Supplemental Data

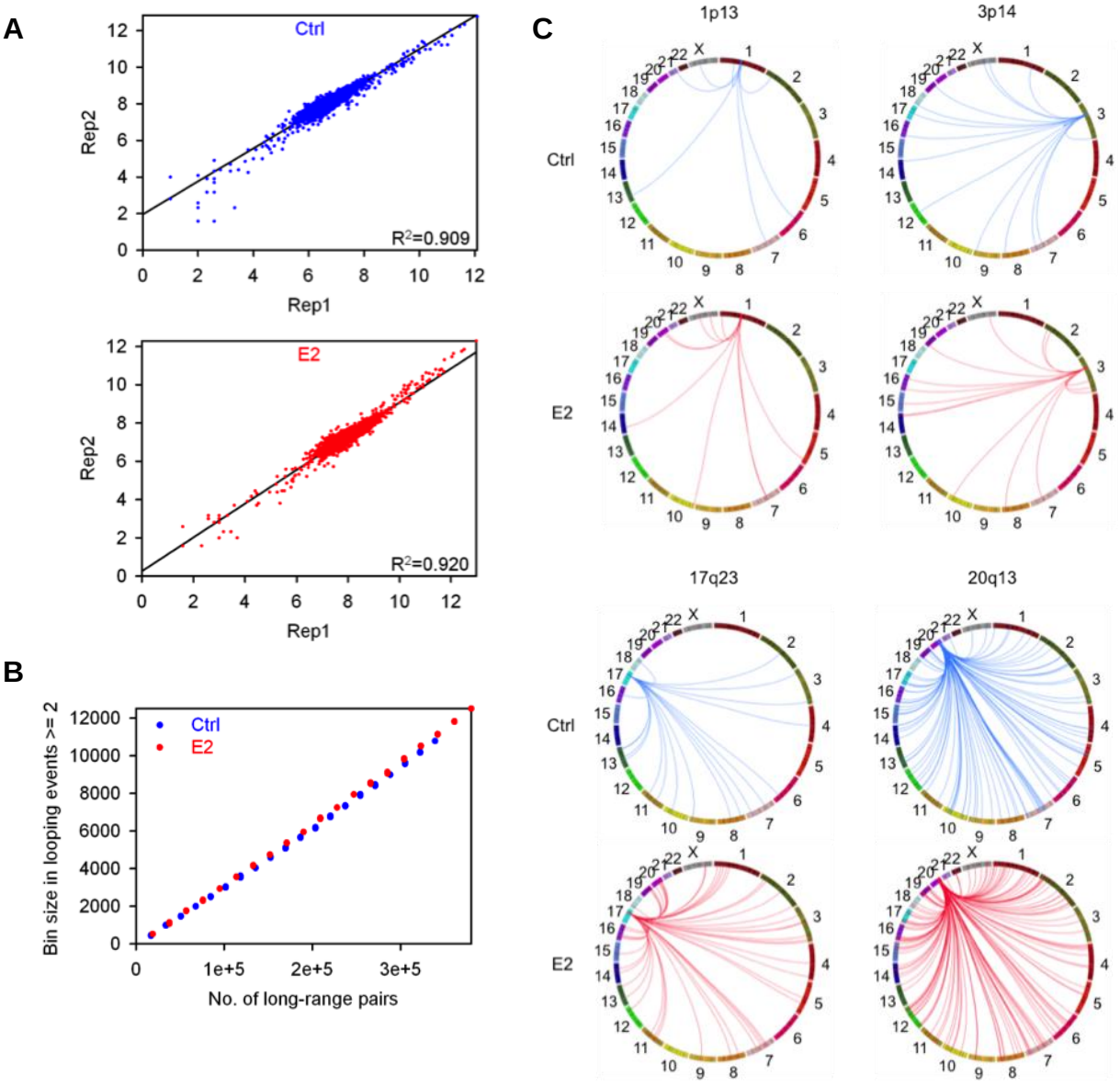


Figure S1, related to Figure 1. Genome-wide Mapping of Chromatin Interactions by Chromosome Conformation Capture Coupled with Paired-end Sequencing (3C-seq)

(A-B) Statistical analyses of genome-wide chromatin interactions by 3C-seq. (A) Examination of reproducibility. 3C assay coupled with paired-end sequencing was performed on two biological replicates of control (blue) and E2-treated (red) MCF-7

cells, respectively. Scatter plots were generated based on log2[interaction events per Mb]. The value of linear regression plots in both Ctrl (R^2=0.909) and E2 (R^2=0.920) datasets support the reproducibility of this approach. (B) Saturation analysis of genome-wide chromatin interaction dataset. To determine whether sequence coverage provided by 3C assay coupled with paired-end sequencing is sufficient to identify chromatin interactions in a genome-wide scale, differential reduced-scale datasets were generated by randomly extracting 5%, 10%,...,95% of the original data for determining the maxima number of positive interaction events. As shown in the plot, the maxima number of positive events was reduced following the extraction scale, indicating that this dataset is sufficient for globally screening chromatin interactions in a genome.
(C) Circular maps of clustered DERE-involved chromosomal looping upon estrogen stimulation. Chromatins looping events associated with clustered DEREs before (blue) and after (red) estrogen stimulation were depicted as circular maps using the Circos software (http://mkweb.bcgsc.ca/circos/). Chromosomes are individually colored for circular visualization.

Table S1, related to Figure 1: Sequencing statistics of mapping genome-wide chromatin interactions

Sample	Lane ID		No. Reads	*No. Aligned Reads	%	No. Bases (Gb)	*No. Aligned Bases (Gb)	*Genome Coverage
Ctrl	Rep1	#1	33,240,584	27,607,307	83.05	1.20	0.99	2.64X
		#2	33,383,384	27,386,580	82.04	1.20	0.99	
		#3	42,681,244	34,176,497	80.07	1.54	1.23	
	Rep2	#4	50,718,124	25,788,183	50.85	1.83	0.93	
		#5	51,813,594	27,938,517	53.92	1.87	1.01	
		#6	51,956,014	32,354,461	62.27	1.87	1.16	
		#7	52,971,562	34,359,214	64.86	1.91	1.24	
		Total	316,764,506	209,610,759	66.17	11.40	7.55	
E2, 24 hr stimulation	Rep1	#1	46,232,094	35,752,846	77.33	1.66	1.29	3.02X
		#2	46,864,612	36,585,449	78.07	1.69	1.32	
		#3	49,243,110	36,241,568	73.60	1.77	1.30	
		#4	48,880,210	35,408,643	72.44	1.76	1.27	
	Rep2	#5	53,262,334	34,425,348	64.63	1.92	1.24	
		#6	53,145,252	33,538,479	63.11	1.91	1.21	
		#7	51,885,506	28,054,649	54.07	1.87	1.01	
		Total	349,513,118	240,006,982	68.67	12.58	8.64	

*Sequences were aligned to the Human Genome Assembly (NCBI build 36.1/hg18).

Table S2, related to Figure 1: Sensitivity Analysis by Mix Negative Binomial Distribution

Sample	Distance of Pairs	n.True	n.False	ps.True	ps.False	r.True	*Sensitivity	#*FDR*
0 hr	1K ~ 5K	233624.02	0.64502	0.99999	0.02979	0.72395	0.77304	0.29084
	5K ~ 10K	135930.23	0.37209	0.99998	0.02136	0.82749	0.78985	0.15101
	> 10K	364787061.14	0.00042	1.00000	0.01288	0.80405	0.71575	0.00048
	Inter-chromosomal	207298.68	0.50442	0.99999	0.03066	0.78868	0.71201	0.21857
24 hr	1K ~ 5K	14915386.29	0.74795	1.00000	0.04145	0.70561	0.75232	0.31812
	5K ~ 10K	2059971.42	0.53276	1.00000	0.03105	0.79897	0.76859	0.19955
	> 10K	873491.26	0.31159	1.00000	0.02156	0.81435	0.72647	0.15961
	Inter-chromosomal	6329178.78	0.48269	1.00000	0.03042	0.79221	0.71973	0.20967

*Sensitivity: Pr(counts>=2 | true signal).
**FDR*: Pr(false signal | counts>=2).

Table S3, related to Figure 1: Mate-pair sequencing statistics

Lane ID	No. Reads	*No. Aligned Reads	%	No. Bases (Gb)	*No. Aligned Bases (Gb)	*Genome Coverage
#1	75,259,850	55,001,116	73.08	3.84	2.81	0.98
#2	73,726,690	54,376,881	73.75	3.76	2.77	0.97
#3	75,375,318	56,094,659	74.42	3.84	2.86	1.00
#4	74,366,524	55,751,490	74.97	3.79	2.84	0.99
#5	75,295,422	56,577,933	75.14	3.84	2.89	1.01
#6	73,467,116	52,506,244	71.47	3.75	2.68	0.94
#7	74,246,026	52,839,223	71.17	3.79	2.69	0.94
#8	70,653,096	49,204,131	69.64	3.60	2.51	0.88
#9	72,738,184	51,126,582	70.29	3.71	2.61	0.91
#10	72,182,080	51,034,656	70.70	3.68	2.60	0.91
#11	71,516,634	50,583,122	70.73	3.65	2.58	0.90
#12	70,402,582	49,874,552	70.84	3.59	2.54	0.89
#13	71,697,496	50,694,766	70.71	3.66	2.59	0.90
#14	87,642,016	63,055,858	71.95	4.47	3.22	1.13
#15	84,604,606	60,202,242	71.16	4.31	3.07	1.07
#16	84,369,284	62,102,585	73.61	4.30	3.17	1.11
#17	83,563,498	61,516,634	73.62	4.26	3.14	1.10
#18	85,435,444	62,432,062	73.08	4.36	3.18	1.11
#19	83,459,818	61,416,264	73.59	4.26	3.13	1.10
#20	85,208,642	62,378,791	73.21	4.35	3.18	1.11
#21	81,277,408	59,496,934	73.20	4.15	3.03	1.06
#22	58,315,936	42,117,447	72.22	2.97	2.15	0.75
#23	72,529,374	42,979,879	59.26	3.70	2.19	0.77
#24	60,789,654	42,384,004	69.72	3.10	2.16	0.76
#25	60,980,534	45,833,205	75.16	3.11	2.34	0.82
#26	62,470,194	46,843,143	74.98	3.19	2.39	0.84
#27	62,543,874	46,922,928	75.02	3.19	2.39	0.84

(continued)

Lane ID	No. Reads	*No. Aligned Reads	%	No. Bases (Gb)	*No. Aligned Bases (Gb)	*Genome Coverage
#28	60,580,630	45,520,915	75.14	3.09	2.32	0.81
#29	59,508,310	33,288,108	55.94	3.03	1.70	0.59
#30	67,754,368	46,109,264	68.05	3.46	2.35	0.82
#31	68,026,638	46,250,965	67.99	3.47	2.36	0.83
#32	71,663,088	48,028,796	67.02	3.65	2.45	0.86
#33	68,637,566	45,841,546	66.79	3.50	2.34	0.82
#34	70,245,348	52,199,892	74.31	3.58	2.66	0.93
Total	2,470,533,248	1,762,586,817	71.34	126.00	89.89	31.45

*Sequences were aligned to the Human Genome Assembly (NCBI build 36.1/hg18).

Table S4, related to Figure 1: List of Translocation Fusions in the MCF-7 Genome

Hsu-20q13_Table S4.xls

Table S5, related to Figure 1: List of Chromatin Interaction Events in Untreated MCF-7 Cells

Hsu-20q13_Table S5.xls

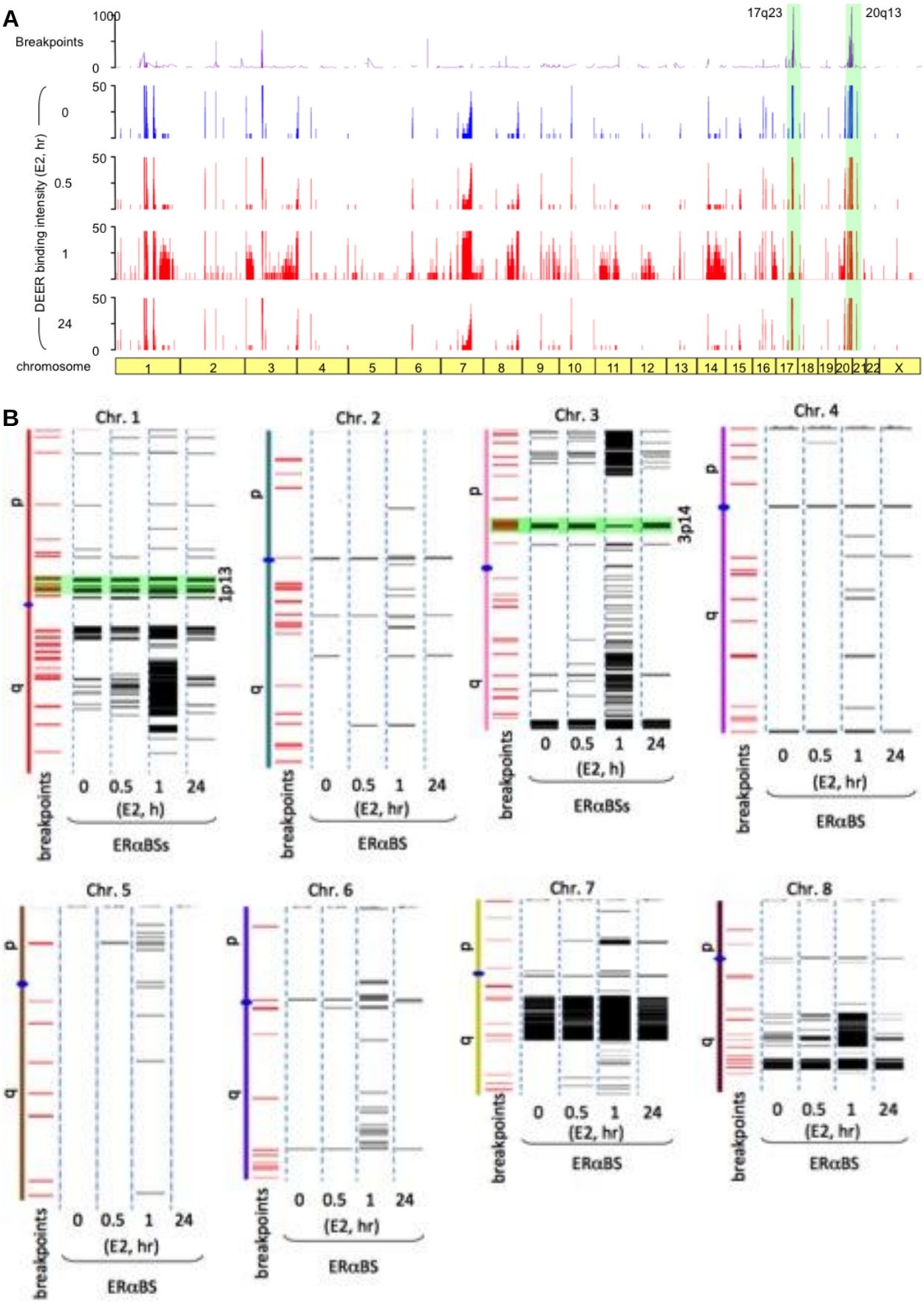
Table S6, related to Figure 1: List of Chromatin Interaction Events in E2-treated MCF-7 Cells

Hsu-20q13_Table S6.xls

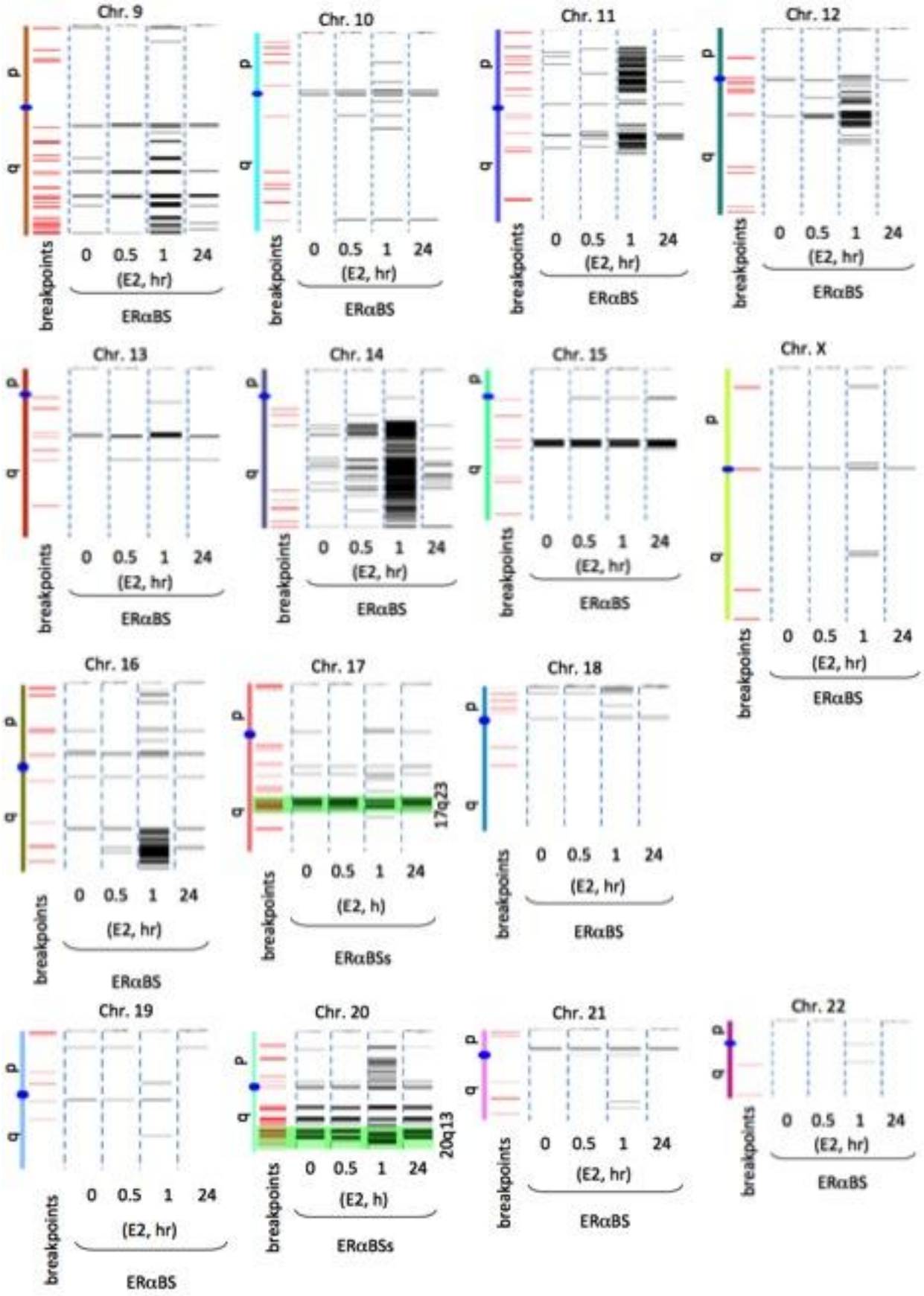
Table S7, related to Figure 1: Statistics of ERα ChIP-seq in E2-treated MCF-7 Cells

Sample	Lane ID	No. of reads	*No. of aligned reads	%	No. of bases (Gb)	*No. of mapped bases (Gb)
0 hr	#1	37,427,797	21,958,082	58.67	1.91	1.12
	#2	39,554,721	22,039,376	55.72	2.02	1.12
0.5 hr	#3	45,582,024	24,219,056	53.13	2.32	1.24
	#4	45,235,626	24,449,046	54.05	2.31	1.25
1 hr	#5	36,727,295	24,953,252	67.94	1.87	1.27
	#6	33,565,499	21,436,148	63.86	1.71	1.09
24 hr	#7	40,761,726	24,546,099	60.22	2.08	1.25
	#8	40,241,912	24,239,655	60.23	2.05	1.24

*Sequences were aligned to the Human Genome Assembly (NCBI build 36.1/hg18).


A
Breakpoints
1000
0
17q23
20q13
DEER binding intensity (E2, hr)
0
50
0.5
1
24
chromosome
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
B
Chr. 1
1p13
Chr. 2
Chr. 3
3p14
Chr. 4
Chr. 5
Chr. 6
Chr. 7
Chr. 8
p
q
breakpoints
0 0.5 1 24
(E2, h)
(E2, hr)
ERαBSs
ERαBS

B (continued)


Chr. 9
Chr. 10
Chr. 11
Chr. 12
Chr. 13
Chr. 14
Chr. 15
Chr. X
Chr. 16
Chr. 17
Chr. 18
Chr. 19
Chr. 20
Chr. 21
Chr. 22
p
q
breakpoints
0 0.5 1 24
(E2, hr)
(E2, h)
ERαBS
ERαBSs
17q23
20q13

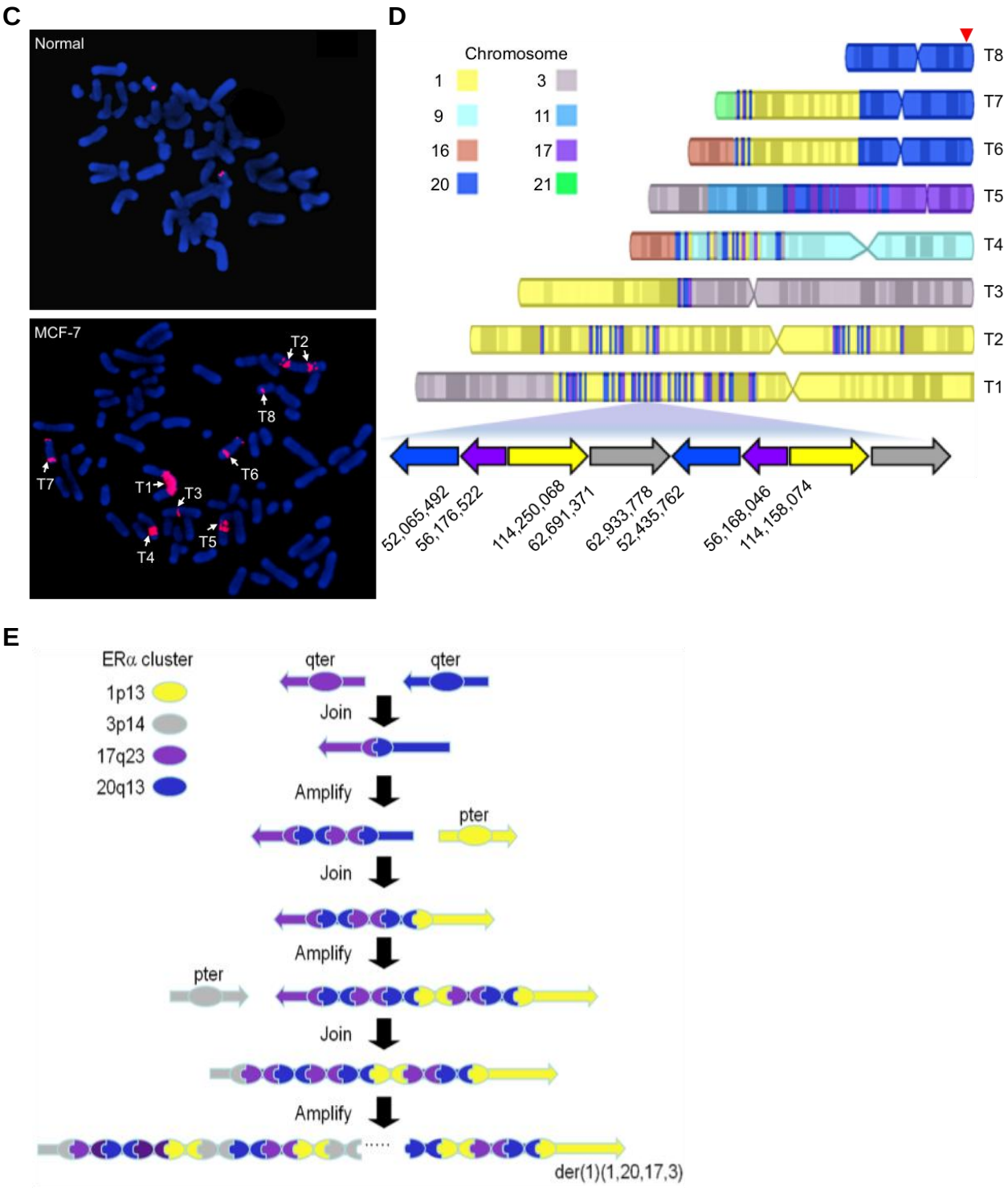


Figure S2, related to Figure 2. Genomic Location of DEREs and Translocation Breakpoint Sites in MCF-7 Cells

(A-B) Genomic distribution of DEREs and translocation breakpoints in MCF-7 cells.

Mate-pair sequencing was conducted to explore breakpoint sites of translocations across the MCF-7 genome. Charcoal-stripped MCF-7 cells were stimulated with E2 (70 nM) in four time points (0, 0.5, 1, and 24 hr) and then subjected to ERα ChIP-seq (see also Table S7). (A) Genome-wide location map. Fusion frequencies of breakpoint sites were plotted in purple and binding intensities of DERE in blue (untreated) and red (E2-treated). Green highlighted regions indicated co-localization of breakpoint clusters and densely ERα-bound DERE regions. (B) Distribution maps of DEREs and breakpoints in individual chromosomes. The maximum binding events is found at 1 hr of stimulation. Pre-existing binding of ERα is observed in some chromosome regions prior to estrogen stimulation, consistent with previous reports (Carroll et al., 2006; Lupien et al., 2010). Increased binding intensities events of these regions were observed within 24 hr of stimulation.

(C-E) Proposed Model of DERE-DERE fusions and amplification. (C) Metaphase FISH analysis of 20q13 DERE region in normal breast epithelial cells (*upper*) and MCF-7 cancer cells (*lower*). Arrowheads indicate eight derivative chromosomes, T1 to T8. (D) Translocation patterns of 20q13 DERE region in derivative chromosomes, reconstructed based on mate-pair sequencing data. Using the T1 derivative chromosome as an example, densely ERα-bound DERE regions were fused and amplified as a concatenate unit (20q13-17q23-1p13-3p14) (*lower*). A putative deletion as shown in red triangle was found in Chromosome 20 (T8). (E) Proposed 20q13-17q23 DEREs as a unit participating in DERE-DERE fusions and amplification. Based on mate-pair sequencing data and metaphase FISH results, amplification events of 20q13 and 17q23 DERE regions occurred more frequently than other amplified regions, and the junctures

between 20q13 and 17q23 segments were the most common events observed in MCF-7 cells (see also Figure 2C). Therefore, it is assumed that rearrangements in the MCF7 genome aberrantly started from the fusion fragments between 20q13 and 17q23 DERE regions. After a few steps of amplification, part of this region is translocated onto chromosome 1 and fused with 1p13. Subsequently, the short arm of chromosome 3 joins this the amplification. The rearranged chromosome T1 (see Figure S2D) is derived from multiple steps of amplification and translocation.

Table S8, related to Figure 2: Genomic Characteristics of 17q23 and 20q13 DEREs

Chromosome Region	Coordinates	Length (Mb)	Breakpoints (No. per Mb)	No. ERα Binding Sites (No. per Mb)		Chromatin Loops		
				Ctrl	E2	Ctrl	E2	Common
17q23	chr17:55,268,000-59,758,000	4.49	46 (10.24)	81 (18.04)	207 (46.10)	68	98	29
20q13	chr20:42,060,000-61,180,000	19.12	98 (5.13)	328 (17.15)	974 (50.94)	174	242	65

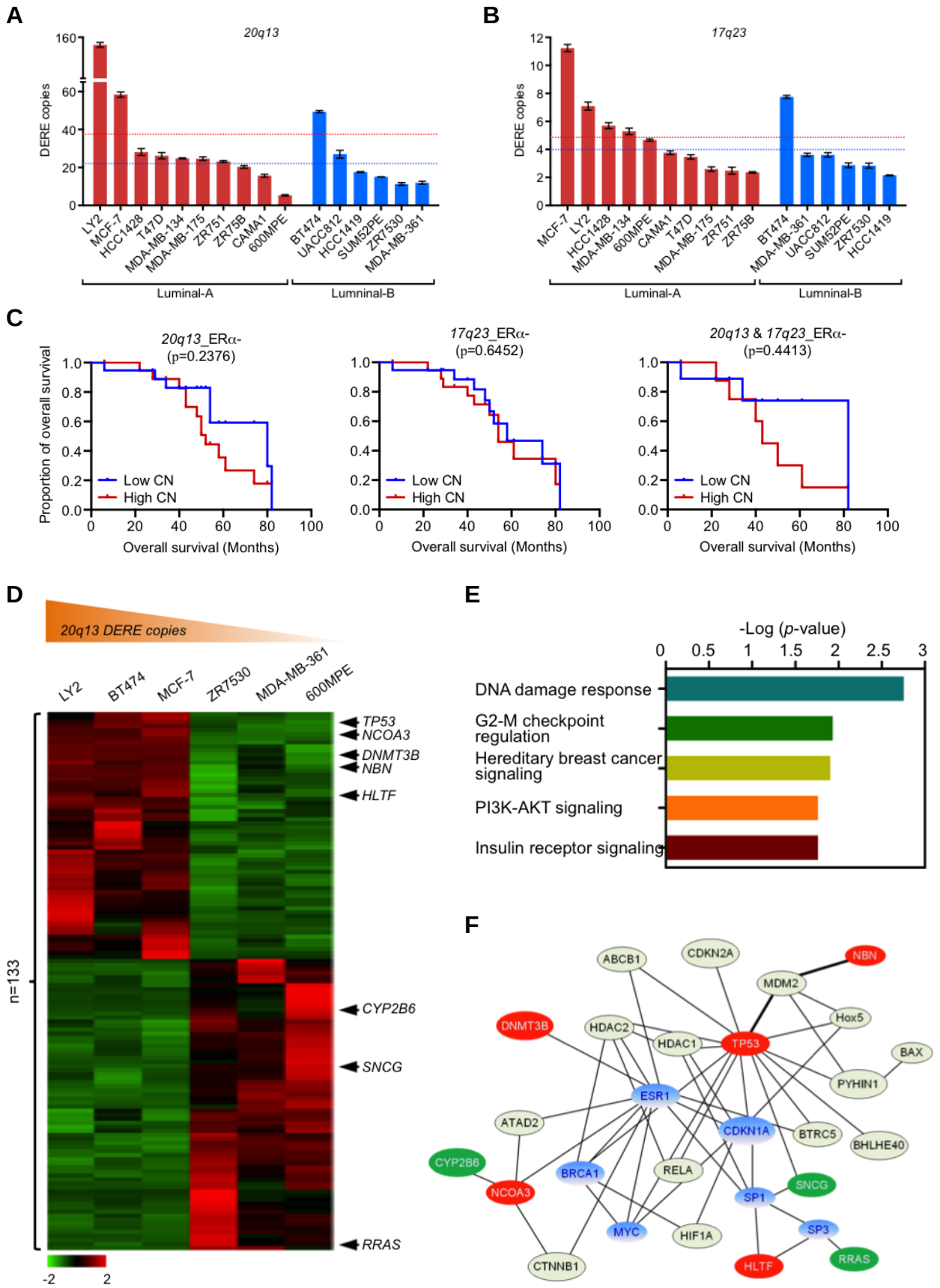

A
20q13
DERE copies
LY2
MCF-7
HCC1428
T47D
MDA-MB-134
MDA-MB-175
ZR751
ZR75B
CAMA1
600MPE
BT474
UACC812
HCC1419
SUM52PE
ZR7530
MDA-MB-361
Luminal-A
Lumninal-B
B
17q23
DERE copies
MCF-7
LY2
HCC1428
MDA-MB-134
600MPE
CAMA1
T47D
MDA-MB-175
ZR751
ZR75B
BT474
MDA-MB-361
UACC812
SUM52PE
ZR7530
HCC1419
Luminal-A
Lumninal-B
C
20q13_ERα-
(p=0.2376)
17q23_ERα-
(p=0.6452)
20q13 & 17q23_ERα-
(p=0.4413)
Proportion of overall survival
Low CN
High CN
Overall survival (Months)
D
20q13 DERE copies
LY2
BT474
MCF-7
ZR7530
MDA-MB-361
600MPE
TP53
NCOA3
DNMT3B
NBN
HLTF
CYP2B6
SNCG
RRAS
n=133
-2
2
E
-Log (p-value)
DNA damage response
G2-M checkpoint regulation
Hereditary breast cancer signaling
PI3K-AKT signaling
Insulin receptor signaling
F
ABCB1
CDKN2A
NBN
MDM2
DNMT3B
HDAC2
Hox5
HDAC1
TP53
BAX
PYHIN1
ESR1
ATAD2
CDKN1A
BTRC5
BHLHE40
CYP2B6
BRCA1
RELA
SNCG
NCOA3
SP1
SP3
MYC
HIF1A
CTNNB1
HLTF
RRAS

Figure S3, related to Figure 4. Amplified DEREs are Profoundly Observed in ERα-positive Breast Cancers and Association of TP53-involved Signaling Network in 20q13 DERE Amplification

(A-B) Copy-number variation of 20q13 and 17q23 DEREs in ERα-positive breast cancer cell lines. Histograms depicting copy-number variation of 20q13 (A) and 17q23 (B) DEREs in ERα-positive breast cancer cell lines (n=16) were plotted with individual cell lines and categorized into two groups based on subtypes (Heiser et al., 2009). Mean values of DERE copies in two groups were shown in dashed lines. Red, Luminal-A; Blue, Luminal-B.

(C) Correlation between amplified DERE copies and overall survival of ERα-negative breast cancer patients. Kaplan-Meier survival curves of ERα-negative breast cancer patients (n=19) harboring either high (n>2) or low CN (n<2) of the 20q13 (*left*) or 17q23 (*middle*) or both (*right*) DEREs. Wilcoxon test was used to determine statistical significance.

(D-F) Association of TP53-involved signaling network in 20q13 DERE amplification. (D) Heat map of 133 differentially expressed genes in three ERα-positive cell lines (LY2, BT-474, and MCF-7) with high CN (n>50) *versus* three low-CN (n<15) cell lines (600MPE, MDA-MB-361, and ZR-7530). Arrow heads, genes involved in the identified TP53/BRCA1-associated signaling network (see Figure S3F). (E) The Ingenuity Pathway Analysis (IPA) version 8.8 (Ingenuity Systems, http://www.ingenuity.com) was used to identify different signaling networks. The aforementioned 133 differentially expressed genes were subjected to and eligible for the analysis (see also Table S9).

(F) IPA was applied to determine networks associated with DNA break-repair

mechanisms. The results indicated that five up-regulated (red ovals) and three down-regulated (green ovals) genes were significantly involved in TP53/BRCA1-associated signaling network (p=0.00743)

Table S9, related to Figure 4: Differentially Expressed Genes Associated with 20q13 DERE Copy Variation

Gene	Description	Cellular Location
ABCC4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	Plasma Membrane
ABLIM1	actin binding LIM protein 1	Cytoplasm
ADAMTSL3	ADAMTS-like 3	unknown
APOL1	apolipoprotein L, 1	Extracellular Space
ARAP2	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2	Cytoplasm
ARL6IP5	ADP-ribosylation-like factor 6 interacting protein 5	Cytoplasm
ASNS	asparagine synthetase (glutamine-hydrolyzing)	unknown
ATP8A1	ATPase, aminophospholipid transporter (APLT), class I, type 8A, member 1	Cytoplasm
AZGP1	alpha-2-glycoprotein 1, zinc-binding	Extracellular Space
BASP1	brain abundant, membrane attached signal protein 1	Plasma Membrane
BMP7	bone morphogenetic protein 7	Extracellular Space
C10ORF81	chromosome 10 open reading frame 81	unknown
C1ORF59	chromosome 1 open reading frame 59	unknown
C20ORF114	chromosome 20 open reading frame 114	Extracellular Space
C20ORF95	Rho GTPase activating protein 40	unknown
C2ORF54	chromosome 2 open reading frame 54	unknown
C4ORF19	chromosome 4 open reading frame 19	unknown
CALD1	caldesmon 1	Cytoplasm
CAPN9	calpain 9	Cytoplasm
CCNG1	cyclin G1	Nucleus
CDC25B	cell division cycle 25 homolog B (S. pombe)	Nucleus
CLN5	ceroid-lipofuscinosis, neuronal 5	Cytoplasm
CYBRD1	cytochrome b reductase 1	Cytoplasm
CYP2B6	cytochrome P450, family 2, subfamily B, polypeptide 6	Cytoplasm
CYR61	cysteine-rich, angiogenic inducer, 61	Extracellular Space
DAK	dihydroxyacetone kinase 2 homolog (S. cerevisiae)	unknown
DCAF4	DDB1 and CUL4 associated factor 4	unknown
DHCR7	7-dehydrocholesterol reductase	Cytoplasm
DNAJB4	DnaJ (Hsp40) homolog, subfamily B, member 4	Nucleus
DNAJC6	DnaJ (Hsp40) homolog, subfamily C, member 6	Cytoplasm
DNALI1	dynein, axonemal, light intermediate chain 1	Cytoplasm

(continued)

Gene	Description	Cellular Location
DNMT3B	DNA (cytosine-5-)-methyltransferase 3 beta	Nucleus
DOCK3	dedicator of cytokinesis 3	Cytoplasm
DUS4L	dihydrouridine synthase 4-like (S. cerevisiae)	unknown
EGF	epidermal growth factor	Extracellular Space
ELP2	elongation protein 2 homolog (S. cerevisiae)	Cytoplasm
EPHB4	EPH receptor B4	Plasma Membrane
FADS1	fatty acid desaturase 1	Plasma Membrane
FAM114A1	family with sequence similarity 114, member A1	unknown
FAT1	FAT tumor suppressor homolog 1 (Drosophila)	Plasma Membrane
FMOD	fibromodulin	Extracellular Space
GATM	glycine amidinotransferase (L-arginine:glycine amidinotransferase)	Cytoplasm
GBP2	guanylate binding protein 2, interferon-inducible	Cytoplasm
GBP3	guanylate binding protein 3	Cytoplasm
GCH1	GTP cyclohydrolase 1	Cytoplasm
GCLM	glutamate-cysteine ligase, modifier subunit	Cytoplasm
GCNT2	glucosaminyl (N-acetyl) transferase 2, I-branching enzyme (I blood group)	Cytoplasm
GPR180	G protein-coupled receptor 180	unknown
GPR98	G protein-coupled receptor 98	Plasma Membrane
HEATR6	HEAT repeat containing 6	unknown
HES1	hairy and enhancer of split 1, (Drosophila)	Nucleus
HFE	hemochromatosis	Plasma Membrane
HLTF	helicase-like transcription factor	Nucleus
HNRNPAB	heterogeneous nuclear ribonucleoprotein A/B	Nucleus
IL20RA	interleukin 20 receptor, alpha	Plasma Membrane
INPPL1	inositol polyphosphate phosphatase-like 1	Cytoplasm
KCNS3	potassium voltage-gated channel, delayed-rectifier, subfamily S, member 3	Plasma Membrane
LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein	Plasma Membrane
LMO7	LIM domain 7	Cytoplasm
LPGAT1	lysophosphatidylglycerol acyltransferase 1	Cytoplasm
LXN	latexin	Cytoplasm
MAPT	microtubule-associated protein tau	Cytoplasm
MMRN2	multimerin 2	Extracellular Space

(continued)

Gene	Description	Cellular Location
MOCS3	molybdenum cofactor synthesis 3	Cytoplasm
MST1R	macrophage stimulating 1 receptor (c-met-related tyrosine kinase)	Plasma Membrane
NBN	nibrin	Nucleus
NCOA3	nuclear receptor coactivator 3	Nucleus
NTN4	netrin 4	Extracellular Space
ORC5L	origin recognition complex, subunit 5-like (yeast)	Nucleus
PARD6B	par-6 partitioning defective 6 homolog beta (C. elegans)	Plasma Membrane
PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	Cytoplasm
PDGFRL	platelet-derived growth factor receptor-like	Plasma Membrane
PFDN4	prefoldin subunit 4	Cytoplasm
PLCE1	phospholipase C, epsilon 1	Cytoplasm
PLS3	plastin 3	Cytoplasm
PMPCB	peptidase (mitochondrial processing) beta	Cytoplasm
PNMA1	paraneoplastic antigen MA1	Nucleus
PNPLA8	patatin-like phospholipase domain containing 8	Cytoplasm
PPP1R3C	protein phosphatase 1, regulatory (inhibitor) subunit 3C	Cytoplasm
PRMT6	protein arginine methyltransferase 6	Nucleus
PRTG	protogenin	unknown
PSCA	prostate stem cell antigen	Plasma Membrane
PSMG1	proteasome (prosome, macropain) assembly chaperone 1	Plasma Membrane
PTCD1	pentatricopeptide repeat domain 1	unknown
PUS7	pseudouridylate synthase 7 homolog (S. cerevisiae)	unknown
RINT1	RAD50 interactor 1	Nucleus
RNASEH2A	ribonuclease H2, subunit A	Nucleus
RNF217	ring finger protein 217	unknown
RPP40	ribonuclease P/MRP 40kDa subunit	Nucleus
RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1	Cytoplasm
RRAS	related RAS viral (r-ras) oncogene homolog	Cytoplasm
SCGB1D2	secretoglobin, family 1D, member 2	Extracellular Space
SCGB2A2	secretoglobin, family 2A, member 2	Extracellular Space
SCNN1A	sodium channel, nonvoltage-gated 1 alpha	Plasma Membrane
SGCE	sarcoglycan, epsilon	Plasma Membrane

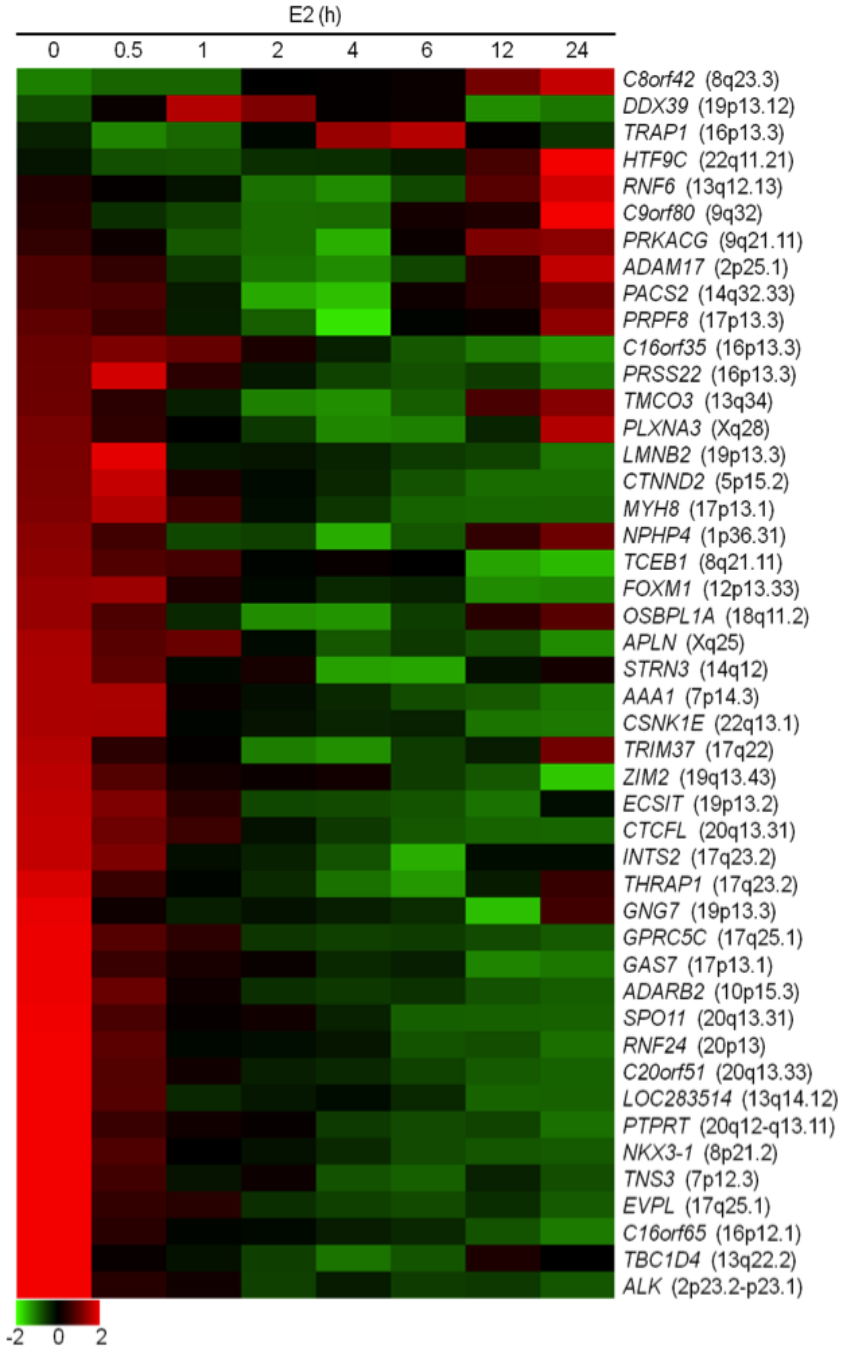
(continued)

Gene	Description	Cellular Location
SIM2	single-minded homolog 2 (Drosophila)	Nucleus
SLC15A2	solute carrier family 15 (H+/peptide transporter), member 2	Plasma Membrane
SLC41A2	solute carrier family 41, member 2	unknown
SLC44A3	solute carrier family 44, member 3	unknown
SLC44A5	solute carrier family 44, member 5	unknown
SLC46A3	solute carrier family 46, member 3	Extracellular Space
SNCG	synuclein, gamma (breast cancer-specific protein 1)	Cytoplasm
SNX10	sorting nexin 10	Cytoplasm
SPATA2	spermatogenesis associated 2	Extracellular Space
STAT6	signal transducer and activator of transcription 6, interleukin-4 induced	Nucleus
STC2	stanniocalcin 2	Extracellular Space
STX16	syntaxin 16	Cytoplasm
SUSD2	sushi domain containing 2	Extracellular Space
TCEAL8	transcription elongation factor A (SII)-like 8	unknown
TFAP2B	transcription factor AP-2 beta (activating enhancer binding protein 2 beta)	Nucleus
THOC3	THO complex 3	Nucleus
TLR1	toll-like receptor 1	Plasma Membrane
TLR3	toll-like receptor 3	Plasma Membrane
TMCO3	transmembrane and coiled-coil domains 3	unknown
TMEM40	transmembrane protein 40	unknown
TMPRSS3	transmembrane protease, serine 3	Plasma Membrane
TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	Extracellular Space
TLR1	toll-like receptor 1	Plasma Membrane
TLR3	toll-like receptor 3	Plasma Membrane
TMCO3	transmembrane and coiled-coil domains 3	unknown
TMEM40	transmembrane protein 40	unknown
TMPRSS3	transmembrane protease, serine 3	Plasma Membrane
TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	Extracellular Space
TNFSF15	tumor necrosis factor (ligand) superfamily, member 15	Extracellular Space
TP53	tumor protein p53	Nucleus
TPST1	tyrosylprotein sulfotransferase 1	Cytoplasm
TRIM29	tripartite motif-containing 29	Cytoplasm

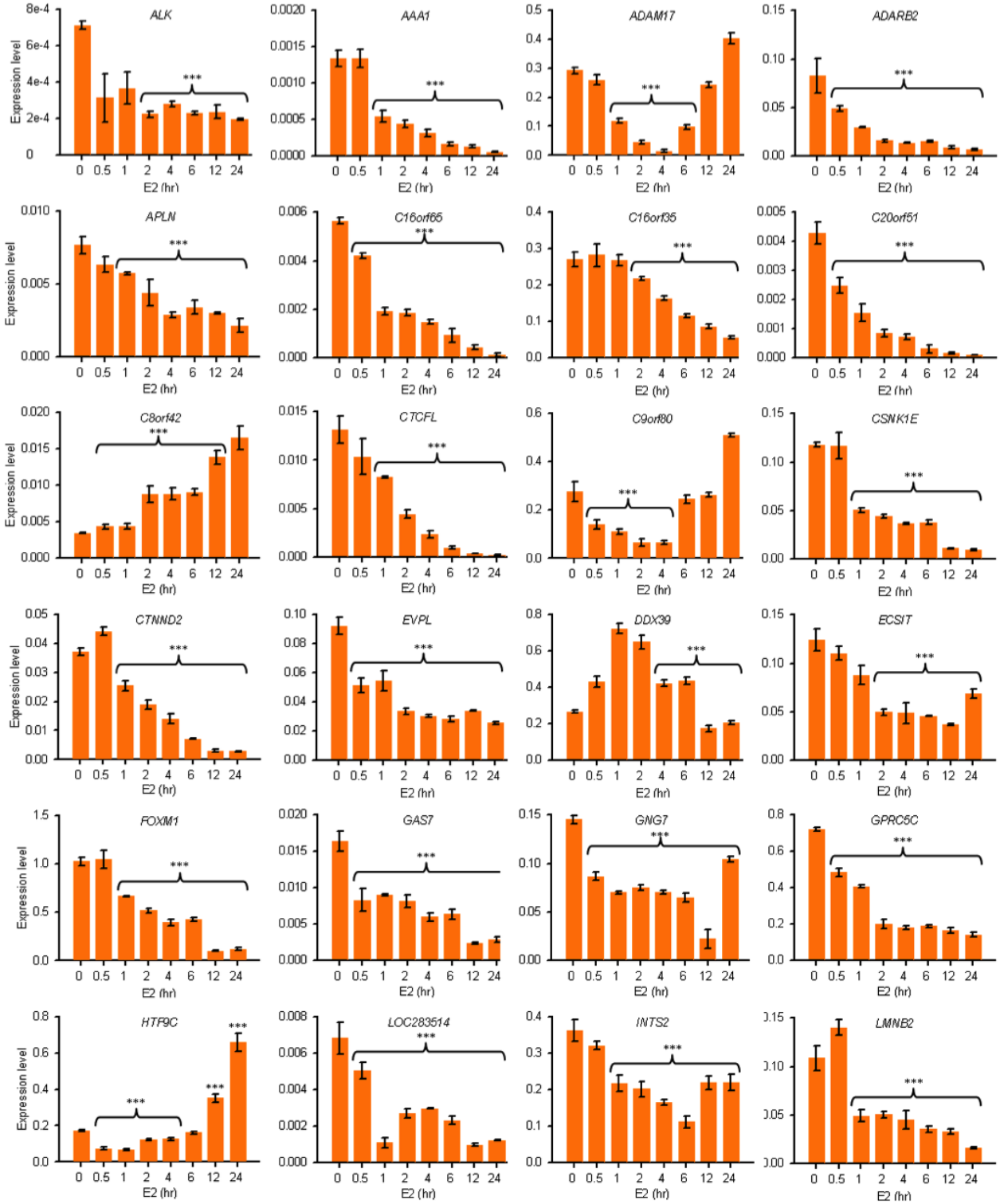
(continued)

Gene	Description	Cellular Location
TRIM38	tripartite motif-containing 38	unknown
TSPYL5	*TSPY-like 5*	*unknown*
VHL	*von Hippel-Lindau tumor suppressor*	*Nucleus*
WBP5	*WW domain binding protein 5*	*unknown*
ZFP37	*zinc finger protein 37 homolog (mouse)*	*Nucleus*
ZNF114	*zinc finger protein 114*	*unknown*
ZNF425	*zinc finger protein 425*	*unknown*
ZNF75A	*zinc finger protein 75a*	*Nucleus*
ZNF786	*zinc finger protein 786*	*Nucleus*
ZSCAN21	*zinc finger and SCAN domain containing 21*	*Nucleus*

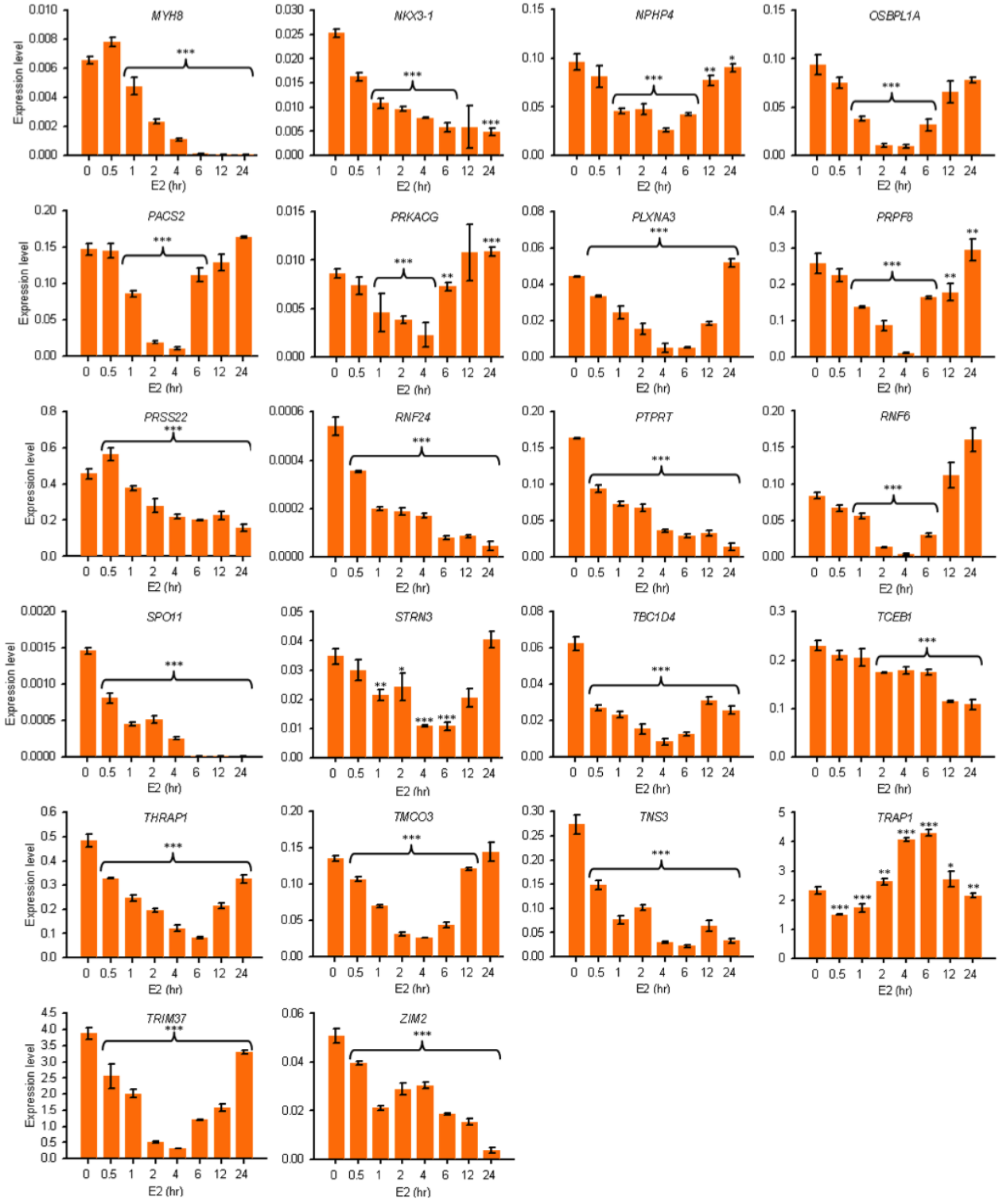
A


E2 (h)
0
0.5
1
2
4
6
12
24
C8orf42 (8q23.3)
DDX39 (19p13.12)
TRAP1 (16p13.3)
HTF9C (22q11.21)
RNF6 (13q12.13)
C9orf80 (9q32)
PRKACG (9q21.11)
ADAM17 (2p25.1)
PACS2 (14q32.33)
PRPF8 (17p13.3)
C16orf35 (16p13.3)
PRSS22 (16p13.3)
TMCO3 (13q34)
PLXNA3 (Xq28)
LMNB2 (19p13.3)
CTNND2 (5p15.2)
MYH8 (17p13.1)
NPHP4 (1p36.31)
TCEB1 (8q21.11)
FOXM1 (12p13.33)
OSBPL1A (18q11.2)
APLN (Xq25)
STRN3 (14q12)
AAA1 (7p14.3)
CSNK1E (22q13.1)
TRIM37 (17q22)
ZIM2 (19q13.43)
ECSIT (19p13.2)
CTCFL (20q13.31)
INTS2 (17q23.2)
THRAP1 (17q23.2)
GNG7 (19p13.3)
GPRC5C (17q25.1)
GAS7 (17p13.1)
ADARB2 (10p15.3)
SPO11 (20q13.31)
RNF24 (20p13)
C20orf51 (20q13.33)
LOC283514 (13q14.12)
PTPRT (20q12-q13.11)
NKX3-1 (8p21.2)
TNS3 (7p12.3)
EVPL (17q25.1)
C16orf65 (16p12.1)
TBC1D4 (13q22.2)
ALK (2p23.2-p23.1)
-2
0
2

B


ALK
AAA1
ADAM17
ADARB2
APLN
C16orf65
C16orf35
C20orf51
C8orf42
C7CFL
C9orf80
CSNK1E
CTNND2
EVPL
DDX39
ECSIT
FOXM1
GAS7
GNG7
GPRC5C
HTF9C
LOC283514
INTS2
LMNB2
Expression level
E2 (hr)
0 0.5 1 2 4 6 12 24
***

B (continued)


MYH8
NKX3-1
NPHP4
OSBPL1A
PACS2
PRKACG
PLXNA3
PRPF8
PRSS22
RNF24
PTPRT
RNF6
SPO11
STRN3
TBC1D4
TCEB1
THRAP1
TMCO3
TNS3
TRAP1
TRIM37
ZIM2
Expression level
E2 (hr)
0 0.5 1 2 4 6 12 24

C

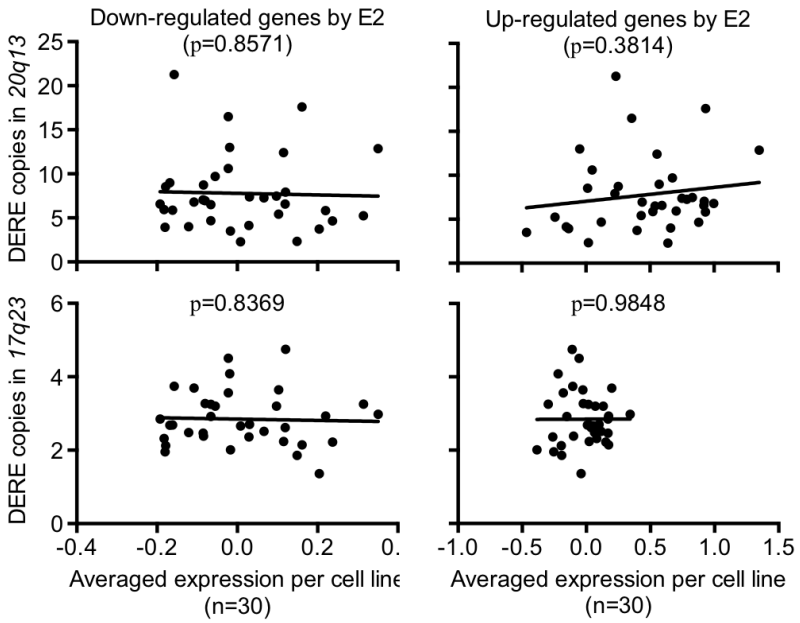


Figure S4, related to Figure 5. Expression Analysis of 46 Target Genes Interacting with 20q13 DEREs and Correlation Analysis of DERE Copy Changes and DERE-regulated Target Gene Expression in ERα-negative Breast Cancer Cell Lines

(A-B) Quantitative RT-PCR analysis was performed on MCF-7 cells treated with E2 (70 nM) in eight time periods (0, 0.5, 1, 2, 4, 6, 12, and 24 hr). Data were summarized in a heat map (A) and shown in 46 bar charts (B) with individual genes. Mean ± SD (n=6). ***, $p<0.001$; **, $p<0.01$; *, $p<0.05$ (Student's *t* test), comparing to control cells (time point "0").

(C) Expression microarray data of ICBP cell lines (Heiser et al., 2009) was integrated with experimental copy-number results of 20q13 and 17q23 DEREs to study the correlation between DERE amplification and transcription control in ERα-negative breast cancer cell lines (n=30).

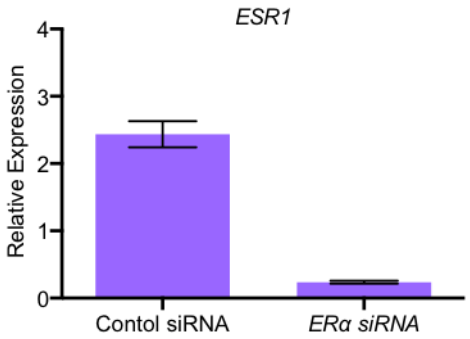


Figure S5, related to Figure 6. Quantitative RT-PCR Analysis of *ESR1* Expression in MCF-7 Transfectants

After transient transfection of either ERα siRNA or scramble siRNA (Control siRNA) into MCF-7 cells, total RNA was collected for expression analysis.

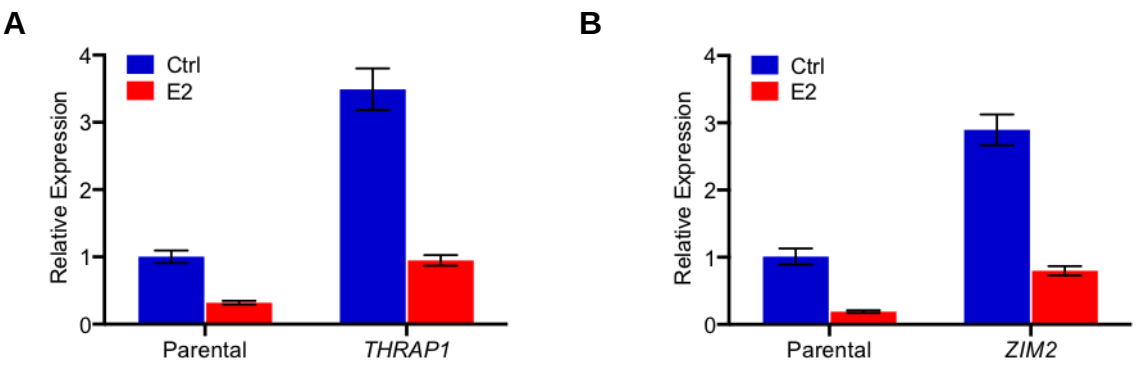
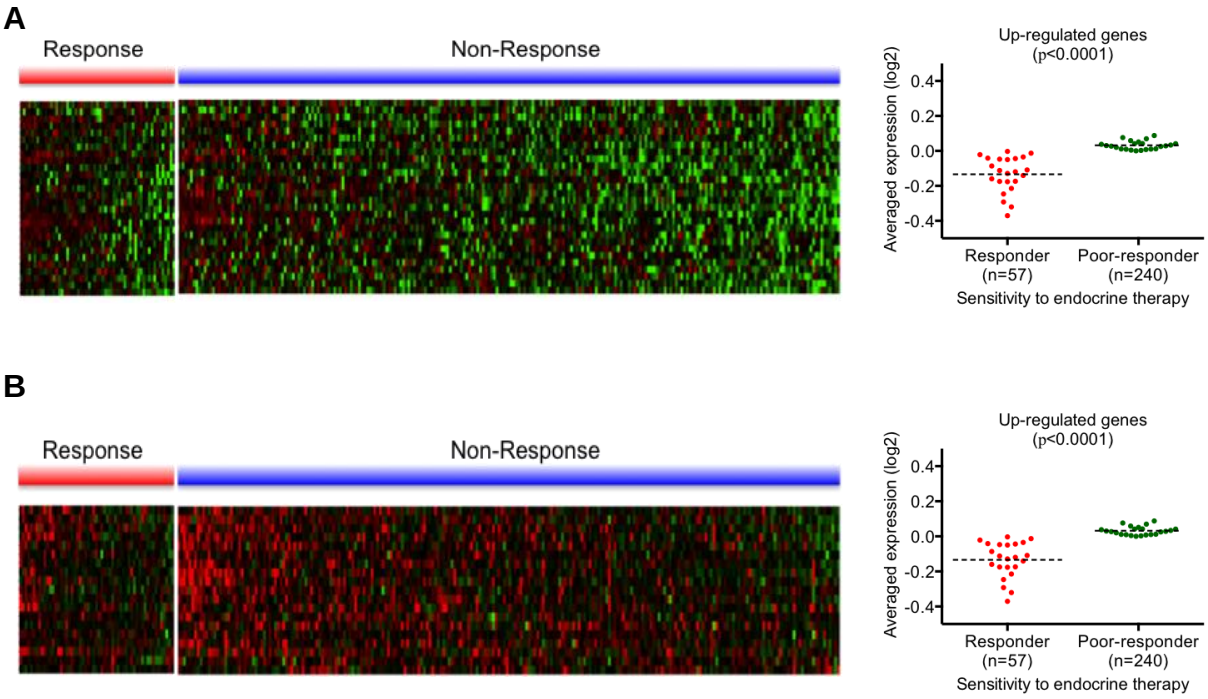
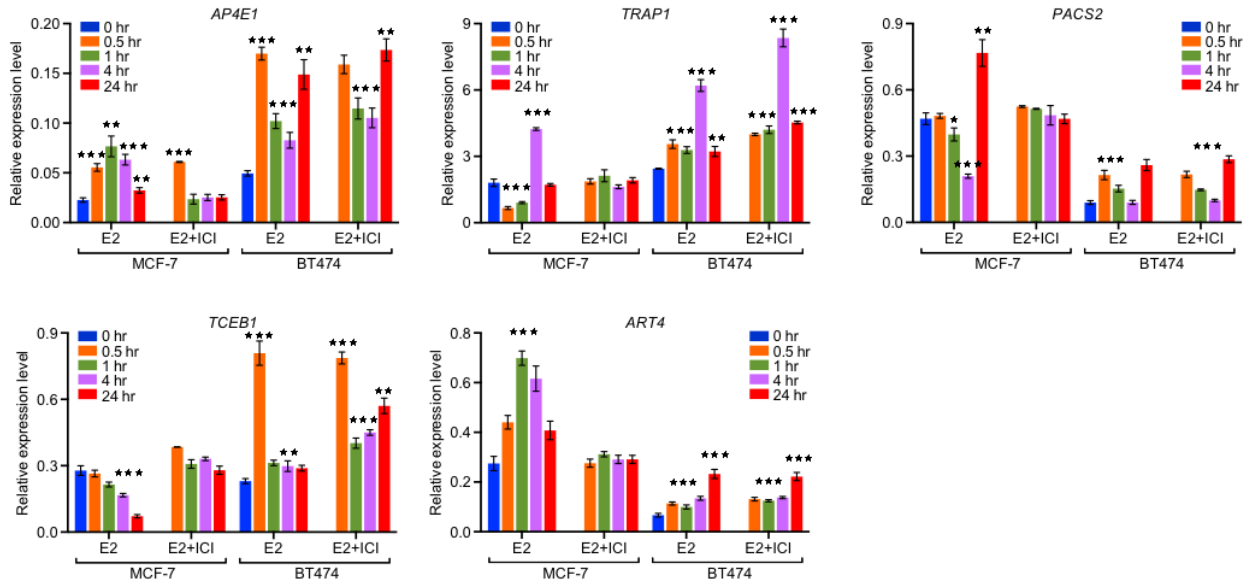


Figure S6, related to Figure 7. Quantitative RT-PCR Analysis of *THRAP1* and *ZIM2* Expression in MCF-7 Transfectants

After transient transfection of *THRAP1* (A) or *ZIM2* (B) plasmids into MCF-7 cells, total RNA from both untreated and E2-treated (24hr) cells was collected and then subjected to quantitative RT-PCR analysis.


A
Response
Non-Response
Up-regulated genes
(p<0.0001)
Averaged expression (log2)
0.4
0.2
0.0
-0.2
-0.4
Responder
(n=57)
Poor-responder
(n=240)
Sensitivity to endocrine therapy
B
Response
Non-Response
Up-regulated genes
(p<0.0001)
Averaged expression (log2)
0.4
0.2
0.0
-0.2
-0.4
Responder
(n=57)
Poor-responder
(n=240)
Sensitivity to endocrine therapy


C
AP4E1
TRAP1
PACS2
TCEB1
ART4
Relative expression level
0 hr
0.5 hr
1 hr
4 hr
24 hr
E2
E2+ICI
MCF-7
BT474

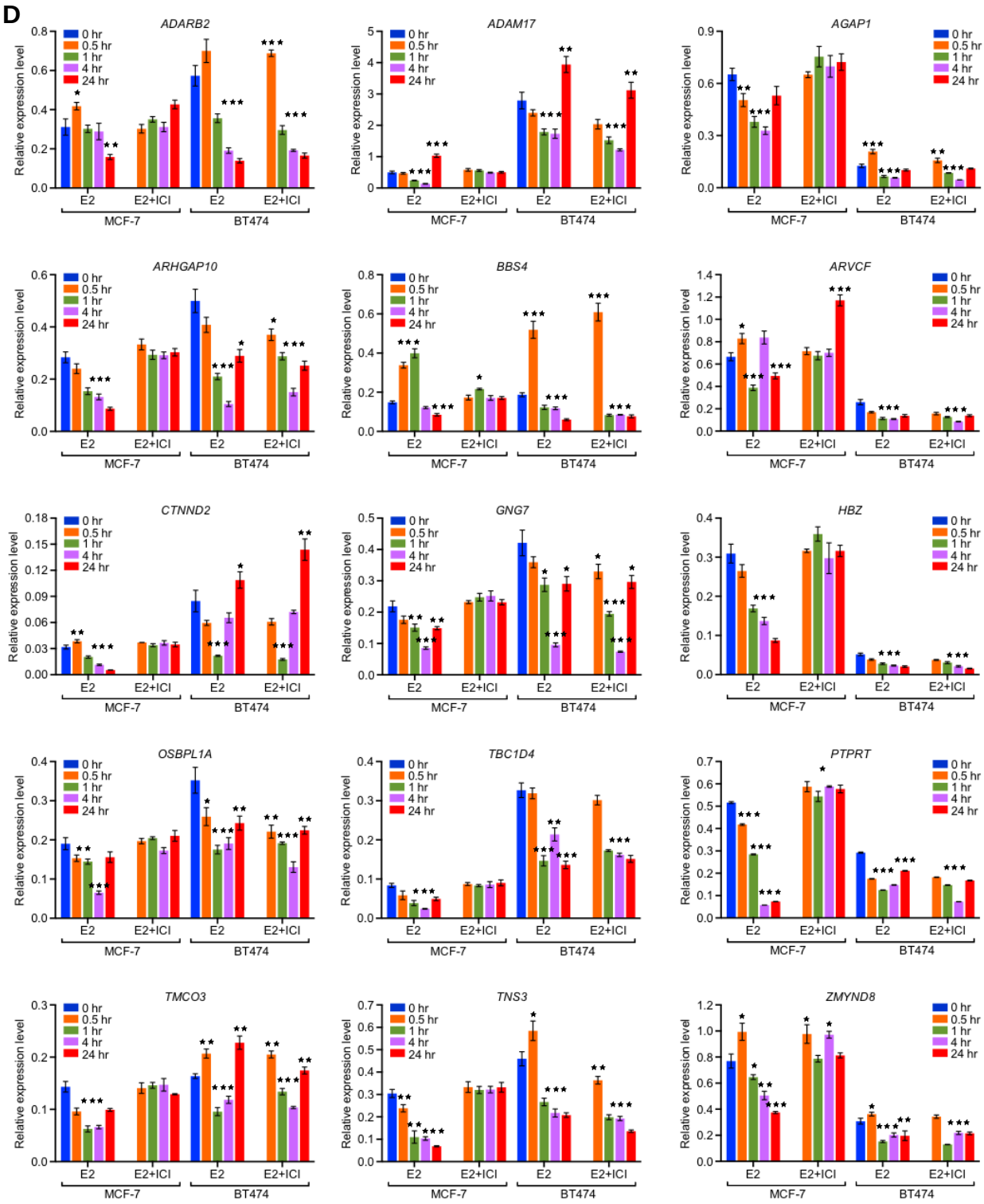


Figure S7, related to Figure 8. Expression Signature of DERE-interacting Genes Correlates with Relapse after Hormone Therapy

(A-B) Expression signature of DERE-interacting targets is associated with relapse after endocrine therapy. A case-control breast tumor cohort was applied to interrogate whether the gene signature identified in Figure 8A-B is associated with the resistant response of patients after endocrine therapy. This cohort (n=508) comprises of 439 patients with low-sensitivity to endocrine therapy and 69 with either high- or intermediate-sensitivity to treatment (Hatzis et al., 2011). Expression profiling from a total of 297 patients with ERα-positive breast tumors was analyzed. (A), Down-regulated genes. (B) Up-regulated genes.

(C-D) Expression analysis of 26 DERE-regulated target genes involved in tamoxifen resistance. Quantitative RT-PCR analysis was performed on MCF-7 and BT474 cells treated with E2 (70 nM) and ERα antagonist, ICI-182-7820 (ICI, 1 μM) in five time periods (0, 0.5, 1, 4, and 24 hr). (C) Up-regulated genes. (D) Down-regulated genes. Data were shown in 26 bar charts with individual genes (see also Figure 8C-D). Mean ± SD (n=6). ***, $p<0.001$; **, $p<0.01$; *, $p<0.05$ (Student's *t* test), comparing to control cells (time point "0 hr").

Supplemental Experimental Procedures

Chromosome Conformation Capture Assay Coupled with Paired-end Sequencing (3C-seq)

MCF-7 cells were stimulated with E2 (70 nM) or DMSO (Ctrl) for 24 hr. The rationale of choosing this 24-hr time point for 3C-seq was based on our recent observation that progressive increases in the frequency of ERα-mediated looping might preferentially occur during the period of estrogen stimulation (Hsu et al., 2010). Two biological replicates of treated cells were then subjected to chromosome conformation capture assay as previously described (Hagège et al., 2007). Briefly, cells were fixed with 1% formaldehyde. Chromatin was digested with *Bam*HI, and then ligated by T4 DNA ligase in diluted conditions. Ligated DNA was then de-crosslinked, purified by phenol extraction procedures, and subjected to paired-end sequencing using the Illumina sequencing technology platform. Sample preparation for paired-end sequencing was performed following the manufacturer's instructions. Briefly, ligated DNA (5 µg) was randomly sheared by a nebulizer supplied with the Illumina paired-end sample preparation kit. Fragmented DNA was end-paired using T4 DNA polymerase and Klenow polymerase with T4 polynucleotide kinase to phosphorylate the 5' ends. A 3' overhang was created using a 3'-5' exonuclease-deficient Klenow fragment, and Illumina paired-end adaptor oligonucleotides were ligated to the sticky ends thus created. The ligation mixtures were electrophoresed on E-gel® SizeSelect™ 2% pre-cast agarose gels (Invitrogen) to collect 250-bp fragments. Size-selected DNA fragments were enriched with Illumina paired-end primers by a 12-cycle PCR reaction. DNA samples (20 nM per sample), quantified by an Agilent Bioanalyzer, were loaded

onto the paired-end flowcell of GAIIx in the supplied cluster station according to the manufacturer's protocol. Clusters of PCR colonies were then sequenced on GAIIx with 36-bp per read.

Mate-pair Sequencing

Genomic DNA of untreated MCF-7 cells was subjected to mate-pair sequencing using the Illumina sequencing technology platform. We constructed mate-pair paired end libraries, prepared flowcells, and generated clusters following the manufacturer's instructions. Briefly, genomic DNA (10 μg) was randomly sheared by a nebulizer supplied with the Illumina mate-pair preparation kit. Fragmented DNA was end-paired using T4 DNA polymerase and Klenow polymerase, and 5' ends phosphorylated by T4 polynucleotide kinase. Biotinylated nucleotides were then incorporated to the repaired ends of the DNA. Biotin-labeled DNA was electrophoresed on E-gel® SizeSelect™ 0.8% pre-cast agarose gels to collect 2.5- to 4.5-kb fragments and quantified by an Agilent Bioanalyzer. Size-selected DNA (600 ng) was circularized by circularization ligase supplied by the kit. Circularized DNA was randomly fragmented by a nebulizer and then end-repaired by kinases as mentioned above. A 3' overhang was created using a 3'-5' exonuclease-deficient Klenow fragment, and Illumina paired-end adaptor oligonucleotides were ligated to the sticky ends thus created. The ligation reactions were carried out on the biotinylated DNA immobilized to the streptavidin beads. Adaptor-modified DNA fragments were enriched by an 18-cycle PCR reaction with Illumina paired-end primers. DNA fragments (300-bp) were collected by E-gel® SizeSelect™ 2% pre-cast agarose gels. DNA (20 nM per sample), quantified by an

Agilent Bioanalyzer, was subjected to the paired-end flowcell of GAIIx on the supplied cluster station. Clusters of PCR colonies were then sequenced on GAIIx with 51-bp per read.

Chromatin Immunoprecipitation Coupled with Single-end Sequencing (ChIP-seq)

After culture in media containing charcoal-stripped serum for 24 hr, MCF-7 cells were stimulated with E2 for varying lengths of time (0, 0.5, 1, and 24 hr) and subjected to ChIP assays using ERα monoclonal antibody (sc-8005X, Santa Cruz Biotechnology). Immunoprecipitated DNA was then applied to single-end sequencing using the Illumina sequencing technology platform following the manufacturer's instructions. Briefly, ChIP DNA (10 ng, quantified by the Qubit™ quantitation platform) was end-paired using T4 DNA polymerase and Klenow polymerase with T4 polynucleotide kinase to phosphorylate the 5' ends. A 3' overhang was created using a 3'-5' exonuclease-deficient Klenow fragment, and Illumina single-end adaptor oligonucleotides were ligated to the sticky ends thus created. The ligation mixtures were electrophoresed on E-gel® SizeSelect™ 2% pre-cast agarose gels (Invitrogen) to collect 150-bp fragments. Size-selected DNA fragments were enriched with Illumina single-end primers by a 15-cycle PCR reaction. DNA samples (20 nM per sample; 2 lanes per sample) quantified by an Agilent Bioanalyzer were subjected to the single-end flowcell of GAIIx on the supplied cluster station. Clusters of PCR colonies were then sequenced on GAIIx with 51-bp per read. ERα ChIP-seq data were analyzed as previously described (Hsu et al., 2010).

Genomic Mapping of Sequence Reads

Reads generated from the Illumina GAIIx pipeline from mate-pair seq, ChIP-seq, and 3C-seq datasets were aligned to the Human Genome Assembly (NCBI build 36.1/hg18) using ELAND algorithm. Redundant tags were removed to reduce DNA amplification and system biases. Fragments with one/both end(s) uniquely mapped to the human genome were kept for further processing.

Statistical Analyses of Next-generation Sequencing Data

Latent Class Poisson Regression Modeling. In addition to the biases presented in sequencing data (e.g., unequal efficiency of DNA amplification, copy number differences, existence of amplicon, sequencing bias, image processing, and matching errors), self- or random-ligation that also occurs during genomic library preparation also give rise to false-positive results. A latent class Poisson regression model was built to define true-positive genomic fusions and chromatin interactions (Wedel et al., 1993; Yang et al., 2005).

The proximate ligation and random ligation events were defined as two independent Poisson distributions. A latent class model with hidden variable showed the overall ligation event. Hence, the probability of Y_i from a particular class *k* is given by

$$f_k(y_i|\lambda_{i|k}) = \frac{e^{-\lambda_{i|k}}(\lambda_{i|k})^{y_i}}{y_i!}, \qquad i = 1, \dots, n, k = 1, 2.$$

where $\lambda_{i|k}$ is the mean rate of individual ligation event (denoted as *i*) in either proximate or random ligation class (denoted as *k*), and n is the total number of ligation events. The canonical log link function was then used to transform the mean of the Poisson

distribution to linear predictor $\boldsymbol{\beta_k}$ as

$$\ln(\lambda_{i|k}) = \beta_{0k} + \sum_{l=1}^{L} x_{il}\,\beta_{0k}$$

where $x_{i1}, \ldots, x_{iL}$ represents the explanatory variables. Expectation Maximization (EM) algorithm (Dempster et al., 1977; Wedel et al., 1993; Yang et al., 2005) was applied to estimate the unknown parameter $\boldsymbol{\beta_k}$ as well as α_k which is the proportion of *k*th class in all ligation events with $\sum_{k=1}^{2} \alpha_k = 1,\ 0 < \alpha_k < 1$. Given the threshold enrichment of fusion fragments as *t*, the false discovery rate (*FDR*) of proximate ligation could be estimated using the following formula,

$$FDR = \frac{\alpha_2[1 - F(t-1, \lambda_2)]}{\alpha_1[1 - F(t-1, \lambda_1)] + \alpha_2[1 - F(t-1, \lambda_2)]}$$

where $F(t-1, \lambda)$ is the cumulative distribution function of the Poisson distribution.

Identification of Translocation Fusion Fragments Based On Mate-Pair Sequencing Data.

The mixture Poisson regression model was utilized to reduce the false-discovery rate (*FDR*). Since the sequencing depth of the mate-pair sequencing dataset (34 lanes) is high, we set a higher threshold (t = 10) to filter out randomly ligated fusions and define the fusion events with $t \geq 10$ as positive translocation fusions. This criterion resulted in a *FDR* of close to 0 for identifying translocation fusions from the mate-pair sequencing dataset.

Another potential factor contributing to a *FDR* is mapping errors caused by translocations of small DNA elements, resulting in individual variations of DNA sequence. To increase the specificity on defining translocation fusion fragments, we

excluded all translocated fragments from small regions (<300bp). In contrast to the ordered orientation of tags derived from translocation breakpoints, the repetitive regions, especially centromeric regions, often show a random orientation of tags. Therefore, we removed all potential translocation fusion fragments with random orientation of tags to further reduce the *FDR*.

Re-construction of 20q13-associated Derivative Chromosomes Harboring DERE-DERE Fusions. Mate-pair sequencing data indicated that the 17*q23*::*20q13* fusion was the most frequent event (Figure 2C), suggesting that complex rearrangements start from the formation of this translocation in MCF-7 cells. The probable sequence of multiple translocations was then reconstructed for the seven derivative chromosomes (T1 to T7), based on the assumption that these translocations were originally derived from a normal chromosome 20 through minimum steps of rearrangements including breaks, intra-/inter-chromosomal fusions, and amplification (Figure S2C-E). Moreover, structural fusions found in MCF-7 cells suggest that breaks, fusions and amplification are driving forces for this complex aberration.

1) Determination of the orientation of "joint DNA". Structural fusion fragments identified by mate-pair sequencing were termed as "*joint DNA*". Tags from sequencing results were used to determine the orientation of two joining DNA fragments. If two tags of a single joint DNA are mapped to a different strand (forward or reverse), this joint DNA has an identical orientation.

2) Definition of "the chronological sequence" in amplification. If a juncture of joint DNA caused by fusions or amplifications in the early steps is located in amplification-

susceptible regions, it may have multiple copies by subsequent amplifications in the MCF-7 genome. Based on the above assumption, we proposed that rearrangement events easily occurred in the aforementioned joint DNA to produce higher copies of the junctures in a certain genomic region. Therefore, “the chronological sequence” in rearrangements could be determined based on the copy numbers of the junctures.

Reproducibility, Saturation, and Sensitivity Tests of 3C-seq. 3C-seq was performed in two biological replicates to determine whether these datasets can reliably survey chromatin interaction events in a genome-wide scale. The interaction events were categorized into control (Ctrl) and treated (E2) in each replicate. Scatter plots were then conducted based on log2 value of chromatin interaction events per Mb. The linear regression analysis was made using SigmaPlot 11. The entire genome was divided into bins 250-bp in length. Pairs of bins which contain two or more long-range (>1-kb or inter-chromosome) read-pairs were counted as positive chromatin interaction events and the maxima number of chromatin interaction events was estimated based on the above criterion. Oppositely, pairs of bin that contain only one long-range read-pair were considered as errors or random events. To examine whether our 3C-seq data provide sufficient sequencing depth for identifying chromatin interactions in a genome-wide scale, we tried to generate reduced-scale datasets by randomly extracting 5%, 10%, ... 95% data from all the read-pairs to investigate if the aforementioned criterion was applied in these reduced datasets. Each data scale was examined twice to increase the statistical confidence.

The entire genome was divided into ~3 million bins (100-bp in length). Pairs of bins

were further categorized into five groups based on the distance of the pairs: <1, 1-5, 5-10, >10-kb and inter-chromosome. We analyzed interactions of bin pairs for the four groups (1-5, 5-10, >10-kb, and inter-chromosome) separately. Specifically, for two bins A and B, we define the "frequency of interaction" as the count of sequence pairs such that one member of the pair overlaps with A and the other overlaps with B. We fitted a mix negative binomial model to the distributions of the count, and used the approach developed by Efron and his coworkers (Efron et al., 2001) to compute sensitivity and *FDR* associated with the detection criterion of ≥2 counts.

Identification of Significant Chromatin Interaction Events from 3C-Seq Data

1) Definition of inter- and intra-chromatin Interactions. The aforementioned Poisson regression model was not able to eliminate self-ligation events, which possess spatial proximity on the two ends of a single DNA fragment. To further classify proximate ligation events based on the 3C-seq dataset, a high-stringent filtering criterion was applied. Briefly, a fusion fragment with two ends mapped to different chromosomes was defined as an inter-chromatin interaction fragment. If 1) the coordinates of the strand end from paired read #1 is larger than that of the strand end from paired read #2; 2) the distance between these two ends is less than 20-kb (Lieberman-Aiden et al, 2009), the fusion fragment was counted as a structural interaction. In addition, a non-self-ligated fusion fragment with both ends aligned to the same chromosome was classified as an intra-chromatin fusion fragment. Since the sequencing depth of the 3C-seq data is relatively lower (7 lanes per treatment, totally of 14 lanes) compared to the mate-pair sequencing dataset (a total of 34 lanes), we lowered the threshold *t* to 2 (*FDR*=8.35%).

If the numbers of fragments, which have two ends mapped to the same two loci exceed the threshold *t*, these two loci were defined as interactive loci.

2) DERE-mediated chromatin interaction sites. With integration of the ERα ChIP-seq and the 3C-seq datasets, we define an ERα-bound DERE-interacting locus as follows: 1) at least one DERE is located within 10-kb of an interactive site; 2) if an DERE is located between 10-kb upstream to the transcriptional start site (TSS) and the 3' end of a gene, the gene is considered to be regulated by ERα. Two main regions localized in 17q23 and 20q13 regions containing a high density of ERα-bound DEREs in response to estrogen stimulation were defined as DERE clusters. Their genomic characteristics were described in Table S8.

Fluorescence *In Situ* Hybridization (FISH)

The probe mapped to the 20q13 region was prepared from a BAC (Invitrogen). The BAC clone was purified using a large-construct DNA kit (Qiagen) and labeled by nick translation using the Nick Translation kit (Vysis, Downer Groves, IL) following the manufacturer's recommendations. Briefly, 1 μg of the BAC clone was conjugated with either SpectrumGreen- or SpectrumRed-labeled dUTP, co-precipitated with 10X (v/v) human Cot-1 (Invitrogen), and dissolved in Hybridization Buffer (Sigma). The reaction was carried out for 8 hr at 15^{o} C and stopped by heating the sample to 70°C for 10 min.

Cells were harvested after mitotic arrest with 2 hr incubation of colcemid (1 μg/mL, Invitrogen) for metaphase FISH. In metaphase and interphase FISH analyses, cells were hypotonized with a 0.075 M KCl buffer, fixed by cold Carnoy's fixative, and spread on pre-cleaned slides. Hybridization was performed according to standard cytogenetic

methods; metaphase images were captured by Zeiss fluorescence microscope (Zeiss Axioscope 40) and analyzed with the Applied Imaging System. Interphase images were captured by Olympus fluorescence microscope (Olympus BX53TF) and analyzed with CellSens Dimension Imaging System.

For three-dimensional interphase FISH (3D-FISH), treated MCF-7 cells were subjected to 3D-FISH according to the protocol published by Steinhaeuser and coworkers (Steinhaeuser et al., 2002). The entire 3D-FISH procedure was conducted in cell suspension. Briefly, fixation was done in cold Carnoy's fixative and 0.005% pepsin solution was used for digestion. Hybridization was performed overnight at 37°C with pre-hybridized labeled-probes (80 ng per sample) and slides were washed in 0.4x (68°C for 2 min) and 4x SSC (42°C for 2 min) respectively. Nuclei counterstained with DAPI (0.1 µg/mL) were placed on a polished concave slide with Vectashield Mounting Medium (Vector Laboratories, Burlingame, CA). An Olympus FV1000 confocal microscope using single interference filter sets for red (SpectrumRed), and blue (DAPI), as well as dual (red/blue) filters, was utilize to capture images. Since images contain background noise, the standard 3D median filtering implemented in MATLAB software was used for denoising. First, two image stacks in single color were generated from the original dual-color images. The standard 3D median filtering was then applied on the single-color image stack. The neighborhood window for the median filter was set to be 5*5*3 voxels. It is noticed that the window size of x and y axes was larger than that of z axis, but the anisotropy sampling resolutions in x, y and z directions were balanced. Denoised images were re-constructed using a free bioimage visualization toolkit, called V3D (http://penglab.janelia.org/proj/v3d/V3D).

Reverse Transcription-quantitative PCR (RT-qPCR)

Total RNA (2 μg) was reversely transcribed to cDNA with oligo-dT (SuperScript III; Invitrogen). RT-qPCR was performed by using SYBR Green dye chemistry (Applied Biosystems) on a StepOnePlus™ Real-Time PCR System apparatus (Applied Biosystems). Gene expression was measured by the ΔΔCt method using *β-actin* as the internal control. Details of primer sequence for quantifications are provided in the following table.

Name	Forward	Reverse
AAA1	TACCTCCCTTGCTGAACGTC	CAGTGCTCCTGCACATGACT
ADAM17	CCTATTCCTGACCAGCGTG	ACCGAATGCTGCTGGATATT
ADARB2	GAGGGCTGAGCAGTCAACTC	TTTGTGATGCCAGGACTCAG
AGAP1	GCACGGCGCTGCATCT	GCGTAGGCCAGAGCTGTGTT
AJAP1	TGCCCGTGTACACCGATGA	CGAAGGCCACAGGAATAAGATG
ALK	CTTCCAGATCTTCGGGACTG	TGCGATGAGACAGGAAAGG
APLN	CTGCTCTGGCTCTCCTTGAC	CCATTCCTTGACCCTCTGG
AP4E1	CTCTGCAACAGAAACTAAGACTCCATA	ACTCGGCAATGCAGTAAGCA
ARHGAP10	CCCTTTTCTCCTCCTGCTACTGT	AGGTTCCCTGGAGGTTTGTACATC
ART4	TGGGAACCTGAGCACATATAACTG	CACTGGTCAAAAAGGAGAGAGATG
ARVCF	AACTGGCAGCCGGGATGT	GTCGAGTTTCAAGGGTTCCTTCTC
BBS4	CAAGCTCTAGGACAGGCAATGTC	GGGCTTTGTGAACTGGGATGT
BRF1	AGCGGGCCCGTGTCATA	CATCGCCATCACAGCCATAG
C16orf35	ACAAGTAAGCCGCGTAGCAG	TTTTGGCCACACATTTCAGA
C16orf65	CTCATCAGGAAGGGGGCT	TGGCATGGCCAACACTAAT
C20orf51	AAGCTGTCTTTCCATGGTGG	TGAGAGTAAAATTCCGCAATCA
C8orf42	CCTGACCATGTGGAAGCTG	TTTCCAACCTCGGAAACTTG
C9orf80	GGCAGCAAACTCTTCAGGAC	AGGGTCTCGAGAGTGCAATG
CSNK1E	AGCTACGTGTGGGGAACAAG	AGCTTGATGGCGACTTCCT
CTCFL	GCAAATTTCATCCCGACTGT	TTCCCCTGATCCACACTTCT
CTNND2	CTCTTCCCGGACTGAGGAG	GCTCAGGGAACTCGTCTTCT
DDX39	AGCAACTCGTGTCTGAGCG	CTGGGGCTCTTCCTCTTCAT
ECSIT	GCCCAGCAACAAGACAGAG	CTCCCCAGGCCCTACAGA
EVPL	CAAGGGGTCCCCCAAAG	CTCTTCTGCGTCTCCAGGAT
FOXM1	ATAGCAAGCGAGTCCGCATT	TCTCCTCTTTCCCTGGTCCT
GAS7	AGCAAAAAGCAGAGCAAGGA	CCTTGGGGGTCCTTCTTATC
GNG7	AGGGAGTGGAGCTTGGTTTC	CCCGTTGTTCAGAGAGCTTTT
GPRC5C	ACCCTTGAAGACAGCCATTG	AGGAAGAGAGGCAGTCCCAG
HBZ	GGTGAAGAGCATCGACGACATC	GCAGTGGGACAGGAGCTTGA
HTF9C	GATCTCTCCACACGCCTTCT	ACAGCACACGTCCAGCAC
INTS2	GGTGCCTGGTTCTTGTGTCT	AGCGTAAGGCATCTGTCCTG
LMNB2	GGCTCCTGCTCAAGATCTCA	CCAATCTCTATCTGCAGCCG
LOC283514	GCAACTTGTCTGTGTCGTCAA	GAGAGATGGTGAGGGTGGAA
MYH8	TGCCTACAGAGGCAAAAAGC	ATTCTCCGGTGATCAGGATG

(continued)

Name	Forward	Reverse
NKX3-1	CAGAGCCAGAGCCAGAGG	CAGATAAGACCCCAAGTGCC
NPHP4	CTTCCAGGAAACCACCCAC	CAGCTGGGAAATTGACAACC
OSBPL1A	GCCTTAATCTCTTCACCAAACAA	CGTCTCCAGTGCTTCTGACA
PACS2	ATGAACCTGTTCGCCACCT	TCCTTGAAGACCACCAGCTT
PLXNA3	AGCCAGCGGAGGGACAG	ACCACGAAGGCACGGAAG
PRKACG	TATTCAGGCCAGGTCTTGCT	TCCAAAGGGTAGGTGTGCTT
PRPF8	CGTGTGGACGATGAGTCAAT	AGCCTCCCGAAAGTATGTCA
PRSS22	AACACCCACTGCTGGATCTC	CTGCAGACTTCCGAGTCGAT
PTPRT	CCCTCAGCCTGCTCCTG	AACCCATTGGTCCCTAGAGC
RGN	CCGTGGATGCCTTTGACTATG	CCCCTCAGCATCAATACACATTC
RNF24	GAGGGTGAGTCCCTCACAAA	AGGCAGATTCTGGAATCCAAT
RNF6	GATGGCGTCAAGGAACAACT	TTCGAGTTGCATTTCCTGTG
SPO11	GGCTTGGTCTTCTCCCTTCT	GGTTGGCAGGTAACATAAGGTC
STRN3	CCGGGGATACTGCACTACAT	TGACCTTTTCTTTCGCCTTG
TBC1D4	CAGCCACGACCTCACCTACT	TGGCCGCTTTAGATAATTGC
TCEB1	GTGGCTGCGGGACTGAC	CAATTTGACATACATGGCATCAG
THRAP1	TCCTGGGATTTTGTTGAAGC	AAGATGGTGCCTGTCCTTGT
TMCO3	AGGCGACATTGACTACAGCA	CAGGATTCGGAGAACTTCCA
TNS3	CAGCATTTCCTGTGTCTCCA	CGTCCTCCTTTGTGCTGACT
TRAP1	TTGGAAAAACTGCGTCACAA	CACCAGCTCTTCCTGTGTCA
TRIM37	CCCCGGATGAAGATACACAT	TTAACCCATCAGGTGCAGTG
ZIM2	CAAAATGTGGTGGACCTCG	AGGGACAGAGTCCTGAGCAA
ZMYND8	AGGAGACGTCAGCTGAGAAAAGC	AGCCTTGGTTGGAGCCTAAGAG
β-actin	CAACTGGGACGACATGGAGA	ACGTACATGGTGGGGTGTTG

Chromosome Conformation Capture-Quantitative PCR

Charcoal-stripped MCF-7 cells stimulated with E2 (70 nM) were collected at different time-points of treatment (0, 0.5, 1, 4, and 24 hr). Treated cells were then subjected to 3C-qPCR analyses as previously described (Hagège et al., 2007). Briefly, fixed chromatin by 1% formaldehyde was digested using *Bam*HI (for inter-chromosomal interactions) or *Hin*dIII (for intra-chromosomal interactions at *THRAP1*), and then ligated by T4 DNA ligase in a diluted condition. Ligated DNA was then de-crosslinked and purified by classical phenol extraction procedures. Real-time PCR was performed on a 7500 Real-Time PCR System apparatus using the TaqMan technology (QuantiTect Probe PCR Master Mix, Qiagen). We used a 5'FAM-3'BHQ1 oligonucleotidic probe

(Invitrogen). To rule out the possibility of false-negative looping occurrence caused by unsuccessful 3C assay, we pooled two human bacterial artificial clones (BAC), mapping the interested regions as the positive control of the 3C-qPCR assays. These BACs were also used to examine the primer efficiency. For data analysis, the Ct obtained for each chimerical ligation fragment was processed using parameters of a standard curve (slope and intercept) from BAC to obtain quantification values that were normalized to a *GAPDH* loading control. Details of primer and probe sequence for quantifications are provided in the following table.

Name	Sequence
THRAP1-C1	GTTACCCACCCTTGACAGGAA
THRAP1-C2	GAAGGACACAGAACGGAGAGCTT
THRAP1-C3	AATTCAATTCACCTCCTGCAACTT
*THRAP1*_+52K	CGATCCTACACGGATATTTAGAGATGA
*THRAP1*_+48K	ATCGATCCTACACGGATATTTAGAGATG
*THRAP1*_+0.2K	CGGACTGCGGATCCTTCTC
*THRAP1*_-2.9K	GATCCTTCTCTCTGGGCCACTT
THRAP1-P1	TTCTTCTCCAAATGCAGCAGCTTCTA
THRAP1-P2	ACTACATAGTTCTATCTGATTCTAG
ZIM2-C1	AGGGTTTCATGAGGCATGTC
ZIM2-C2	CCCAAGGGATCCTCCTCTT
ZIM2-C3	GGGTTTCATGAGGCATGTCT
*ZIM2*_-1.7K	CCCAAGGGATCCTCCTCTT
*ZIM2*_-4.2K	GGGTTTCATGAGGCATGTCT
*ZIM2*_-4.4K	GCTATTGCTTTTTATACGCTGACT
ZIM2-P1	TTCCAAAGTCAACCCAAAAAATTAGTTTAGGCCA
ZIM2-P2	TCTTGGCCGAGGACTTTCCAA

Copy Number-Quantitative PCR (CN-qPCR)

Genomic DNA of 47 breast cancer cell lines and 5 immortalized cell lines was a generous gift from Dr. Joe Gray at the Lawrence Berkeley National Laboratory.

Additional DNA samples from breast tumors, breast epithelial progeny of E2-exposed breast progenitor cells, and treated MCF-7 cells were extracted by PureLink™ genomic DNA purification kit (Invitrogen) following the manufacturer's protocol. Quantitative PCR was performed to measure copy-number by using SYBR Green dye chemistry using a StepOnePlus™ Real-Time PCR System apparatus (Applied Biosystems). Copy-number of the 20q13 enhancer cluster was estimated by the ΔΔCT method normalized to the internal control, diploid albumin gene. Details of primer and probe sequences are provided in the following table.

Name	Forward	Reverse
17q23	GGCCTGACACAGTGGCTCAT	CTGGTCTTGAACTCCTGACCTTGT
20q13	AAATCTCAATCTCCCTGCGCTAAG	CACACGAGGGCTGGGTTTC
Albumin	TTGTGGGCTGTAATCATCG	TGCTGGTTCTCTTTCACTGAC

Small Interfering RNA Transfections

MCF-7 cells (10^6 cells) were seeded in a 10-cm dish and transfected the next day with small interfering RNA (siRNA) against ERα (SMARTpool siRNA, Dharmacon) using the Fugene HD transfection reagent (Roche) according to manufacturer's protocol. Culture media were switched to contain 5% charcoal-dextran serum. After 48 hr, the cells were treated with E2 (70 nM) in different time periods (0, 0.5, 1, 4, and 24 hr.). Scrambled oligonucleotide (Dharmacon) was used as a negative control.

Foci Formation Assay

MCF-7 cells (10^6 cells) were seeded in a 10-cm dish and transfected the next day with *THRAP1* (OriGene) and *ZIM2* (OriGene) plasmids using the Fugene HD transfection reagent (Roche) according to manufacturer's protocol. Vector plasmid was used as a

negative control. After 48 hours, transiently transfected cells (500 cells per dish) were re-seeded in 10-cm dish within culture media containing 5% charcoal-dextran serum, irrespective of E2 (70 nM) treatment. After 20 days, cells were fixed with 4% paraformaldehyde (Sigma) and stained by Gemisa solution (Sigma). Universal Hood II Image System (Bio-Rad) was used to capture images.

PCR Method and Gel Electrophoresis for Detecting *17q23::20q13* Fusion

Genomic DNA from MCF-7 cells and 126 clinical breast tissues, including 106 primary tumors and 20 normal ones, was prepared using PureLink™ genomic DNA purification kit following the manufacturer's protocol. PrimeSTAR GXL DNA Polymerase (TaKaRa, Japan) was used for detection of the *17q23::20q13* fusion fragment in both clinical specimen and untreated MCF-7 samples. The following program was applied in a Veriti PCR system apparatus (Applied Biosystems) for 35 cycles: 98°C for 10s, 59 °C for 15s, and 68°C for 35s. Aliquots of the PCR reaction mix were electrophoretically separated in a 1.2% agarose gel in TBE buffer. The electrophoresis runtime was optimized to best distinguish fragments between 200 bp and 500 bp.

Name	Forward	Reverse
Fusion-1	TCTAGGAGAGCATTCCGGTCAT	TGTGTCCCCCACAAATCTCA
Fusion-2	GGAGAGCATTCCGGTCATCA	AGGATGTGTCCCCCACAAATC

Supplemental References

Dempster, A.P., Laird, M.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B. *39*, 1-38.

Efron, B., Tibshirani, R., Storey, J.D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. J. Am. Stat. Assoc*. 96*, 1151-1160.

Lupien, M., Meyer, C.A., Bailey, S.T., Eeckhoute, J., Cook, J., Westerling, T., Zhang, X., Carroll, J.S., Rhodes, D.R., Liu, X.S., et al. (2010). Growth factor stimulation induces a distinct ER(alpha) cistrome underlying bresat cancer endocrine resistance. Genes Dev. *24*, 2219-2227.

Steinhaeuser, U., Starke, H., Nietzel, A., Lindenau, J., Ullmann, P., Claussen, U., and Liehr, T. (2002). Suspension (S)-FISH, a new technique for interphase. J. Histochem. Cytochem. *50*, 1697-1698.

Wedel, M., Desarbo, W.S., Bult, J.R., and Ramaswamy, V. (1993). A latent class Poisson regression model for heterogeneous count data. J. App. Econometrics *8*, 397-411.

Yang, M., and Lai, C. (2005). Mixture Poisson regression models for heterogeneous count data based on latent and fuzzy class analysis. Soft Comput. *9*, 519-524.