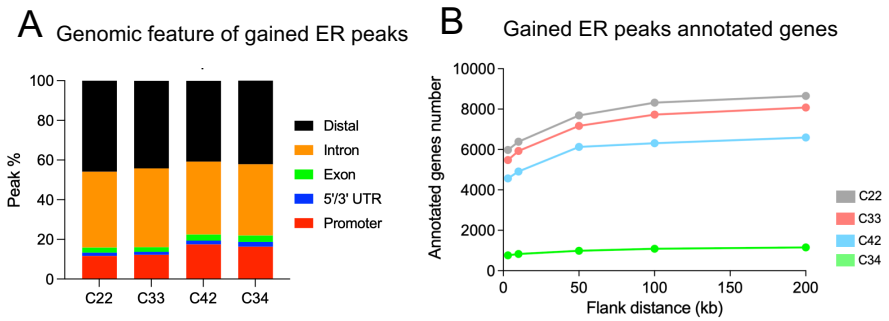
The EstroGene database reveals diverse temporal, context-dependent, and bidirectional estrogen receptor regulomes in breast cancer

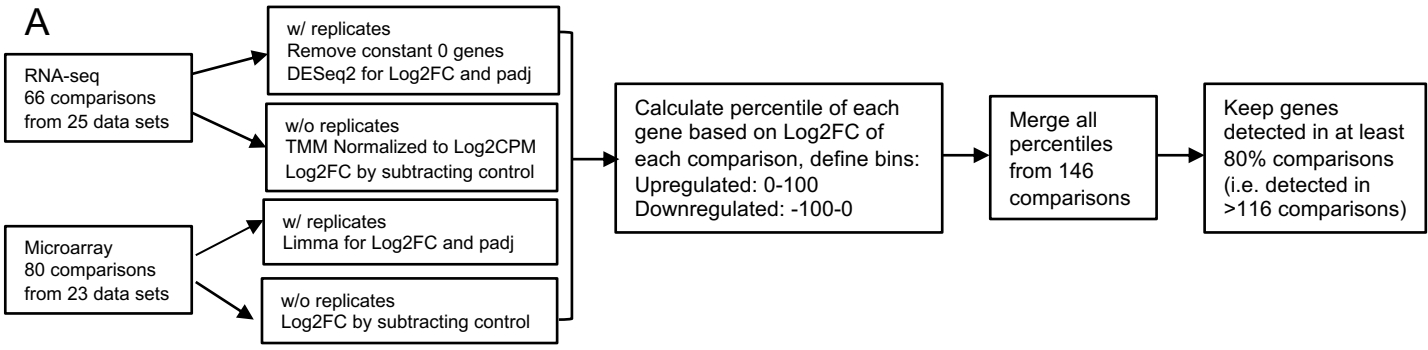
Supplementary Figures

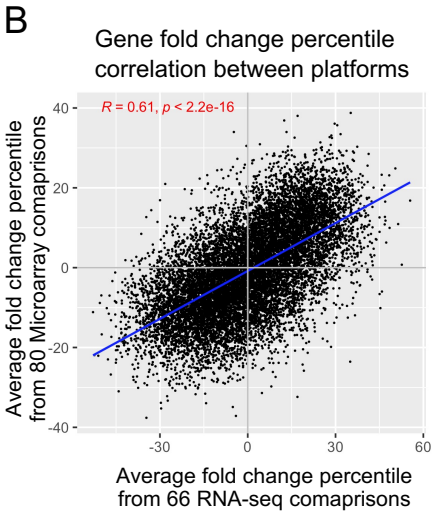
Supplementary Figure S1

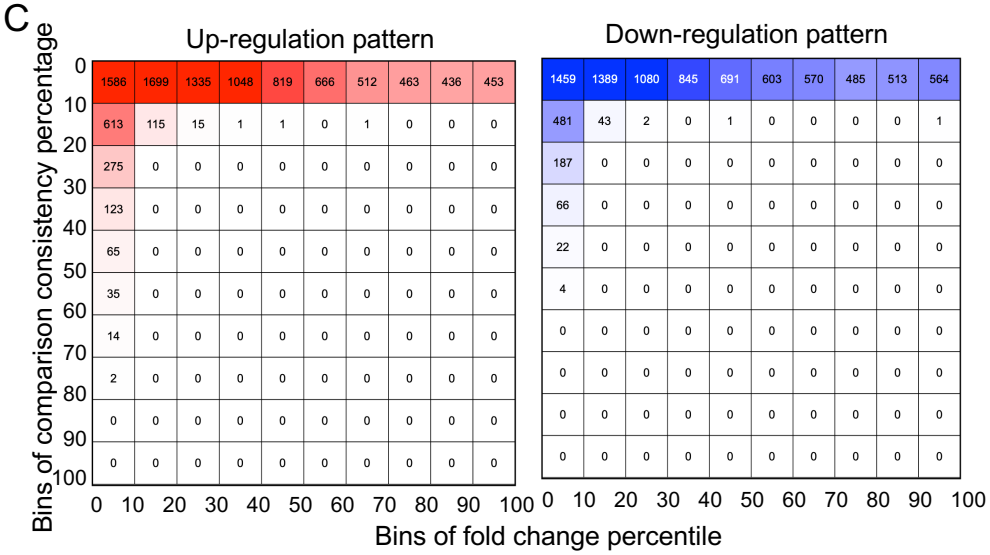


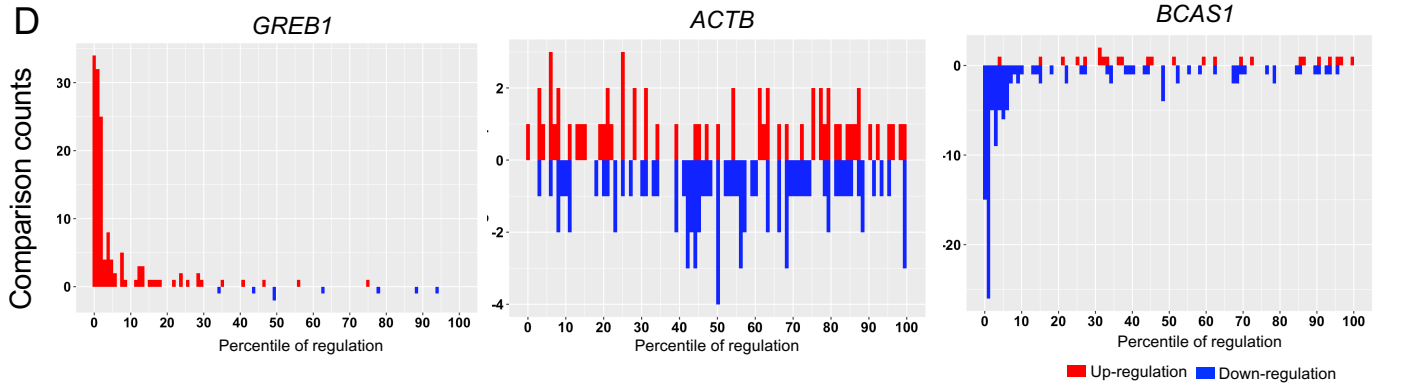
Supplementary Fig. S1. (Related to Fig. 4)
A. Stacked plots showing the genomic feature distributions of gained ER peak derived from four independent experiments.
B. Line plot illustrating the annotated gene numbers with increasing flank distance for associated gene calling of the gained ER peaks from the four studies.

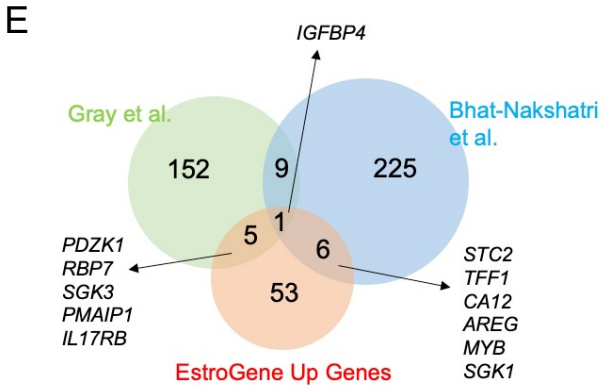
Supplementary Figure S2



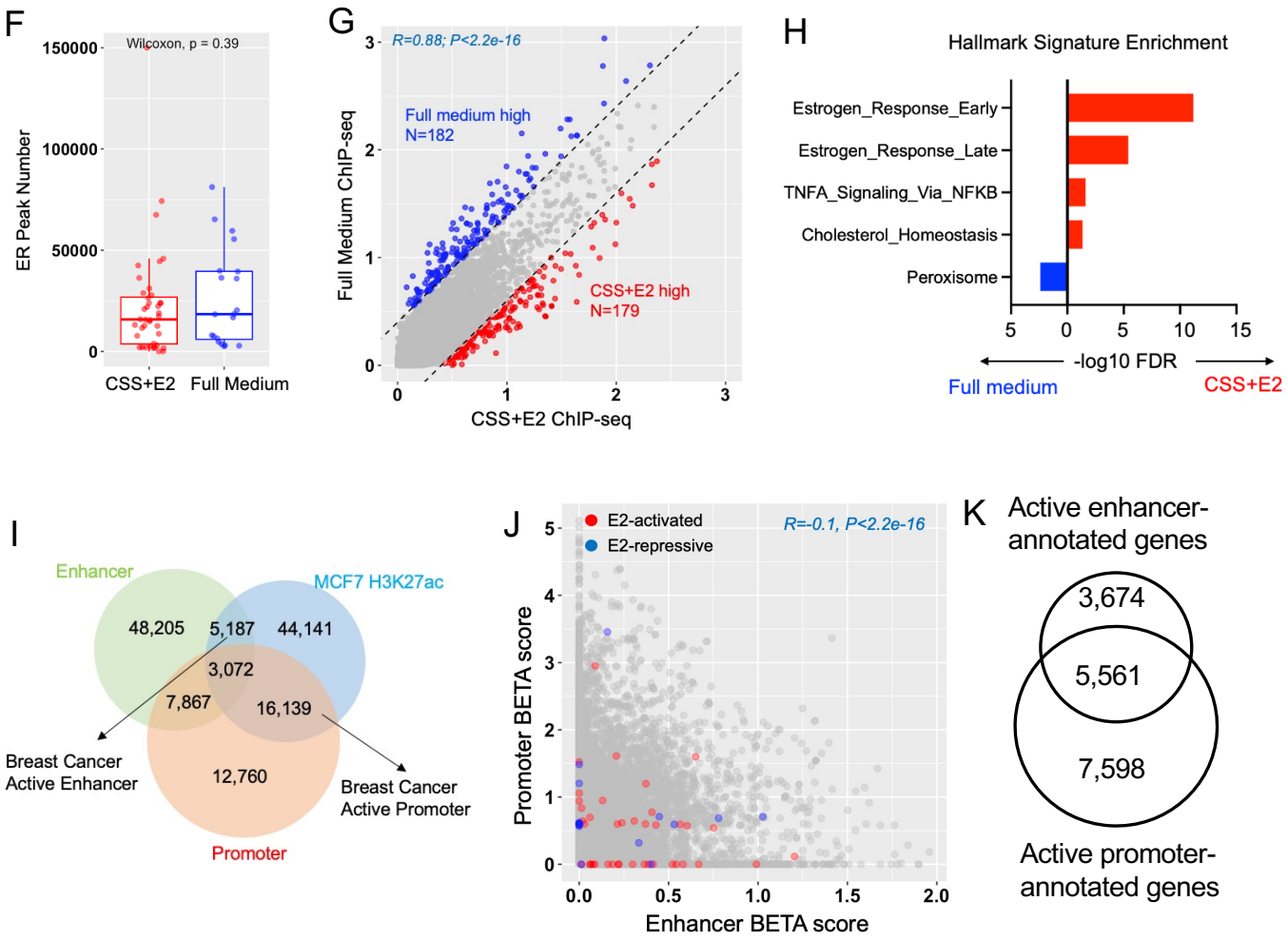








Supplementary Figure S2-Continued



Supplementary Fig. S2 (Related to Fig. 5)

A. A flow chart illustrating the procedure of merging 146 transcriptomic comparisons by percentile-based method.

B. Scattered plot depicting the average fold change percentile correlation of each gene between RNA-seq and microarray platforms. P value was calculated based on Pearson correlation.

C. Heatmaps showing the number of up (Left panel) and down (Right panel) regulated genes in different fold change percentiles and consistency percentages across all the comparisons.

D. Histograms showing distribution of E2 regulation percentile 146 comparisons towards an E2-induced gene *GREB1*, E2-repressed gene *BCAS1* and non-E2 regulated gene *ACTB*. Comparisons showing upregulation and downregulations were shown in upper panel in red and bottom panel in blue respectively. Small percentiles indicate strong regulations.

E. Venn diagram showing the intersection of enriched genes in *ESR1*+ normal epithelial mammary gland cells from two studies (GSE180878 and GSE164898) and E2-induced genes in breast cancer from this study (Fig. 5A).

F. Box plot comparing the called ER peak numbers from 40 and 20 ChIP-seq samples collected in CSS+E2 and full medium condition. Mann Whitney U test was used.

G. Scattered plot depicting the correlation of average BETA scores between CSS+E2 and full medium conditions of each gene. Pearson correlation was performed. Genes with BETA score differences above 0.4 between the two conditions were highlighted.

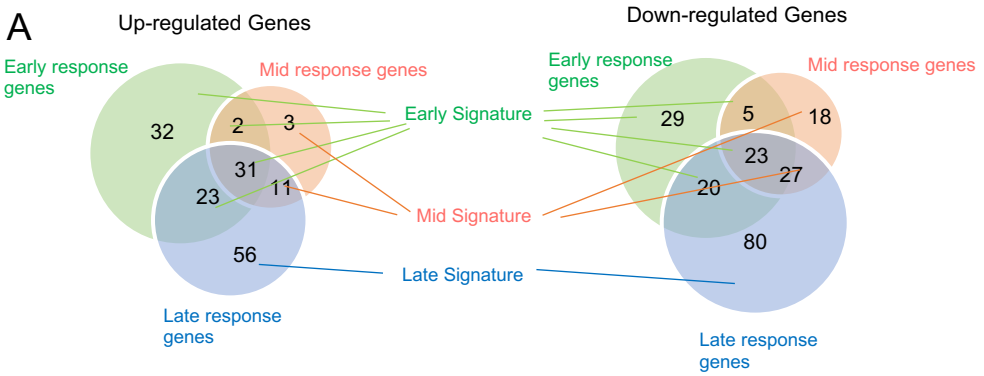
H. Bar plot representing the enrichment level of the Hallmark pathways significantly enriched in full medium and CSS+E2 ER binding sites annotated genes.

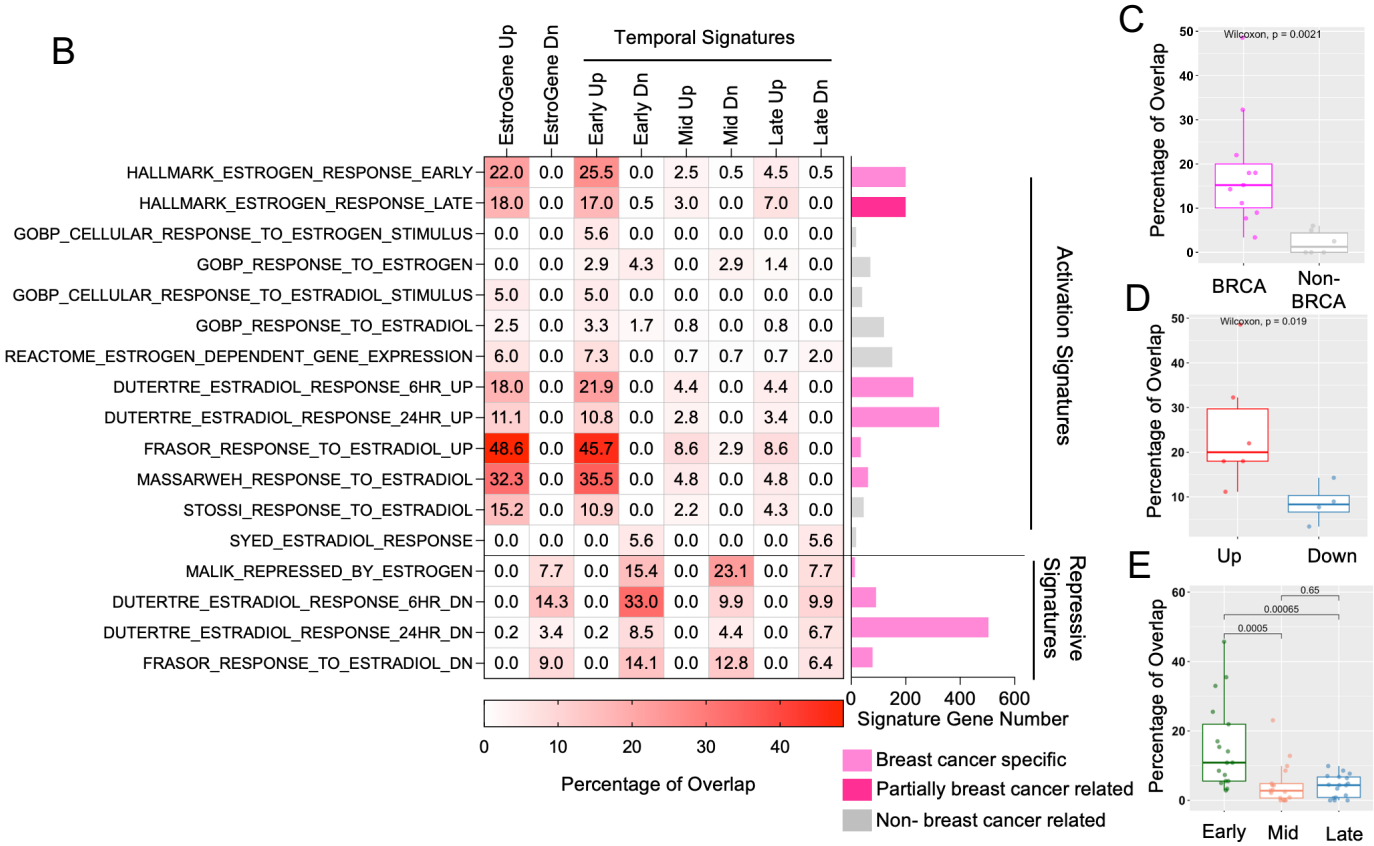
I. Venn diagram showing the intersection of promoter and enhancer regions of human genome derived from FANTOM5 project with MCF7 H3K27ac ChIP-seq binding sites from GSE85158.

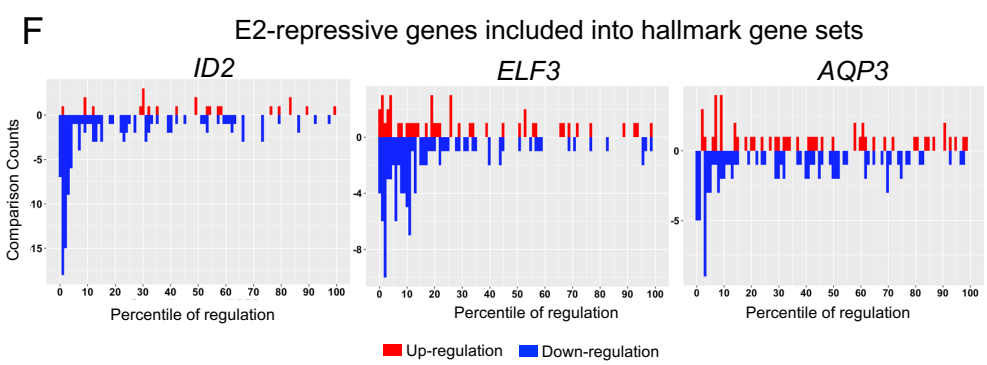
J. Scattered plot showing the correlation between promoter and enhancer region derived BETA score of each genes. High-confident E2-induced and repressed genes were labelled with red and blue. Pearson correlation was applied.

K. Venn diagram illustrating the intersection of genes that annotated by active enhancer and promoter regions in I.

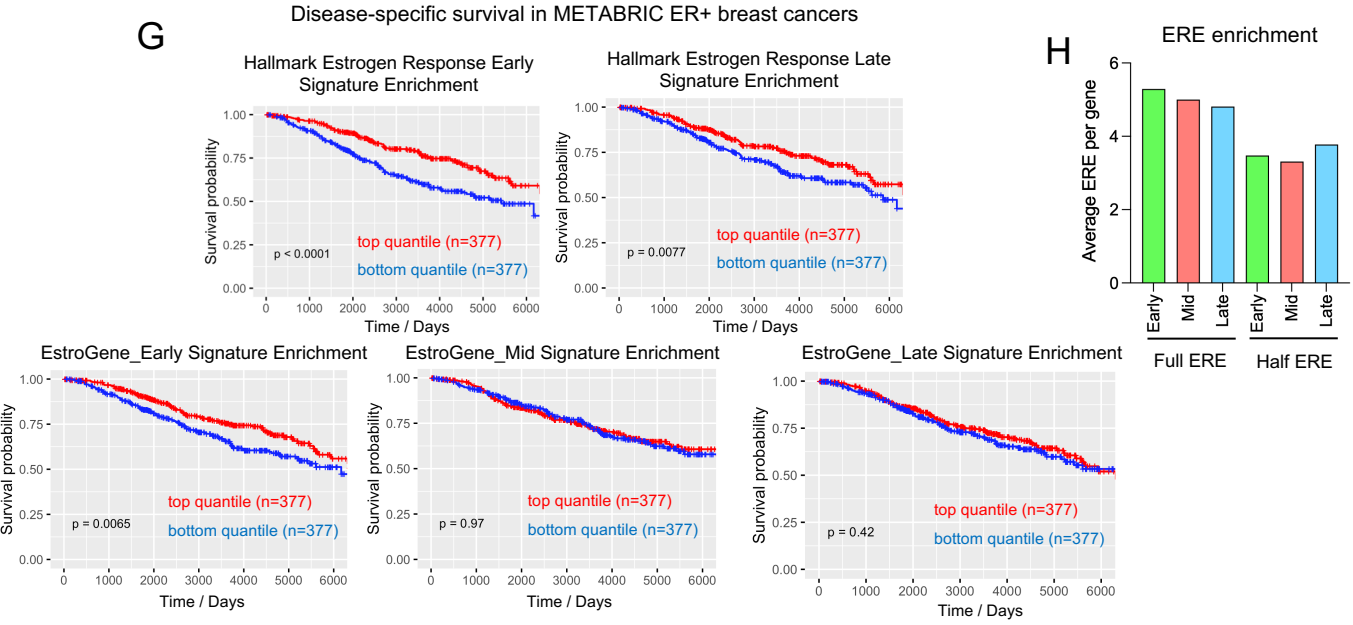
Supplementary Figure S3







Supplementary Figure S3-Continued



Supplementary Fig. S3 (Related to Fig. 6)

A. Venn diagrams demonstrating the overlap between E2 response genes extracted from comparisons of early (<6h), mid (6-24 h) and late (>24h) time points with top 10% percentile alterations and consistent across at least 50% of each comparison sub-collection.

B. Heatmap summarizing the overlap ratios of 17 MSigDB curated E2 response signatures with EstroGene broad and temporal signatures (separated by up and down) generated from this study. Gene set size of MSigDB signatures and types of the sources were shown in the bar graph in the right panel.

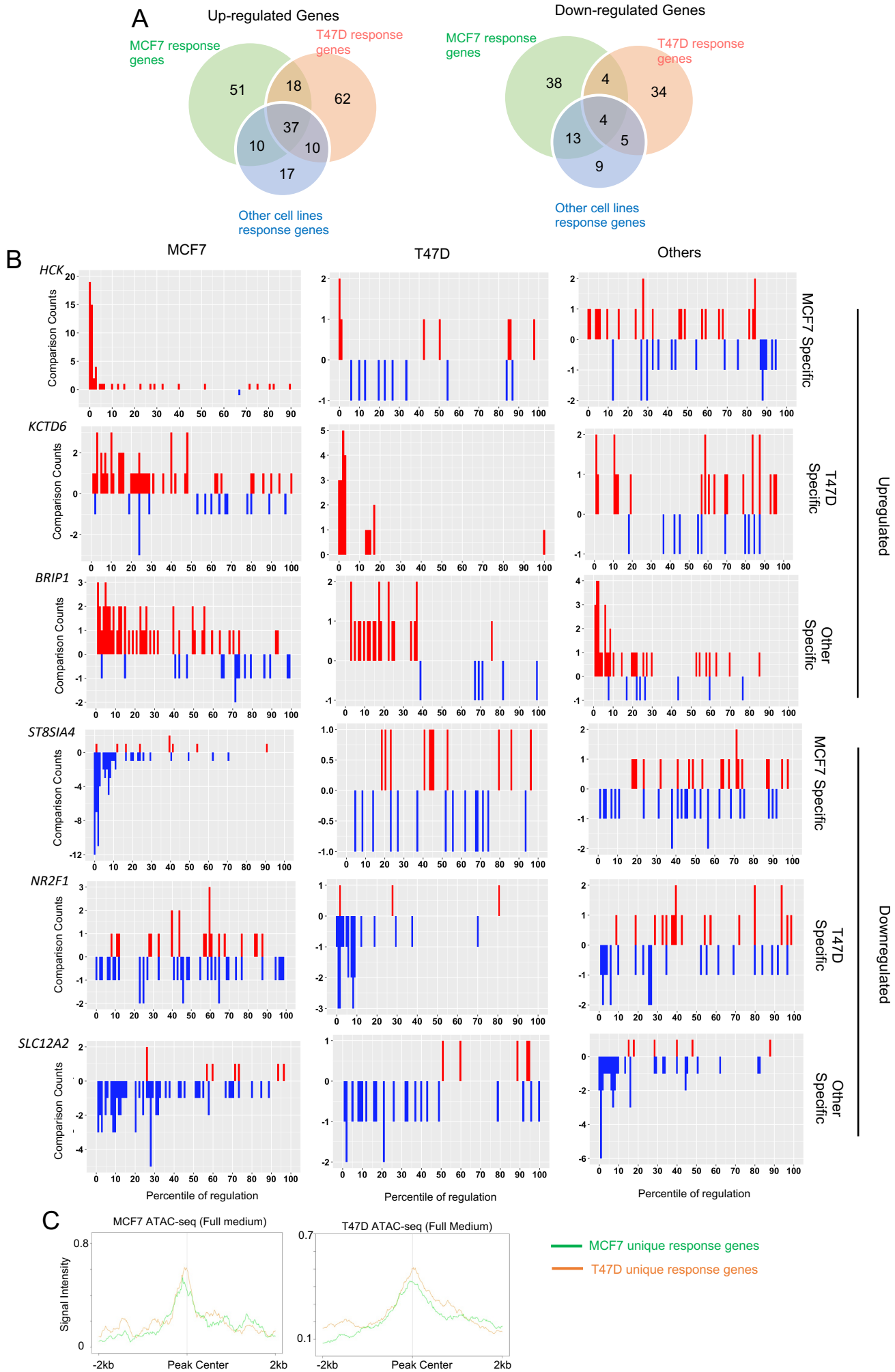
C-E: Box plots depicting the overlapping ratios with the broad EstroGene signatures between breast cancer related and unrelated signatures (C), breast cancer related upregulated and downregulated signatures (D) and temporal signatures (E). Mann Whitney U test was applied.

F. Histograms showing distribution of E2 regulation percentile 146 comparisons towards three E2-repressive genes included into the Hallmark Estrogen Response signatures. Comparisons showing upregulation and downregulations were shown in upper panel in red and bottom panel in blue respectively. Small percentiles indicate strong regulations.

G. Kaplan-Meier plots showing the disease-specific survival (DSS) (METABRIC) comparing patients with tumors with high and low enrichment for each indicated gene sets. High and low were defined by the upper and bottom quartiles of each subset. Censored patients were labelled in cross symbols. Log rank test was used.

H. Histogram showing the full ERE and half ERE motif enrichment per gene in EstroGene Early, Mid and Late subgroups.

Supplementary Figure S4



Supplementary Figure S4-Continued

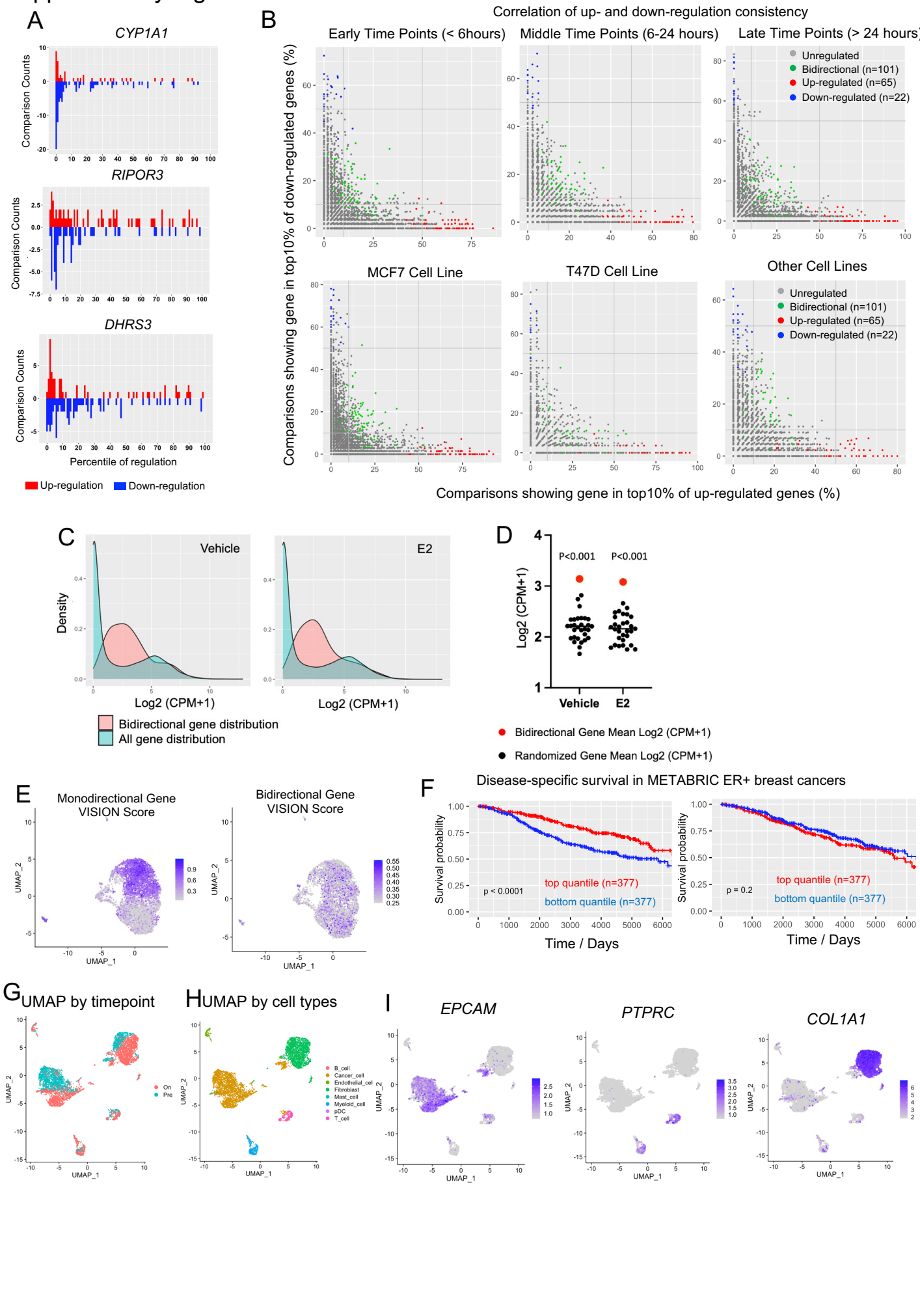
Supplementary Fig. S4 (Related to Fig. 7)

A. Venn diagrams demonstrating the overlap between E2 response genes extracted from comparisons limited to MCF7, T47D and non-MCF7/T47D experiments with top 10% percentile alterations and consistent across at least 50% of each comparison sub-collection.

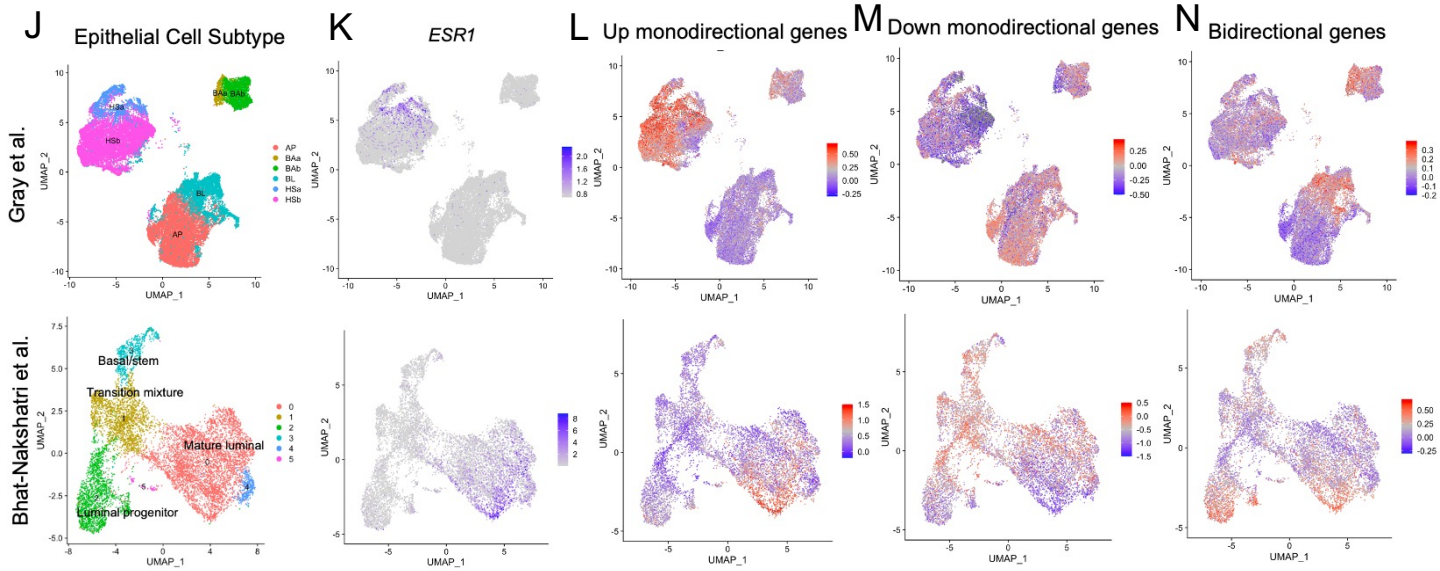
B. Histograms showing distribution of E2 regulation percentile 146 comparisons towards representative up- and down-regulation genes identified from each context. Comparisons showing upregulation and downregulations were shown in upper panel in red and bottom panel in blue respectively. Small percentiles indicate strong regulations.

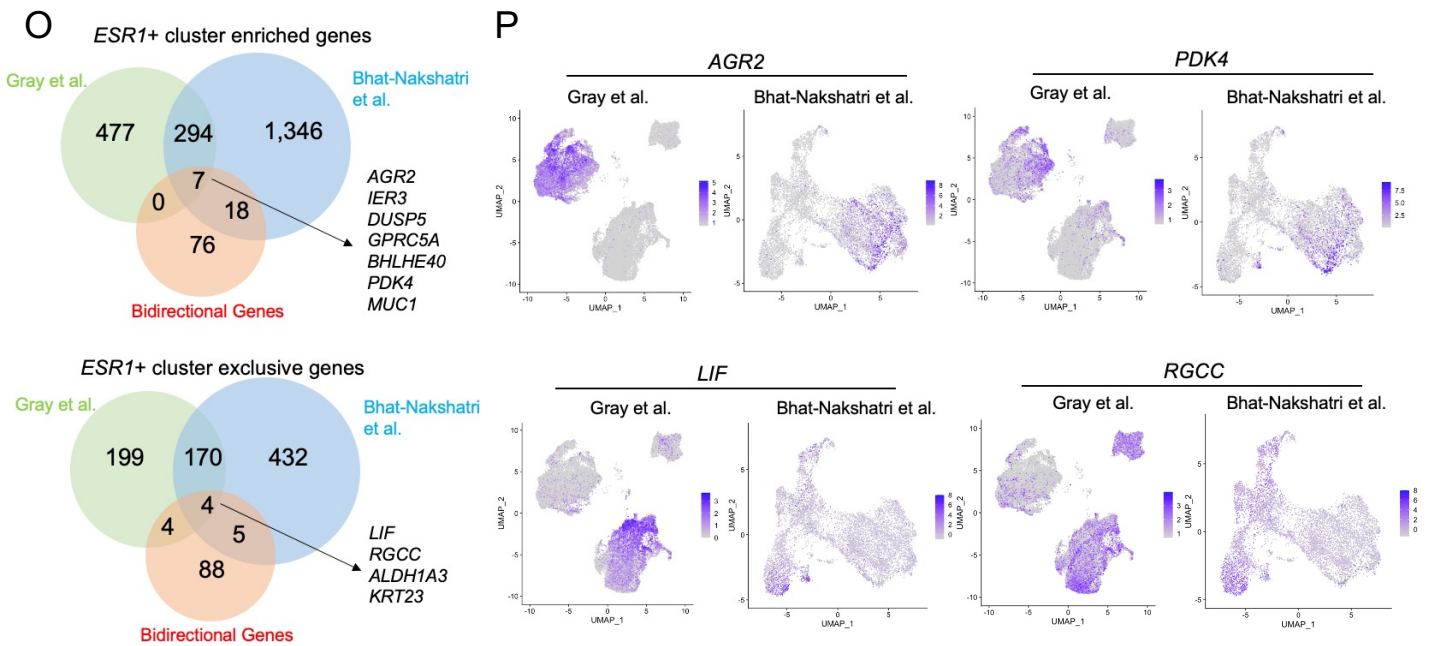
C. Intensity plot showing the signals from ATAC-seq of MCF7 and T47D cells on -/+ 2kb region of TSS of MCF7 and T47D-unique E2 response genes. Epigenetic profiles are downloaded from GSE102441 and GSE84515.

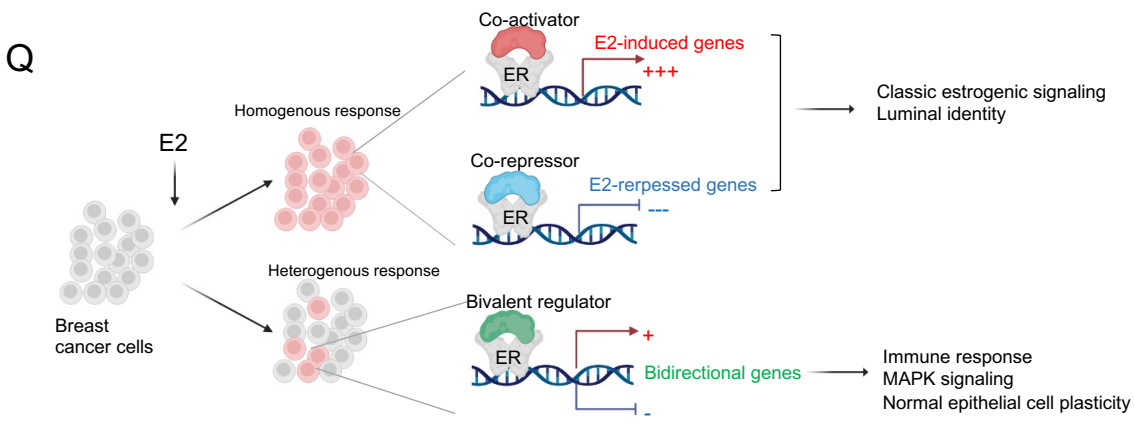
Supplementary Figure S5



Supplementary Fig. S5-Continued







Supplementary Fig. S5-Continued (Related to Fig. 8)

A. Histograms showing distribution of E2 regulation percentile 146 comparisons towards three representative bidirectional E2 response genes *CYP1A1*, *RIPOR3*, *DHRS3*. Comparisons showing upregulation and downregulations were shown in upper panel in red and bottom panel in blue respectively. Small percentiles indicate strong regulations.

B. Scattered plot showing the correlation of each individual gene's percentage falling into top 10% up and down altered targets by fold changes comparisons limited to specific time courses (Top panel) and cell lines (Bottom panel). Monodirectional genes are labelled in red (up) and blue (down). Bidirectional genes are labelled in green.

C. Density plots comparing the log2(CPM+1) value distribution of the 101 E2 bidirectional response gens and all the genes from the merged transcriptomic analysis in vehicle (left, n=166) and E2 (right, n=208)-treated samples from RNA-seq.

D. Dot plot showing the comparison of the mean log2(CPM+1) value of the 101 E2 bidirectional genes to 30 times of 101-gene random sampling in vehicle and E2 groups. One sample t-test was applied.

E. UMAP showing enrichment score of mono- and bidirectional E2 response genes in the MCF7 single cell RNA-seq data sets described in Fig. 5C.

F. Kaplan-Meier plots showing the disease-specific survival (DSS) (METABRIC) comparing patients with tumors with high and low enrichment for monodirectional or bidirectional gene sets. High and low were defined by the upper and bottom quartiles of each subset. Censored patients were labelled in cross symbols. Log rank test was used and hazard ratio with 95% CI were labelled.

G-I. UMAP showing all cell distributions (G), cell subtype assignment (H) and epithelial (*EPCAM*), lymphocytes (*PTPRC*) and fibroblast (*COLA1A*) marker expression (I) from two biopsies of an ER+ patient separated by anti-PD1 treatment status from the BIOKEY cohort.

J-N. UMAPs showing cell subtype assignment (J), *ESR1* expression (K), enrichment scores of up- (L) and down-regulated (M) monodirectional and bidirectional (N) E2 response signatures in the re-clustered normal mammary epithelial cells from two different studies (GSE180878 and GSE164898).

O. Venn diagrams showing the intersection of enriched (top) and exclusive (bottom) genes in *ESR1*+ normal epithelial mammary gland cells from two studies and the bidirectional E2 response genes in breast cancer from this study.

P. UMAPs showing expression of two positively (*AGR2, PDK4*) and two negatively (*LIF, RGCC*) associated bidirectional E2 response genes in the re-clustered normal mammary epithelial cells from two different studies.

Q. A schematic illustration for mechanistic and functional perspectives regarding to mono- and bidirectional estrogen response genes. (Created with BioRender.com)