Supplemental information

Systems consequences of amplicon formation in human breast cancer
Inaki K. et al.

Supplemental Tables
Supplemental Figures
Supplemental Methods
Supplemental Results
Supplemental References

Clinical Data					Copy number ratio (mutation[1])			*ERBB2* Gene expression[2]	
Sample ID	**Breast Cancer Subtype**	**Tumor Grade**	**Clinical Stage**	**TNM**	***TP53***	***BRCA1***	***ERBB2***	***RNAseq-based***	***Array-based***
BT47	Luminal B	II	IIA	pT2 pN0 M0	0.76 (WT)	0.78	11.96	na	**1.79**
BT55	Luminal B	III	I	pT1c pN0 M0	0.74 (WT)	0.76	6.92	**5.3**	na
BT58	ERBB2	I	IIA	T1c N0 M0	0.96 (na)	0.98	3.52	na	**1.71**
BT71	Luminal A	II	IIIA	pT1c pN2 M0	0.77 (na)	0.8	1.42	na	**1.5**
BT73	Luminal B	II	IIA	pT1c pN2 M0	0.61 (I195T)	0.61	1.45	**4.2**	na
BT122	TNB	III	IIA	pT1c pN1a M0	0.58 (G245S)	0.59	0.59	**2.71**	na
BT146	ERBB2	II	IIB	pT2 pN1a M0	0.73 (K132N)	0.69	13.09	**7.13**	na
BT191	ERBB2	II	IIB	T2 N1 MX	0.69 (WT)	0.69	10.90	na	**1.82**

Supplemental Table 1. Clinical and molecular attributes of the eight sequenced breast tumor samples

TNB, Triple Negative Breast cancer; WT, wild type.

[1] Based on previously published data (Banerji et al. 2012)

[2] Log2 Ratio *vs* per-sample median expression value

SampleID	Origin	Library preparation[1]	Sequencing Platform	Physical Coverage [2]	No. somatic SVs[3]
BT47	Tumor	Nextera Mate Pair library prep	Illumina HiSeq 2500	25	150
BT47_N	Peripheral Blood Lymphocytes	Nextera Mate Pair library prep	Illumina HiSeq 2500	128	na
BT55	Tumor	DNA-PET library prep	AB SOLiD 3 Plus System	131	381
BT55_N	Peripheral Blood Lymphocytes	DNA-PET library prep	AB SOLiD 3 Plus System	117	na
BT58	Tumor	Nextera Mate Pair library prep	Illumina HiSeq 2500	59	89
BT58_N	Peripheral Blood Lymphocytes	Nextera Mate Pair library prep	Illumina HiSeq 2500	209	na
BT71	Tumor	Nextera Mate Pair library prep	Illumina HiSeq 2500	181	386
BT71_N	Peripheral Blood Lymphocytes	Nextera Mate Pair library prep	Illumina HiSeq 2500	199	na
BT73	Tumor	DNA-PET library prep	AB SOLiD 4 System	165	362
BT73_N	Peripheral Blood Lymphocytes	DNA-PET library prep	AB SOLiD 4 System	128	na
BT122	Tumor	DNA-PET library prep	AB SOLiD 3 Plus System	222	390
BT122_N	Peripheral Blood Lymphocytes	DNA-PET library prep	AB SOLiD 3 Plus System	252	na
BT146	Tumor	DNA-PET library prep	AB SOLiD 3 Plus System	294	296
BT146_N	Peripheral Blood Lymphocytes	DNA-PET library prep	AB SOLiD 3 Plus System	308	na
BT191	Tumor	Nextera Mate Pair library prep	Illumina HiSeq 2500	111	322
BT191_N	Peripheral Blood Lymphocytes	Nextera Mate Pair library prep	Illumina HiSeq 2500	196	na

Supplemental Table 2. DNA sequencing statistics

[1] described in Supplemental Methods
[2] (number of non redundant concordant mate pair reads * Median library span)/bps of mappable genome
[3] Structural Variations (SVs) of support ≥ 6 for samples sequenced on a SOLiD platform, ≥ 10 for samples sequenced on Illumina HiSeq 2500

SampleID	No. mapped reads	No. uniquely mapped reads (%)
BT55	65,942,980	46,901,945 (71.1%)
BT73	67,845,400	43,592,403 (64.3%)
BT122	70,187,839	52,360,392 (74.6%)
BT146	63,457,125	45,931,897 (72.4%)

Supplemental Table 3. RNA sequencing statistics

	NUMBER (PERCENTAGE)							
	T47	**BT55**	**T58**	**T71**	**BT73**	**BT122**	**BT146**	**T191**
Deletion	15 (10)	49 (12.9)	12 (13.48)	34 (8.81)	45 (12.4)	23 (5.9)	42 (14.2)	25 (7.76)
Tandem duplication	21 (14)	32 (8.4)	10 (11.24)	38 (9.84)	39 (10.8)	287 (73.6)	24 (8.1)	61 (18.94)
Unpaired inversion	45 (30)	89 (23.4)	17 (19.1)	84 (21.76)	67 (18.5)	33 (8.5)	55 (18.6)	135 (41.93)
Inversion	0 (0)	2 (0.5)	0 (0)	0 (0)	0 (0)	0 (0)	2 (0.7)	0 (0)
Intra-chromosomal insertion	3 (2)	7 (1.8)	0 (0)	7 (1.81)	4 (1.1)	2 (0.5)	0 (0)	13 (4.04)
Inter-chromosomal insertion	0 (0)	5 (1.3)	1 (1.12)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Isolated translocation	29 (19.33)	91 (23.9)	38 (42.7)	74 (19.17)	47 (13)	26 (6.7)	42 (14.2)	52 (16.15)
Balanced translocation	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Intra-chromosomal complex	21 (14)	62 (16.3)	4 (4.49)	85 (22.02)	136 (37.6)	11 (2.8)	117 (39.5)	20 (6.21)
Inter-chromosomal complex	16 (10.67)	44 (11.5)	7 (7.87)	64 (16.58)	24 (6.6)	8 (2.1)	14 (4.7)	16 (4.97)
Total	**150 (100)**	**381 (100)**	**89 (100)**	**386 (100)**	**362 (100)**	**390 (100)**	**296 (100)**	**322 (100)**

Supplemental Table 4. Number and percentage of the different types of somatic structural variations detected in the eight sequenced breast tumors

	Number (percentage)						
	Fusion Gene	**3' Truncation**	**5' Truncation**	**Intragenic Rearrangement**	**Intergenic**	**Total Intragenic**	**Total**
BT55	30 (7.9)	89 (23.4)	104 (27.3)	5 (1.3)	**153 (40.2)**	**228 (59.8)**	**362 (100)**
BT73	42 (11.6)	90 (24.9)	77 (21.3)	15 (4.1)	**138 (38.1)**	**224 (61.9)**	**381 (100)**
BT122	39 (10)	74 (19)	106 (27.2)	31 (7.9)	**140 (35.9)**	**250 (64.1)**	**296 (100)**
BT146	30 (10.1)	83 (28)	89 (30.1)	7 (2.4)	**87 (29.4)**	**209 (70.6)**	**390 (100)**

Supplemental Table 5. Number and percentage of somatic structural variations which caused gene fusions, truncations or intragenic rearrangements in the four breast tumors examined using DNA-PET. The numbers presented here are based on the analysis of DNA sequencing data.

	SM	5' gene	5' chr	3' gene	3' chr	Fused exons	Frame	5' domain (Pfam)	3' domain (Pfam)
BT55	Del	*MYH9*	Chr 22	*TTC28*	Chr 22	E22-E14	Out	Myosin_N, Myosin_head, IQ	None
		CTTN	Chr 11	*RBM4*	Chr 11	E17-E3	In	HS1_rep, SH3_1 / SH3_2 (truncated)	RRM_1 (truncated), zf-CCHC
	TD					E5/ins/mid of E17	Out		
		AATF	Chr 17	*KSR1*	Chr 17	E5/E17	In	None	Pkinase / Pkinase_Tyr (truncated)
						E4/E17	Out		
	U-Inv	*CABIN1*	Chr 22	*BRD1*	Chr 22	E1-E2	5'UTR of 5' gene	None	Bromodomain, PWWP
		DESI1	Chr 22	*EIF3L*	Chr 22	E1-E13	Out	DUF862 (truncated)	None
	Inter-Ins	*RBFOX2*	Chr 11	*MAP6*	Chr 11	E4-E2	Out	None	STOP (truncated)
BT146	Del	*ASB3*	Chr 2	*NRXN1*	Chr 2	E1-E12	5'UTR of 5' gene	None	Laminin_G_1 / Laminin_G_2 (truncated)
		ZMAT3	Chr 3	*PLCH1*	Chr 3	E1-E3	5'UTR of 5' gene	None	efhand_like, PI-PLC-X, PI-PLC-Y, C2
	U-Inv	*FAM107B*	Chr 10	*FRMD4A*	Chr 10	E3-E3	Out	DUF1151 (truncated)	FERM_N, FERM_M, FERM_C, DUF3338
						E3-(intron+)E3	Out		
		CPT1A	Chr 11	*FAM111B*	Chr 11	E13-E4	5'UTR of 3'gene	Carn_acyltransf (truncated)	None
						E13(+intron)-E4	5'UTR of 3'gene		
		LRRC27	Chr 10	*UROS*	Chr 10	E13-E2	5'UTR of 3'gene	LRR_1	HEM4
	Transloc	*USP6NL*	Chr 10	*GALNTL6*	Chr 4	E2-E3	Out	None	Glycos_transf_2, Ricin_B_lectin
						E1-E3	5'UTR of 5' gene		
		MYPOP	Chr 19	*COL14A1*	Chr 8	E2/E15	Out	None	fn3 / VWA (truncated), Collagen
	Complex-Intra (TD)	*MSI2*	Chr 17	*UBE2Z*	Chr 17	E7-E6	Out	RRM_1 (truncated)	UQ_con (truncated)
						E7-E5 (extendend to 5')	Out		
						E6-E6	Out		
BT122	Del	*SKI*	Chr 1	*MACF1*	Chr 1	E1-E2	Out	Ski_Sno, c-SKI_SMAD_bind	CH, CAMSAP_CH, Spectrin, GAS2
						E1(+intron)-(intron+)E2	Out		
						E1-E3	In		CH (truncated), CAMSAP_CH, Spectrin, GAS2
		TOR1AIP2	Chr 1	*ACBD6*	Chr 1	E2-E8	5'UTR of 5' gene	None	Ank (truncated)
						E2-(intron+)E8			
		MAP4	Chr 3	*COL7A1*	Chr 3	E2-E14	In	None	VWA / fn3 (truncated), Collagen, Kunitz_BPTI
						E2-E15	In		
						E2(+intron)-E14	In		
						E2(+intron)-E14	Out		
		FAM13A	Chr 4	*SNCA*	Chr 4	E20-E6	In	RhoGAP	Synuclein (truncated)
		IL6ST	Chr 5	*ANKRD55*	Chr 5	E15-E8	Out	Lep_receptor_Ig, EpoR_lig-bind / IL6Ra-bind, fn3	Ank (truncated)
		STYXL1	Chr 7	*YWHAG*	Chr 7	E4-E2	Out	Rhodanese (truncated), DSPc	14-3-3 (truncated)
		DCAF10	Chr 9	*ZCCHC7*	Chr 9	E2-E6	Out	WD40 (truncated)	zf-CCHC (truncated)
		PPP6C	Chr 9	*MAPKAP1*	Chr 9	E1-E2	5'UTR of 3' gene	None	SIN1
	TD	*DDX50*	Chr 10	*CCAR1*	Chr 10	E6-E16	Out	ResIII, DEAD	SAP (truncated)
						E7-E16	In		
		FAM222B	Chr 17	*SSH2*	Chr 17	E1b-E3		None	DEK_C, DSPc
						E1b(+intron)-E3	5'UTR of 5' gene		
						E1b(+intron)-(intron+)E3			
		FBF1	Chr 17	*SRP68*	Chr 17	E7-E9	Out	None	None
		FDXR	Chr 17	*SRP68*	Chr 17	E2-E5	In	Pyr_redox_2 (truncated)	None
		KIAA1328	Chr 18	*ELP2*	Chr 18	E4-E6	Out	None	WD40 (truncated)
						E4-E8	Out		
		ELAVL1	Chr 19	*FBN3*	Chr 19	E4(+intron)-E22	Out	RRM_1 (truncated)	EGF_CA / TB (truncated)
		LOC100131691	Chr 19	*ZSWIM4*	Chr 19	E1-E10	5'gene = non-coding	None	None
						E1-E12			
		RALY	Chr 20	*CBFA2T2*	Chr 20	E2-E11	5'UTR of 5' gene	None	zf-MYND
		MED8	Chr 1	*CDC20*	Chr 1	E5-mid of E4	Out	Med8 (truncated)	WD40
		RNF214	Chr 11	*SIK3*	Chr 11	E5-E5	Out	None	Pkinase / Pkinase_Tyr (truncated)
	U-Inv					E9-mid of E15	In		
		TP53	Chr 17	*POLR2A*	Chr 17			P53, P53_tetramer (truncated)	RNA_pol_Rpb1_4 (truncated), RNA_pol_Rpb1_5, RNA_pol_Rpb1_6, RNA_pol_Rpb1_7, RNA_pol_Rpb1_R
						E10a-mid of E15	Out		
		ZNF740	Chr 3	*RASSF1*	Chr 12	E6-E8	In	zf-C2H2	None
		ATAD2	Chr 8	*BFAR*	Chr 16	E20-E6	Out	AAA, Bromodomain (truncated)	None
						E21-E6	In		
	Transloc	*SMARCA4*	Chr 10	*CHST15*	Chr 19	E30-E2	5'UTR of 3' gene	QLQ, HAS, BRK, SNF2_N, Helicase_C	Sulfotransfer_1
						E30-(intron+)E2			
		RSU1	Chr 10	*ROBO3*	Chr 11	E2-E20	Out	None	None
						E2-insertion(chr11)-E18	Out		

Supplemental Table 6. List of the 37 fusion gene transcripts detected by RT-PCR

ID	SV Type	ChrL	Left BP: Median[1]	Left BP: SD[1]	StrandL	ChrR	Right BP: Median[1]	Right BP: SD[1]	StrandR	Span: Median[2]	Span:SD[2]	No SVs[3]	Samples (Highest Cluster Size) [4]	Consequence
1	DEL	Chr 17	35,327,718	901,170	+	Chr 17	44,040,832	1,555,393	+	9,229,089	1,956,113	11	BT47(53); BT55(51); BT58(28); BT146(285)	Loss of the *BRCA1 locus*
2	DEL	Chr 17	26,477,558	365,257	+	Chr 17	53,167,545	1,642,549	+	26,689,987	1,277,292	2	BT47(13); BT55(22)	
3	DEL	Chr 17	44,577,064	529,252	+	Chr 17	47,997,384	554,462	+	3,420,320	25,210	2	BT47(12); BT55(126)	
4	TD	Chr 17	43,382,084	702,632	+	Chr 17	33,466,706	242,212	+	9,915,378	460,419	2	BT146(11); BT191(15)	
5	TD	Chr 3	187,084,139	242,560	+	Chr 3	130,357,857	264,454	+	56,726,282	21,893	2	BT146(56); BT191(66)	
6	TD	Chr 17	35,484,752	502,337	+	Chr 17	32,471,649	2,890,927	+	3,140,266	2,578,085	8	BT55(49); BT71(141); BT73(44); BT146(359)	Gain of *ERBB2*
7	TD	Chr 17	48,761,811	343,544	+	Chr 17	35,187,023	138,359	+	13,574,789	481,902	2	BT47(45); BT55(58)	
8	TD	Chr 1	13,598,917	39,324	+	Chr 1	13,372,169	39,878	+	226,748	554	2	BT47(10); BT191(11)	
9	TD	Chr 1	146,015,135	35,645	+	Chr 1	144,709,478	35,995	+	1,305,554	646	3	BT47(11); BT191(11)	
10	TD	Chr 17	35,511,432	35,780	+	Chr 17	35,114,153	93,482	+	397,871	57,705	3	*BT55(207); BT191(625)*	*Gain of ERBB2*
11	TD	Chr 17	48,218,299	1,919,047	+	Chr 17	19,265,756	1,656,736	+	28,952,543	3,575,783	2	BT47(10); BT55(124)	17q21.3 amplification
12	TD	Chr 11	62,009,478	147,749	+	Chr 11	61,632,172	33,204	+	377,306	114,545	2	BT73(6); BT122(11)	
13	TD	Chr 16	18,112,266	34,249	+	Chr 16	15,195,421	39,056	+	2,916,846	4,806	2	BT58(10); BT191(14)	
14	TD	Chr 10	88,893,085	35,442	+	Chr 10	46,306,886	39,748	+	42,586,199	4,306	2	BT71(11); BT191(11)	
15	TD	Chr 11	12,498,043	1,801,332	+	Chr 11	6,614,704	1,267,526	+	5,883,339	533,806	2	BT71(27); BT73(34)	
16	INV	Chr 17	43,485,798	891,783	+	Chr 17	46,036,716	562,884	-	3,493,024	830,501	3	BT55(189); BT71(14)	
17	INV	Chr 17	35,336,616	1,303,691	+	Chr 17	45,874,398	2,065,841	-	10,957,045	2,264,413	6	BT47(43); BT55(75); BT58(10); BT146(88)	Co-amplification of *ERBB2* and 17q21.3 amplicons
18	INV	Chr 17	34,388,439	1,216,754	-	Chr 17	44,664,869	2,226,404	+	9,990,810	1,870,681	8	BT47(42); BT55(331); BT71(12); BT146(852)	Co-amplification of *ERBB2* and 17q21.3 amplicons
19	INV	Chr 2	109,922,834	42,154	-	Chr 2	111,030,974	185,568	+	1,108,140	225,167	6	BT47(11); BT71(10); BT191(13)	
20	INV	Chr 7	149,443,301	151,750	-	Chr 7	153,238,864	218,928	+	3,795,563	370,558	6	BT71(11); BT191(11)	
21	INV	Chr 16	16,276,197	600,804	-	Chr 16	18,403,415	127,326	+	2,154,658	653,012	6	BT71(12); BT191(20)	
22	INV	Chr 7	56,865,867	144,479	-	Chr 7	62,839,319	124,243	+	5,973,452	268,701	3	BT47(12); BT71(12)	
23	INV	Chr 3	130,546,270	206,557	+	Chr 3	186,945,403	1,628,443	-	56,667,233	1,776,278	3	BT146(46); BT191(89)	
24	INV	Chr 7	66,329,672	81,827	-	Chr 7	71,789,290	147,033	+	5,567,923	131,645	3	BT71(11); BT191(12)	
25	INV	Chr 17	48,259,177	184,392	-	Chr 17	68,990,010	1,650,813	+	20,733,827	1,466,428	3	BT47(27); BT146(35)	
26	INV	Chr 15	80,407,106	225,677	-	Chr 15	83,597,596	560,427	+	3,190,490	786,094	3	BT71(10); BT191(18)	
27	INV	Chr 17	14,998,735	525,647	-	Chr 17	19,088,765	616,620	+	4,090,031	1,142,267	2	BT71(11); BT146(11)	
28	INV	Chr 17	46,780,169	18,938	+	Chr 17	48,815,995	141,184	-	2,035,826	122,247	2	BT47(21); BT55(23)	
29	INV	Chr 3	130,301,332	207,504	-	Chr 3	137,442,289	1,303,769	+	7,140,958	1,511,273	2	BT73(32); BT191(29)	
30	INV	Chr 17	32,025,103	706,304	-	Chr 17	33,861,224	366,022	+	1,836,121	340,282	2	BT146(9); BT191(14)	
31	INV	Chr 5	68,989,006	169,133	-	Chr 5	70,195,527	310,008	+	1,206,521	479,141	2	BT47(12); BT191(11)	
32	INV	Chr 9	68,304,589	227,677	-	Chr 9	70,168,483	57,461	+	1,863,894	285,138	2	BT47(13); BT71(10)	
33	INV	Chr 16	28,273,350	18,978	-	Chr 16	28,614,736	61,798	+	341,386	42,820	2	BT47(10); BT191(15)	
34	INV	Chr 17	33,719,586	506,828	-	Chr 17	35,186,945	219,338	+	1,467,359	287,490	2	BT47(33); BT191(66)	
35	INV	Chr 9	89,735,239	7,038	-	Chr 9	89,920,610	5,434	+	185,371	12,472	2	BT71(10); BT191(10)	
36	INV	Chr 8	60,878,089	291,100	-	Chr 8	76,281,297	1,152,573	+	15,403,209	861,473	2	BT71(46); BT73(108)	
37	INV	Chr 8	82,011,394	570,480	+	Chr 8	86,729,758	1,649,851	-	4,718,364	2,220,331	2	BT55(6); BT73(132)	
38	INTER	Chr 10	39,124,979	1,289,820	+	Chr 20	29,296,124	2,691	-	NA	NA	4	BT47(39); BT71(20); BT191(12)	
39	INTER	Chr 10	39,163,542	1,135,045	-	Chr 20	29,293,988	2,028	-	NA	NA	5	BT47(13); BT71(35)	
40	INTER	Chr 10	40,414,572	1,821,801	+	Chr 21	9,850,007	11,483	-	NA	NA	2	BT47(38); BT71(15)	
41	INTER	Chr 8	102,055,262	960,605	+	Chr 12	53,859,523	630,340	-	NA	NA	2	BT73(23); BT191(10)	
42	INTER	Chr 13	58,728,264	104,773	+	Chr 17	45,198,558	2,018,617	-	NA	NA	2	BT55(35); BT71(43)	
43	INTER	Chr 4	48,847,023	4,137	-	Chr 20	29,293,328	17	-	NA	NA	2	BT47(14); BT71(13)	
44	INTER	Chr 1	714,797	2	-	Chr 10	41,699,578	428	+	NA	NA	2	BT71(20); BT191(15)	
45	INTER	Chr 11	51,424,279	639	-	Chr X	61,628,923	57	-	NA	NA	2	BT71(11); BT191(11)	

Supplemental Table 7. List of the recurrent somatic structural variations observed across the eight breast tumors analyzed.

Each row corresponds to a recurrent SV event, as detected based on the analysis of 8 breast cancer samples.

[1] Genomic coordinates for the recurrent event: median and standard deviation values of the left and right break point coordinates correspond to the distributions of the individual SVs comprising the recurrent event.

[2] Median and standard deviation values relative to the span distribution of the individual structural variations comprising the recurrent event.

[3] Number of individual structural variations comprising the recurrent event.

[4] List of the tumor samples contributing to each recurrent event.

In blue: recurrent structural variation events which contribute to Chr 17 oncogenic rearrangements as described in Figure 4 and Supplemental Figure 5.

SV, structural variation; SD, standard deviation; BP, break point; DEL, deletion; TD, tandem duplication; INV, unpaired inversion; INTER, isolated translocation.

	CN ratio				Tumorscape[1]	*ERBB2*+ tumors[2]	LSVD[3]	RNA-seq (FPKM)				Oncomine[4]			Microarray co-expression with ERBB2 (P-value, student's t-test)						KM plotter[5]		DMFS		DMFS based on *ERBB2* status			
												Number of cohorts	TCGA									RFS				*ERBB2*+		*ERBB2*-	
	BT122	BT146	BT55	BT73	CN peak	CN peak		BT122	BT146	BT55	BT73		P-value	Fold change	UPP	Van	Chin	Wang	Sto	Oxf	P-value	HR	P-value	log(HR)	P-value	HR	P-value	HR
PNPO	0.6	3.8	0.8	0.6				16	44	11	12	0	1.1E-14	1.8	5.9E-01		5.4E-01	6.0E-03	2.3E-03	1.4E-02	4.8E-01	1.0	6.E-01	-0.1	7.2E-01	0.9	4.3E-01	0.9
PRR15L	0.6	3.8	0.8	0.6				7	496	51	20		1.1E-07	2.9							5.6E-03	0.8	3.E-02	0.3				
CBX1	0.6	3.3	6.8	0.6				4	21	34	4	0	7.9E-13	2.2	2.1E-01	7.0E-01		5.0E-01	4.9E-01	1.7E-02	2.4E-13	1.7	7.E-02	0.3	3.0E-01	1.3	1.9E-02	1.6
HOXB7	0.6	3.0	10.9	0.6			LSVD	8	9	72	14	2	NS	NS	3.5E-03	9.3E-01	1.9E-01	8.6E-01	3.3E-01	6.4E-01	1.1E-03	1.3	3.E-01	0.2	6.4E-01	1.1	9.9E-01	1.0
UBE2Z	0.6	3.6	10.7	0.6				43	266	426	19	3	5.9E-20	2.0	1.5E-04	4.9E-01	2.0E-01	5.4E-02	1.8E-02	7.0E-06	1.0E-01	0.9	2.E-01	0.2	9.9E-02	0.7	1.8E-02	1.8
SNF8	0.6	3.6	10.7	0.6				22	226	553	20	0	NS	NS	2.0E-04	5.7E-03	6.4E-02	1.3E-02	1.5E-02	1.6E-02	9.5E-01	1.0	1.E-02	0.4	1.7E-01	0.7	3.5E-03	2.0
PHB	0.6	4.9	11.5	0.6	peak		LSVD	51	506	659	51	5	NS	NS	1.4E-06	2.7E-01	3.0E-02	1.4E-05	8.1E-04	3.1E-02	0.0E+00	1.8	5.E-03	0.4	1.4E-01	0.8	4.0E-04	1.8
SLC35B1	0.6	5.3	12.3	0.6			LSVD	36	301	709	57	5	1.5E-29	1.6	1.8E-07	1.5E-02	3.5E-01	2.9E-02	1.3E-02	1.1E-02	1.3E-02	1.2	0.E+00	0.5	5.7E-02	0.7	1.6E-05	2.0
KAT7	0.6	5.3	12.3	0.6			LSVD	5	22	25	1	1	NS	NS	4.5E-03	5.7E-01	1.3E-01	9.0E-03	4.4E-02	5.5E-03	5.8E-02	0.9	3.E-01	-0.2	2.4E-02	0.6	1.6E-02	1.7
PDK2	0.6	5.3	10.8	0.6		peak		10	177	12	13	0	NS	NS	6.3E-01	9.3E-01	3.0E-01	8.8E-03	2.4E-03	5.8E-03	4.3E-04	1.3	7.E-01	0.1	9.6E-01	1.0	8.1E-01	1.0
PPP1R9B	0.6	5.3	10.8	0.6		peak		11	64	142	18	1	3.7E-19	1.7	6.8E-07	5.0E-02			1.9E-01	2.2E-02			4.E-01	-0.1	5.6E-02	0.3	8.1E-01	1.1
MRPL27	0.6	5.6	10.8	0.6	peak		LSVD	12	402	601	37	1	NS	NS	1.5E-06	1.9E-02			8.5E-03	1.6E-02	2.5E-04	0.8	8.E-03	0.4	9.3E-01	1.0	1.9E-01	1.4
EME1	0.6	5.6	10.8	0.6	peak		LSVD	2	11	21	1	1	2.1E-27	2.2	4.3E-07				1.2E-01	1.2E-01			0.E+00	0.7	6.4E-01	1.3	1.7E-01	1.8
LRRC59	0.6	5.6	10.8	0.6	peak		LSVD	26	258	419	24	2	2.1E-27	2.2	3.7E-07	1.4E-04	4.4E-02	1.5E-02	4.3E-02	4.3E-02	4.8E-08	1.5	1.E-03	0.5	9.0E-01	1.0	1.7E-07	2.2
EPN3	0.6	5.6	9.1	0.6	peak		LSVD	1	103	94	4	2	8.0E-21	2.2	2.3E-05	2.7E-03	4.5E-01	1.2E-06	6.3E-02	5.8E-03	6.1E-01	1.0	1.E-03	0.5	3.6E-01	1.2	3.6E-04	1.5
SPATA20	0.6	5.6	9.1	0.6	peak		LSVD	7	143	131	15	0	NS	NS	1.2E-01	2.0E-04	8.5E-01	1.0E-03	6.4E-01	6.3E-02	2.3E-06	0.7	3.E-01	0.2	1.5E-01	0.8	2.4E-02	1.3
ANKRD40	0.6	5.0	11.7	0.6				8	54	81	7	1	NS	NS	3.8E-01		1.5E-01	7.5E-01		6.4E-02	2.1E-01	0.9	8.E-03	0.4	9.4E-01	1.0	8.2E-02	1.4
LUC7L3	0.6	5.0	11.7	0.6				35	201	382	40		NS	NS							1.1E-06	1.4	2.E-01	0.2				
TOB1	0.6	5.0	13.6	0.6				16	572	1243	145	0	NS	NS	2.8E-01	1.7E-01	5.3E-01	8.5E-05	1.1E-02	3.3E-04	2.7E-01	1.1	3.E-01	0.2	2.9E-01	0.8	2.3E-01	1.2
NME1	0.6	0.7	11.4	0.8			LSVD	28	31	850	73	5	6.8E-31	1.7	5.5E-06	9.8E-02	3.2E-02	2.3E-01	2.6E-01	3.3E-04	7.7E-14	1.7	4.E-03	0.4	7.3E-01	0.9	1.2E-04	2.3
NME2	0.6	0.7	11.4	0.8			LSVD	69	64	1468	235		NS	NS							7.0E-06	1.4	4.E-03	0.4				

Supplemental Table 8. Selection of oncogene candidates on 17q21.3.

1: n=240 (Tumorscape data set).

2: n=200, Staaf et al., 2010.

3: LSVD: Local Singular Value Decomposition, Zhang et al. 2009.

4 :Ductal breast carcinoma vs. normal breast, Over-expression, P-value < 0.05, fold change > 1.5 (www.oncomine.org).

5: n=2422, median separation (http://kmplot.com/analysis/).

CN: copy number. UPP: Uppsala cohort (Miller et al. 2005) . Van: Van cohort (van de Vijver et al. 2002). Chin: Chin cohort (Ivshina et al. 2006).
Wang: Wang cohort (Wang et al. 2005). Sto: Stockholm cohort (Pawitan et al. 2005). Oxf: Oxford cohort (Loi et al. 2007).
HR: hazard ratio. RFS: relapse free survival. DMFS: distant metastasis free survival.

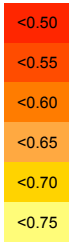
	ER/PR (-)									ER/PR (+)									Less sensitive to Lapatinib (#)				17q21.3 non-amplified			*ERBB2* non-amplified	
	HCC202			UACC893			HCC1419			HCC2218			ZR75-30			BT474			MB361		CAMA1		SKBR3			MCF7	
Lapatinib	(-)	(+)	CN	(-)	(+)	CN	(-)	(+)	CN	(-)	(+)	CN	(-)	(+)	CN	(-)	(+)	CN	(-)	CN	(-)	CN	(-)	(+)	CN	(-)	CN
ERBB2	0.683*	0.865*	4.9	0.677*	0.771*	21	0.496*	0.561*	19.2	0.433*	0.383*	12.8	0.374*	0.404*	18.8	0.295*	0.427*	14.1	0.946	7.3	0.794*	6.7	0.144*	0.219*	9.4	0.754*	1.4
PNPO	0.800*	0.781*	1.5	0.779*	0.791*	1.8	0.812*	0.841*	7.5	0.845*	0.956	1.7	0.845*	0.881*	1.5	0.832*	0.773*	1.9	0.777*	3.8	0.511*	1.6	0.803*	0.992	1.6	0.584*	2.9
PRR15L	0.772*	0.712*	1.5	0.683*	0.730*	1.8	0.996	1.164*	7.5	0.627*	0.818*	1.7	1.047	1.03	1.5	0.955	1.139	1.9	0.934	3.8	1.008	1.6	0.973	1.065*	1.6	0.781*	2.9
CBX1	0.644*	0.723*	1.5	0.781*	0.723*	1.8	0.848*	0.868*	7.5	0.560*	0.709*	1.7	0.918*	0.928	17.8	0.909*	0.973	1.9	1.019	3.8	0.787*	1.6	0.986	1.015	1.6	0.736*	2.9
HOXB7	0.817*	0.896	8.3	0.743*	0.752*	1.8	0.612*	0.768*	7.5	0.595*	0.662*	14.9	0.733*	0.915	11.1	0.614*	0.642*	14.4	0.567*	3.8	0.550*	6.8	0.756*	0.809*	1.6	0.544*	2.9
UBE2Z	0.655*	0.607*	8.3	0.799*	0.749*	1.8	0.963	0.951*	7.5	0.646*	0.709*	14.9	0.844*	0.883*	6.9	1.06	0.919	5.8	1.094*	3.8	0.885*	4.5	1.117*	1.130*	1.6	0.741*	2.9
SNF8	0.759*	0.663*	4.7	0.880*	0.440*	1.8	1.003	0.894*	7.5	0.923	0.933	14.9	0.972	0.926	6.9	0.996	0.687*	5.8	0.813*	3.8	0.473*	4.5	0.898*	0.782*	1.6	0.834*	2.9
PHB	0.885*	0.894*	4.7	0.918*	0.838*	1.8	0.956*	0.984	7.5	0.847*	0.893	5.7	0.876*	0.837*	3.9	0.992	0.979	5.8	1.053	3.8	0.587*	4.5	0.765*	0.904	1.6	0.947	2.9
SLC35B1	1.008	0.956	4.7	0.856*	0.874*	3.1	0.97	1.004	7.5	0.909	0.903	5.7	0.867*	0.891*	3.9	1.073	1.081	5.8	1.001	3.8	0.812*	4.5	1.255*	1.350*	1.6	0.881*	2.9
KAT7	0.642*	0.574*	4.7	0.870*	0.692*	3.1	1.068*	0.993	7.5	0.649*	0.715*	5.7	0.954	0.93	3.9	1.05	1.194*	5.8	0.834*	3.8	0.815*	4.5	1.139*	1.217*	1.6	0.693*	2.9
PDK2	0.955	0.931	4.7	0.575*	0.614*	3.1	0.901*	0.882*	7.5	0.717*	0.885	5.7	0.786*	0.834*	3.9	0.883*	0.855	5.8	0.604*	3.8	0.802*	4.5	0.817*	1.049	1.6	0.872*	2.9
PPP1R9B	0.789*	0.981	4.7	0.760*	0.772*	3.1	0.914*	0.992	7.5	0.739*	0.885	5.7	1.076	1.232*	3.9	0.743*	0.731*	5.8	0.884*	3.8	0.672*	4.5	0.964	1.041	1.6	0.673*	2.9
MRPL27	0.722*	0.757*	4.7	0.832*	0.847*	3.1	0.941*	0.942	4	0.840*	0.877*	5.7	0.890*	0.884*	2.4	1.112*	0.851*	5.8	0.841*	3.8	1.232*	4.5	0.938	0.839*	1.6	0.807*	2.9
EME1	0.727*	0.544*	4.7	0.775*	0.403*	3.1	1.044	0.920*	4	0.964	0.933	5.7	0.897*	0.874*	2.4	0.958	0.653*	5.8	0.955	3.8	0.519*	4.5	1.131*	1.257*	1.6	0.908	2.9
LRRC59	1.209*	1.099	4.7	0.962	0.802*	3.1	1.003	1.013	4	0.966	0.947	5.7	0.989	0.969	2.4	0.983	0.880*	5.8	0.986	3.8	0.678*	4.5	0.994	0.898*	1.6	0.965	2.9
EPN3	1.225*	1.245*	4.7	1.076	0.966	3.1	1.065*	1.016	2.1	1.018	0.981	5.7	0.974	0.942	2.4	1.102	1.154*	5.8	1.017	3.8	0.981	4.5	0.983	0.873*	1.6	1.257*	2.9
SPATA20	0.794*	0.767*	4.7	0.726*	0.507*	3.1	0.870*	0.830*	2.1	0.677*	0.687*	5.7	0.774*	0.760*	2.4	0.880*	0.669*	5.8	0.857*	3.8	0.826*	4.5	0.770*	0.683*	1.6	0.943	2.9
ANKRD40	1.057	1.045	6.8	0.827*	0.821*	3.1	0.967	0.974	2.1	0.95	1.025	5.7	0.99	1.072	2.4	1.096*	1.067	5.8	1.037	3.8	0.631*	4.5	1.125*	1.287*	1.6	0.968	2.9
LUC7L3	0.978	0.885*	6.8	0.827*	0.544*	3.1	0.941*	0.841*	2.1	0.974	0.974	5.7	0.892*	0.825*	2.4	1.072*	0.934	5.8	0.924	3.8	0.685*	4.5	0.824*	0.838*	1.6	0.772*	2.9
TOB1	0.743*	0.674*	6.8	0.653*	0.493*	3.1	1.001	0.907*	2.1	0.862*	0.869*	5.7	0.861*	0.814*	2.4	0.932	0.614*	2.9	0.843*	3.8	0.495*	4.5	0.828*	0.818*	1.6	0.805*	2.9
NME1	1.013	1.162*	6.8	1.323*	1.169*	3.1	1.065*	1.077*	2.1	0.873*	0.965	2.4	1.103*	1.096*	2.4	1.146*	1.117*	2.9	1.154*	3.8	0.988	2.3	1.164*	1.247*	1.6	1.109*	2.9
NME2	0.536*	0.443*	6.8	0.897	0.622*	3.1	0.930*	0.912*	2.1	0.410*	0.461*	2.4	1.076	1.152*	2.4	0.978	0.707*	2.9	1.061	3.8	0.820*	2.3	0.996	0.874*	1.6	0.851*	2.9

Supplemental Table 9. Cell proliferation assay read-outs (relative to the negative control) for 10 breast cancer cell lines treated with gene-specific siRNAs in the presence or in the absence of Lapatinib.

*: P-value < 0.05 (Student's t-test).

#: Konecny et al. 2006.

CN: copy number of the gene in each cell line (modified from Kao et al. 2009 and Tumorscape (CAMA1))



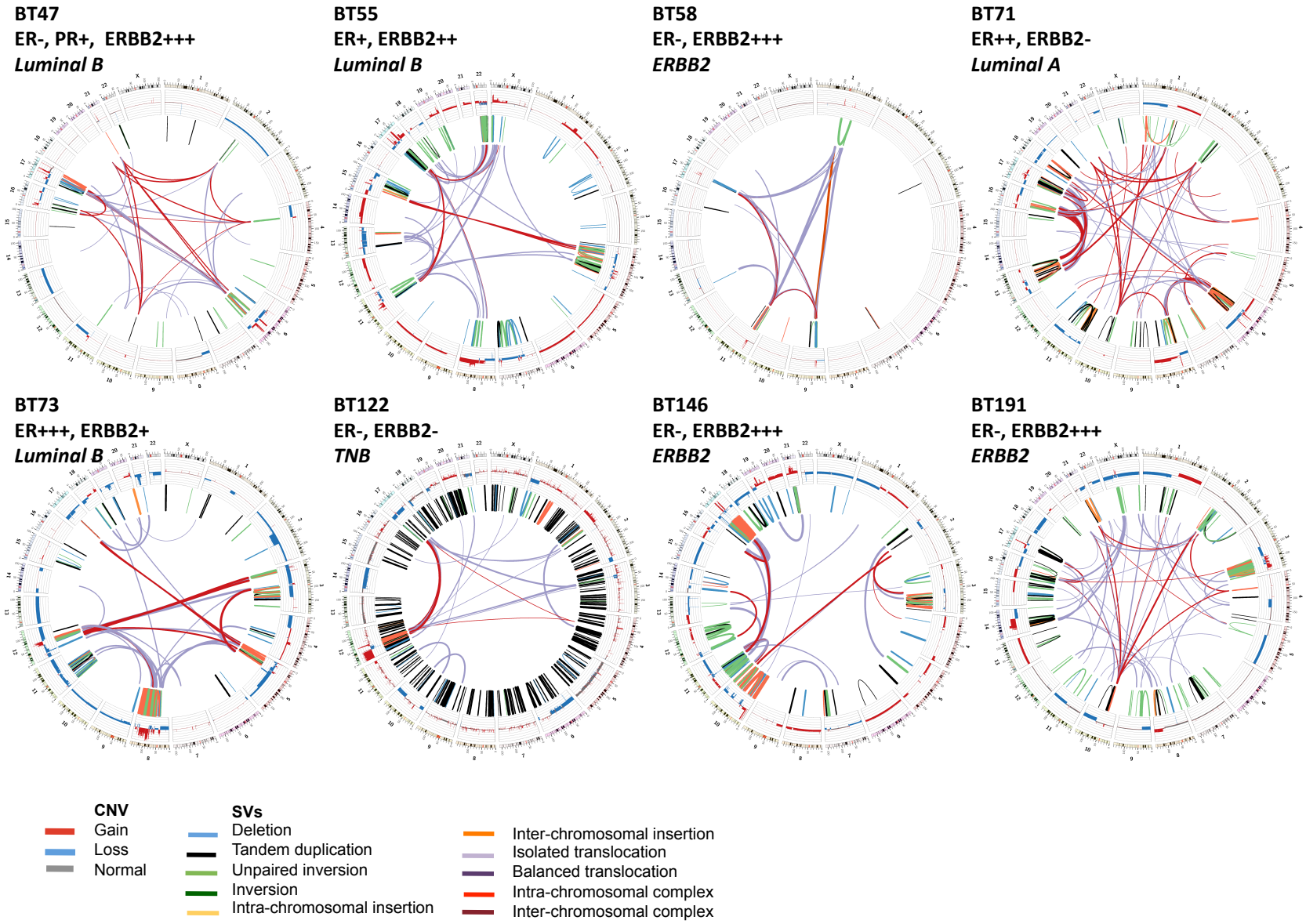
		HCC202	UACC893	HCC1419	HCC2218	ZR75-30	BT474
	ERBB2	1.27	1.14	1.13	0.89	1.08	1.45
	PNPO	0.98	1.02	1.04	1.13	1.04	0.93
	PRR15L	0.92	1.07	1.17	1.31	0.98	1.19
	CBX1	1.12	0.93	1.02	1.27	1.01	1.07
	HOXB7	1.10	1.01	1.26	1.11	1.25	1.05
	UBE2Z	0.93	0.94	0.99	1.10	1.05	0.87
	SNF8	0.87	0.50	0.89	1.01	0.95	0.69
	PHB	1.01	0.91	1.03	1.05	0.96	0.99
	SLC35B1	0.95	1.02	1.04	0.99	1.03	1.01
	KAT7	0.89	0.80	0.93	1.10	0.98	1.14
	PDK2	0.98	1.07	0.98	1.23	1.06	0.97
17q21.3	***PPP1R9B***	1.24	1.02	1.09	1.20	1.15	0.98
	MRPL27	1.05	1.02	1.00	1.04	0.99	0.77
	EME1	0.75	0.52	0.88	0.97	0.97	0.68
	LRRC59	0.91	0.83	1.01	0.98	0.98	0.90
	EPN3	1.02	0.90	0.95	0.96	0.97	1.05
	SPATA20	0.97	0.70	0.95	1.02	0.98	0.76
	ANKRD40	0.99	0.99	1.01	1.08	1.08	0.97
	LUC7L3	0.90	0.66	0.89	1.00	0.92	0.87
	TOB1	0.91	0.76	0.91	1.01	0.95	0.66
	NME1	1.15	0.88	1.01	1.10	0.99	0.98
	NME2	0.83	0.69	0.98	1.12	1.07	0.72
	RAB10	1.05	1.02	1.08	1.07	1.09	1.14
	EPT1	1.00	0.91	1.06	0.98	1.01	1.01
	YPEL5	1.06	0.98	0.82	0.89	1.10	1.03
	TTC27	1.09	1.06	1.14	0.96	0.97	0.97
2p23	***HADHB***	0.97	0.99	1.06	0.96	1.03	1.02
	OST4	1.00	0.88	1.03	0.92	1.03	0.95
	ATRAID	0.83	0.71	0.85	0.95	0.99	0.94
	SLC4A1AP	1.03	1.09	1.01	1.08	1.05	1.04
	MRPL33	0.98	1.02	0.86	0.98	1.05	0.94

Supplemental Table 10. Relative cell proliferation ratios between cells cultured in the presence and in the absence of Lapatinib ((+) / (-)) following treatment with gene targeting siRNAs.

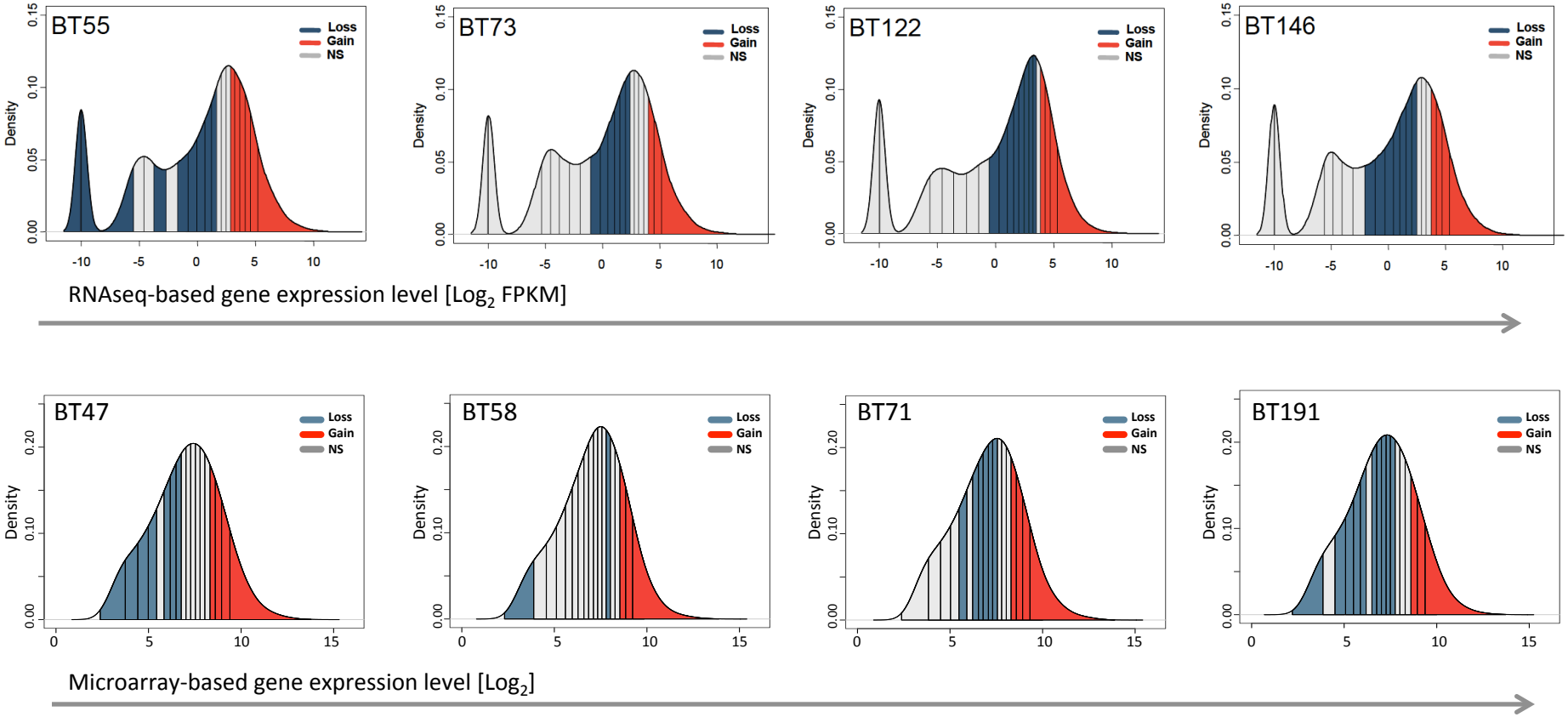
Yellow: synergistic effect of siRNA and Lapatinib (ratio < 1.0, P-value < 0.05, Student's t-test)

Blue: antagonistic effect of siRNA and Lapatinib (ratio > 1.0, P-value < 0.05, Student's t-test)

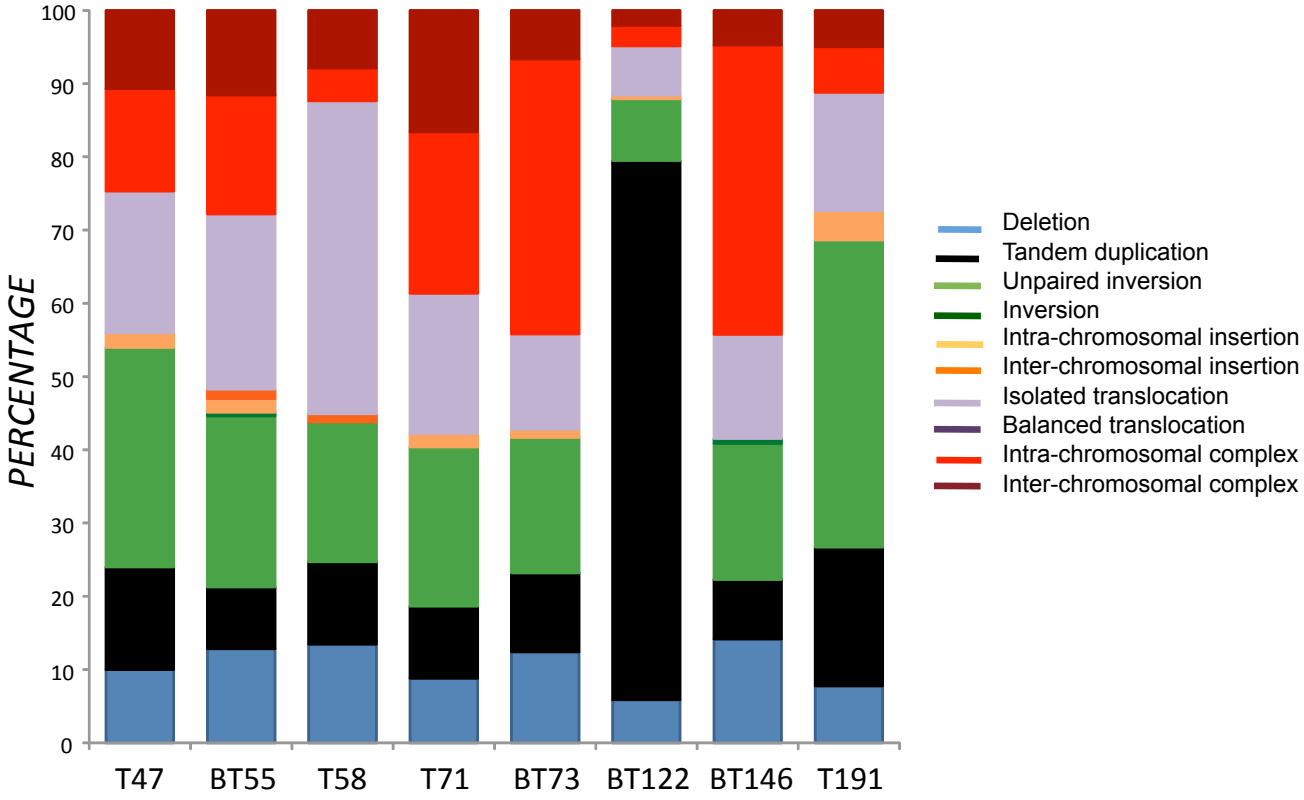
Supplemental Figures



Supplemental Figure 1. Genomic organization of eight breast tumors.
Circos plots representing the somatic genomic structure of the eight breast cancer studied. Tracks from the outermost to the innermost: 1) *Hg18* cytogenetic banding; 2) Copy number changes in the tumour genome compared to its normal counterpart; 3) Somatic structural variations as defined by DNA-PET analysis (described in Hillmer *et al*. 2011). CNV, Copy Number Variations; SV, Structural Variations. ER, PR and ERBB2 immunohistochemistry scores, as well as the breast cancer subtype classification, are indicated next to the sample ID.

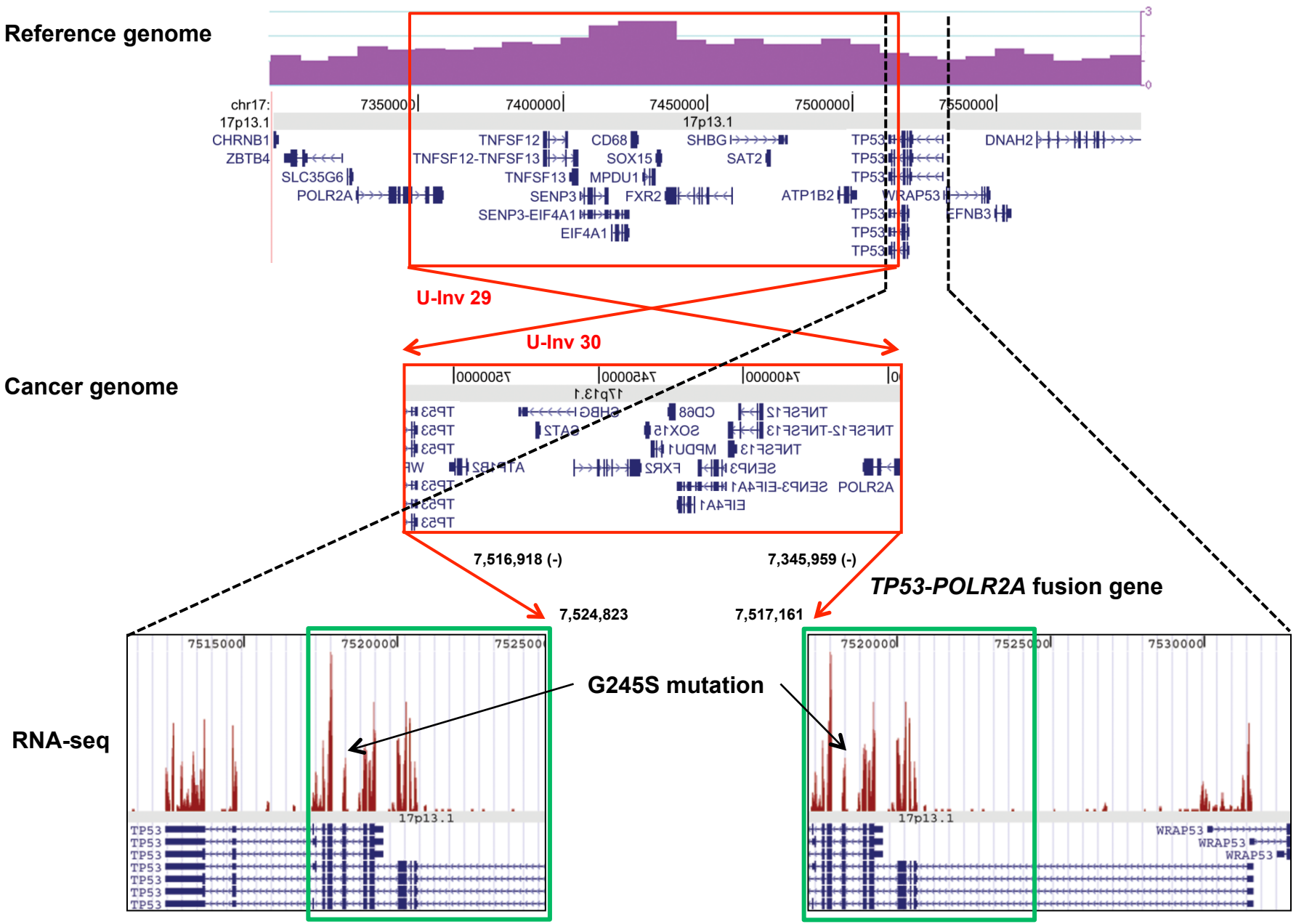


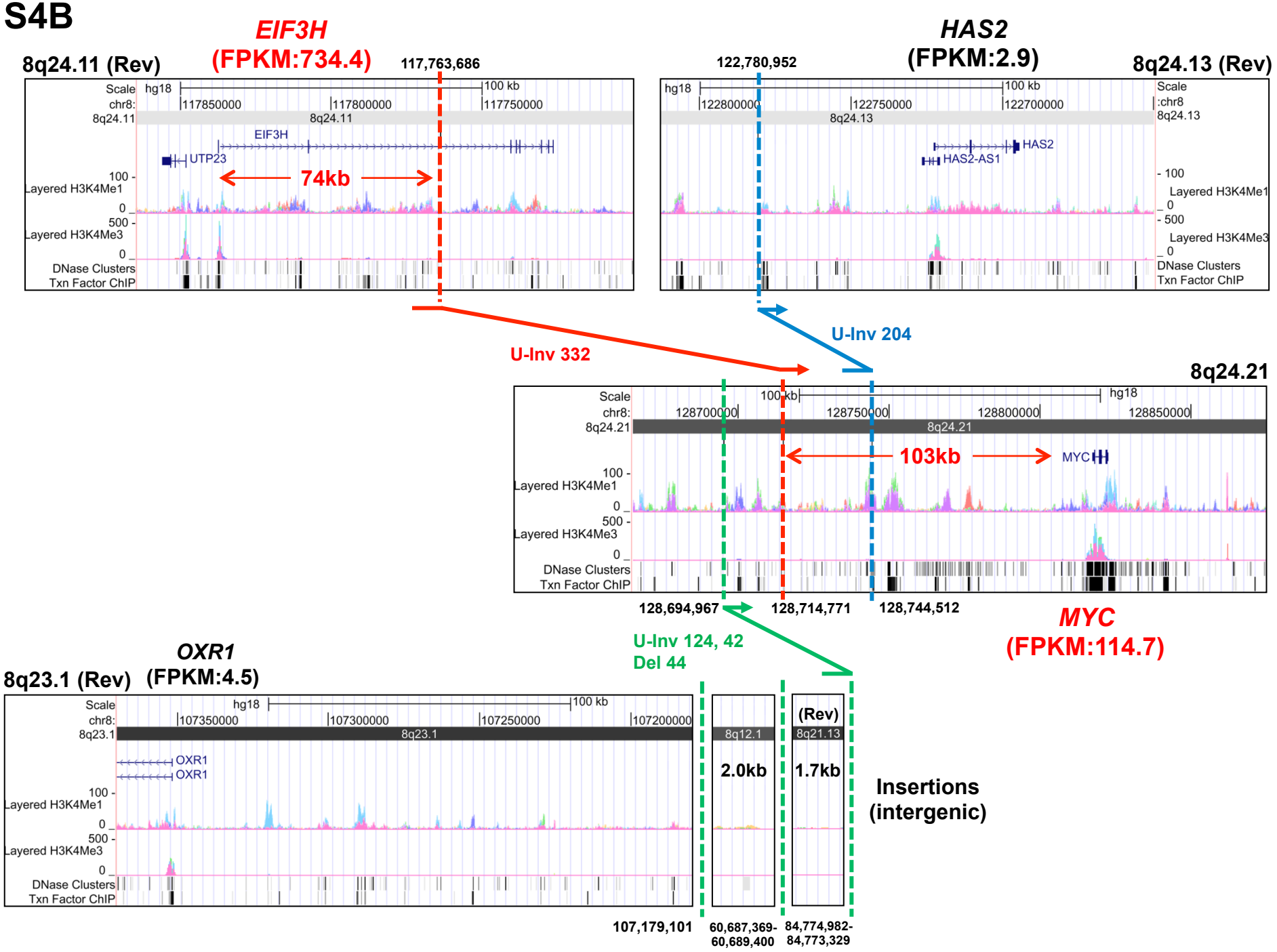
Supplemental Figure 2. Copy Number Variations affect the overall transcriptional output of cancer samples.
For each tumor examined, genes affected by copy number gain and loss associate with high and low absolute levels of gene expression, respectively. For each sample, genes were ranked based on their level of expression and binned accordingly into 2,000-gene bins, with contiguous bins overlapping by 1,000 genes. For each bin, the Fisher's exact test was used to test the significance of the enrichment in genes affected by copy number change. Significant bins (P-value < 0.05) are colorized in the density plot according to the direction of the enrichment: gain (red) or loss (blue). NS, not significantly enriched.



Supplemental Figure 3. Percentage of somatic structural variations in the eight breast tumors analyzed.

S4A

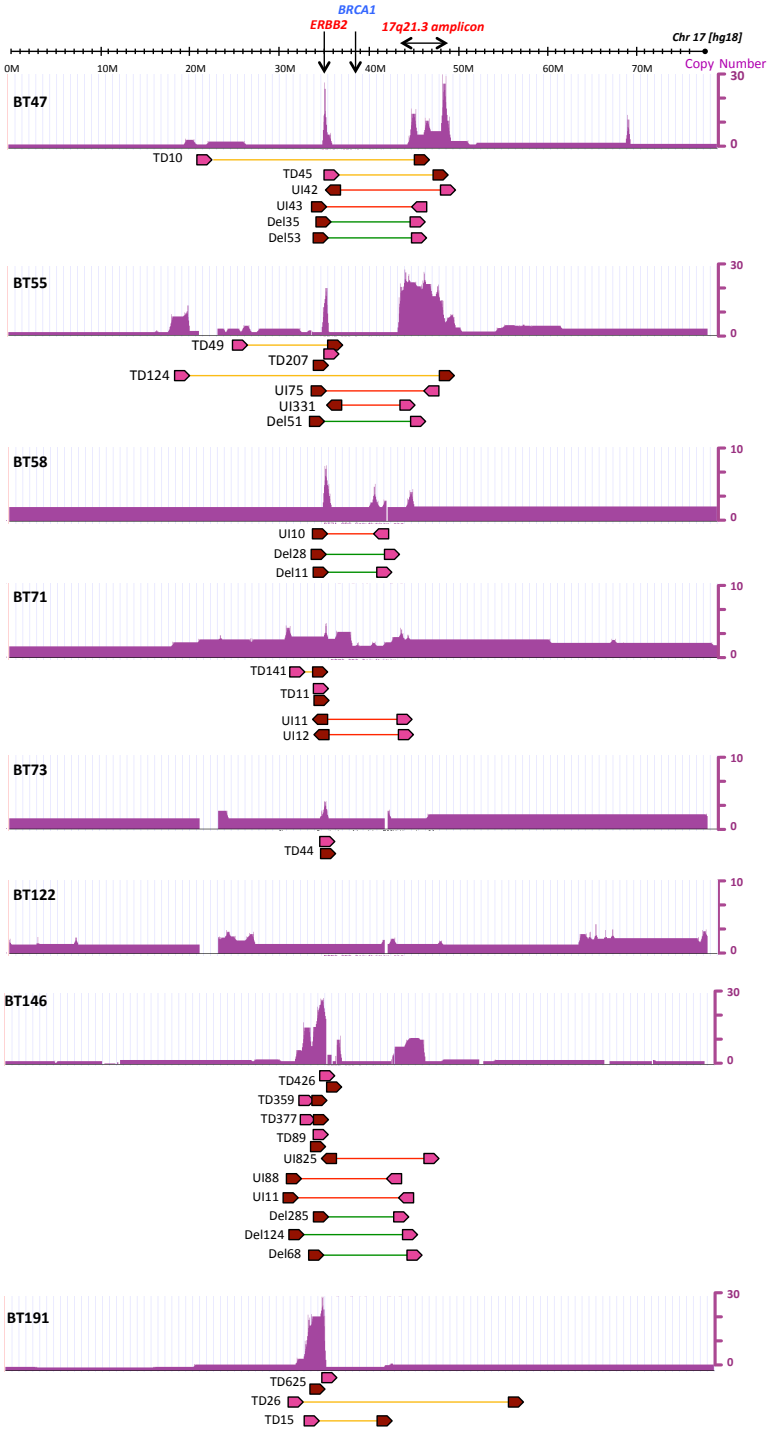

Reference genome
chr17:
7350000
7400000
7450000
7500000
7550000
17p13.1
CHRNB1
ZBTB4
SLC35G6
POLR2A
TNFSF12
TNFSF12-TNFSF13
TNFSF13
SENP3
SENP3-EIF4A1
EIF4A1
CD68
SOX15
MPDU1
FXR2
SHBG
SAT2
ATP1B2
TP53
WRAP53
EFNB3
DNAH2
U-Inv 29
U-Inv 30
Cancer genome
7,516,918 (-)
7,345,959 (-)
TP53-POLR2A fusion gene
7,524,823
7,517,161
RNA-seq
G245S mutation
7515000
7520000
7525000
7530000
WRAP53


S4B
EIF3H
(FPKM:734.4)
8q24.11 (Rev)
117,763,686
122,780,952
HAS2
(FPKM:2.9)
8q24.13 (Rev)
Scale
hg18
100 kb
chr8:
117850000
117800000
117750000
8q24.11
EIF3H
UTP23
74kb
100
Layered H3K4Me1
0
500
Layered H3K4Me3
DNase Clusters
Txn Factor ChIP
122800000
122750000
122700000
8q24.13
HAS2
HAS2-AS1
U-Inv 332
U-Inv 204
8q24.21
128700000
128750000
128800000
128850000
103kb
MYC
128,694,967
128,714,771
128,744,512
MYC
(FPKM:114.7)
U-Inv 124, 42
Del 44
8q23.1 (Rev)
OXR1
(FPKM:4.5)
107350000
107300000
107250000
107200000
8q23.1
OXR1
8q12.1
2.0kb
(Rev)
8q21.13
1.7kb
Insertions
(intergenic)
107,179,101
60,687,369-
60,689,400
84,774,982-
84,773,329

Supplemental Figure 4. Somatic structural variations affecting gene integrity

(*A*) In the BT122 sample, an inverted duplication on 17p causes the truncation of the *TP53* gene and its fusion to the *POLR2A* gene located upstream. This event corresponds to low levels of *TP53* expression (6.4 FPKM in BT122, 3 times lower than the average expression across the four breast cancer samples for which RNAseq data is available). (*B*) In the BT73 sample, an intra-chromosomal inversion causes the translocation of the highly expressed *EIF3H* gene promoter to ~100 Kbps upstream of the *MYC* locus on Chr 8q.

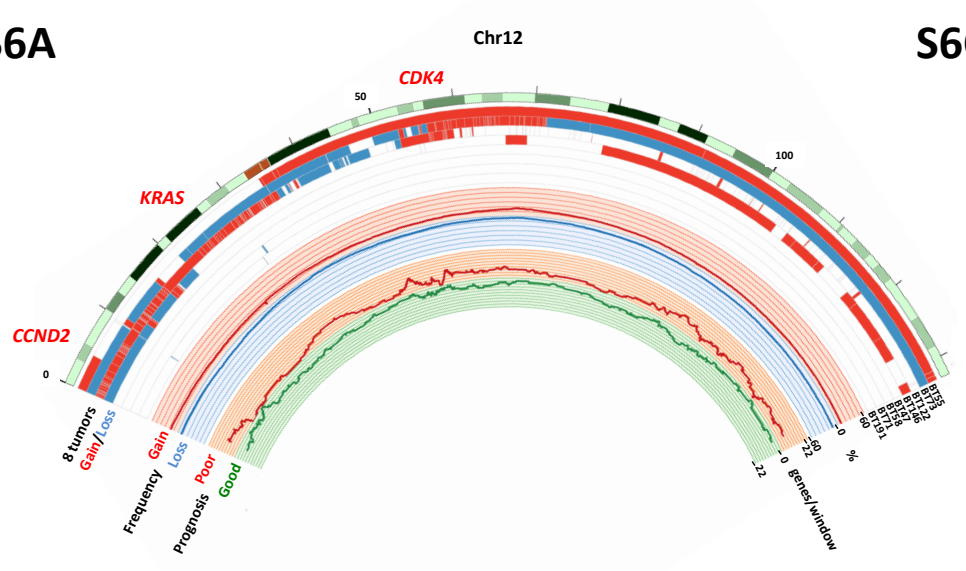
S5


BRCA1
ERBB2
17q21.3 amplicon
Chr 17 [hg18]
0M
10M
20M
30M
40M
50M
60M
70M
Copy Number
BT47
30
0
TD10
TD45
UI42
UI43
Del35
Del53
BT55
30
0
TD49
TD207
TD124
UI75
UI331
Del51
BT58
10
0
UI10
Del28
Del11
BT71
10
0
TD141
TD11
UI11
UI12
BT73
10
0
TD44
BT122
10
0
BT146
30
0
TD426
TD359
TD377
TD89
UI825
UI88
UI11
Del285
Del124
Del68
BT191
30
0
TD625
TD26
TD15

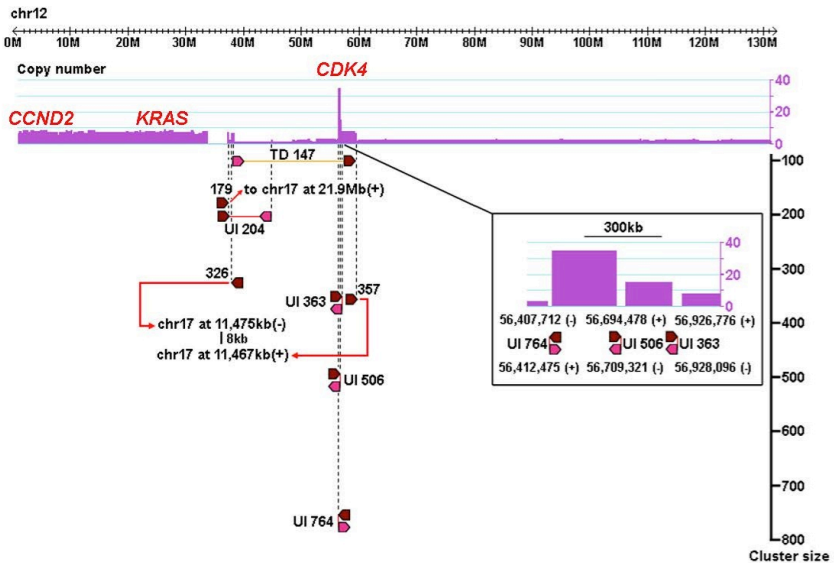
Supplemental Figure 5. Recurrent structural variations affecting Chr 17.

Copy number profile and somatic recurrent structural variations affecting Chr 17 in the eight analyzed breast cancers. A more detailed description of the somatic events affecting Chr 17 in BT55, BT146, BT47 and BT73 can be found in Fig. 4 and Supplemental Figs. 9A-D. For each structural variation, the brown and pink arrowheads correspond to the left and right mate pair read clusters, respectively. SV break points map to the end of the brown arrowhead and to the start of the pink arrowhead (as detailed in Hillmer *et al.* 2011). TD, tandem duplication (yellow); UI, unpaired inversion (red); Del, deletion (green). The number associated with each structural variation corresponds to its cluster size.

S6A

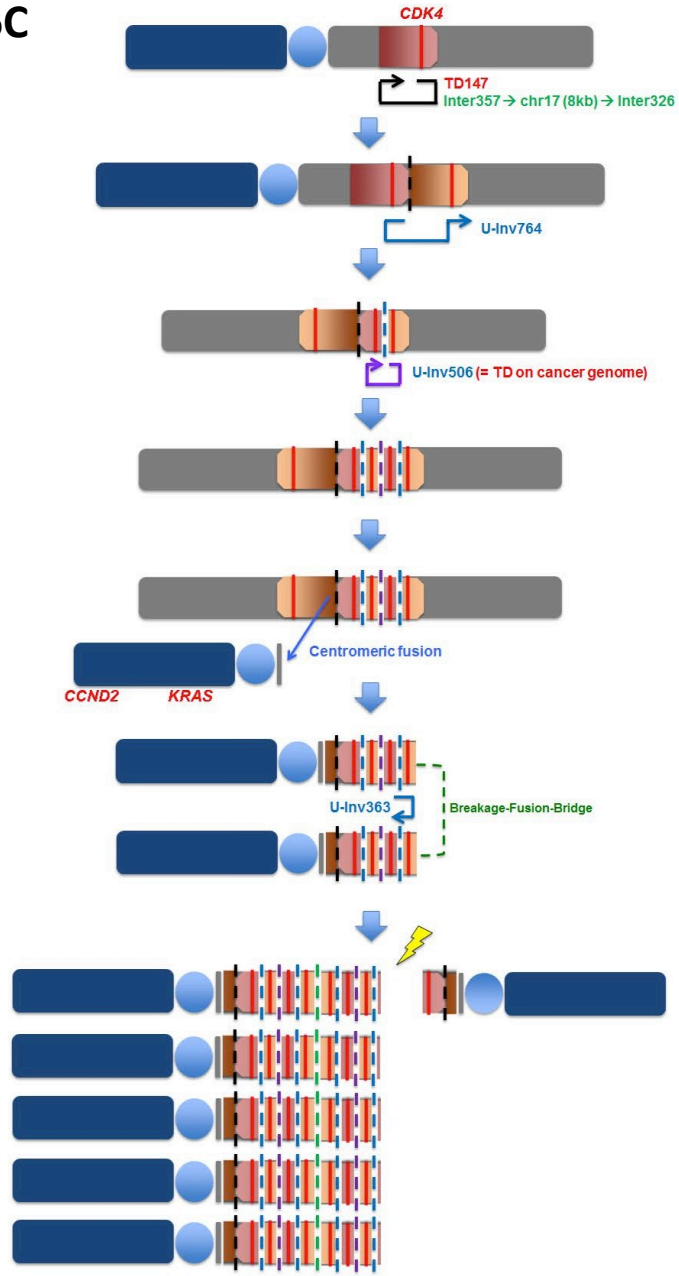


S6B



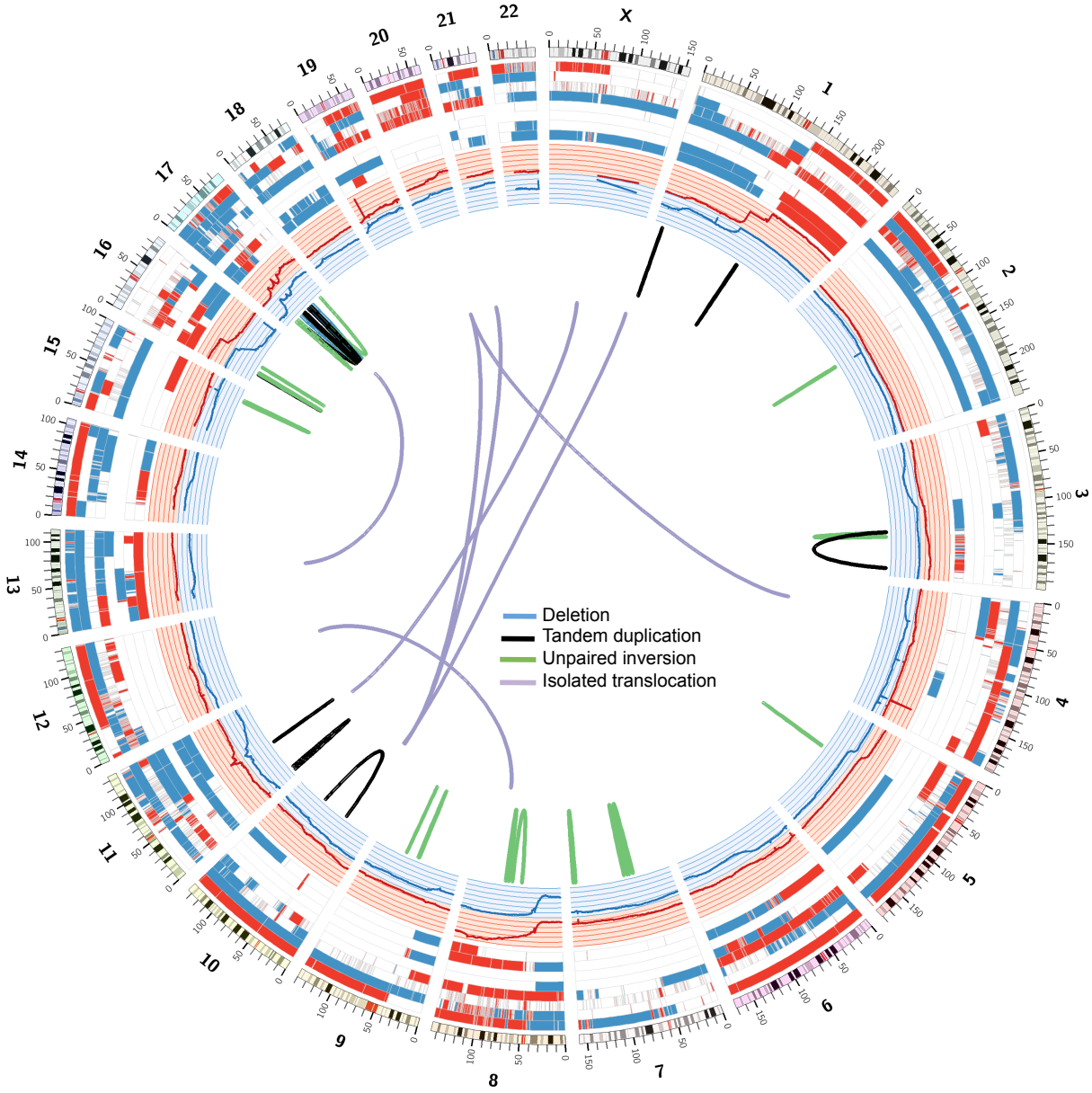
	Copy number ratio (vs. normal)				RNA-seq (FPKM)			
	BT122	BT146	BT55	BT73	BT122	BT146	BT55	BT73
CCND2	3.3	0.7	1.5	0.6	5.1	1.5	4.7	3.7
KRAS	3.0	1.0	1.0	0.6	12.6	4.7	2.3	1.4
CDK4	16.8	1.0	1.3	1.1	2501.6	88.6	114.0	167.7

S6C



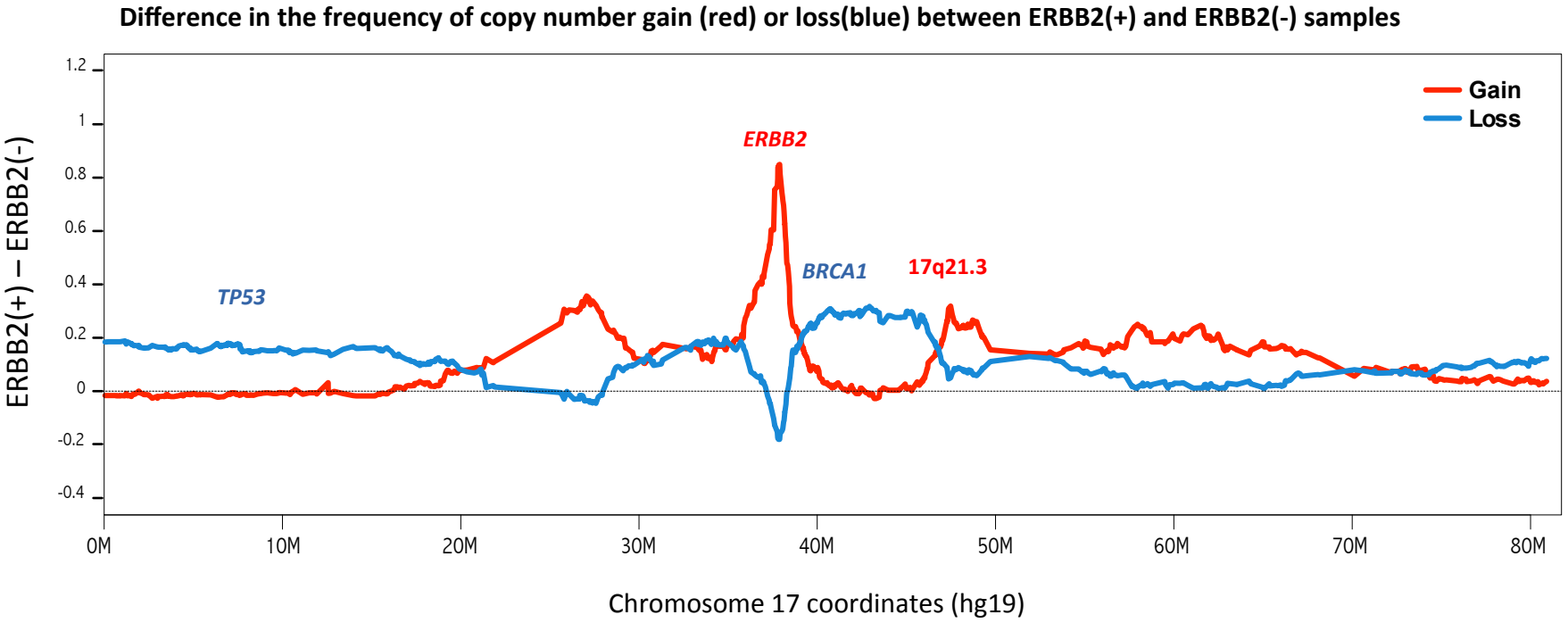
Supplemental Figure 6. Organization of Chr 12 in breast cancer

(*A*) Chromosomal map of Chr 12, showing the copy number state of the eight sequenced breast tumors, the frequency of copy number gain and loss across 1,174 breast cancer samples, and the number of genes associated with good and poor breast cancer patients' prognosis. (*B*) (upper panel) copy number and structural variations on Chr 12 in BT122. The inset represents a magnified view of the *CDK4* locus. TD, tandem duplication. UI, unpaired-inversion. Translocations are indicated by arrows. The number associated with each structural variation corresponds to its cluster size. Structural variations are depicted as in Supplemental Fig. 5. (lower panel) copy number ratios (vs. the paired normal genome) and levels of gene expression (FPKM by RNA-seq) of the *CCND2*, *KRAS*, and *CDK4* genes in the four tumors for which RNAseq expression data is available. (*C*) Plausible mechanism of Chr 12 structural rearrangement in BT122. The sequence of structural variations described in the figure converge to a tandem duplication encompassing the *CDK4* gene. The segmental duplication was followed by unpaired inversions UI764 and UI506, which corresponds to a tandem duplication in the cancer genome, leading to an increase in the copy number of the *CDK4* locus. The amplified DNA fragment presumably connects to the short arm of Chr 12, which harbors the *KRAS* and *CCND2 genes*, via a centromeric fusion, followed by unpaired inversion UI363, indicative of a breakage-fusion-bridge event. Finally, the rearranged Chr 12 is further amplified through aneuploidy.



Supplemental Figure 7. Circos plot of recurrent structural variations.
The figure shows the copy number state of eight tumors (red, copy number gain; blue, copy number loss; from the outmost to the innermost: BT55, BT73, BT122, BT146, BT47, BT58, BT71, BT191), the frequency of copy number gain (red) and loss (blue) across a data set of 1,174 breast tumors, and the recurrent structural mutations observed across the eight breast tumors analyzed (as in Supplemental Table 7).

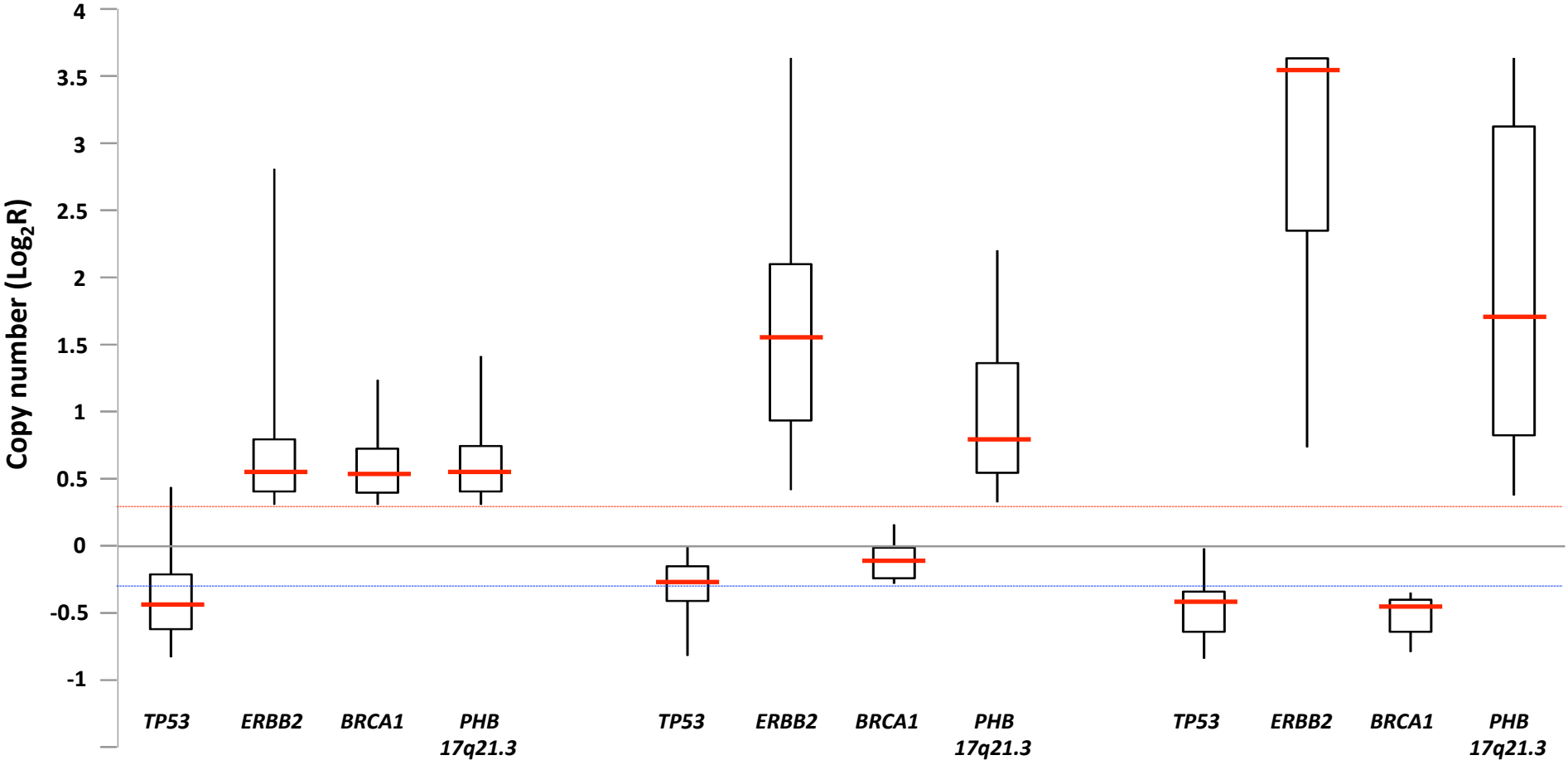
S8A



ERBB2 status	Number of samples	Frequency of TP53 loss	Frequency of BRCA1 loss	Frequency of 17q21.3 gain
ERBB2-amp (+)	178	0.58	0.47	0.54
ERBB2-amp (-)	996	0.41	0.19	0.22
P-value*		1.777e-05	4.134e-14	< 2.2e-16

* Fisher's Exact Test

S8B

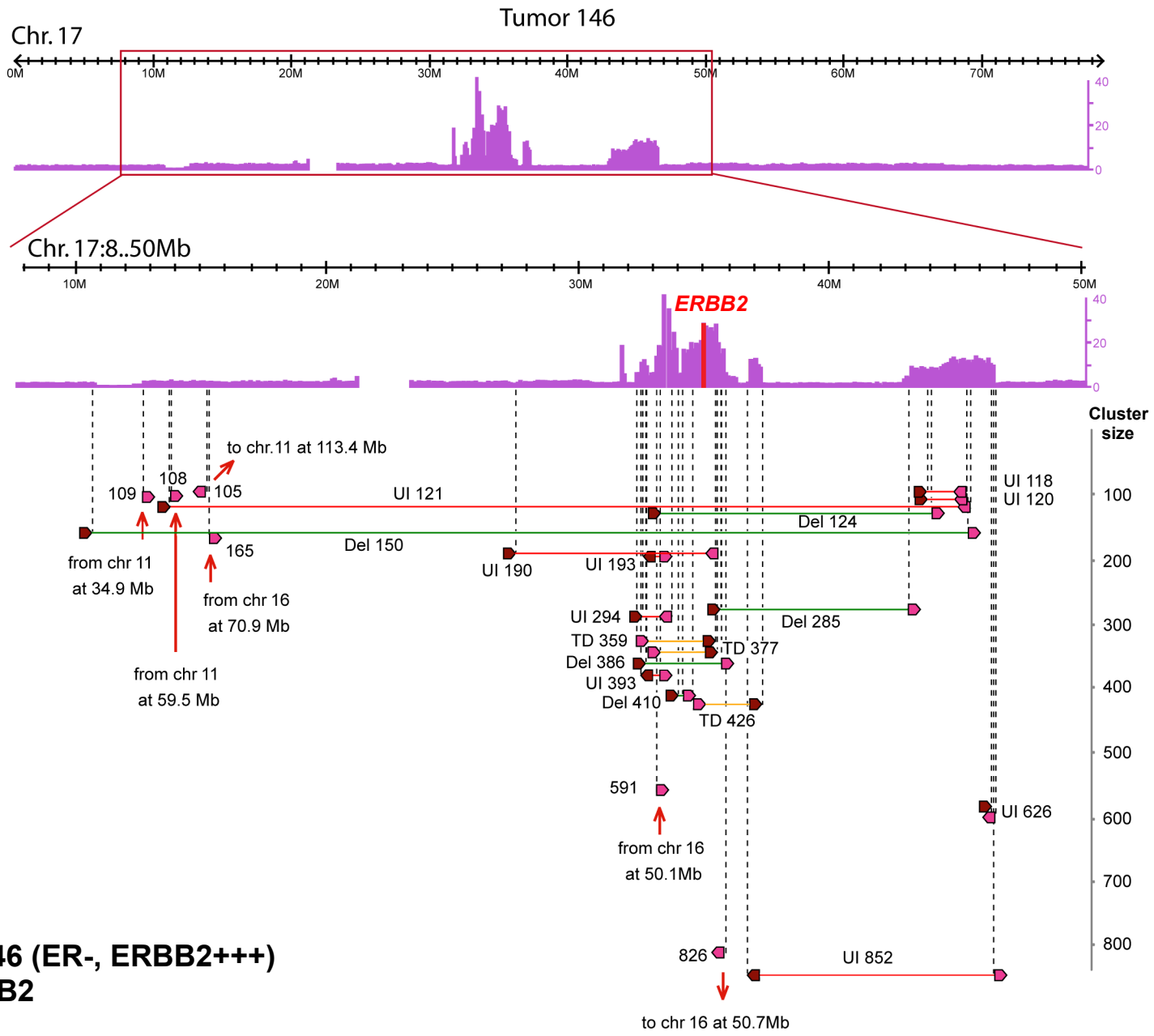


Number	125				39				53			
Median Log_2R	-0.46	0.55	0.53	0.55	-0.27	1.55	-0.1	0.77	-0.43	3.66	-0.48	1.68
Median copy number	1.5	2.9	2.9	2.9	1.7	5.6	1.9	3.4	1.5	25.3	1.4	6.4
P-value (vs. Gain)					NS	4.8E-06		4.0E-03	NS	<2.2E-16		1.7E-09

Supplemental Figure 8. Structure of Chr 17 in breast cancer

(*A*) Difference in the frequency of copy number gain (red) or loss(blue) between ERBB2(+) and ERBB2(-) samples (n=1,174 breast cancer samples). (B) Copy number log2 ratio distributions of critical Chr 17 genes, based on a total of 217 breast cancer samples characterized by additional copies of both the *ERBB2* and the 17q21.3 (*PHB*) loci. Box plots represent the 0.25 and 0.75 percentiles, with vertical lines corresponding to 0.05 to 0.95 percentiles. Red horizontal lines indicate median values. Samples were separated into three groups based on the *BRCA1* gene copy number state. P-values are based on the Student's t-test. NS, not significant.

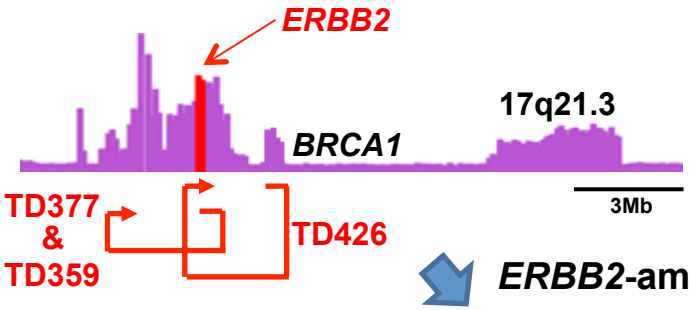
S9A



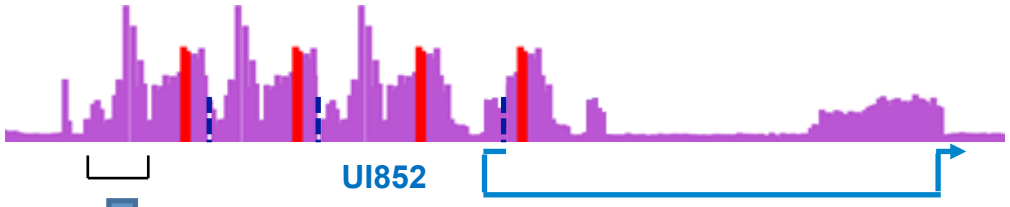
BT146 (ER-, ERBB2+++)
ERBB2

S9B

BT146 (ER-, ERBB2+++)
ERBB2

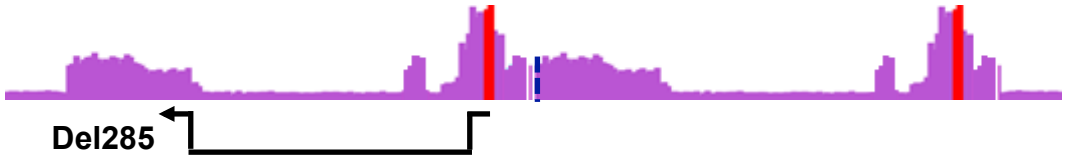


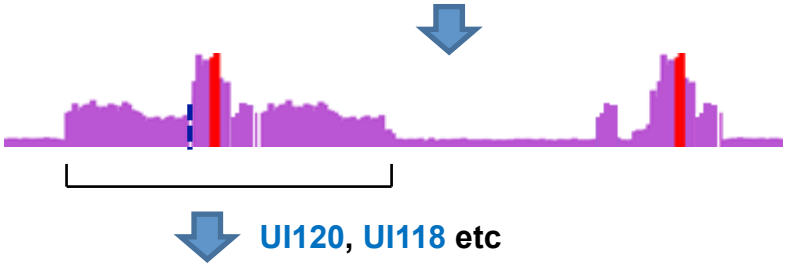
ERBB2-amplification



UI393, UI294 etc

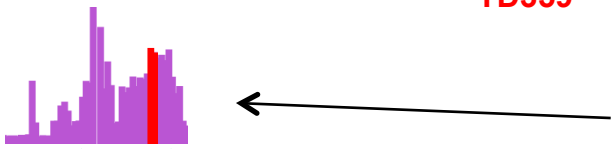
Amplification



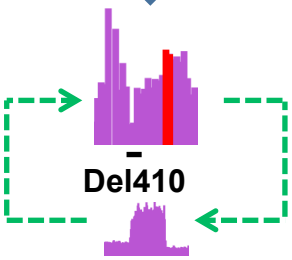


UI120, UI118 etc

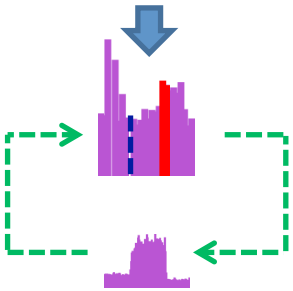
Co-amplification



Inter591 Inter826



Chr 16: 50.1Mb-50.7Mb

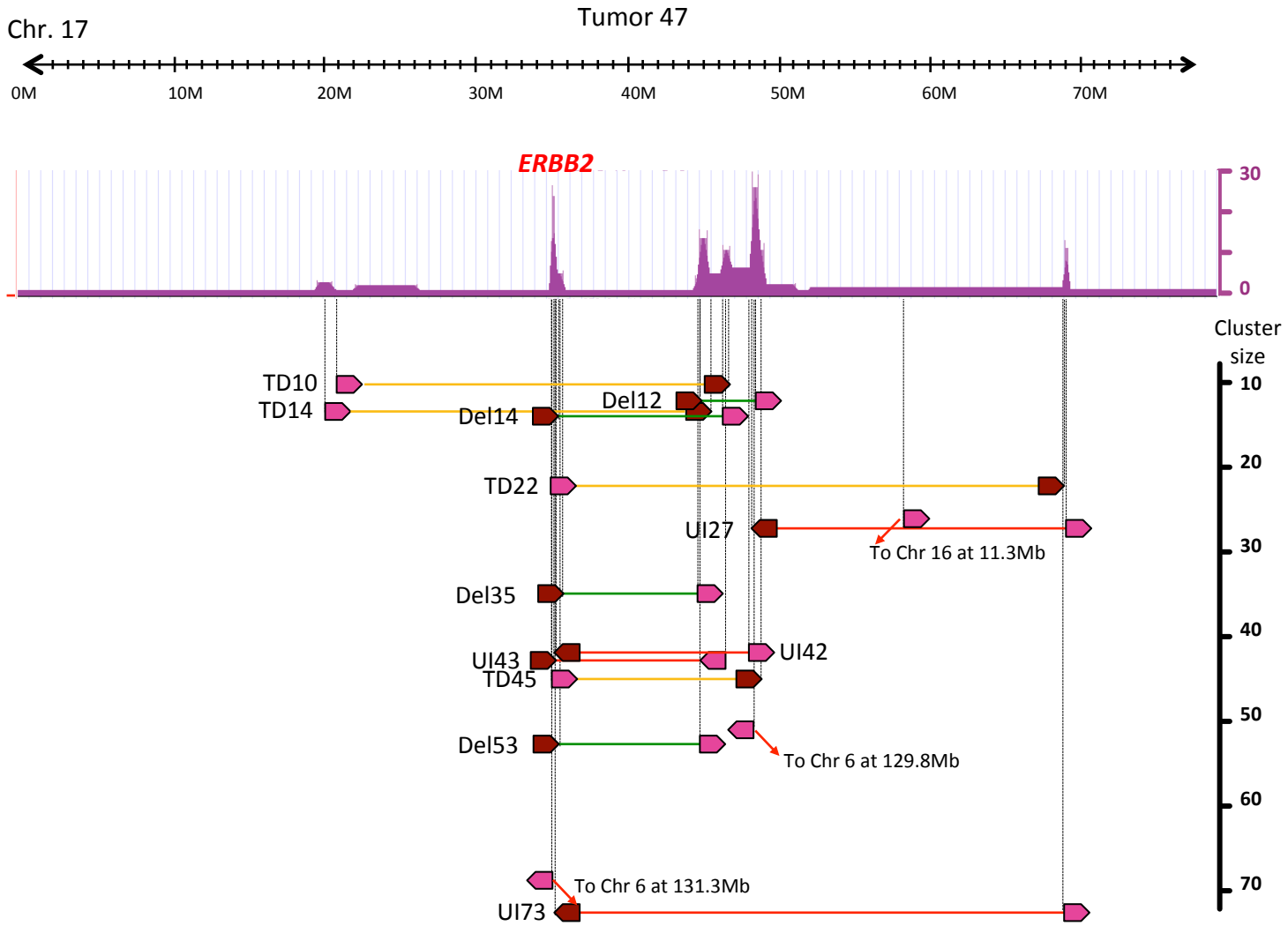


Chr 16: 50.1Mb-50.7Mb

Episomal ?

ERBB2-amplification

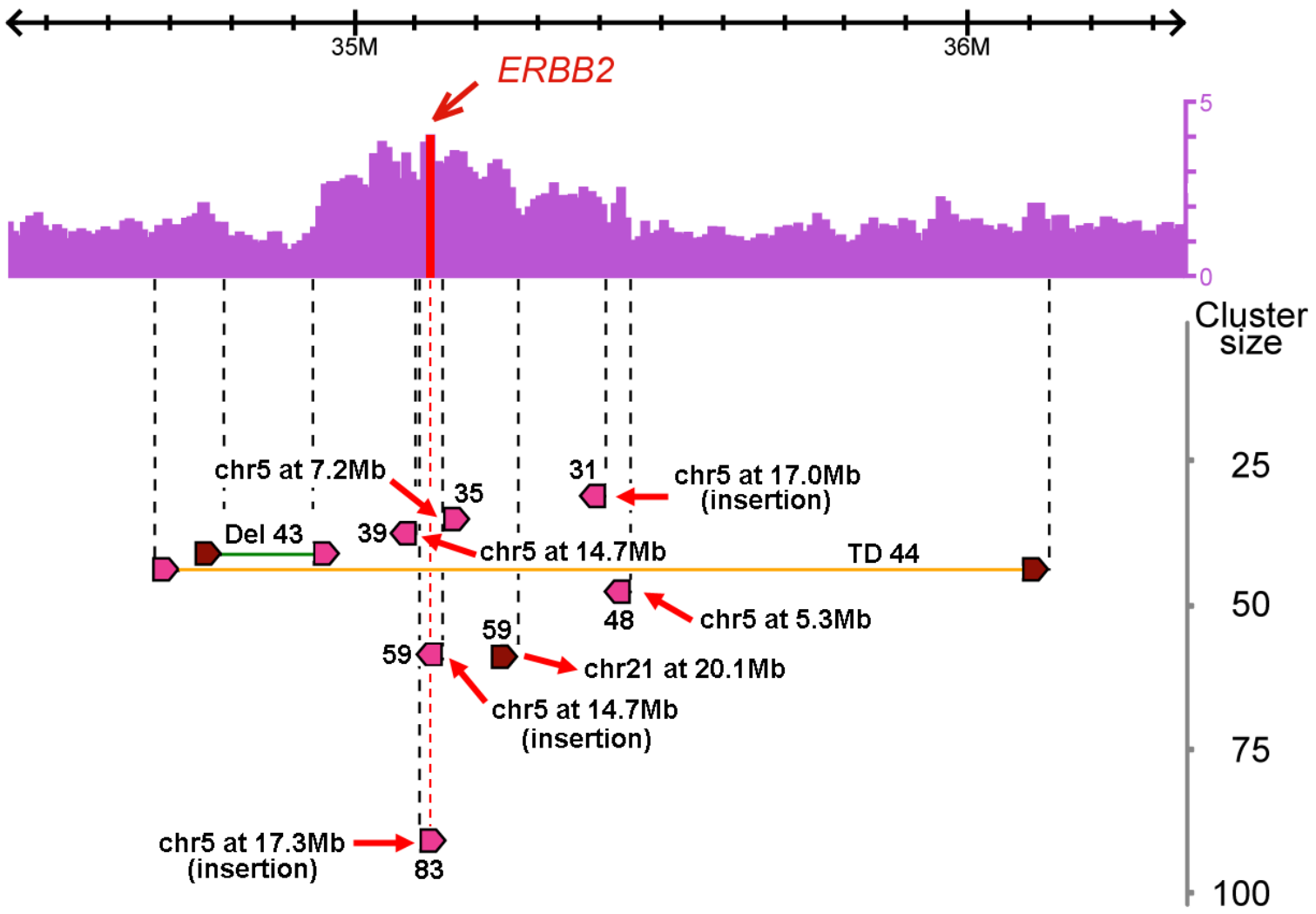
S9C



BT47 (ER-, PR+, ERBB2+++)

Luminal B

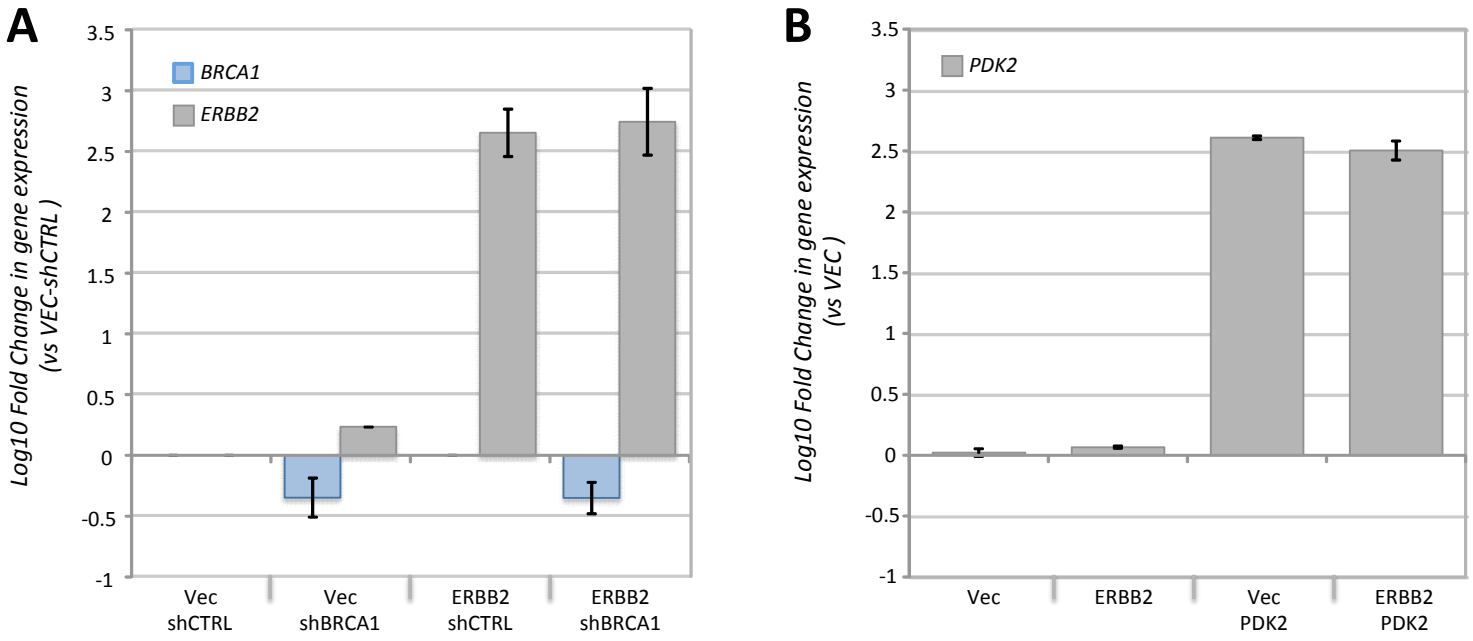
S9D



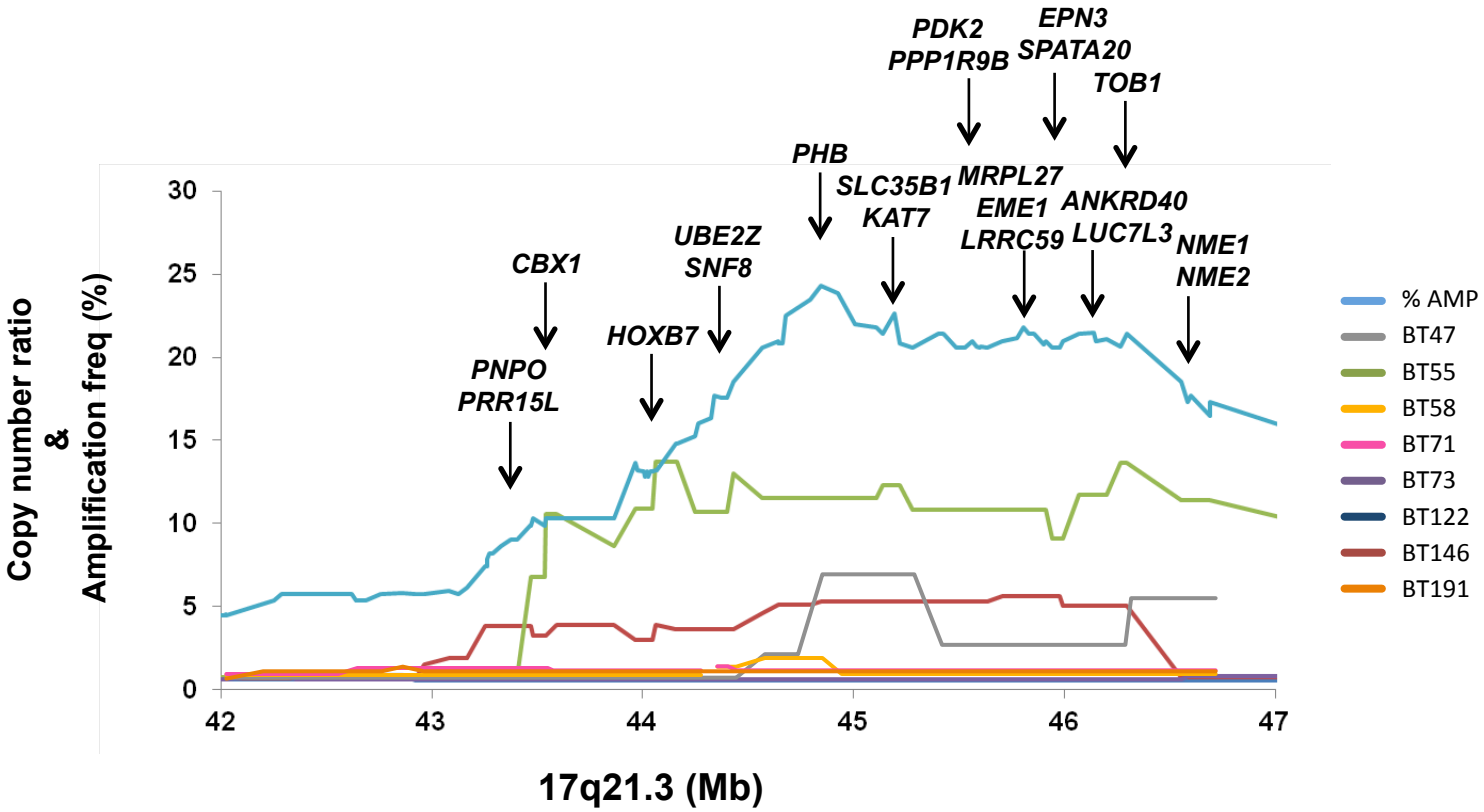
BT73 (ER+++, ERBB2+)
Luminal B

Supplemental Figure 9. Detailed structure of Chr 17 in three *ERBB2*-positive breast cancers

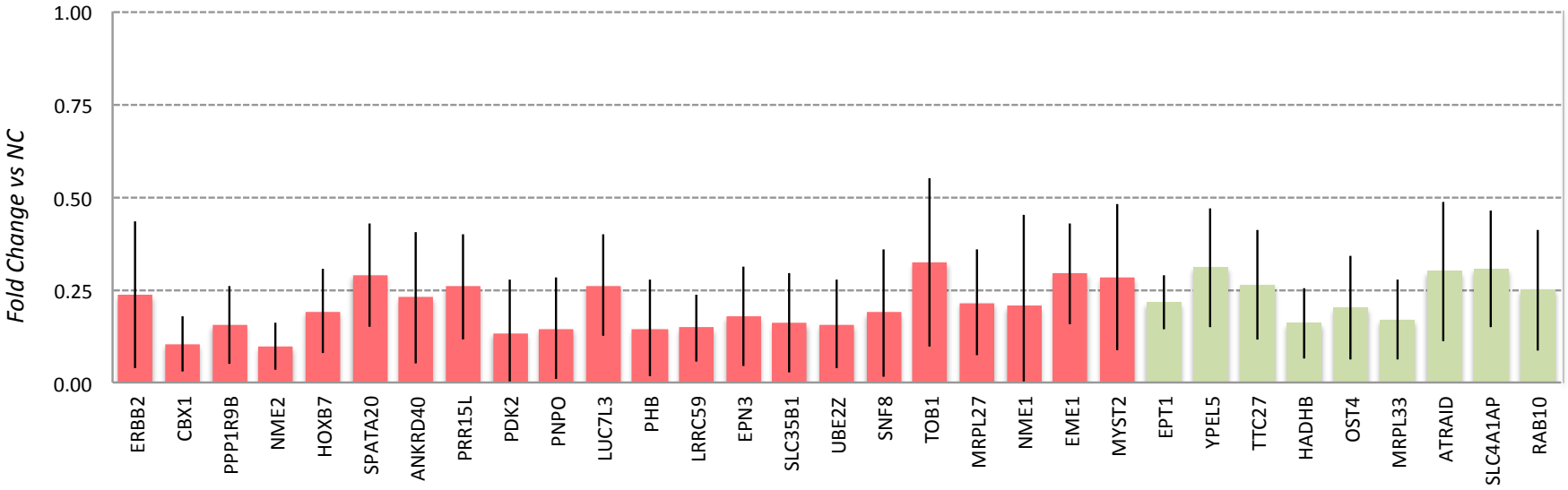
(A) Copy number and structural variations on Chr 17 (8Mb-50Mb) for the BT146 sample. (*B*) Plausible mechanism of Chr17 structure of BT146. (*C*) Copy number and structural variations on Chr 17 in the BT47 sample, showing tandem duplications spanning *ERBB2* (TD45, TD22, TD14 and TD10), unpaired-inversions connecting *ERBB2* and 17q21.3 (UI42 and UI43), and deletions affecting *BRCA1* (Del12, Del35 and Del53). (D) Copy number and structural variations on Chr 17 (34.5Mb-36.3Mb) in the BT73 sample. Del, deletion; TD, tandem duplication; UI, unpaired-inversion; Inter-chromosomal translocations are indicated by arrows. The number associated with each structural variation corresponds to its cluster size. structural variations are depicted as in Supplemental Fig. 5.



Supplemental Figure 10. Fold changes in gene expression assessed by Real Time PCR.
(A) Real Time PCR analysis of *BRCA1* and *ERBB2* gene expression in human mammary epithelial cells (HMECs) transduced with a combination of a scrambled shRNA (shCTRL) or an shRNA targeting the *BRCA1* mRNA (shBRCA1) and a mock vector (Vec) or a vector expressing the human *ERBB2* gene (ERBB2). The *ACTB* gene was used as housekeeping control. Bars correspond to averages of duplicate (for *BRCA1*) or triplicate (for *ERBB2)* experiments +/- standard deviations. (B) Real Time PCR analysis of *PDK2* gene expression in HMECs transduced with a combination of a mock vector (Vec) or a vector expressing the *ERBB2* gene (ERBB2) and a vector expressing the *PDK2* gene (PDK2). The *ACTB* gene was used as housekeeping control. Bars correspond to averages of duplicate experiments +/- standard deviations.

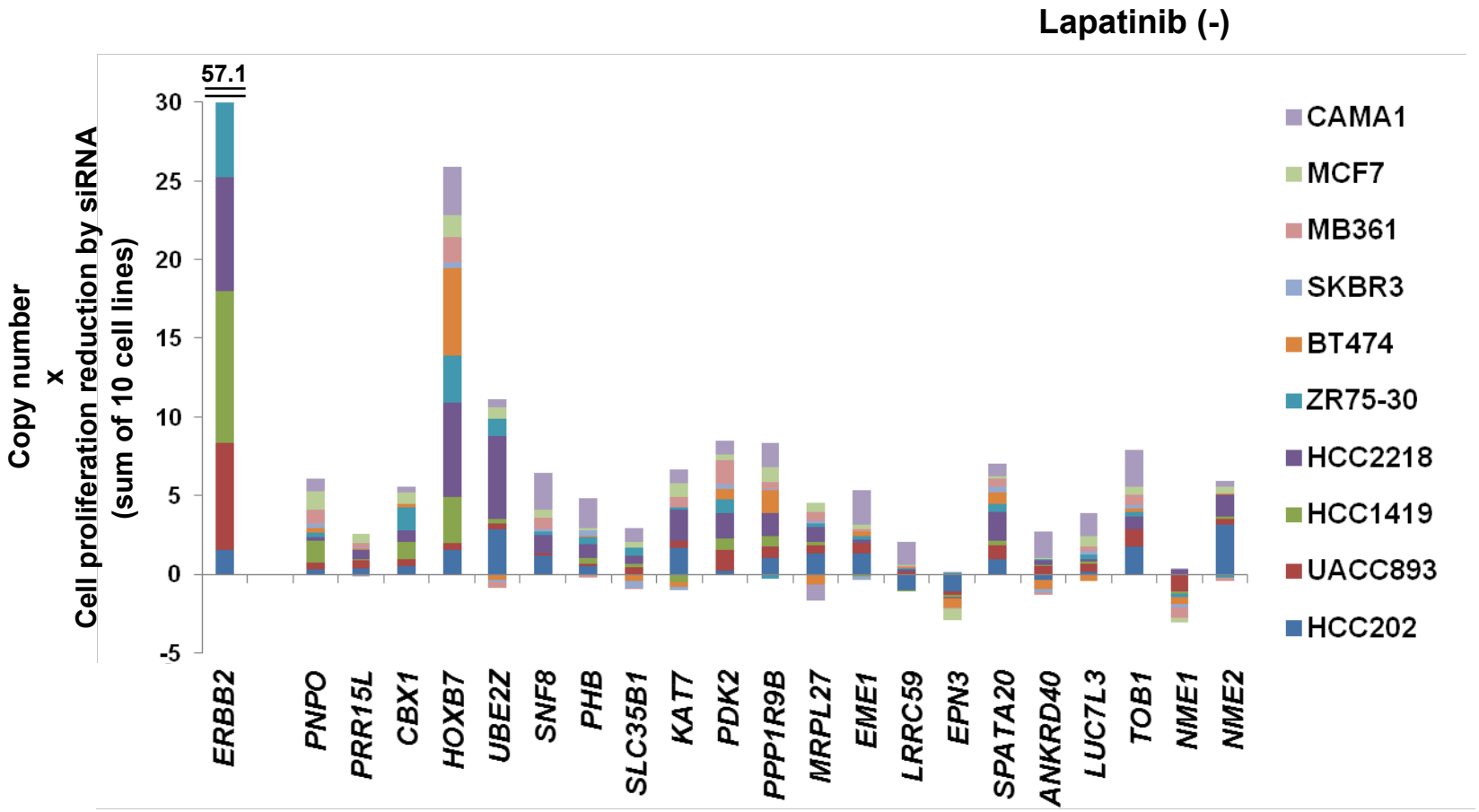


Supplemental Figure 11. Putative oncogene cluster on 17q21.3
Copy number ratio for the eight tumor samples examined and frequency of copy number gain along the 17q21.3 locus, based on the Tumorscape breast cancer copy number data set. Oncogene candidates are indicated by arrows.

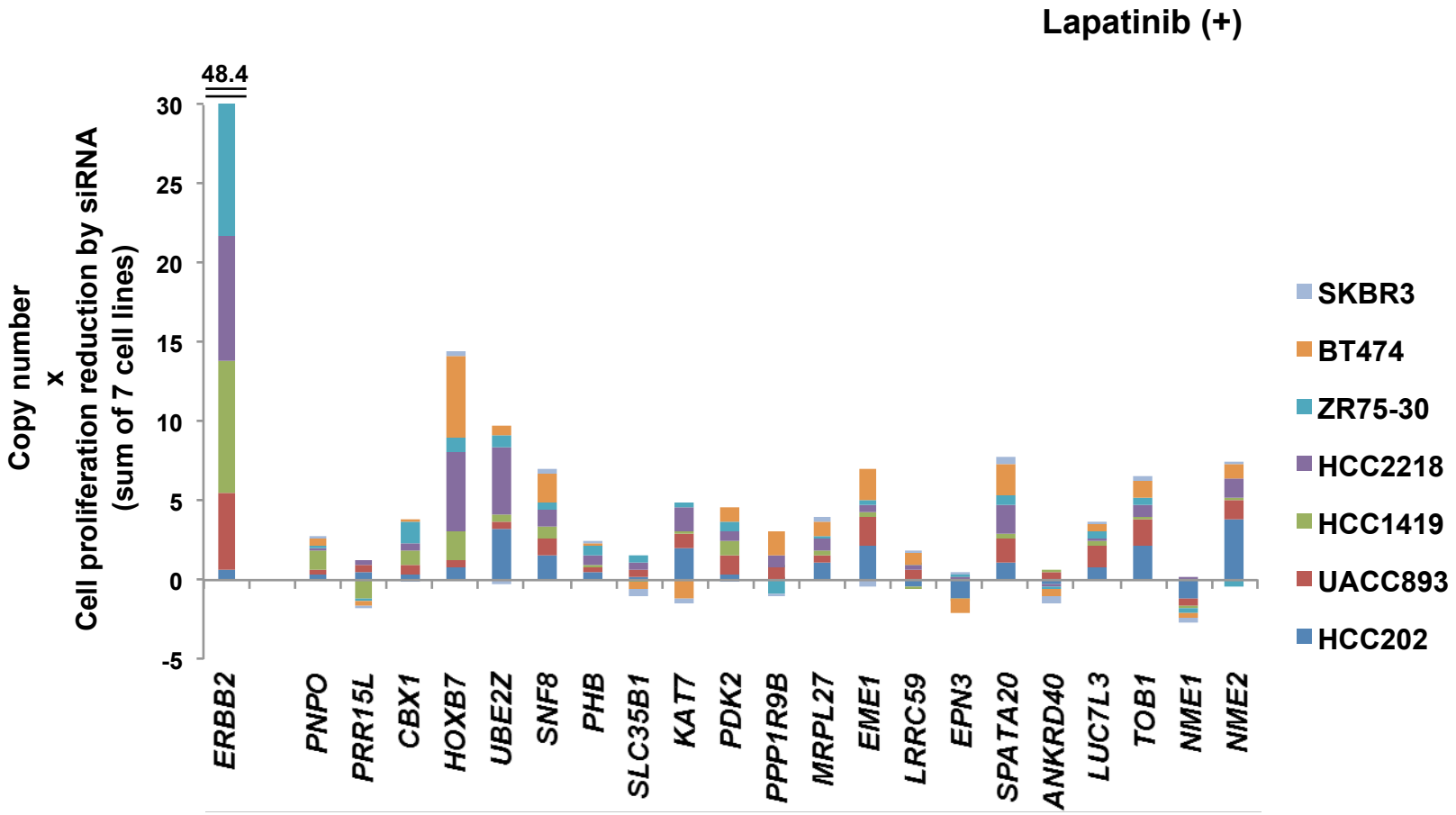


Supplemental Figure 12. Fold changes in gene expression following siRNA treatment.
Fold changes in gene expression values of siRNA targeted genes in the siRNA-treated cell lines relative to non-targeting siRNA treated cells (NC). The bars correspond to average values across the different cell lines examined with standard deviations represented by the error bars . Red, genes mapping to Chr 17, n=10 cell lines; green, genes mapping to Chr 2, n=6.

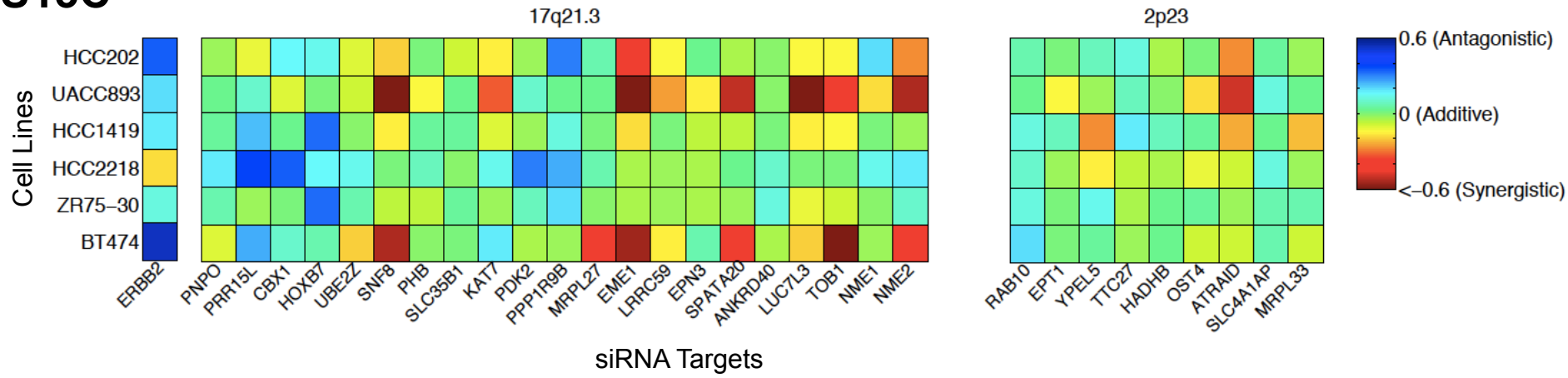
S13A


Lapatinib (-)
Copy number
x
Cell proliferation reduction by siRNA
(sum of 10 cell lines)
57.1
30
25
20
15
10
5
0
-5
CAMA1
MCF7
MB361
SKBR3
BT474
ZR75-30
HCC2218
HCC1419
UACC893
HCC202
ERBB2
PNPO
PRR15L
CBX1
HOXB7
UBE2Z
SNF8
PHB
SLC35B1
KAT7
PDK2
PPP1R9B
MRPL27
EME1
LRRC59
EPN3
SPATA20
ANKRD40
LUC7L3
TOB1
NME1
NME2

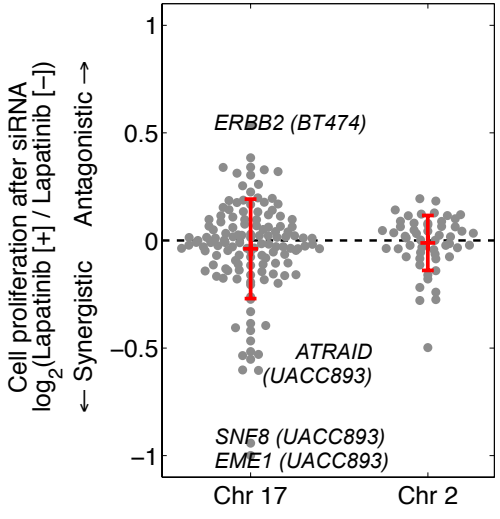
S13B


Lapatinib (+)
Copy number
x
Cell proliferation reduction by siRNA
(sum of 7 cell lines)
48.4
30
25
20
15
10
5
0
-5
ERBB2
PNPO
PRR15L
CBX1
HOXB7
UBE2Z
SNF8
PHB
SLC35B1
KAT7
PDK2
PPP1R9B
MRPL27
EME1
LRRC59
EPN3
SPATA20
ANKRD40
LUC7L3
TOB1
NME1
NME2
SKBR3
BT474
ZR75-30
HCC2218
HCC1419
UACC893
HCC202

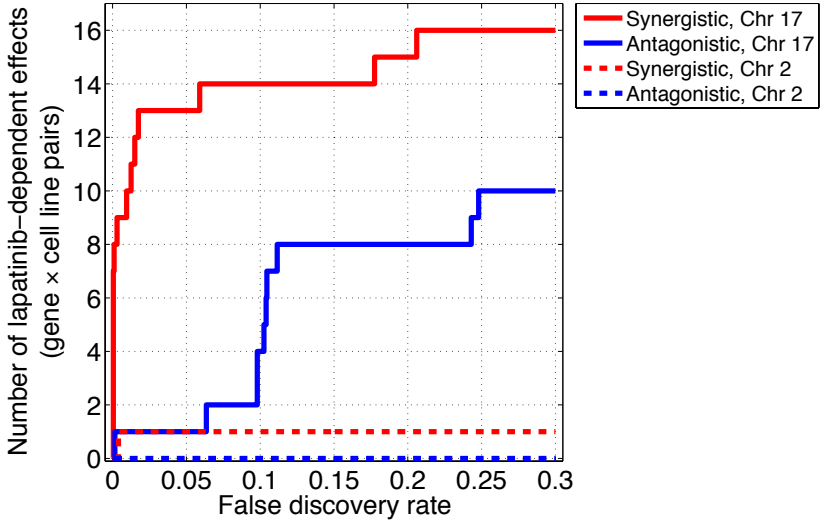
S13C



S13D



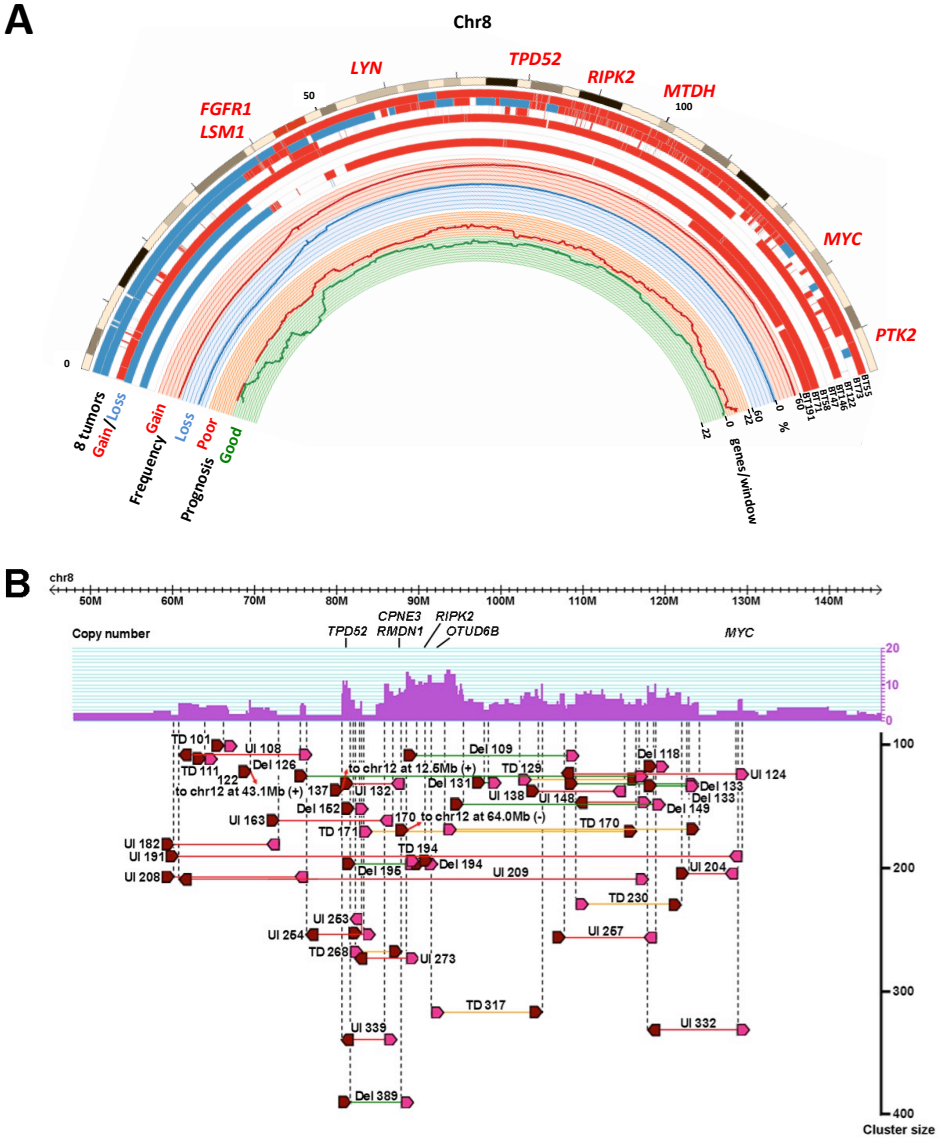
S13E



Supplemental Figure 13. Functional consequences of Chr 17 structure of breast cancer

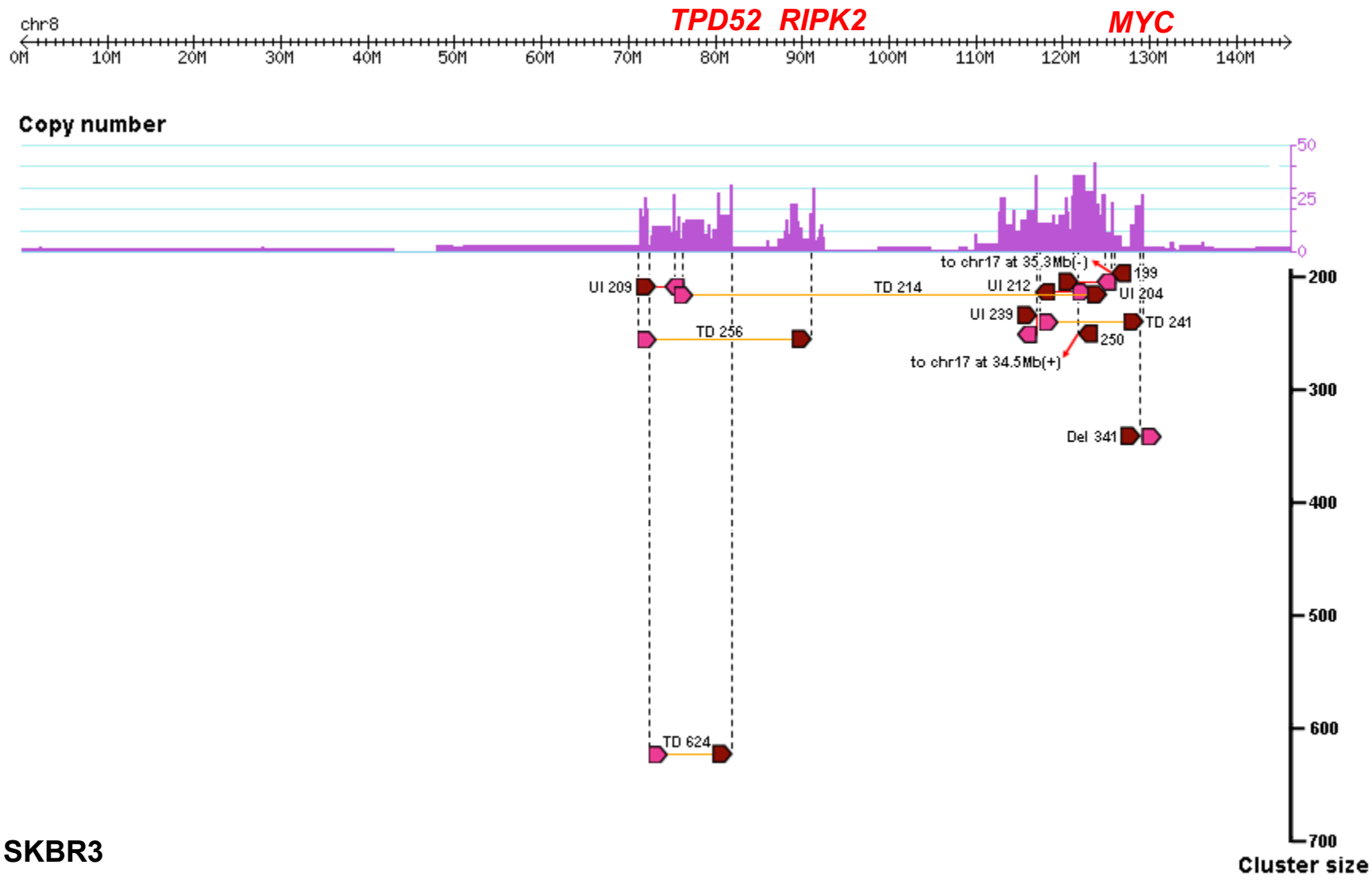
(*A*) Cell proliferation assay for 10 breast cancer cell lines treated with siRNAs against 17q21.3 candidate oncogenes and *ERBB2*. The reduction in cell proliferation following siRNA gene targeting is multiplied by the corresponding copy number in each cell line (copy number dependency). Results for the 10 cell lines are summed and shown in the bar chart. Gene symbols are ordered based on chromosomal location (lower coordinates on the left). (*B*) Cell proliferation assay for seven breast cancer cell lines treated with both siRNAs and Lapatinib (at the appropriate IC50). Copy number dependencies are shown as in (*A*). (*C*) Heatmaps showing the interactions between the pharmacological inhibition of ERBB2 with Lapatinib and siRNA-mediated knockdown of genes on Chr 17q21.3 and Chr 2p23. Positive values (blue) indicate antagonistic effects on cell proliferation between the gene knockdown and Lapatinib, whereas negative values (red) indicate synergistic effects. Values are on a $\log_2$ scale and represent the ratio of each siRNA's effect on cell proliferation with/without Lapatinib. Rows indicate cell lines and columns indicate genes targeted by siRNA. (*D*) Scatter plot showing values from heatmaps in (*C*). Each data point represents one cell line × gene pair. Error bars represent the mean ± standard deviation within each chromosome. Four example cell line × gene pairs are shown. (*E*) The number of significant Lapatinib-dependent effects (among cell line × gene pairs) that have synergistic or antagonistic interactions with Lapatinib as a function of the false discovery rate used to denote significance. At a 5% false discovery rate, the synergistic pairs on Chr17 are EME1 (in BT474, HCC202, and UACC893 cells), LUC7LC (in UACC893 cell), MRPL27 (in BT474 cell), and NME2, SNF8, SPATA20, and TOB1 (in BT474 and UACC893 cells), whereas the antagonistic pair on Chr17 is ERBB2 (in BT474 cell). The synergistic pair on Chr 2p23 is ATRAID (in UACC893 cell). At 5% false discovery rate, 17q21.3 is significantly enriched (hypergeometric test, P = 7.14E-07) for Lapatinib-dependent cell line × gene pairs compared to 2p23 (counting both synergistic and antagonistic effects together). Considering only unique genes among the Lapatinib-dependent cell line × gene pairs, at a 5% false discovery rate, 17q21.3 (containing 8 unique genes) is still enriched for Lapatinib-dependent genes compared to Chr 2p23 (containing 1 unique gene) (hypergeometric test, P = 0.0059).

S14



Gene	Location (Mb)	DMFS		Copy number ratio				RNA-seq (FPKM)				Gene Ex (TCGA)	
		Log (HR)	P-value	BT122	BT146	BT55	BT73	BT122	BT146	BT55	BT73	FC (log2)	P-value
TPD52	81.2	0.43	0.005	1.0	1.3	2.2	4.2	25.2	53.3	96.6	156.6	1.66	5.1E-34
RMDN1	87.6	0.51	0.001	0.9	1.3	1.9	3.8	9.2	16.2	12.1	30.3	0.15	7.9E-02
CPNE3	87.6	0.39	0.009	0.9	1.3	1.9	3.8	23.7	39.5	51.5	86.7	0.68	2.9E-12
RIPK2	90.8	0.47	0.002	1.1	1.3	1.8	4.3	18.6	8.7	24.2	60.3	0.57	5.7E-10
OTUD6B	92.1	0.33	0.026	1.1	1.3	2.1	4.7	6.4	2.7	3.0	11.6	0.40	8.0E-05

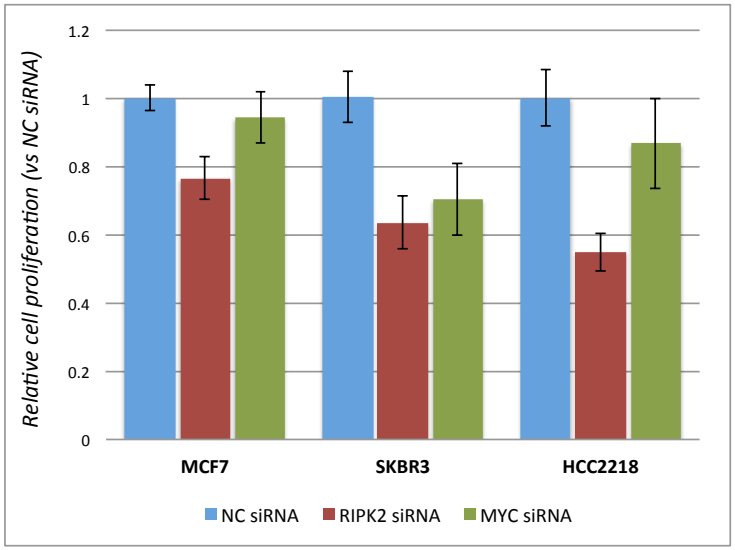
S14C


chr8
0M
10M
20M
30M
40M
50M
60M
70M
80M
90M
100M
110M
120M
130M
140M
TPD52
RIPK2
MYC
Copy number
50
25
0
to chr17 at 35.3Mb(-)
199
UI 209
TD 214
UI 212
UI 204
UI 239
TD 241
TD 256
250
to chr17 at 34.5Mb(+)
Del 341
TD 624
200
300
400
500
600
700
Cluster size

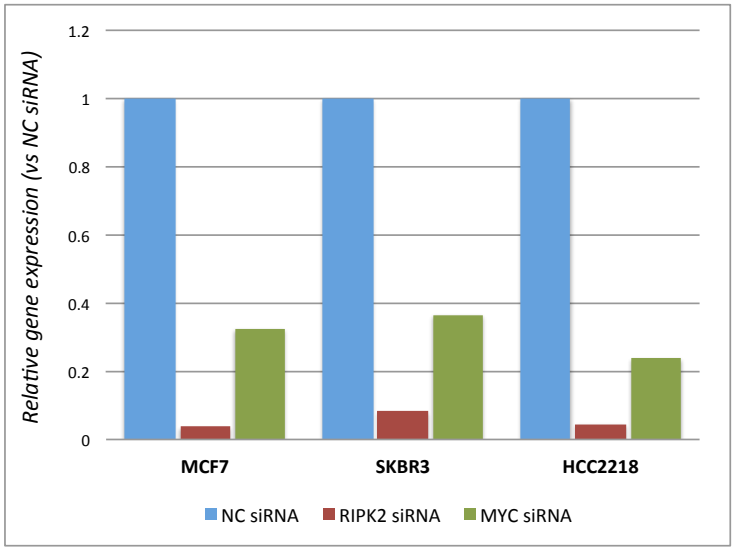
SKBR3

S14

D



E



	RIPK2		*MYC*	
	Copy number	Relative expression	Copy number	Relative expression
SKBR3	9.4	2.59	4.5	5.36
MCF7	2.9	3.89	2.5	2.99
HCC2218	2.6	3.08	3.4	5.37

Supplemental Figure 14. Structure and functional consequence of Chr 8 in breast cancer

(A) Chromosomal map of Chr 8, showing copy number state of the eight breast tumors, the frequency of copy number gain and loss, and the number of genes associated with breast cancer prognosis. Clusters of poor and good prognosis associated genes are located in copy number gain (8q) and loss (8p) regions, respectively. (*B*) Copy number and structural variations relative to BT73, showing a complex structure caused by massive parallel rearrangements. Del, deletion. TD, tandem duplication. UI, unpaired inversion. Inter-chromosomal translocations are indicated by arrows. structural variations are depicted as in Supplemental Fig. 5. The lower panel reports chromosomal location, association of gene expression with disease metastasis free survival (DMFS) of breast cancer patients, copy number ratio (vs. normal control) and gene expression (FPKM by RNA-seq) in the four tumors for which RNAseq-based expression data is available, and fold changes in gene expression of tumor vs normal breast (TCGA data set) for five candidate genes on 8q. HR, hazard ratio. FC, fold change. (*C*) Copy number and structural mutations on Chr 8 for the SKBR3 breast cancer cell line. structural variations are depicted as in Supplemental Fig. 5. (D) Three different cell lines were transfected with 2-pmol of siRNAs targeting either the *RIPK2* or the *MYC* genes. Five days after transfection, the siRNA effect on cell proliferation was measured using the WST-1 assay. Averages +/- standard deviations of two sets of triplicates from two independent experiments are shown in the bar plot. NC, negative control. Copy number and gene expression level (relative to the ACTB gene, assessed by qPCR) for the *RIPK2* and *MYC* genes in the three cell lines examined are reported in the table. (E) Relative level of target gene expression in cells treated with siRNAs against either the *RIPK2* or the *MYC* genes (as shown in (D)), as measured by real time PCR. The *ACTB* gene was used as housekeeping control.

Supplemental Results

Structure of Chr 12p in BT122

In the triple-negative breast cancer BT122, the *CDK4* locus on Chr 12 shows a narrow and high peak of amplification (copy number ~34, Supplemental Figs. S6A and B). Intriguingly, this *CDK4* amplification and overexpression co-exists with a lesser degree of amplification and overexpression of the *KRAS* and *CCND2* oncogenes on Chr 12p (copy number ~6) (Supplemental Fig. S6B). Since *CCND2* is the activator of *CDK4* (Musgrove et al. 2011), this represents a co-gain of two members of the same oncogenic complex. Analysis of the precise boundaries of the *CDK4* amplicon allowed for a detailed reconstruction of the amplification process. We posit that the first structural variation is a ~21.7 Mb tandem duplication encompassing the *CDK4* locus (Supplemental Fig. S6C), which was followed by an unpaired inversion (UI764) that generated another copy of *CDK4*. This locus was further amplified by a smaller rearrangement (UI506 corresponding to a tandem duplication in the cancer genome). Such progressive internal "folding" generated high copy numbers for the *CDK4* locus (34 copies), but lower at the edges of the amplicon (6 copies). Subsequently, a copy number increase of the KRAS and CCND2 genes occurs via Chr 12p aneuploidy. A plausible reconstruction can be hypothesized in which Chr 12p and the associated remnant 12q join the *CDK4* amplicon and generate a rearranged Chr 12, which is then subjected to chromosomal copy number increase to a final number of six copies (Supplemental Fig. S6C).

Structure of Chr 8 and its functional consequences in breast cancer

In the luminal B breast cancer BT73, the most prominent genomic change is the amplification of Chr 8q, which appears to result from the accumulation of smaller amplicons generated by a multitude of overlapping tandem duplications and unpaired inversions(Supplemental Figs. 14A and B). Though MYC is the validated and most well known oncogene on 8q, it localizes at the edge of the region of copy number increase and is clearly less amplified than other centromeric loci. Based on copy number frequencies from the 1,174 breast cancer data set, MYC is located at the telomeric end of a very broad 8q amplicon, which extends from the centromere to near the telomere (Supplemental Fig. 14A). This suggested that perhaps other oncogenes exist on 8q that might be as important as MYC in breast cancer. In BT73, five genes show greater copy number than MYC, high gene expression, and association with poor prognosis in breast cancer patients*: TPD52, RMDN1, CPNE3, RIPK2,* and *OTUD6B* (Supplemental Fig. 14A). DNA-PET-based analysis of the breast cancer cell line SKBR3, which also harbors an amplification of 8q, showed a tandem duplication-driven amplification pattern on 8q, which includes massive co-amplification of *TPD52* (13.2 copies)*, RIPK2* (13.2 copies), and *MYC* (16.6 copies) (Supplemental Fig. 14C). *TPD52* has been reported as an oncogene in breast cancer (Shehata et al. 2008). *RIPK2* is a kinase with a CARD domain and a component of the canonical NF-kB activating pathway by TNF (Inohara et al. 1998; McCarthy et al. 1998; Thome et al. 1998) and is frequently amplified in the breast cancer (38% of breast cancer samples in the Tumorscape data set). The NF-kB pathway is known to be highly activated in a wide variety of cancers and involved in cell proliferation (Chaturvedi et al. 2011; Perkins 2012), as well as MYC signaling in breast cancer (Yeh et al. 2011). We examined the action of *RIPK2* by using siRNA in 3 breast cancer cell lines, which show copy number gain of *RIPK2* (SKBR3, MCF7, and HCC2218), and found that *RIPK2* silencing significantly decreases cell proliferation (by 53%, 42%, and 40%, respectively), confirming its oncogenic role in breast cancer (Supplemental Figs. 14D and E). Based on the Tumorscape copy number data set, *TPD52, RIPK2,* and *MYC* are frequently amplified (34.7%, 38.0%, and 45.9%, respectively) and significantly co-amplified (29.8%, P = 2.62E-30). These results indicate that the co-amplification of functional oncogenes, *TPD52, RIPK2,* and *MYC*, frequently observed among breast cancers, is likely to confer selective advantage in the cancer cells over single oncogene overexpression.

Supplemental Methods

DNA-PET library construction and sequencing

Construction of DNA-PET libraries has been described elsewhere (Hillmer et al. 2011). In brief, genomic DNA was hydrosheared and long mate pair (LMP) CAP adaptors with only a single 5' phosphorylated end were ligated, thus creating a nick on each strand after circularization of the DNA. Both nicks were translated >50bp into the circularized genomic DNA fragment by DNA polymerase I, and mate-pair reads of >50bp were released by T7 exonuclease and S1 nuclease. SOLiD sequencing adaptors P1 and P2 were ligated to the library DNA and constructs were then amplified by PCR and sequenced by 2x50bp using the Applied Biosystems SOLiD 3 Plus System for BT55, BT122, and BT146 and SOLiD 4 System for the BT73 sample. Sequence data was mapped in color space to the human reference genome hg18 using BioScope (Applied Biosystems), allowing up to 6 color code mismatches. Data has been processed as described in the Supplemental Methods.

Illumina Nextera mate pair library preparation and sequencing

Mate pair genomic libraries were generated using the Illumina Nextera Mate Pair library preparation method (Illumina, San Diego, CA) with a gel-based size selection step to optimize the genomic fragment size around 7-9 kb. Genomic DNA was treated with a Tagment enzyme which randomly breaks the DNA and adds a biotinylated junction adapter to each end. After gel size selection and circularization, the DNA was fragmented to 400-2,000 bp fragment size using Covaris Adaptive Focused Acoustic technology (Covaris E220, Covaris, Woburn, MA) and the junction fragments from the ends of the 7-9 kb genomic fragments were captured using streptavidin beads (Life Technologies, Grand Island, NY). Standard Illumina TruSeq library preparation was performed on the fragments resulting in a library of genomic DNA junction fragments with Illumina Paired End oligonucleotide adapters ligated at each end. The libraries were amplified by PCR to obtain optimal products for high-quality paired end sequencing on a HiSeq 2500 instrument. Each mate-pair library was sequenced by multiplexing 8 indexed samples on 8 High Output lanes. The fastq files from the mate-pair library sequencing were paired and run through the NGSQCToolkit (v2.3, IlluQC_PRLL.pl) with a quality control cutoff of 30 (-s 30). The filtered data was then aligned to the human reference genome (NCBI Build 37 from the 1000 genomes project) using bwa (v0.7.4) using default parameters (bwa mem) and the result was stored as bam files.

Determination of copy number by DNA-PET

Copy number profiles for the eight breast cancer samples were generated using concordant mat-pair read data as described before (Hillmer et al. 2011). In the current analysis, somatic copy number ratios were calculated as the ratio of GC bias- and sequencing depth-normalized concordant mate-pair reads from the cancer libraries *versus* those from their matched normal libraries in consecutive 10 Kbps windows. Windows with the lowest 1% count of normal concordant mate-pair reads were filtered out as these tend to give spurious copy number predictions. Copy number segments were obtained by smoothing copy number ratios across 10Kb windows using the R package DNAcopy. The mode of the density profile of the segmented copy number ratios in 10kb windows was used to normalize the copy number ratio data based on the assumption that the mode represents the normal copy number state. Thresholds for copy number gain and loss were then selected using minimum points on the density profile to the right and left of the mode respectively. Focal copy number variations were defined as copy number segments which affect less than half of the chromosomal arm where they occur.

Discovery of recurrent cancer structural variations

We compared structural variations across the eight breast tumors analyzed in this study to identify recurrent events. We searched for discordant mate pair read clusters, which we defined as events with more than 50% sequence overlap between the two breakpoint boundaries (with the exception of inter-chromosomal translocations, for which no sequence overlap was requested) and whose breakpoints map within 3 Mbps of each other but more than a minimum distance apart, to avoid the identification of possible recurrent germline events which had been mistakenly classified as somatic. This minimum distance was set as the 5th percentile of the span distribution of each type of somatic structural variation across the eight sequenced cancer samples and equals to 8.2 Kbps for deletions, 42.5.8 Kbps for tandem duplications, and 5.4 Kbps for inverted orientations.

SOLiD RNA-seq

Approximate 500 ng of poly(A) RNA was enriched using µMACS™ mRNA Isolation Kit. Random fragmentation of the enriched mRNA was carried out with RNase III (SOLiD™ Total RNA-Seq Kit), followed by purification of the fragmented mRNA using RiboMinus™ Concentration Module (Invitrogen). Evaluation of the mRNA yield and size distribution was done using Agilent® 2100 Bioanalyzer RNA 6000 Pico Chip Kit. After ligation of SOLiD adapter and reverse transcription (SOLiD™ Total RNA-seq Kit), purification of the cDNA was done using MinElute® PCR Purification Kit (Qiagen). 150-250bp of cDNA was excised from Novex® pre-cast gel (Invitrogen), and the adapter-ligated cDNA template was amplified by 15 PCR cycles (SOLiD™ Total RNA-seq Kit), followed by purification using PureLink™ PCR (Micro Kit, Invitrogen). The final yield and size distribution of the adapter-ligated and PCR-amplified DNA templates were assessed using Agilent 2100 Bioanalyzer with the DNA1000 Kit (Agilent). The SOLiD RNA-sequencing was carried out using SOLiD 4 system and reagents, and obtained reads are 50 bp length.

RNA-seq reads were aligned to the reference human genome hg18 assembly using BioScope v1.3.1. FPKM values for each transcript were derived using Cufflinks (v2.0.0).

Microarray-based gene expression

Total RNA from the tumors was hybridized on human whole-transcript microarrays (Human Gene ST 1.0, Affymetrix), according to the manufacturer's instructions at the microarray core lab unit at the National Institute of Genomic Medicine in Mexico City. Gene expression estimates were obtained from raw CEL files using the Affy package implemented in R.

Breast cancer copy number data set

Copy number data from a total of 1,174 breast cancer samples were collected by joining two independent sources: The Tumorscape breast cancer data set (240 cancer samples, Beroukhim et al. 2010) and the TCGA breast cancer data set (934 cancer samples, The Cancer Genome Atlas Network, 2012). Copy number calls were downloaded from the Tumorscape portal (http://www.broadinstitute.org/tumorscape/pages/portalHome.jsf) and the cBioPortal for Cancer Genomics Website (Cerami et al. 2012), respectively. Gene-centric frequencies of copy number changes, corresponding to the percentage of samples affected by copy number gain (log2 copy number ratio Tumor vs. Normal > 0.3) and copy number loss (<(-0.3)), were computed for a total of 18,026 genes.

TCGA gene expression data set

Gene-based levels of gene expression using AgilentG4502A_07 platform and corresponding to a total of 395 breast cancer samples and 61 normal breast samples were downloaded from The Cancer Genome Atlas Data Portal (https://tcga-data.nci.nih.gov/tcga/).

Breast cancer patients prognosis data set

Gene expression data and patients' clinical data were obtained from 3 breast cancer cohorts comprising a total of 624 breast cancer samples (Miller et al. 2005; Pawitan et al. 2005; Loi et al. 2007). Gene expression was discretized into {0,1}: 1 for above mean and 0 for below mean. Cox Proportional Hazards regression was performed between the distant metastasis free survival (DMFS) and the discretized gene expression using the R function CoxPH and hazards ratio (HR) was obtained from the model. Wald test was conducted to obtain P-value of the survival association, probability of hazards ratio being hazards being equal i.e. HR = 1.

Cancer gene lists

Lists of genes which are differentially activated in human cancer were obtained from The Cancer Gene Census (Futreal et al. 2004), downloaded from the Sanger Cancer Genome Project website. As described in Supek et al.(2014), tumor suppressor genes were defined as genes marked as "Rec" in the "Cancer Molecular Genetics" column, whereas oncogenes as the genes marked as "Dom". Oncogenes were further subdivided into their different mechanisms of activation based on the "Mutation Type" column.

List of genes frequently spanned by tandem duplications.

We combined a data set of tandem duplication coordinates as detected in a collection of 96 breast cancer genomes (Stephens et al., 2009; Hillmer et al., 2011; Natrajan et al., 2012; Nik-Zainal et al., 2012; Yang et al., 2013). A list of genes mapping to regions spanned by a tandem duplication was generated by intersecting gene bodies' coordinates with tandem duplication coordinates, and requiring the overlap to span 100% of each gene feature. A list of genes that are frequently spanned by tandem duplications was generated by setting a gene count threshold equal to the 99^{th} percentile of a cumulative distribution of gene recurrence generated by combining 100 independent random gene samplings of size equal to the total number of genes which are entirely spanned by at least one tandem duplication. Based on this evaluation, 2,587 unique genes are recurrently spanned by tandem duplications (> 17 times across 96 genomes).

Significance of siRNA effects

All siRNA effects were evaluated using $\log_2$(A/B), where A = (Cell proliferation with Lapatinib + siRNA for target gene) / (Cell proliferation with Lapatinib + control siRNA), and B = (Cell proliferation with siRNA for target gene) / (Cell proliferation with control siRNA). We assumed the observed values on Chr 2p23 represented the expected background distribution of siRNA effects in the tested cell lines. Therefore, probabilities, *p*, were used to assigned to each observed siRNA effect value on Chrs 17 and 2p23 using the cumulative distribution function of a normal probability distribution fit to the mean and standard deviation values from Chr 2p23 (using the *normcdf* function in MATLAB). The $\log_2$ values from all cell line × gene pairs on Chr 2p23 were consistent with a normal distribution (Kolmogorov-Smirnov test, p=0.32, using the *kstest* function in MATLAB). Synergistic effects were quantified using *p*, whereas antagonistic effects were quantified using *1–p*. Probabilities were calculated for each Chr (17 and 2p23) separately, providing for four total sets of p-values (*p* and *1–p* on Chr 17; *p* and *1–p* on Chr 2). False discovery rates were estimated for each set of p-values separately (using the Benjamini and Hochberg method with the *mafdr* function in MATLAB).

The significance of enrichment for Lapatinib-dependent interactions on Chr 17 compared to Chr 2p23 was quantified using the hypergeometric distribution (using the *hygepdf* function in MATLAB). At a 5% false discovery rate, there were 14 Lapatinib-dependent cell line × gene pairs on Chr 17 (13 synergistic and 1 antagonistic), and there was one synergistic Lapatinib-dependent pair on Chr 2p23. There were 54 cell line × gene pairs on Chr 2p23 (6 cell lines × 9 genes) and 132 pairs on Chr 17 (6 cell lines × 22 genes) in total, implying that 1/54 pairs on Chr 2p23 and 14/132 pairs on Chr 17 were Lapatinib-dependent. If we assume there are 20,000 genes genome-wide, and that the observed values on Chr 2p23 represent the expected background distribution of siRNA effects in the tested cell lines, then the hypergeometric distribution has population size 6 cell lines × 20,000 genes = 120,000 pairs with 1/54 of all pairs expected to be Lapatinib-dependent. The p-value was determined by calculating the likelihood of observing *n* or more pairs on Chr 17 (*n* = 14 with ERBB2, *n* = 13 without ERBB2). An analogous calculation can be made for unique genes, whereby the hypergeometric distribution has a population size of only 20,000 genes with 1/9 of all genes expected to be Lapatinib-dependent. The p-value was determined by calculating the likelihood of observing *n* or more genes on Chr 17 (*n* = 8 with ERBB2, *n* = 7 without ERBB2). Similar calculations can be done using the binomial distribution, which assumes sampling with replacement, which is not accurate in our case, but does not require an estimate of the total population size (a population size of 20,000 genes was assumed for the hypergeometric distribution). At a 5% false discovery rate, Chr 17 was significantly enriched for Lapatinib-dependent pairs whether ERBB2 was included in the analysis (P = 1.97E-07, using the *binopdf* function in MATLAB) or excluded (P = 7.32E-07). These p-values are very similar to those calculated from the hypergeometric distribution.

PCR validations for structural mutations and fusion genes

To determine the genomic break points of structural mutations, PCR was carried out using the proof reading DNA Polymerase Phusion Flash (Thermo Scientific) and 1 ng of tumor genomic DNA as template in a 10 µl reaction. A second nested PCR was performed, when necessary, using 1.6 µl of the amplification product of the first PCR as a template for a 40 µl PCR reaction. PCR fragments were cloned using StrataClone Blunt PCR cloning kit (Stratagene) and sequenced using standard Sanger sequencing technologies. To test for the expression of fusion genes, 500 ng of total RNA were reverse transcribed using a Maxima First strand cDNA synthesis kit (Fermentas). A PCR was carried out with HotStarTaq (Qiagen) and 5 ng of cDNA as a template. PCR fragments were cloned and sequenced as above.

Real-time PCR

Total RNA was extracted using RNAeasy Mini Kit (Qiagen) according to the manufacturer's protocol. Up to 500 ng of RNA was reverse transcribed using Maxima reverse transcriptase mix (Fermentas). Real-time PCR was performed using 1 ng of cDNA, 0.5 µM forward and reverse primers and SYBR-Green PCR Mix (Fermentas) using the ABI7500 system (ABI). The *ACTB* gene was used as housekeeping control. Primer sequences are listed below.

	F-primer	R-primer
ACTB (1)	GTCACGCACGATTTCCCGCT	AGACCTTCAACACCCCAGCC
ACTB (2)	TCCCTGGAGAAGAGCTACGA	AGGAAGGAAGGCTGGAAGAG
ANKRD40	TTGCTGTGAGCTGGGTGTTA	TCCTGGAAATCTTGGAGTCG
ATRAID	AAAACGACACGAGAGCTAATGC	CAGGGTCCTCCAGAGAACAG
BRCA1	GAGTGTCCCATCTGTCTGGAGTTGA	AGGACACTGTGAAGGCCCTTTCT
CBX1	GGAGCGGATTATTGGAGCTA	ACCAGGTCAGCCTCATCAGA
EME1	AACCGAGTCAGCCTGGAAAT	ACCTGTATGTCTGCGAGCAA
EPN3	CAACCTTGACTCGTTGGTCA	GGTGCGCATCTGGTTTAGC
EPT1	TTTTACTCTGGCCCACATCC	CTGAGATCCCACGTCTGTCA
ERBB2	AGCAGGGCTTCTTCTGTCC	CTCCTCTTCAGAGGGCTCCA
HADHB	GGGCCATGCTATGATAGTGG	GGCAAGGCTTAAGTGCAAAC
HOXB7	CAGAGGGACTCGGACTTGG	TCTGGTAGCGGGTGTAGGTC
KAT7	GACGATCTGCTCGAGTCACC	GCTCCTCAGTGCCAAAAGAC
LRRC59	GGGAGCAGGAGAAGAAACCT	ACAGCCCAGGAACGAGTGT
LUC7L3	AGAAGATGCAGCAGGTCTCG	GATCGATCATGGCTTCTGCT
MRPL27	TCATGCTGGGAACATCATTG	CGGACTATCCCCTCTTCCAG
MRPL33	TGCGGGAAAAACTGACTCTT	TTCCATGCCATTTCCTCTTC
MYC	TGCTCCATGAGGAGACACC	ACTCTGACCTTTTGCCAGGA
NME1	AACCCTGCAGACTCCAAGC	CAAGCCGATCTCCTTCTCTG
NME2	GTGAAGACAGGCCGAGTGAT	TCACTGCCATGAATGATGTTC
OST4	TATCACTACGTGGCCGTCAA	TACTGGGCCCTATTCAGTGG
PDK2	CTTCTCCATGGAAGGCTTTG	CCTGGATGGTCTGGTAGTGG
PHB	GAGTCCATTGGCAAGTTTGG	CACAATGTCCTGCACTCCAC
PNPO	GTGGTCAGCCACCAGAGTTC	CCATCACCTGAGGGTACAGG
PPP1R9B	ATGGACAAGCTCCTGGACAA	ACAGGGGAGGGAGGACAAT
PRR15L	CAGGCAGCTCCTACACTCAA	GTCAGCTTCCACCAACCAAT
RAB10	ATCCTTCGAAAGACCCCTGT	TAGTGGATGGCAACTGATGG
RIPK2	GGGATAGCACCATTTCTGGA	GTTAAGGCAGGCTTCTGTCAT
SLC35B1	TTGGCCTCTGTGATCCTCTT	TGGGATGTCTTCTTAGCTCCT
SLC4A1AP	GCCAGCAGAGATTCCAGAAC	CTCCTTTCATGGCACCAAAT
SNF8	CGGCACTTACCTCATTCAGT	CCCATTTAAGACTGGCTTTG
SPATA20	TGATCTGTGGAGACCGTCAG	CGGAGGGTACTCAGGAAAGG
TOB1	TGAAGGGCACTGGTATCCTG	ATTGCCACGAACATCATCAA
TTC27	GACTCAATTTACGGGGCTTG	TGATTTCGAAACTGGTTGCTT
UBE2Z	AGGGGTGATGGAGAAGTCCT	CGCATCAAGAGGGACTGGTA
YPEL5	ACCCGTCTGTTTTCTTGTGC	GCATGACCCGATCTTGAACT

Supplemental References

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**: 401-4.

Chaturvedi MM, Sung B, Yadav VR, Kannappan R, Aggarwal BB. 2011. NF-κB addiction and its role in cancer: 'one size does not fit all'. *Oncogene* **30:** 1615-1630.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. 2004. A census of human cancer genes. *Nat Rev Cancer*. **4**:177-83.

Inohara N, del Peso L, Koseki T, Chen S, Núñez G. 1998. RICK, a novel protein kinase containing a caspase recruitment domain, interacts with CLARP and regulates CD95-mediated apoptosis. *J Biol Chem* **273:** 12296-12300.

Konecny GE, Pegram MD, Venkatesan N, Finn R, Yang G, Rahmeh M, Untch M, Rusnak DW, Spehar G, Mullin RJ, et al. 2006. Activity of the dual kinase inhibitor Lapatinib (GW572016) against HER-2-overexpressing and trastuzumab-treated breast cancer cells. *Cancer Res* **66:** 1630-1639.

McCarthy JV, Ni J, Dixit VM. 1998. RIP2 is a novel NF-kappaB-activating and cell death-inducing kinase. *J Biol Chem* **273:** 16968-16975.

Musgrove EA, Caldon CE, Barraclough J, Stone A, Sutherland RL. 2011. Cyclin D as a therapeutic target in cancer. *Nat Rev Cancer* **11:** 558-572.

Natrajan R, Mackay A, Lambros MB, Weigelt B, Wilkerson PM, Manie E, Grigoriadis A, A'hern R, van der Groep P, Kozarewa I, et al. 2012. A whole-genome massively parallel sequencing analysis of BRCA1 mutant oestrogen receptor-negative and -positive breast cancers. *J Pathol* **227**:29-41.

Perkins ND. 2012. The diverse and complex roles of NF-κB subunits in cancer. *Nat Rev Cancer* **12:** 121-132.

Shehata M, Bièche I, Boutros R, Weidenhofer J, Fanayan S, Spalding L, Zeps N, Byth K, Bright RK, Lidereau R, Byrne JA. 2008. Nonredundant functions for tumor protein D52-like proteins support specific targeting of TPD52. *Clin Cancer Res* **14:** 5050-5060.

Supek F, Miñana B, Valcárcel J, Gabaldón T, Lehner B. 2014. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**:1324-35.

Thome M, Hofmann K, Burns K, Martinon F, Bodmer JL, Mattmann C, Tschopp J. 1998. Identification of CARDIAK, a RIP-like kinase that associates with caspase-1. *Curr Biol* **8:** 885-888.

Yang L, Luquette LJ, Gehlenborg N, Xi R, Haseley PS, Hsieh CH, Zhang C, Ren X, Protopopov A, Chin L, Kucherlapati R, Lee C, Park PJ. 2013. Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell.* **153**:919-29.

Yeh PY, Lu YS, Ou DL, Cheng AL. 2011. IκB kinases increase Myc protein stability and enhance progression of breast cancer cells. *Mol. Cancer* **10:** 53.