recruitment.

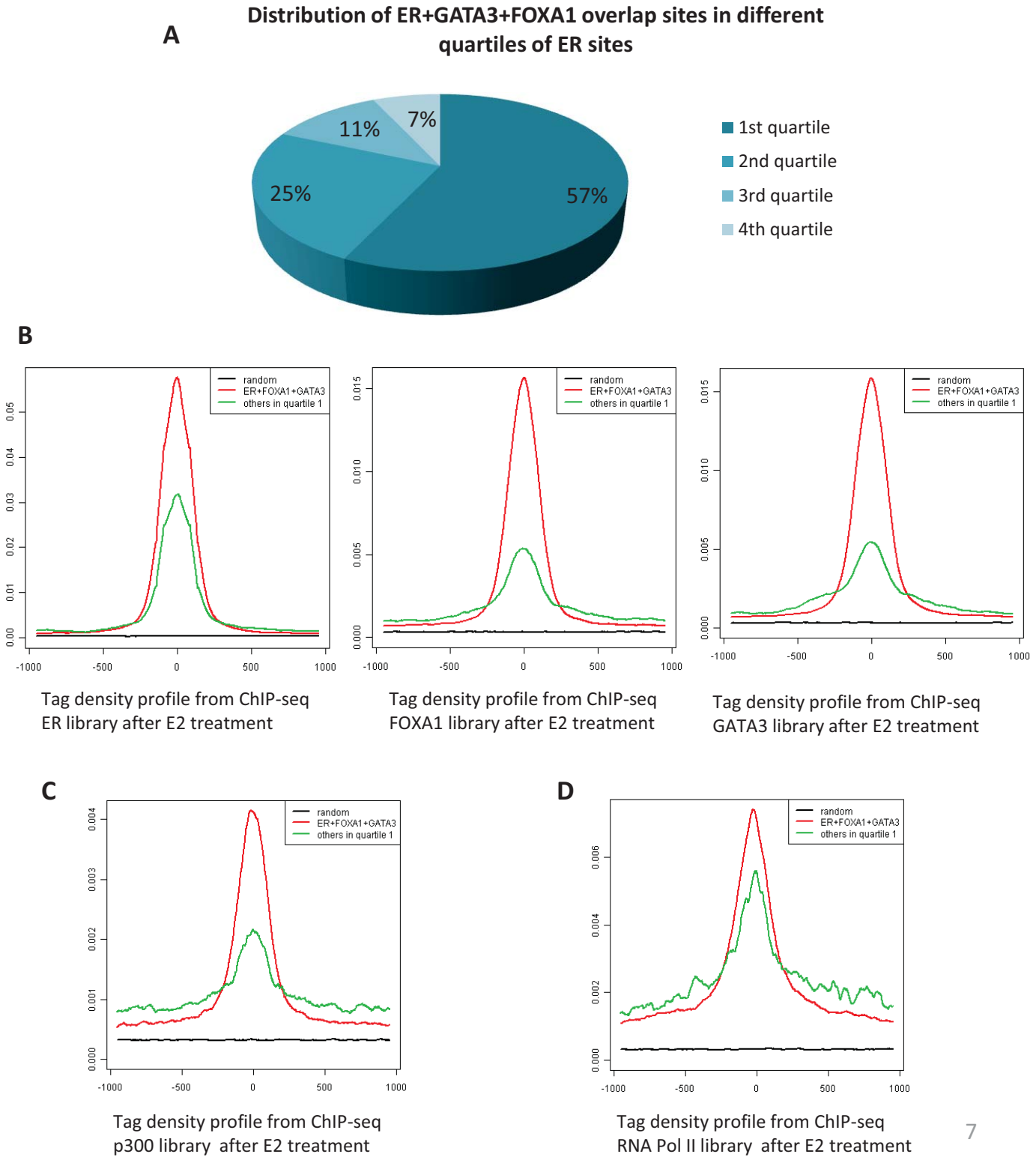
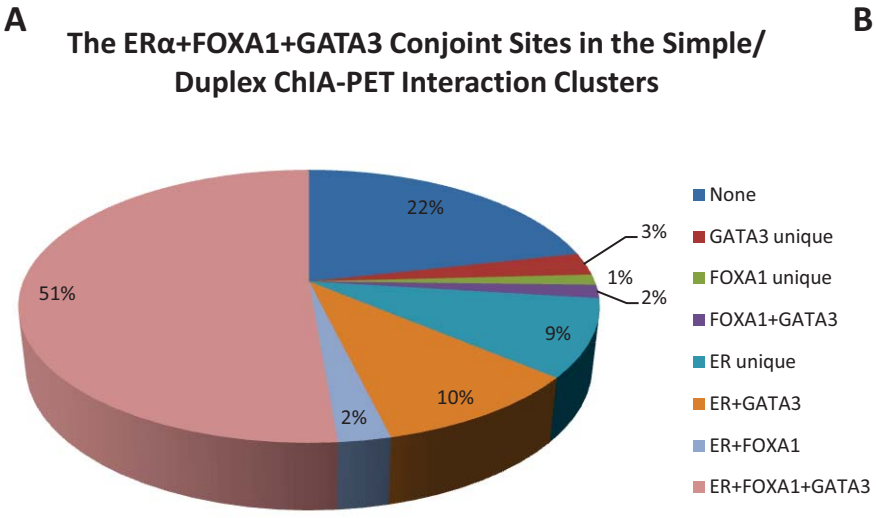
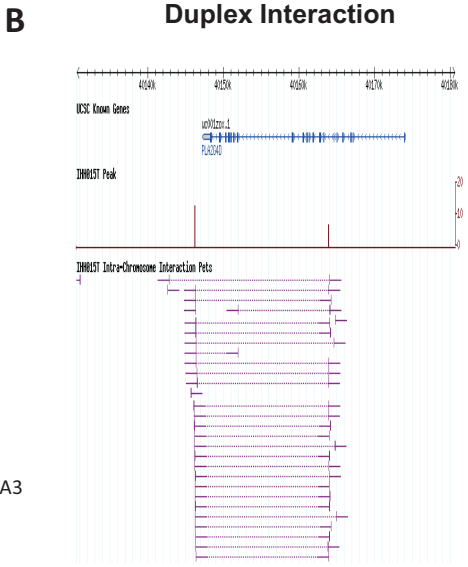
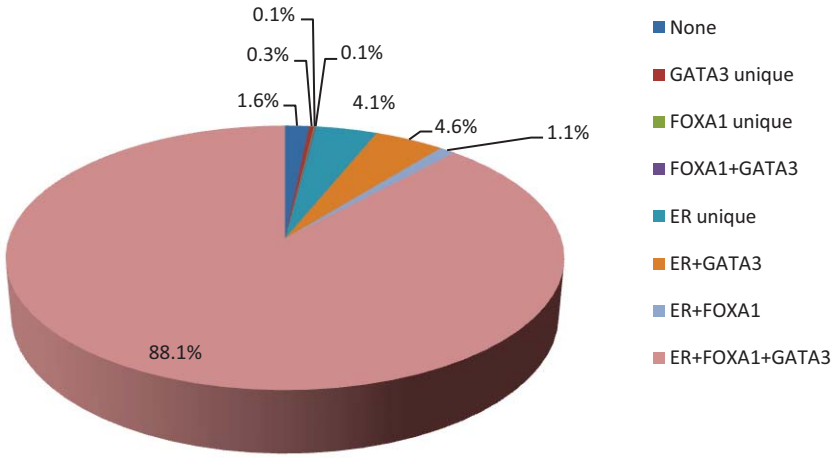
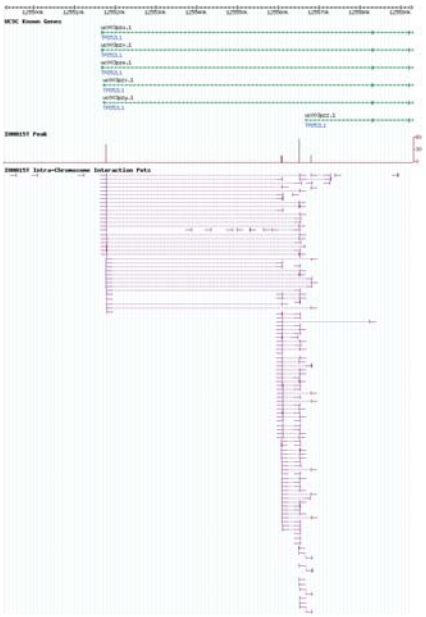


Figure S5. Majority of the ERα interactomes display ERα+FOXA1+GATA3 co-localization. (A) The distribution of simple/ duplex ChIA-PET interaction clusters with ERα+FOXA1+GATA3 conjoint bindings. (B) An example of duplex interaction is shown. (C) The complex ChIA-PET interaction clusters with ERα+FOXA1+GATA3 conjoint bindings. We found that the enhanceosome has better association with complex ERα interactome, indicating that interactome of higher complexity requires multiple TFs. (D) An example of complex interaction is shown. (E) An example of E_2-regulated gene with complex ERα interactome was shown to have co-localization of ERα, FOXA1 and GATA3 bindings.

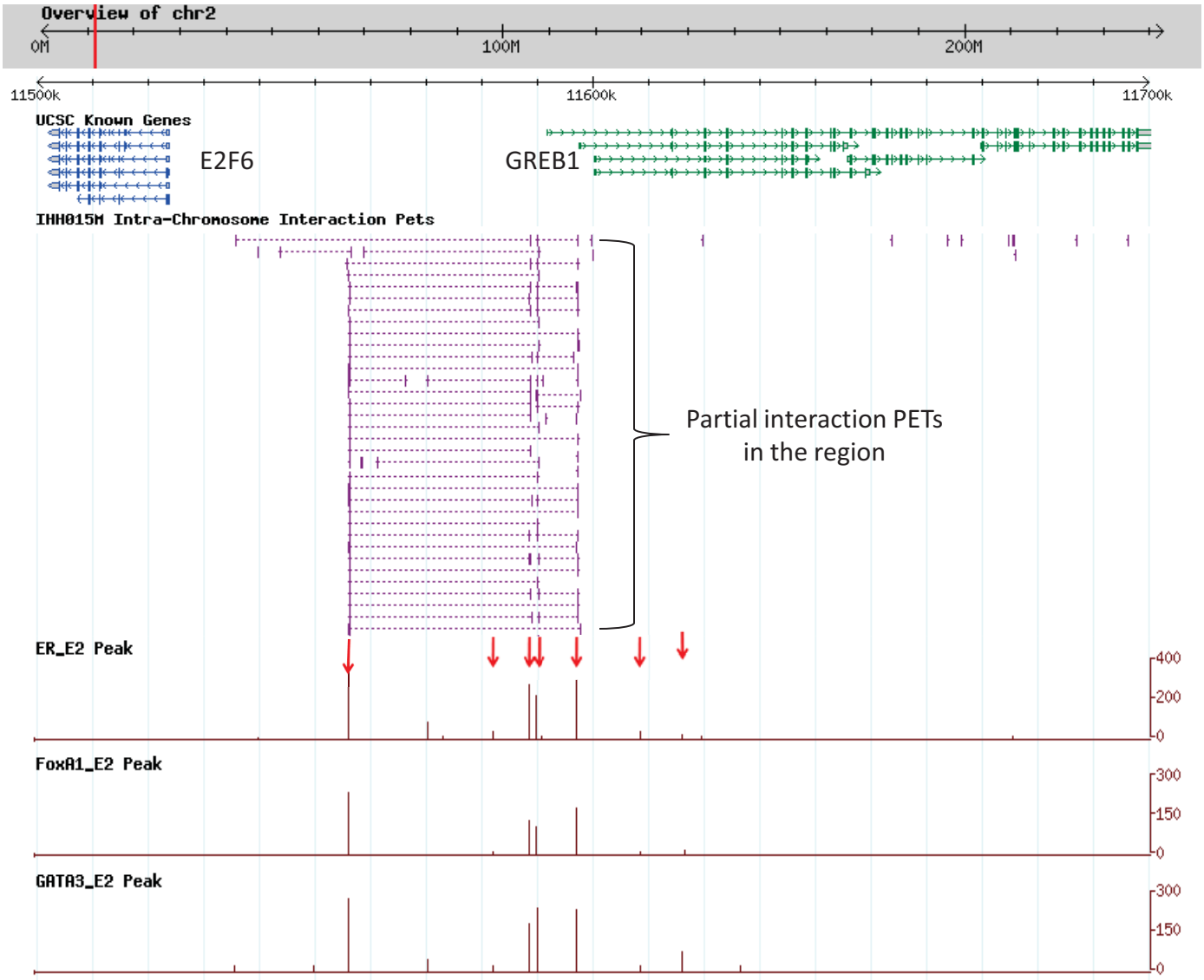








E

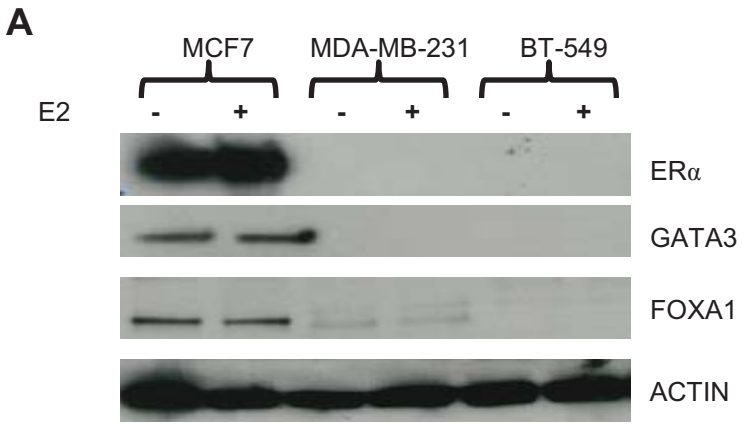


ER ChIA-PET interactions
The thickness of the lines are proportional to the pet counts between the anchors

Legend:

↓ ERα+FOXA1+GATA3 conjoint sites

Supplemental Figure S6.(A) The expression of ERα, GATA3 and FOXA1 in ERα-positive MCF-7 cells and ERα-negative MDA-MB-231 and BT-549 cells prior to transfection experiment. (B) The expression of ERα, GATA3 and FOXA1 in the transfected MDA-MB-231 and BT-549 cells with various TF combination.



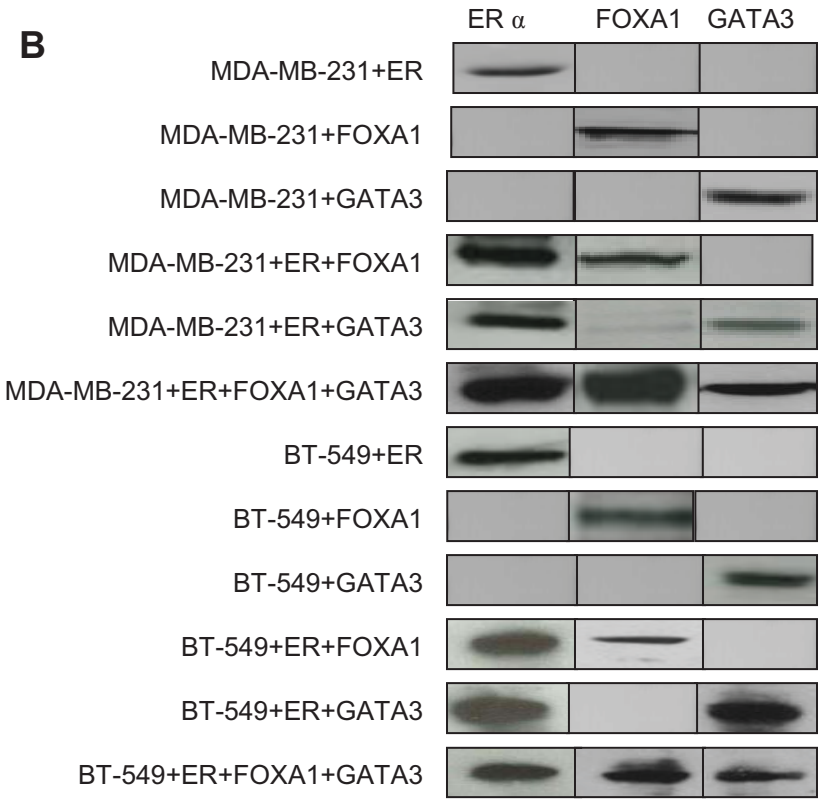


Figure S7. The proliferation of reprogrammed MDA-MB-231 cells assayed by WST-1.
(A) The subtle inhibited growth of MDA-MB-231 cells with the transfection of ERα. (B) The inhibited growth of MDA-MB-231 cells with the transfection of FOXA1. (C) The unaltered growth of MDA-MB-231 cells with the transfection of GATA3. (D) The inhibited growth of MDA-MB-231 cells with co-transfection of ERα+FOXA1. (E) The unaltered growth of MDA-MB-231 cells with co-transfection of ERα+GATA3. (F) The induced cell proliferation in response to E2 stimulation in the MDA-MB-231 cells with the co-transfection of ERα, FOXA1 and GATA3 in combination. For every sub-figures, means and standard errors of three independent experiments are shown.

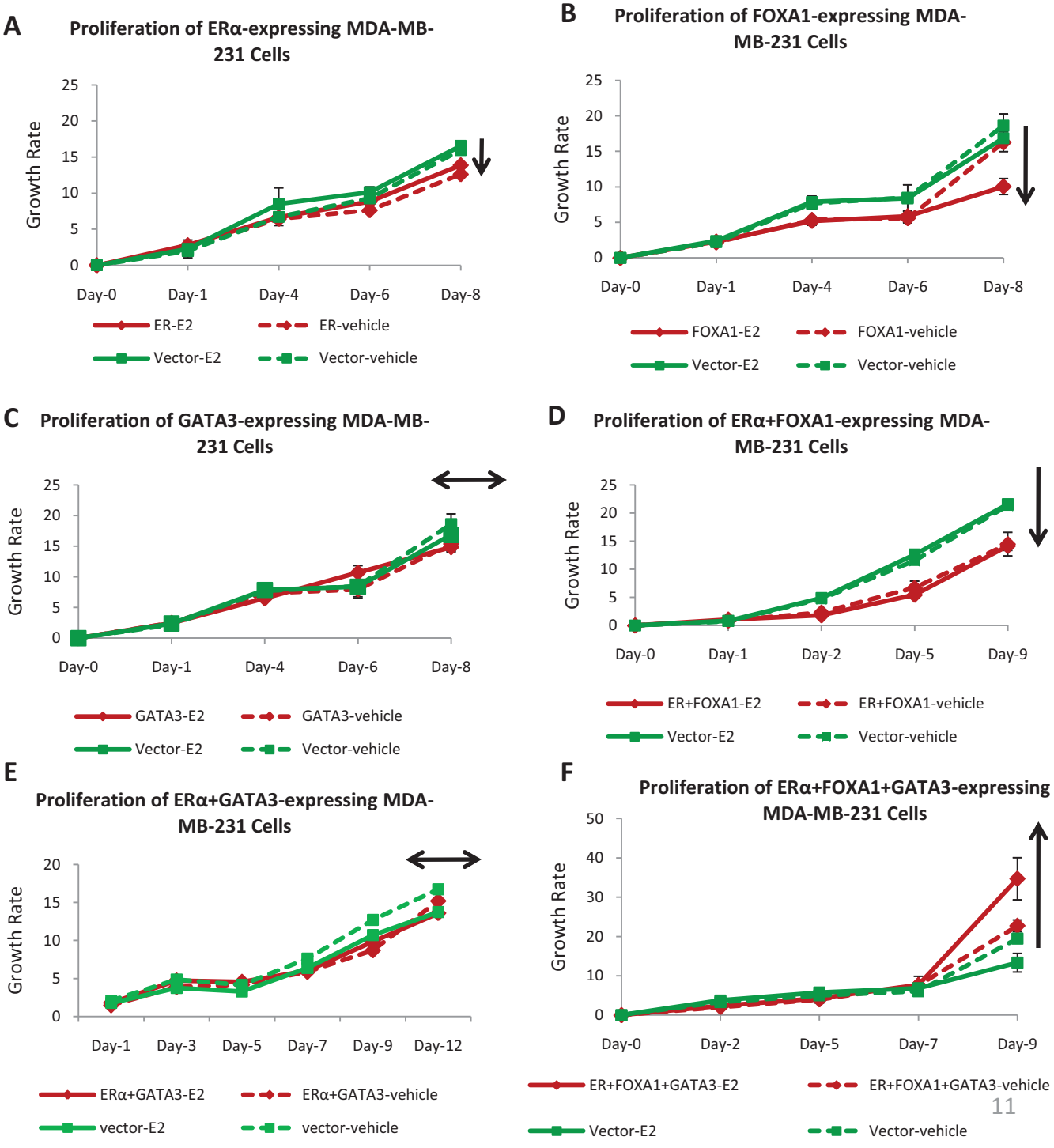


Figure S8. The proliferation of reprogrammed MDA-MB-231 cells assessed by cell number count using Hoechst stain. (A) The subtle inhibited growth of MDA-MB-231 cells with the transfection of ERα. (B) The inhibited growth of MDA-MB-231 cells with the transfection of FOXA1. (C) The unaltered growth of MDA-MB-231 cells with the transfection of GATA3. (D) The unaltered growth of MDA-MB-231 cells with co-transfection of ERα+FOXA1. (E) The inhibited growth of MDA-MB-231 cells with co-transfection of ERα+GATA3. (F) The induced cell proliferation in response to E2 stimulation in the MDA-MB-231 cells with the co-transfection of ERα, FOXA1 and GATA3 in combination. For every sub-figures, means and standard errors of three independent experiments are shown.

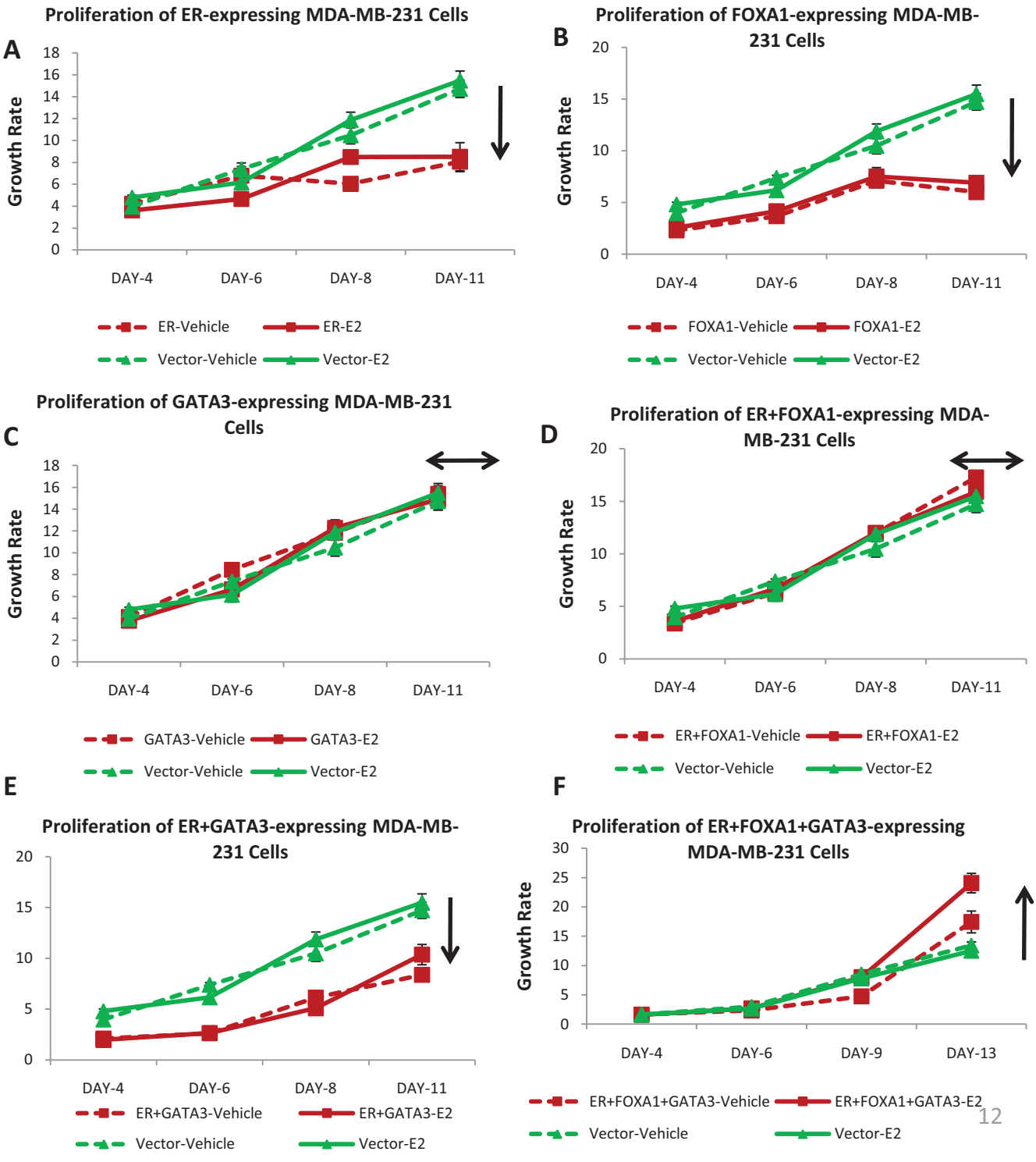


Figure S9. Recapitulating the reprogramming work in another ERα-negative BT-549 cell line. (A) The inhibited growth of BT-549 cells with the transfection of ERα. (B) The inhibited growth of BT-549 cells with the transfection of FOXA1. (C) The subtle induced growth of BT-549 cells with the transfection of GATA3. (D) The unaltered growth of BT-549 cells with co-transfection of ERα+FOXA1. (E) The unaltered growth of BT-549 cells with co-transfection of ERα+GATA3. (F) The induced cell proliferation in response to E2 stimulation in the BT-549 cells with the co-transfection of ERα, FOXA1 and GATA3 in combination. For every sub-figures, means and standard errors of three independent experiments are shown.

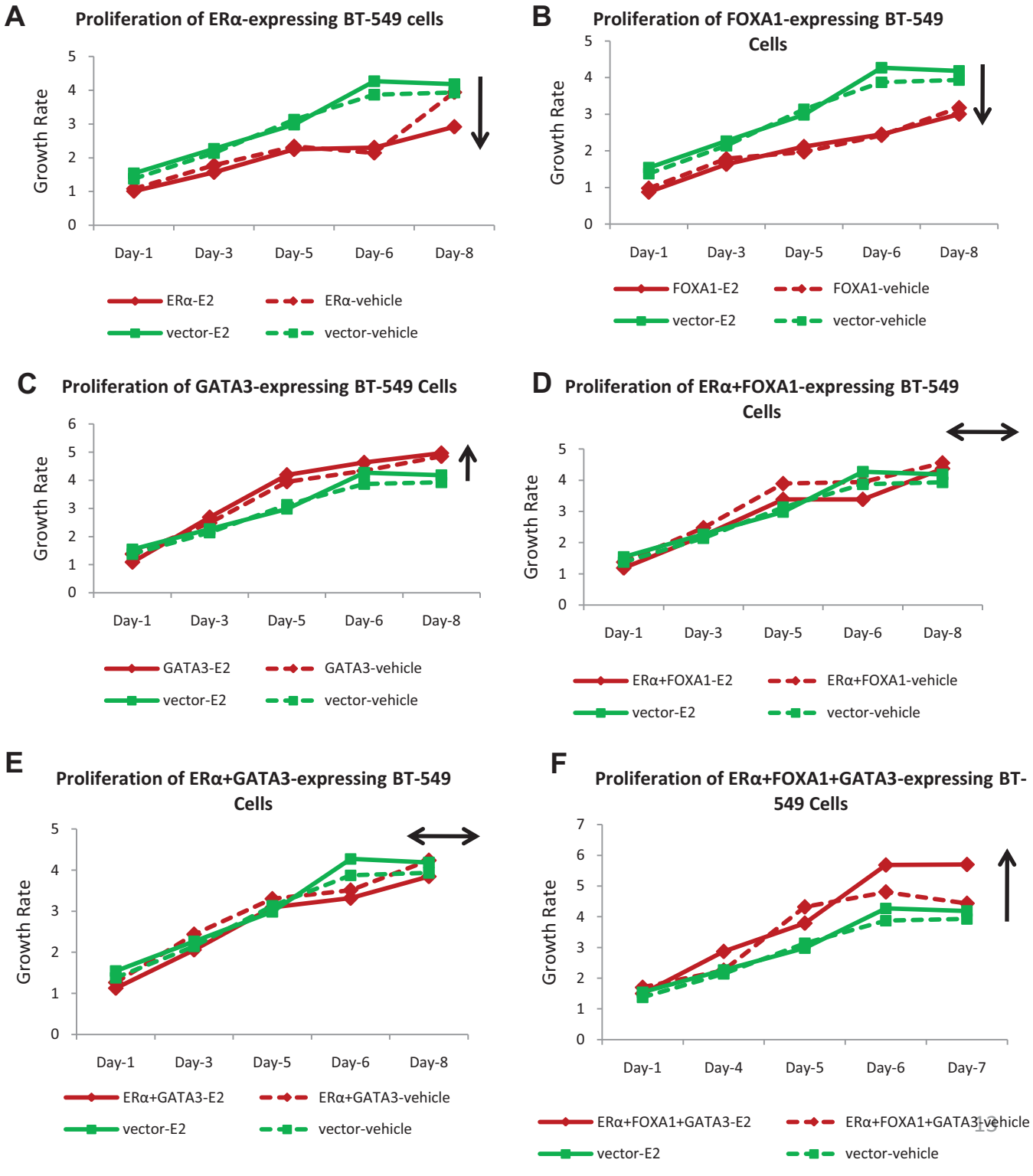


Figure S10. The proliferation of reprogrammed BT-549 cells assessed by cell number count using Hoechst stain. (A) The subtle inhibited growth of BT-549 cells with the transfection of ERα. (B) The inhibited growth of BT-549 cells with the transfection of FOXA1. (C) The marginal induced growth of BT-549 cells with the transfection of GATA3. (D) The inhibited growth of BT-549 cells with co-transfection of ERα+FOXA1. (E) The inhibited growth of BT-549 cells with co-transfection of ERα+GATA3. (F) The induced cell proliferation in response to E2 stimulation in the BT-549 cells with the co-transfection of ERα, FOXA1 and GATA3 in combination. For every sub-figures, means and standard errors of three independent experiments are shown.

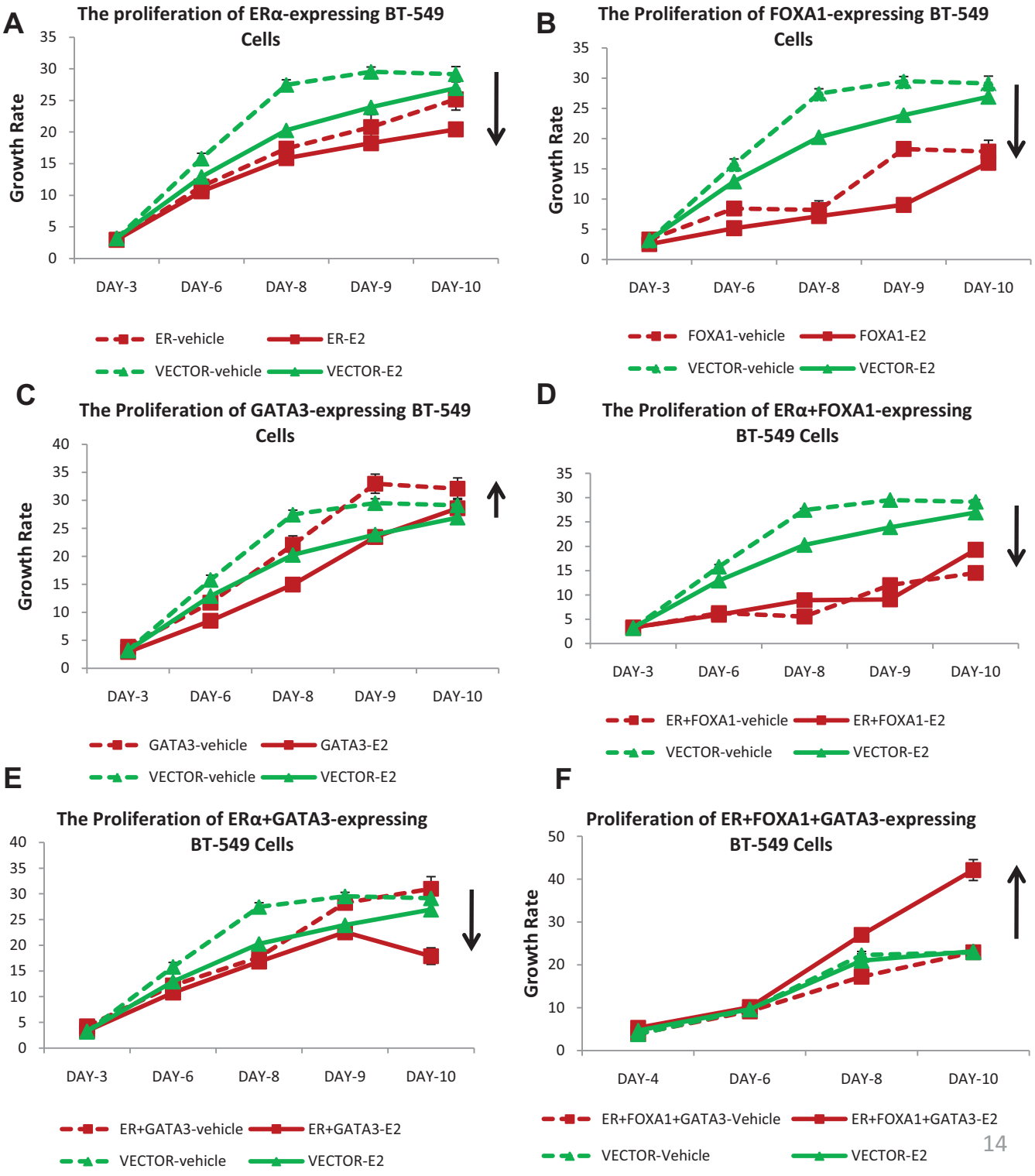
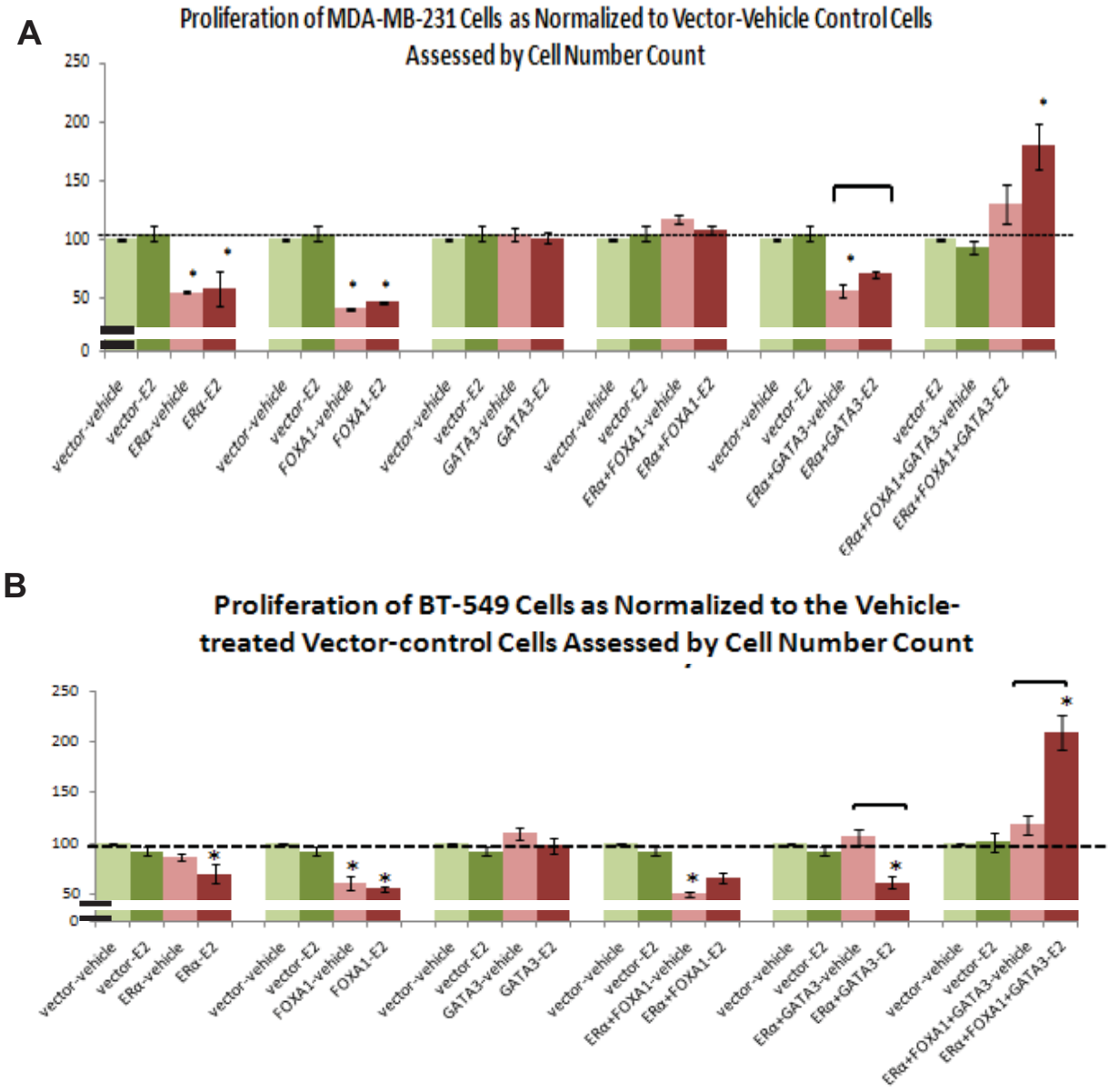


Figure S11. (A) The growth of MDA-MB-231 cells transfected with different combinations of TFs relative to the vehicle-treated MDA-MB-231 vector control cells at the final day of growth assessment by cell number count. (B) The recapitulation of reprogramming work in another ERα-negative BT-549 cells. The growth of BT-549 cells tranfected with different combinations of TFs relative to the vehicle-treated BT-549 vector control cells at the final day of growth assessment by cell number count. For every sub-figures, means and standard errors of three independent experiments are shown.



Note: * means p-value ≤ 0.05 compared to vector-vehicle computed from the Student's t Test
shows significant E2-dependent induced or repressed growth with p-value ≤ 0.05

Figure S12. The correlation of the expression profile of cell cycle, cellular proliferation and DNA replication genes between the MCF-7 cells and the MDA-MB-231 cells transfected with (A) ERα-only, and (B) ERα+FOXA1+GATA3. We observed that there was positive correlation between MDA-MB-231 cells transfected with ERα+FOXA1+GATA3 and MCF-7 cells, no correlation between the ER-only MDA-MB-231 cells and MCF-7 cells. The p-value for the differences between the two correlation coefficient stated in (A) and (B) is ≤ 0.0034.

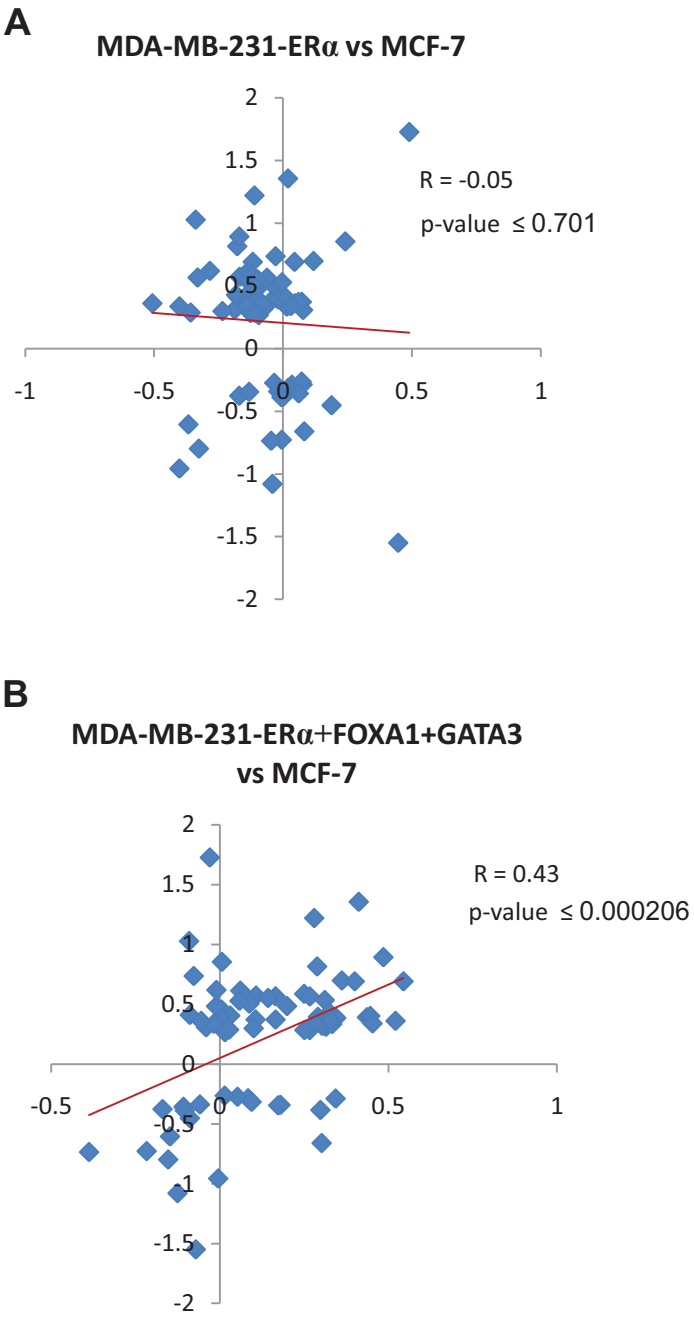
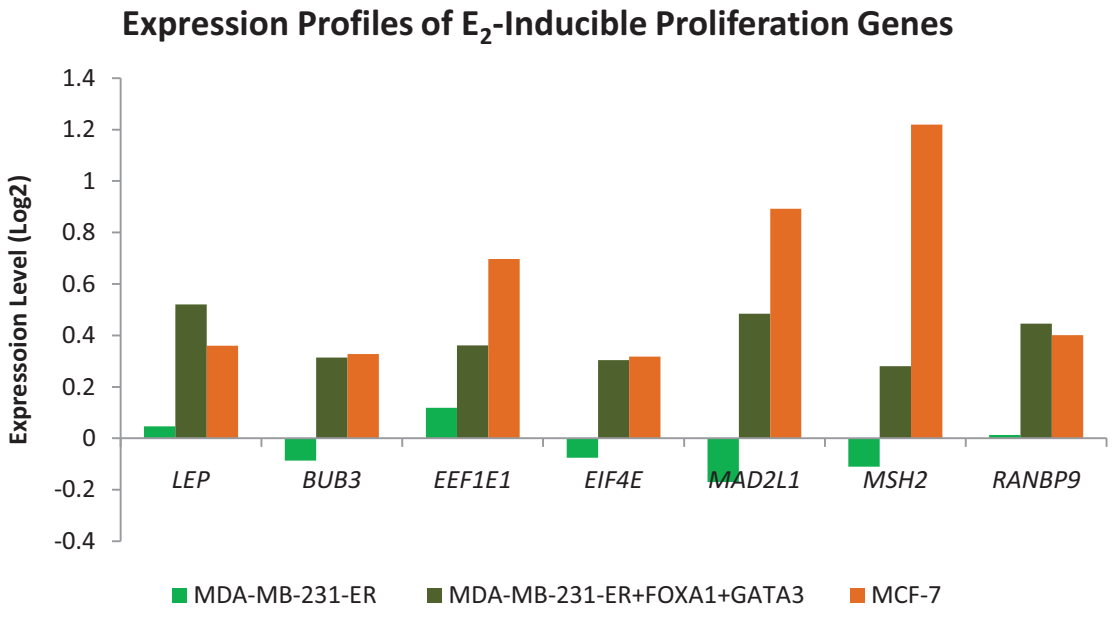


Figure S13. The Ingenuity Pathway Analysis was used to access the functionality of estrogen responsive genes identified in MDA-MB-231 transfectant cells. (A) The estrogen regulated genes identified in the transfected MDA-MB-231 sublines are significantly associated with the cell cycle, cellular proliferation and DNA replication functionalities. (B) The differential regulation of cell cycle and proliferation genes in the MDA-MB-231 transfectants and MCF-7cells in response to estrogen stimulation.

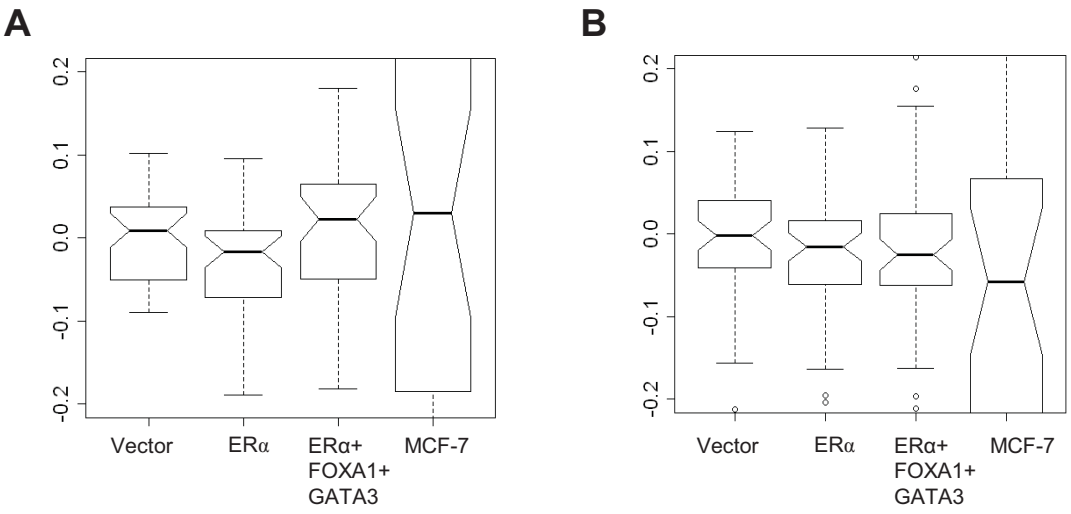
A

Category	p-value
Cell Cycle	7.27E-12-4.69E-04
Cellular Growth and Proliferation	1.84E-08-4.74E-04
DNA Replication, Recombination, and Repair	8.96E-08-1.28E-04

B



Supplemental Figure S14. We compared the expression profile of luminal and basal marker genes defined by Kao J *et al.* in the transfected MDA-MB-231 cells. (A) We observed that there is induced expression of luminal marker genes in the ERα+FOXA1+GATA3-expressing MDA-MB-231 cells as compared to ERα-only or vector-control MDA-MB-231 cells. We included the expression of MCF-7 cells that is defined as luminal subtype as comparison. (B) There is reduced expression of basal marker genes in the ERα+FOXA1+GATA3-expressing MDA-MB-231 cells as compared to ERα-only or vector-control MDA-MB-231 cells. (C) This barplot represented the average expression difference of luminal and basal marker genes in ERα+FOXA1+GATA3-expressing MDA-MB-231 cells as compared to vector-control MDA-MB-231 cells.



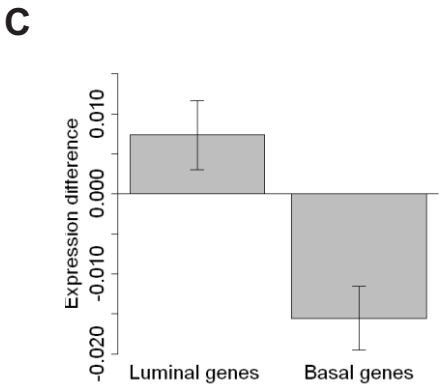
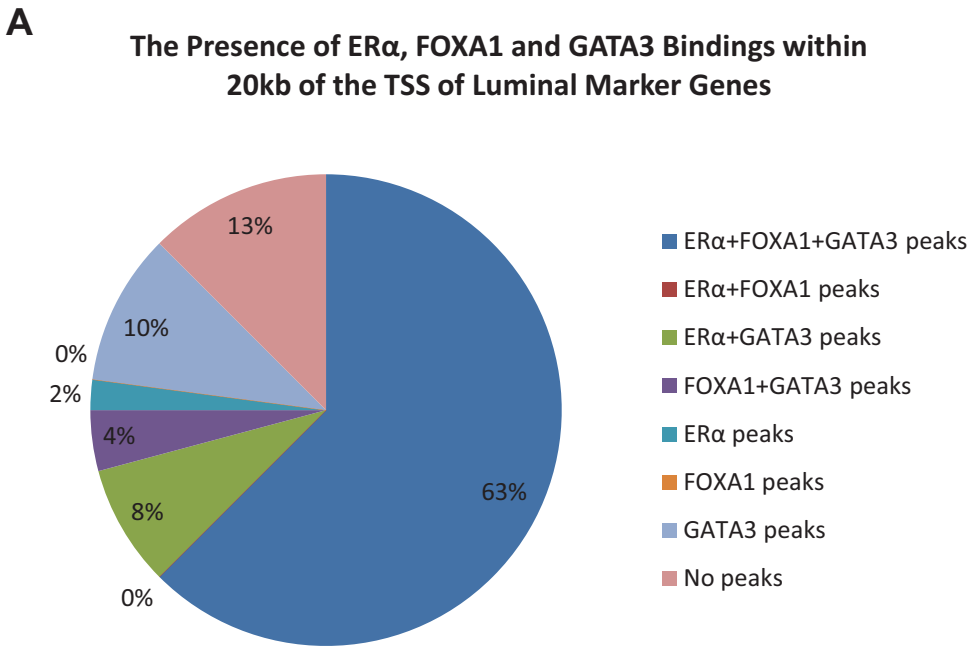


Figure S15. The presence of ERα, FOXA1 and GATA3 bindings at 20kb around the TSS of luminal and basal marker genes defined by Kao *et al.* (A) We demonstrated that ERα, FOXA1 and GATA3 bindings present in 63% of the luminal marker genes, only 13% of these luminal marker genes have no association with either TF binding. (B) The ERα, FOXA1 and GATA3 bindings are only present in 23% of the basal marker genes while there are as many as 40% of these genes have no association with either of these 3 TFs bindings.



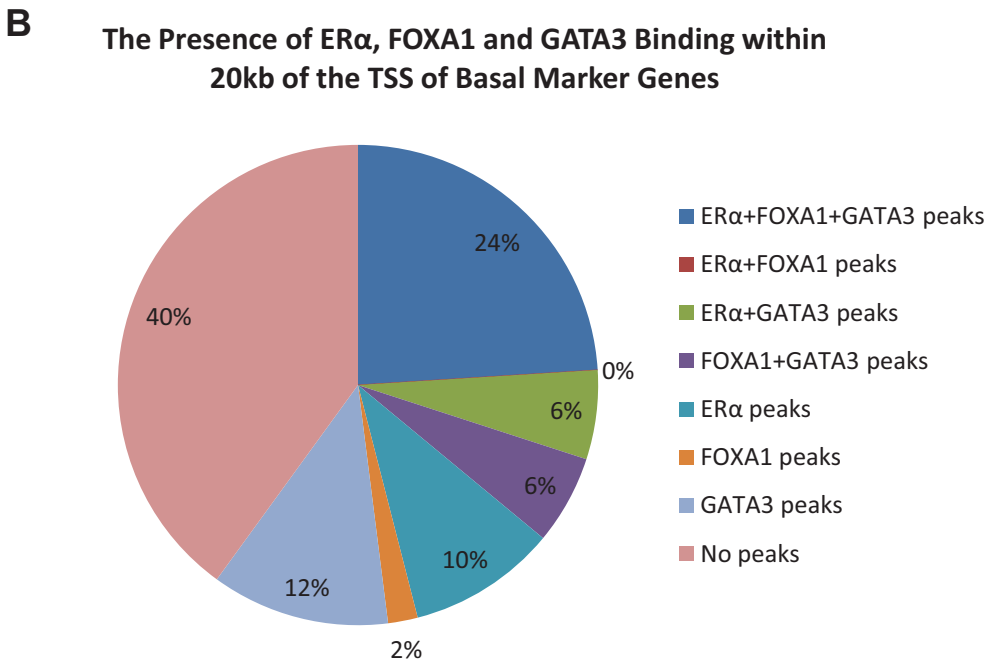


Figure S16. The presence of ERα, FOXA1 and GATA3 recognition motifs. The position weight matrix from de novo motif finding by the Pomoda (Peak Oriented Motif Discovery Algorithm) was used to screen for the co-existing of the TF's motif in the human genome as compared to the random sequences of 200Mb with the cutoff of 1E-4. There is a 64-fold enrichment for the co-localization of ERα+FOXA1+GATA3 motif in the reference human genome as compare to the random nucleotide sequences.