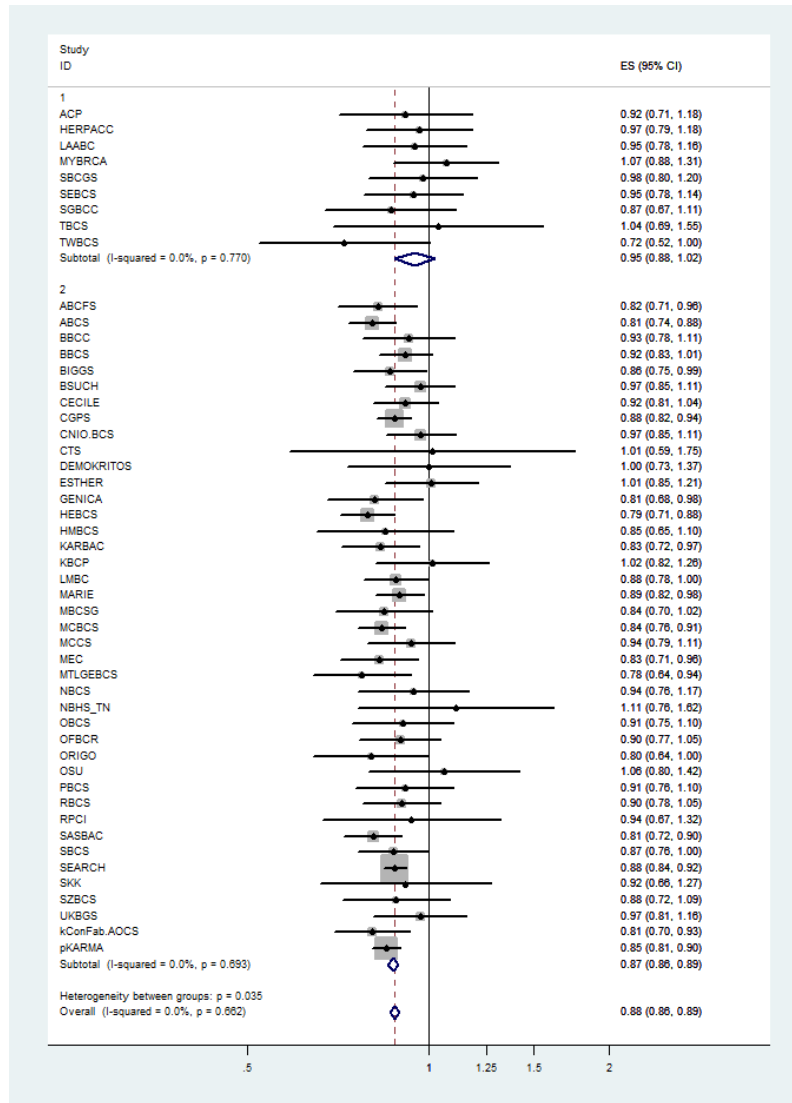
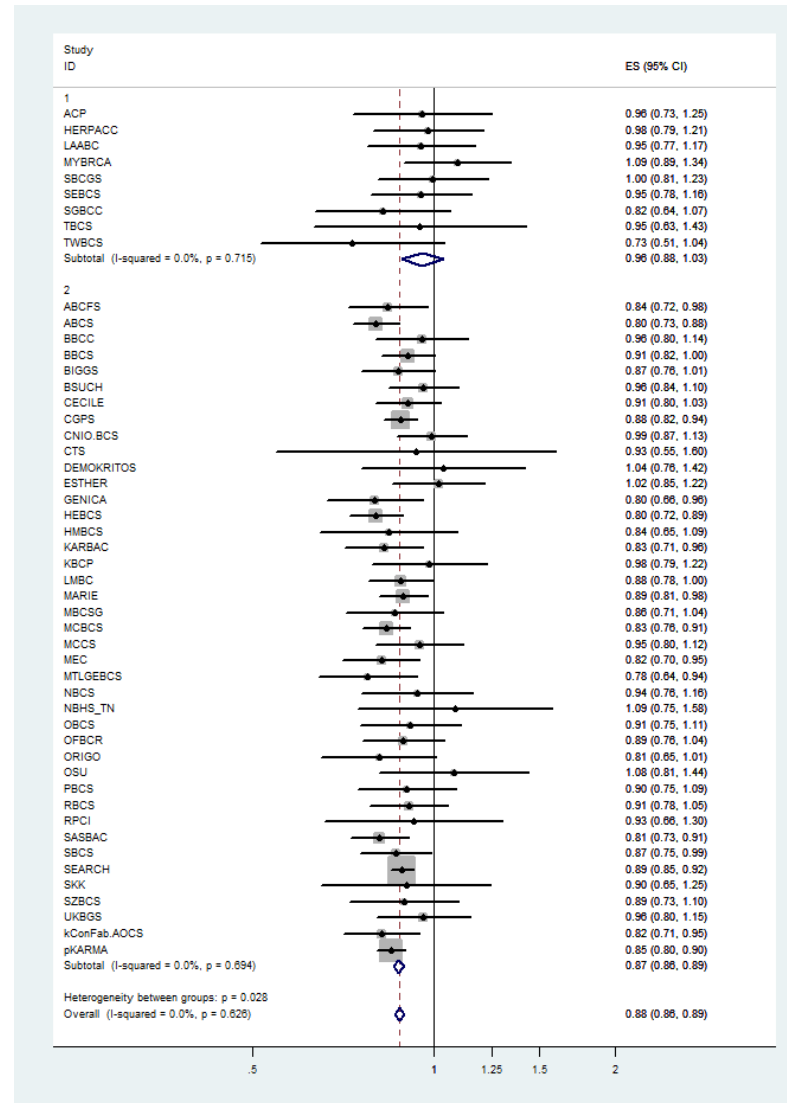


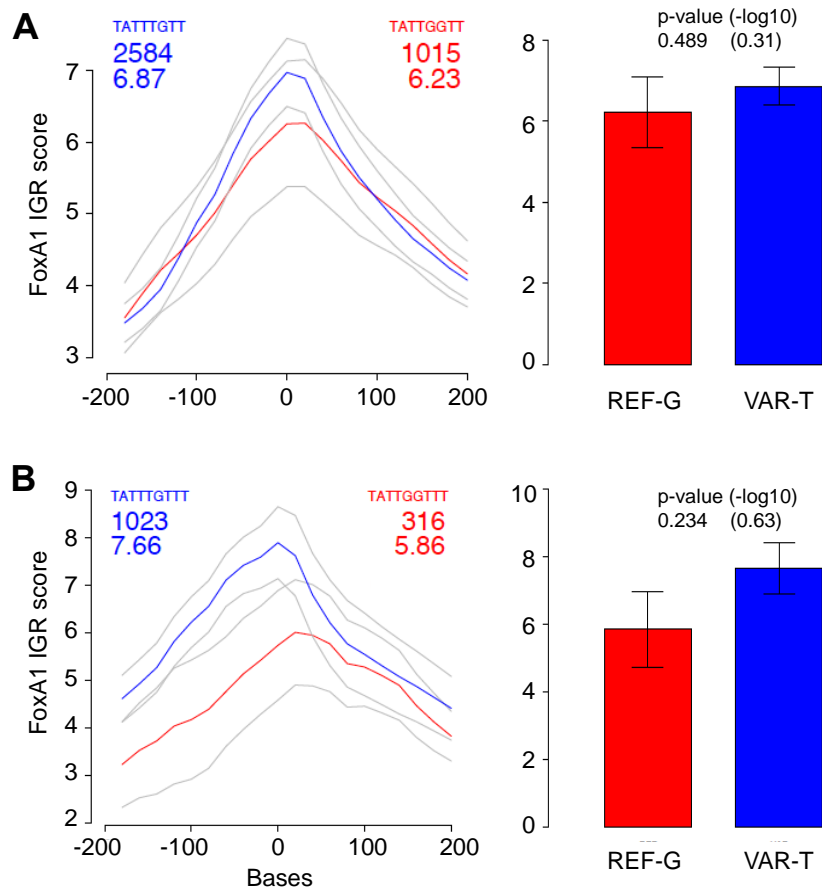
**A**



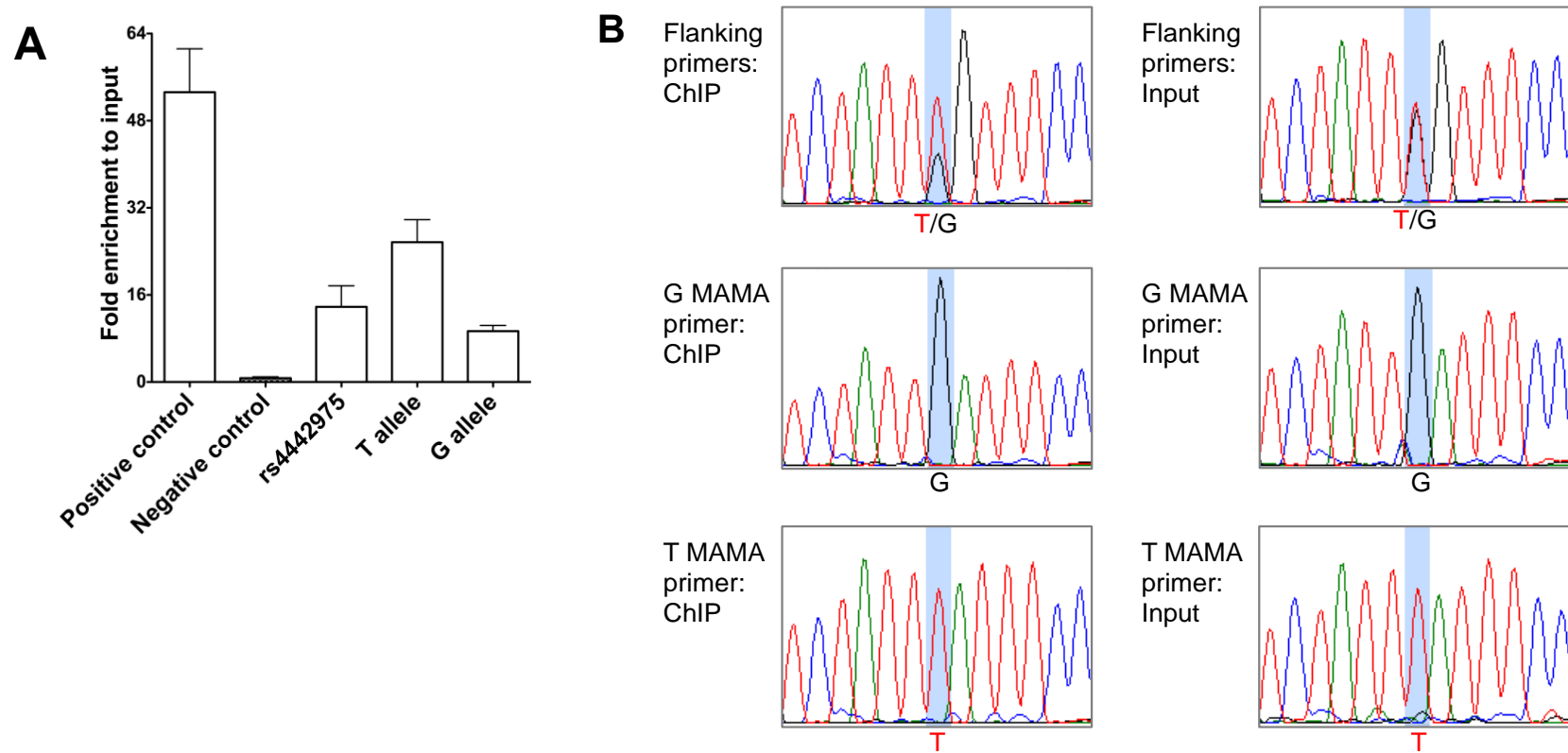
**B**



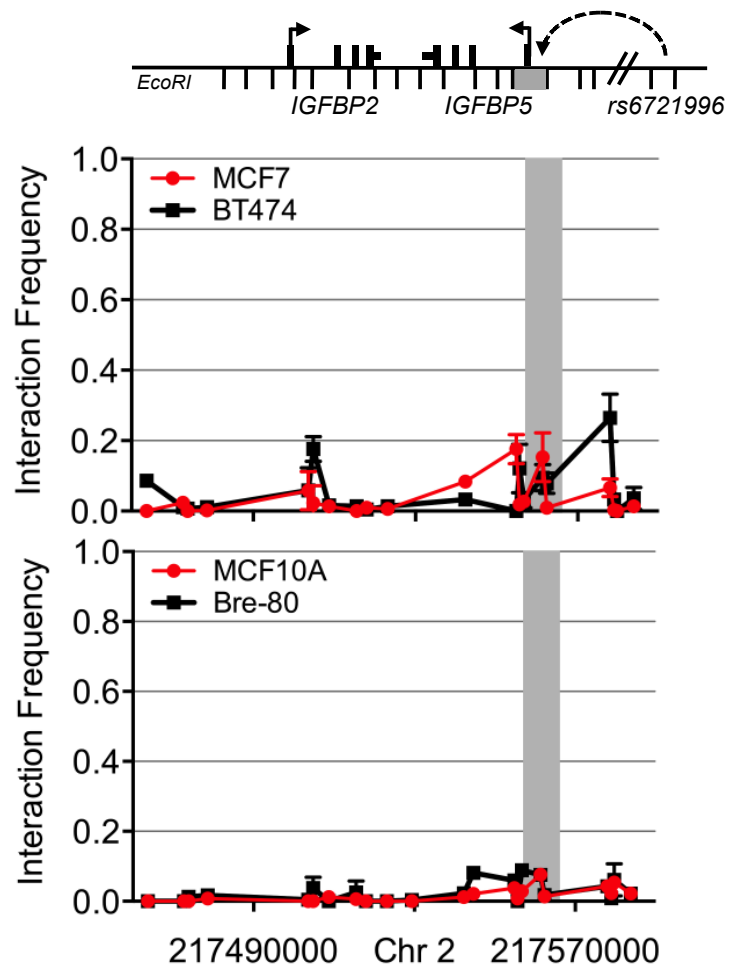
**Supplementary Figure 1.** Forest plot for (A) rs4442975 and (B) rs6721996 in ER+ breast cancer cases versus controls.



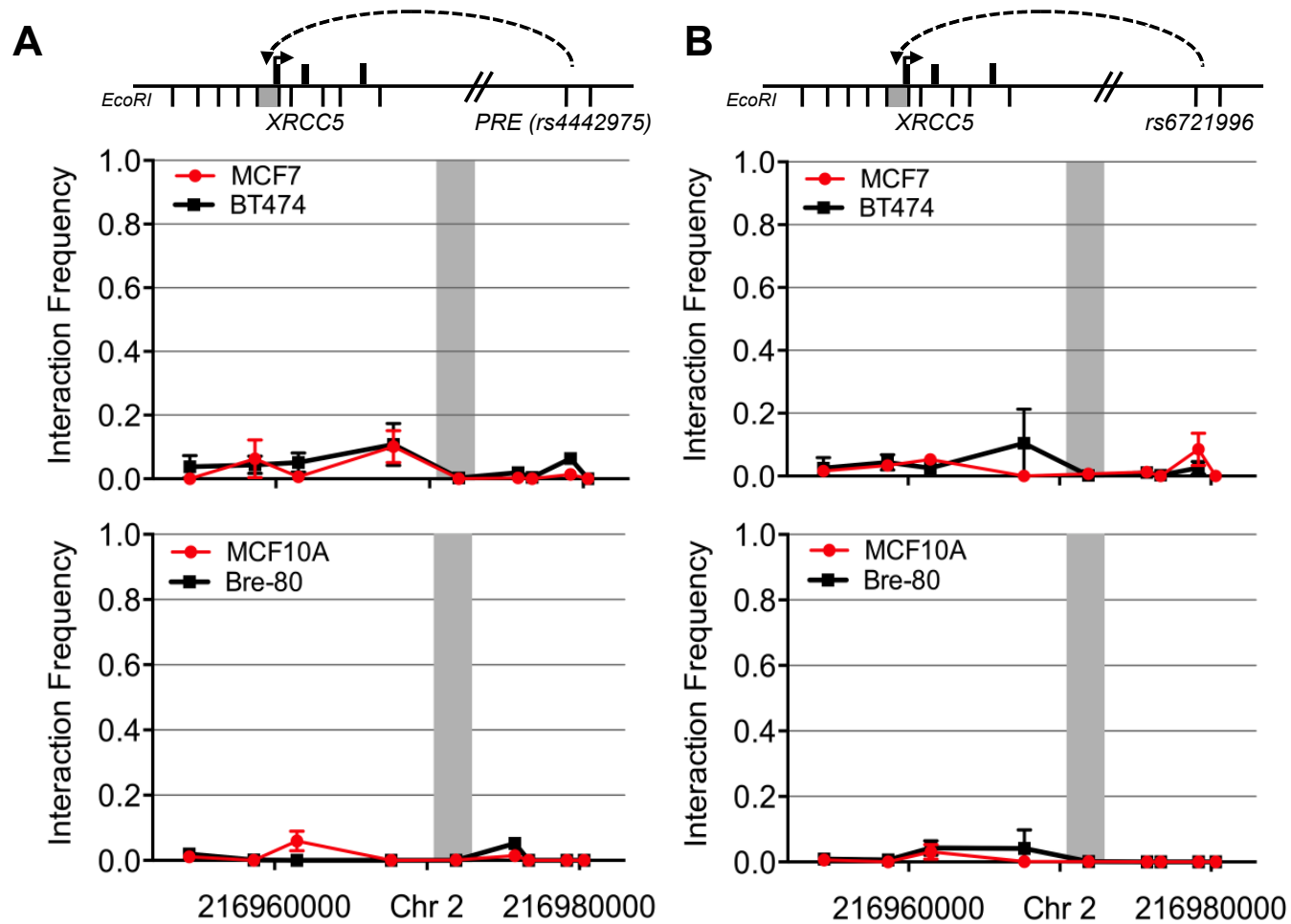
**Supplementary Figure 2. IGR profile plots for SNP rs4442975 showing the interaction effects for FOXA1 using eight (A) and nine (B) nucleotide models.** The top row of colored numbers shows the number of instances for each k-mer found genome wide within H3K4me2 elements in MCF7 cells. The bottom row shows the averaged binding intensities at the k-mers (50 bp window). Control profiles, shown in gray, are generated by scrambling the probed sequence. Bar graphs compare FOXA1 IGR scores between rs4442975 common alleles (REF-G) and minor alleles (VAR-T). The inset shows log10 of the p-values obtained by IGR. Error bars denote SEM (N=3)



**Supplementary Figure 3. Allele-specific FOXA1 ChIP-qPCR at SNP rs4442975 in BT474 cells.** (A) ChIP-qPCR results against FOXA1 in BT474 breast cancer cells. Error bars denote SEM (N=3). (B) Sanger sequencing of PCR fragment generated using MAMA primers against the G or T allele of SNP rs4442975 or primers flanking this SNP following FOXA1 ChIP and their input DNA controls.

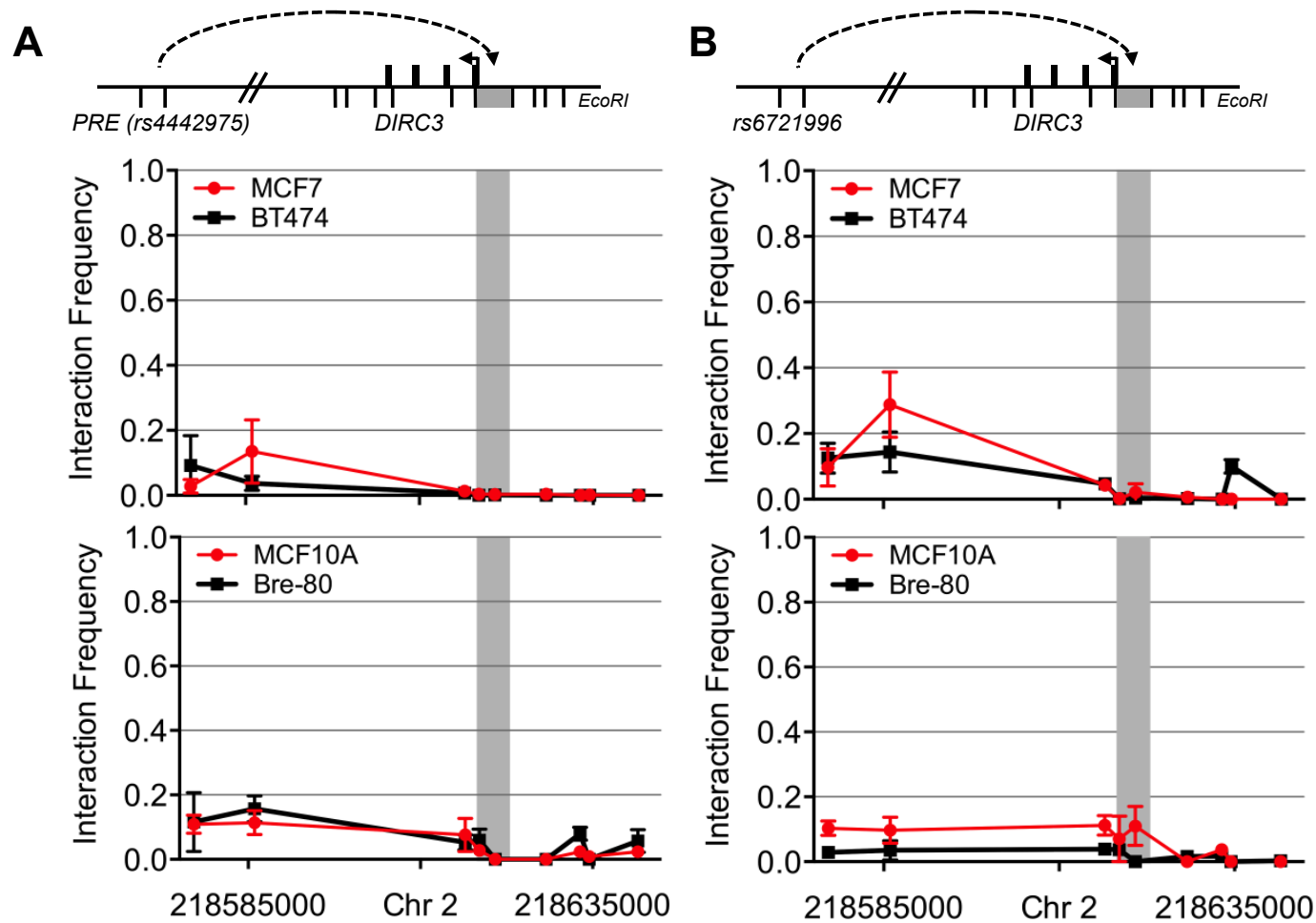


**Supplementary Figure 4. Long-range chromatin interactions between the 2q35 risk region and *IGFBP2* and *IGFBP5* in breast cell lines.** 3C interaction profiles between regions containing *rs6721996* and *IGFBP5* and *IGFBP2*. 3C libraries were generated with *EcoRI*, with the anchor point set at *rs6721996*. A physical map of the region interrogated by 3C is shown above, with grey bars representing the position of the *IGFBP5* promoter (not to scale). Error bars represent SD (N=2).

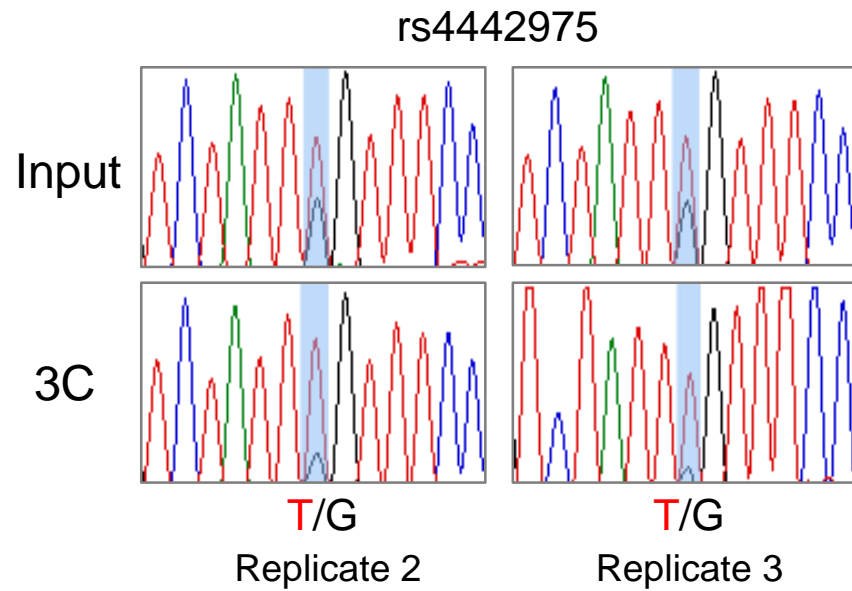


**Supplementary Figure 5. Long-range chromatin interactions between the 2q35 risk region and *XRCC5* in breast cell lines.** 3C interaction profiles between regions containing either (A) PRE (rs4442975) or (B) rs6721996 and *XRCC5*. 3C libraries were generated with *EcoRI*, with the anchor points set at the PRE (rs4442975) or rs6721996. A physical map of the region interrogated by 3C is shown above, with grey bars representing the position of the *XRCC5* promoter (not to scale). Error bars represent SD (N=2).



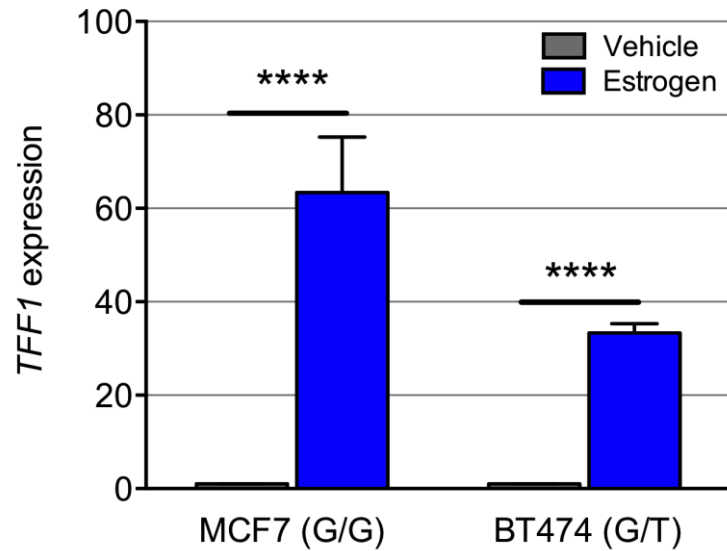


**Supplementary Figure 7. Long-range chromatin interactions between the 2q35 risk region and *DIRC3* in breast cell lines.** 3C interaction profiles between regions containing either (A) PRE (rs4442975) or (B) rs6721996 and the *DIRC3* promoter. 3C libraries were generated with *EcoRI*, with the anchor points set at the PRE (rs4442975) or rs6721996. A physical map of the region interrogated by 3C is shown above, with grey bars representing the position of the *DIRC3* promoter (not to scale). Error bars represent SD (N=2).

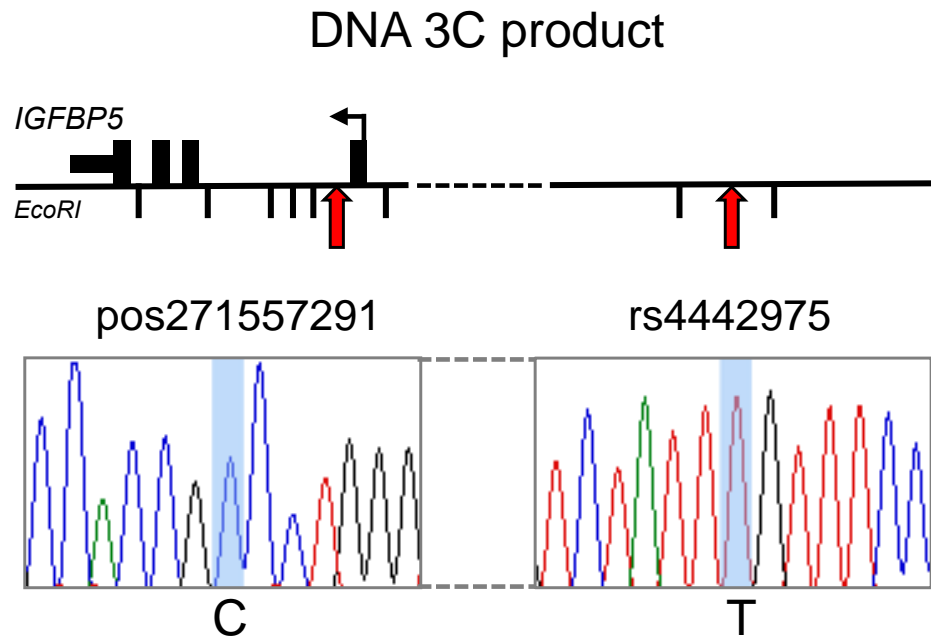


**Supplementary Figure 8. SNP rs4442975 affects chromatin looping between the PRE and *IGFBP5* promoter.** 3C followed by sequencing for the rs4442975-containing region in BT474 breast cancer cells shows allele-specific chromatin looping. Two of three independent 3C libraries are shown which were generated with *EcoRI*, primers used for 3C and sequencing can be found in **Supplementary Table 2**.

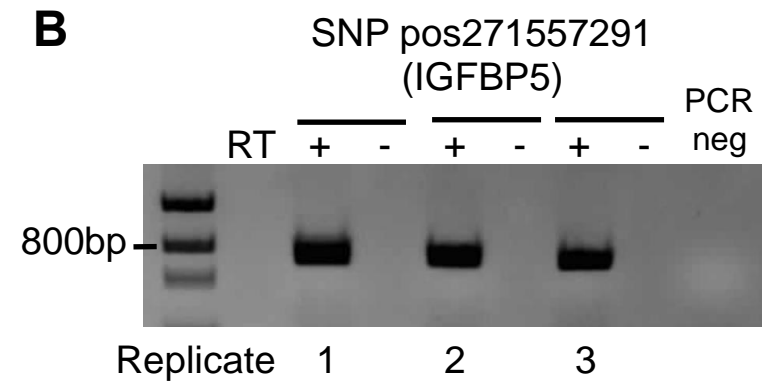
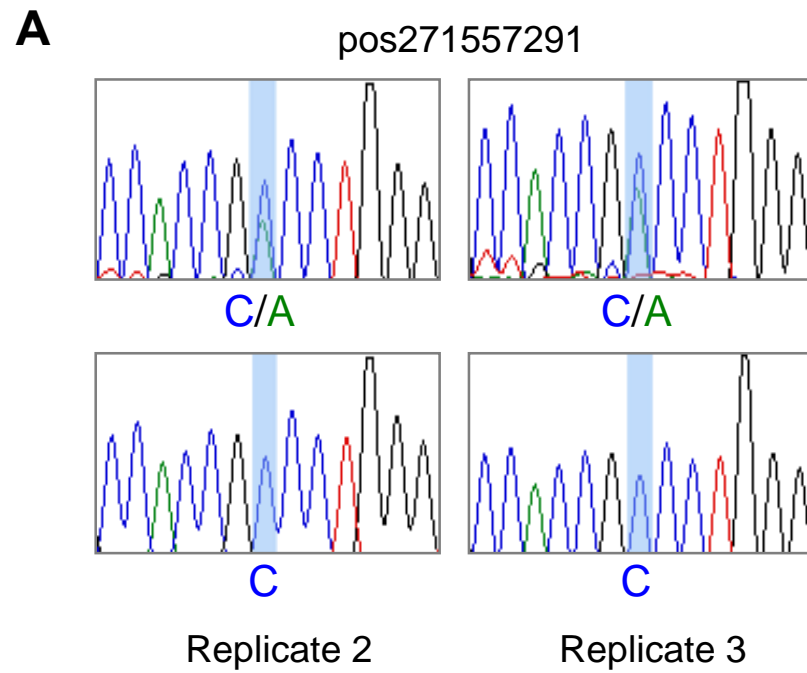




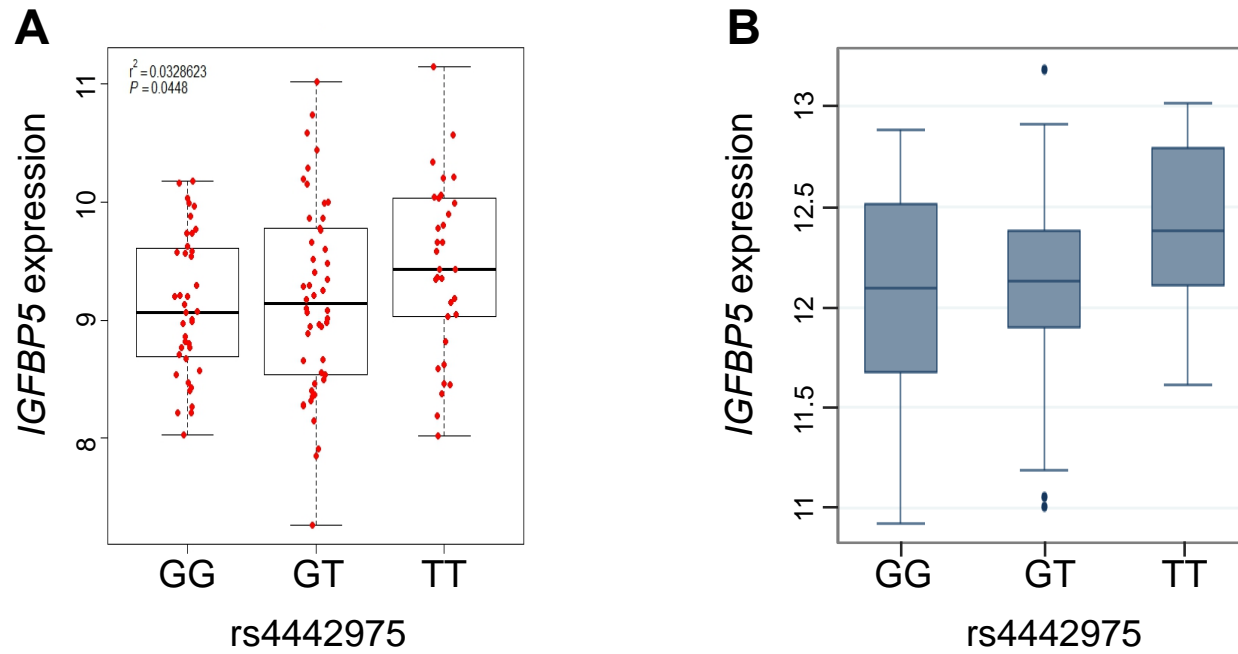
**Supplementary Figure 9. TaqMan qPCR assays confirming estrogen induction in MCF7 and BT474 cells.** MCF7 and BT474 cells were treated with either 10 nM estradiol or vehicle for 24 h. *TFF1* expression was measured by qPCR and normalised using *GUS* expression as an internal control. Error bars denote SEM (N=6). *P* values were determined with a two-tailed t test. \*\*\*\*  $p < 0.0001$ .



**Supplementary Figure 10. 3C sequencing results showing the physical interactions between heterozygous variants at the 2q35 enhancer and *IGFBP5*.** BT474 3C product was PCR amplified and Sanger sequenced, which showed the T allele of rs4442975 is physically linked to the C allele of pos271557291 (within *IGFBP5* intron 1).



**Supplementary Figure 11.** (A) Replicates of Sanger sequencing result indicate DNA and RNA level on each alleles of SNP pos271557291. (B) Gel picture of PCR products of cDNAs generated from BT474 native RNA with or without reverse transcriptase.



**Supplementary Figure 12. eQTL analysis and association of rs4442975 with *IGFBP5* in normal breast tissues. (A) NBCS dataset (N=123) and (B) METABRIC dataset (N=135). The x-axis of each plot corresponds to the three observed SNP genotypes and the y-axis represents log<sub>2</sub>-normalized gene expression values.**

Haplotypes	rs4442975	rs6721996	rs13387042	rs13412666	rs13426489	Haplotype frequency	OR	P-value
1	2	2	1	1	1	0.014	0.84 [0.77-0.91]	3.86 x 10 <sup>-5</sup>
2	2	2	2	2	2	0.469	0.87 [0.86-0.89]	8.99 x 10 <sup>-44</sup>
3	1	1	1	1	1	0.510	-	-
All Others			Rare			0.007	1.02 [0.89-1.16]	0.799

**Supplementary Table 1. Haplotype analysis across the BCAC studies.** Each haplotype was compared to the ancestral haplotype carrying the common alleles of all 5 SNPs (haplotype 3). SNPs rs4442975 and rs6721996 are almost perfectly correlated with each other.

Probe ID	$\beta$ (effect size) of rs4442975 on probe levels	P-value for association of rs4442975	Strongest eSNP at 2q35 with probe	P-value for association of strongest eSNP	R <sup>2</sup> between rs4442975 and strongest eSNP
<b>NBCS study (n=123 normal breast samples)</b>					
A-23_P154115	0.19	0.045	rs2024488	0.0007	0.00009
A_23_P383009	0.14	0.21	rs6716223	0.0002	0.0003
A_23_P376497	0.19	0.52	rs207946	0.004	0.001
<b>METABRIC study (n=135 adjacent normal breast samples)</b>					
Ilmn_1750324	0.13	0.026	rs192116758	0.0004	0.001
Ilmn_2132982	0.10	0.11	rs149640588	0.0004	0.001

**Supplementary Table 2.** Association of rs4442975 and the strongest eSNP at the 2q35 locus with *IGFBP5* in the NBCS and METABRIC studies.

Gene	Probe ID	P-value for association with rs4442975
<i>XRCC5</i>	A_24_P345498	0.622
<i>XRCC5</i>	A_32_P221832	0.620
<i>XRCC5</i>	Ilmn_2105983	0.431
<i>MARCH4</i>	Ilmn_1789991	0.096
<i>SMARCAL1</i>	Ilmn_1752111	0.623
<i>RPL37A</i>	Ilmn_2051519	0.316
<i>IGFBP2</i>	A_23_P119943	0.701
<i>IGFBP2</i>	Ilmn_1715193	0.962
<i>TNPI1</i>	A_23_P165504	0.061
<i>TNPI1</i>	Ilmn_2150797	0.646
<i>DIRC3</i>	No probe available	N/A
<i>IL8RB</i>	Ilmn_1680397	0.719
<i>IL8RB</i>	Ilmn_1691507	0.509

**Supplementary Table 3.** Association of rs4442975 with probes for additional 2q35 locus genes in the NBCS and METABRIC studies. (For comparison, locus SNP rs148028060 is an eQTL for *IL8RB* probes with  $P < 1 \times 10^{-25}$  in the METABRIC study.)

<b>3C <i>Eco</i>RI primers</b>	<b><i>Eco</i>RI fragment (hg19 coordinates)</b>	<b>Sequence (5' to 3')</b>
Bait rs4442975 (PRE)	217,916,637	TGATCCCTGAACTGGGAGAAATGGGTGG
Bait rs6721996	217,909,071	CAGCCTGGGAAGGGTTGGTACTTTCTCTACC
Allele specific 3C FOR1	217,916,637	TTAGCCCTGGAATTTTGGTGATGGGAGG
Allele specific 3C FOR2	217,916,637	GACGTCAAACCTGGCTATTGATTCTCTCCAAAAGG
Spanning <i>IGFBP5</i>	217,583,840	CCAGAGAGAAGAAAGAGGAGTGCAGGACAGC
	217,579,734	GGCTTCCCTAACCCCTCAGGCCCTCTGCC
	217,579,008	CAAAGCTGACTTGCCTCAGAGATGCAAGC
	217,577,955	CATAGGCATCCTCCATAATCCCGTATCCTGC
	217,562,357	ACATGTGCTGTGTTGGTGTGCTAAGCAGC
	217,561,337	TGTGATGTGGCTAGTGTGAAGGGTGTACACG
	217,556,731	TGATGGGCTGGTCCCCAGCATAGATAACC
	217,555,682	CCTGAGGGCTTGAGTTTTAAGTCAGAGCAGC
	217,554,866	CCAGGTTAGGCCAGGTGTCCAAATGC
	217,542,270	GGCTGTGAAGGGACATGAACTGATTTTCAGG
	217,523,181	GACTCAAATCCCGTAGCCAGATCTGAACTCC
	217,517,970	GGAGGGAGGCAGAAGTTAGGAAAGAGGAAGG
	217,515,465	TCTTCCCTTTGACTTCCTGGCTCAGAGAGC
	217,508,733	GGCACAGAAGCCAGTTCATAATAGGGTGAGG
	217,504,807	GTGGTTGACTGGGCTTTGAGGGAAAGG
	217,503,575	ATGGCTGGGTAGTCCACTATGTGCCAGC
	217,478,594	ATGCTTCTTGACCATCAGGAATGGTGG
	217,473,840	CCTGCCTGTGTCTCCCTTTCCCTTCGTATACC
	217,472,670	TGGGAATAGAAGCTACGTCCTCAGAGATTCCC
	217,463,707	CCTTCCCTGGCTACTGATGGAGTGCAGG
Spanning <i>XRCC5</i>	216,954,397	GAAATACTGGGTGAGAGGGCAGCAGAGG
	216,958,636	TACTGCCACATGTACACCATTTGCCATCC
	216,961,479	GAGGCTCTTCAAATTTGCTTTGAATGCTATGG
	216,967,651	CTTCCATGGCACCACAGCCACTTTACC



	216,971,899	GAAACTGGGATGTGCATTTTACTGTTCTCAAAGG
	216,975,759	CCAAGGACTCAGAATAATGGGTAGGAGAGATGG
	216,976,679	ACTTCCGTTGGGCACTAGTACTTCCCACC
	216,979,167	GGGCAAGATTACCTTGCCTTAAGATGGATGC
	216,980,318	GTGCATGTACACATAACCTCTTGCCTGTCAGG
Spanning <i>TNPI</i>	217,706,488	GCAAGAATTAGTGGCAGCAAGGACAATTTAAGC
	217,709,110	AAGTTGACACTGTCATCCTGGTGAGTAGATCAGC
	217,712,044	CCTACCACCCTGGAGCAAGTCTCTCAAGC
	217,712,862	GTTGAGAATGAAGCTACCTTCTGGTTCTGCTGG
	217,717,602	GCGTATGCTCTAACTGCTGCCACAGAAGG
	217,733,066	CCTTCTTCAACTCCATTGCTGGCTTGG
	217,736,813	GTCAAGTGGGAGAAAGCACATGTGTCTGG
	217,746,063	TAAGGTGTAAGCTCAGGGAGGCTCTGACCC
Spanning <i>DIRC3</i>	218,577,073	CTCAGCAGGTCTGCTGTTTTCCAGGAGG
	218,585,922	ATCCTGGGAAATGCTGCAACTCTACCAACC
	218,616,510	CCACGTGCTTCATTCTAAGCATAGAGATCTTGC
	218,618,584	TCTGGAGAGCTGATGAGTGTTATTGTAACCAGAGG
	218,620,848	AGTCTCAAGTCAATGGCCTCATCTTCCACC
	218,628,251	CATCCAAATGGAAACACAGAGGTTAGAGAAATGC
	218,633,195	GACTCGGAGTAGAGGGAAGGGTGTACTGTAATGG
	218,634,491	TCTGATCAAATGCTCCCCAGAATGTATGG
	218,641,600	GAAGGACAAATAGGGAGATGGAACTTTCATCC

Luciferase Cloning Primers	Sequence (5' to 3')
<i>IGFBP5</i> prom FOR	<u>GGTACCAGT</u> GACAGAGAGGGAAGTGTCACTCC
<i>IGFBP5</i> prom REV	CTCGAGTCTTCGAAATTCGCAGGTTCTACG
PRE FOR	<u>ACCGGT</u> TGATTTTCAGTTCCTTGTCTCTCCAGC
PRE REV	<u>CCTGCAGGAGT</u> GTACAGGGACCAATCATCCTCC
rs4442975 FOR	GGTCACTAATATAGTCTATTTGTTTCCCAGG
rs4442975 REV	CCTGGGGAAACA <u>AA</u> ATAGACTATATTAGTGACC

<b>Allele Specific ChIP Primers</b>	<b>Sequence (5' to 3')</b>
<i>FOXAI</i> Positive Control FOR	GAGCTGGCATTCTAGCGAAG
<i>FOXAI</i> Positive Control REV	CCCTCAAGCAGGGTTAGTCA
<i>FOXAI</i> Negative Control FOR	AACCTCGCTATGCTCCCTTC
<i>FOXAI</i> Negative Control REV	TGTGTGCCTTTCATCTGGAG
rs4442975 FOR	CAGCCATGCCTGAACTTTTC
rs4442975 REV	TCAGTCCTATTTGAGAAAAGGAGAA
rs4442975_G_MAMA_REV	AGAACAAGCCTGGGGAAATC
rs4442975_T_MAMA_REV	GAGAACAAGCCTGGGGAAATA
<b>Allele Specific Expression Primers</b>	<b>Sequence (5' to 3')</b>
IGFBP5 allele cDNA REV	TGCGTGTACACCCCTTCACACTAGC
IGFBP5 allele PCR FOR	CAAACCTGGCTATTGATTCTCTCCAAAAGG
IGFBP5 allele PCR REV	TTGGTAAATAATAAGTGGAGCCAGGATGC

**Supplementary Table 4:** Oligonucleotides used in this study.