The role of *ZNF423* and *CTSO* in predicting response to tamoxifen prevention of breast cancer in high-risk women

Running title: *ZNF423* and *CTSO* unlikely to predict tamoxifen response

Authors: Adam R Brentnall[1], Jack Cuzick[1], Helen Byers[2], Ivana Sestak[1], Anthony Howell[3], Trevor J Powles[4], William G Newman[2], Mitchell Dowsett[5]

Affiliations

1. Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, Barts and The London School of Medicine and Dentistry, Charterhouse Square, London, EC1M 6BQ, UK
2. Manchester Centre for Genomic Medicine, University of Manchester and Central Manchester Foundation Trust, Manchester, UK
3. The Christie NHS Foundation Trust, Manchester, UK.
4. Cancer Centre London, London, UK.
5. Institute of Cancer Research, London, UK.

Corresponding author:

Abstract

A case-control study from two randomised breast cancer prevention trials of tamoxifen and raloxifene (P-1 and P-2) identified single nucleotide polymorphisms (SNPs) in or near genes *ZNF423* and *CTSO* as factors which predict which women will benefit most from selective oestrogen receptor modulator (SERM) therapy. In this article we further examine this question by using blood samples from two randomised tamoxifen prevention trials: the International Breast Cancer Intervention Study I (IBIS-I), and the Royal Marsden trial (Tamoplac). A nested case-control study was designed with 2:1 matching in IBIS-I and 1:1 matching in Marsden. The OncoArray was used for genotyping, and included two SNPs previously identified (rs8060157 in *ZNF423* and rs10030044 near *CTSO*), and 102 further SNPs within the same regions. Overall there were 369 cases and 662 controls, with 148 cases and 268 controls from the tamoxifen arms. Odds ratios were estimated by conditional logistic regression, with Wald 95% confidence intervals. In the tamoxifen arms the per-allele odds ratio for rs8060157 was 0.99 (95%CI 0.73–1.34), and 1.00 (95%CI 0.76–1.33) for rs10030044. In the placebo arm, the odds ratio was 1.10 (95%CI 0.87-1.40) for rs8060157 and 1.01 (95%CI 0.79-1.29) for rs10030044. There was no evidence to suggest other SNPs in the surrounding regions of these SNPs might predict response to tamoxifen.

Significance: Results from these two prevention trials do not support the earlier findings. rs8060157 in *ZNF423* and rs10030044 near *CTSO* do not appear to predict response to tamoxifen.

Introduction

Selective oestrogen receptor modulators (SERMs), including tamoxifen and raloxifene, have been shown to reduce the risk of breast cancer. A combined analysis of almost 84 thousand women in nine trials estimated that over ten years approximately two in five breast cancers were prevented in women who had been randomised to receive a SERM[1]. Tamoxifen has been licensed for prevention in women at an elevated risk of breast cancer in the United States of America and approved for this indication by the NICE committee in the United Kingdom (UK), but uptake has been modest[2]. Breast cancer is the most common cancer in women worldwide, with an estimated 1.7 million cases diagnosed and 500,000 deaths in 2012[3]. Improved risk estimates for response to SERMs for an individual woman could have a significant impact on the utility and acceptability of these preventive treatments.

Ingle and colleagues reported that two single nucleotide polymorphisms (SNPs), rs8060157 in *ZNF423* and rs10030044 near *CTSO*, appear to predict response to tamoxifen and raloxifene[4]. Their findings were from a genome-wide association study using DNA from the NSABP P-1 and P-2 (STAR) breast cancer prevention trials[5,6]. Our objective was to assess the value of these SNPs in the IBIS-I and Marsden trials[7,8,9].

Results

Figure 1 shows a CONSORT diagram. A total of 1,276 women from both trials were initially selected for the case-control study. Of these, there was not enough DNA available for assay in 169, so some controls were re-allocated to maintain matching for all cases. For the 80 SNPs in *ZNF423* and 24 near *CTSO,* 35 (89.7%) tissue samples failed more than 5/104 SNPs compared with only 27 (2.6%) of blood samples (Supplementary Table S1). Therefore all tissue samples were excluded from the primary analysis, but a sensitivity analysis was conducted when they were included. This left a total of 369 cases and 662 controls for the main analysis, and the distribution between the trials is shown in Figure 1.

Baseline characteristics for cases and controls are shown in Table 1. In brief, 60% of samples were from women randomised to placebo and the remaining 40% from those randomised to tamoxifen. The majority of breast cancer samples came from ER-positive disease (74%). Age (median 50 years) and BMI were balanced between cases and controls. Approximately half of the women were postmenopausal, but cases were slightly more likely to be pre-menopausal (P=0.06). As expected, cases had a higher Tyrer-Cuzick breast cancer risk at entry compared with controls (P<0.001). The characteristics of the IBIS-I and Marsden sets were broadly similar and therefore are combined for the primary analysis. Trial specific demographics are shown in Supplementary Table S2 and results in Supplementary Table S3.

The SNP genotype results are shown in Table 2. Hardy-Weinberg equilibrium was verified in cases and controls (smallest *P*= 0.2). There were 3/1031(0.3%) failed samples for rs8060157 and 13/1031 (1.3%) for rs10030044. The minor allele frequency (MAF) for rs8060157 was 44% in cases and controls (Table 2), compared with 39% for cases and 47% for controls from [4]; the MAF for and rs10030044 was 41% in cases and controls, compared with 45% and 36% from the Ingle study[4] . This similarity suggests that the two populations are comparable.

Our results show little evidence for an association between the SNPs and case-control status in either the tamoxifen or placebo arms (Table 2). In tamoxifen-treated women, we observed a per-major-allele OR of 0.99 (95%CI 0.73-1.34) for rs8060157, compared with the previously reported OR of 0.70 (95%CI 0.60-0.81) for the same SNP [4]. Similarly, no evidence of tamoxifen benefit was observed in our case-control study for rs11076499 (OR=0.98, 95%CI 0.73-1.32), whereas [4] reported a benefit associated with this allele.

Supplementary Tables S3a and S3b show results for the SNP analysis for each trial separately. Very similar results between the trials were observed, with no evidence of an association between SNPs and case-control status (P_{het} > 0.4, Supplementary Table S4).

Ingle and colleagues suggested using rs8060157 and rs10030044 in combination for individualised breast cancer prevention [4]. However, in our data the estimated odds for these nine genotype combinations were very close to 1.0 in all cells. The joint per-allele effect estimates in a joint model were 1.00 (95% CI 0.74 - 1.36) for rs8060157 and 1.00 (95% CI 0.75 - 1.32) for rs10030044.

A secondary analysis for all nearby SNPs is shown in Figure 2. In total 64 *ZNF423* SNPs and 19 in *CTSO* showed some variation (MAF ≥1%), but there was no evidence of a difference between cases and controls in these regions (tamoxifen P=0.14, untreated P=0.71). Subgroup analyses were used to explore whether there were differences by randomisation arm, trial and other risk factors at baseline. No significant interactions with any baseline factors were found except for parity with ZNF423 and this would not be sustained after adjustment for multiple testing (Table S4). Finally, a sensitivity analysis that included all tissue samples from the Marsden trial had very similar results to those that only included blood samples (Supplementary Table S5).

Discussion

Ingle and colleagues hypothesized that rs8060157 and rs10030044 could be used to predict response to tamoxifen and raloxifene. The hypothesis was driven by a genome-wide association study from the P-1 and P-2 trials. When we examined this issue further in the IBIS-I and Marsden prevention trials, we could not replicate their results. Per-allelle odds ratios were close to unity in both placebo and tamoxifen arms.

A major strength of our study is that two randomised tamoxifen prevention trials were used to assess the hypothesis. In contrast with [4], participants of the placebo arm were also genotyped, which helped to assess whether the SNPs are risk or tamoxifen-response predictors. Another strength of this study is that it is hypothesis driven, and less affected by over-fitting due to multiple comparisons than for the earlier analysis.

A potential weakness of the study is that the ability to detect SNP effects was limited by sample size. Although the confidence intervals rule out effects as large as those observed in [4], a 30% increase or decrease could not be excluded.

In conclusion, this study from the IBIS-I and Marsden tamoxifen-prevention trials has failed to find evidence to support the use of rs8060157 and rs10030044 as biomarkers for tamoxifen response. Different biomarkers for response to tamoxifen might be better prioritised for future research, such as mammographic breast density [10].

Methods

Patients

Healthy women with an increased risk of breast cancer mostly from their family history were recruited to IBIS-I and Marsden trials [7,8,9]. Both trials were double-blind with women randomised to receive tamoxifen (20 mg/day) or placebo for 5 years in IBIS-I, and 5-8 years in Marden. Cases were ascertained during treatment by clinic visits and thereafter by clinic visits of questionnaires. Full details on the trials have been described previously [7,8,9]. They are registered at controlled-trials.com as ISRCTN91879928 (IBIS-I) and ISRCTN07027313 (Marsden).

Specimen characteristics

Blood samples taken at baseline from all women in IBIS-I were stored at -70C. Baseline material at the Marsden was destroyed in a fire and new blood samples were obtained and stored at -70C. Blood samples were not obtainable for 38 women from Marsden,

where paraffin-embedded tissues samples that were obtained from the cancer diagnosis were used in a sensitivity analysis.

Assay methods

Genomic DNA was quantified using the Picogreen protocol (Quant-iT PicoGreen dsDNA Products, Invitrogen, P-7589) and read on SpectraMAX GeminiXS Spectrophotometer. The Illumina OncoArray was used with the HTS method for the microarray data, as described by the manufacturer's protocol (Illumina, San Diego). An Illumina Hybridization oven was used for incubating amplified DNA (37C) and for BeadChips hybridization (48C). A Hybex incubator was used for the fragmentation (37C) and the denaturation (95C) steps. The X-stain step was carried out with a Tecan Freedom evo robot with a Te-Flow module. Arrays were scanned by an Illumina iScan Reader. Data analysis was performed with the Genotyping module (version 1.9.4) of the GenomeStudio software (Illumina; version 2011.1) using Consortium-OncoArray_15047405_A.bpm manifest. Two trios (two parents and their child) of CEPH (Centre de'Etude du Polymorphism Humain) samples were used in continuous rotation as assay controls, and two internal controls were used per plate. The assay controls were used to monitor assay quality and possible sample mismatch between the planned wells and those actually on the plate.

The main focus was on SNPs rs8060157 and rs10030044 that were identified by [4]. We also considered other SNPs on the OncoArray that were in the same regions, with 24 near *CTSO* (between rs7684248 and rs4555581), and 80 in *ZNF423* (between rs10852596 and rs12935130).

Study design

The primary end point was diagnosis of invasive breast cancer or ductal carcinoma in situ (DCIS). A nested case-control study matched all cases with available DNA by trial, follow-up duration, treatment arm (placebo or tamoxifen) and age at entry (±2 years). Samples from the two trials were combined to increase power. Two controls were matched to each case in IBIS-I, and one in Marsden. IBIS-I recruited from 1992-2001, and Marsden from 1986-1996. The end of follow-up for the present article was 2014 and 2010 respectively; median follow-up in IBIS-I was 16.6 years and 8.4 years in Marsden.

Statistical analysis methods

The number of failed SNPs in *ZNF423* and near *CTSO* was examined for quality control of the assay. Hardy-Weinberg equilibrium in cases and controls was tested by assessing the observed number of homozygotes against expected using a binomial distribution. The Tyrer-Cuzick model [11] (v7.02) was used to estimate risk at entry to each trial. This and other baseline characteristics were summarized in a Table, and differences between cases and controls were tested by a likelihood-ratio test from a conditional logistic regression model.

The main analysis estimated the per-allele odds ratios of rs8060157 and rs10030044 via conditional logistic regression, with Wald 95% confidence intervals. Secondary analysis examined the distribution of likelihood-ratio P-values for all nearby SNPs, and

for these the observed P-value was plotted against the expected under a null hypothesis of no effect for any SNP and tested using a Kolmogorov-Smirnov test. Subgroup analyses were used to check heterogeneity by treatment randomisation, trial and other risk factors at baseline, through a likelihood-ratio test of interaction. Analysis was undertaken using the statistical software R 2.15.1 [12].

References

[1] Cuzick J, Sestak I, Bonanni B, Costantino JP, Cummings S, DeCensi A, et al. Selective oestrogen receptor modulators in prevention of breast cancer: an updated meta-analysis of individual participant data. Lancet. 2013;381:1827–1834.

[2]Donnelly LS, Evans DG, Wiseman J, Fox J, Greenhalgh R, Affen J, et al. Uptake of tamoxifen in consecutive premenopausal women under surveillance in a high-risk breast cancer clinic. Brit J Cancer. 2014;110:1681–1687.

[3]Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. 2013. Lyon, France: International Agency for Research on Cancer. Available from http://globocan.iarc.fr, accessed 23-July-2014.

[4]Ingle JN, Liu M, Wickerham DL, Schaid DJ, Wang L, Mushiroda T, et al. Selective Estrogen Receptor Modulators and Pharmacogenomic Variation in ZNF423 Regulation of BRCA1 Expression: Individualized Breast Cancer Prevention. Cancer Discov. 2013;3:812–825.

[5]Fisher B, Costantino JP, Wickerham DL, Redmond CK, Kavanah M, Cronin WM, et al. Tamoxifen for Prevention of Breast Cancer: Report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. J Natl Cancer Inst. 1998;90:1371–1388.

[6]Vogel VG, Costantino JP, Wickerham DL, Cronin WM, Cecchini RS, Atkins JN, et al. Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: the NSABP Study of Tamoxifen and Raloxifene (STAR) P-2 trial. JAMA - J Am Med Assoc. 2006;295:2727–2741.

[7]Cuzick J, Forbes J, Edwards R, Baum M, Cawthorn S, Coates A et al. First results from the International Breast Cancer Intervention Study (IBIS-I): a randomised prevention trial. Lancet. 2002;360:817–824.

[8]Powles TJ, Ashley S, Tidy A, Smith IE, Dowsett M. Twenty-Year Follow-up of the Royal Marsden Randomized, Double-Blinded Tamoxifen Breast Cancer Prevention Trial. J Natl Cancer Inst. 2007;99:283–290.

[9]Cuzick J, Sestak I, Cawthorn S, Hamed H, Holli K, Howell A, et al. Tamoxifen for prevention of breast cancer: extended long-term follow-up of the IBIS-I breast cancer prevention trial. Lancet Oncol. 2015;16:67–75.

[10]Cuzick J, Warwick J, Pinney E, Duffy SW, Cawthorn S, Howell A, et al. Tamoxifen-Induced Reduction in Mammographic Density and Breast Cancer Risk Reduction: A Nested Case-Control Study J Natl Cancer Inst. 2011;103:744–752.

[11]Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Statist Med. 2004;23:1111–1130.

[12]R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria 2012. ISBN 3-900051-07-0.

Tables

Table 1: Baseline characteristics according to case control status.

	Cases	Controls	P
Total	369	662	
Placebo	221	394	
Tamoxifen	148	268	
ER-Positive	273	489	
ER-Negative	65	118	
Unknown	31	55	
Age, median (IQR)	50 (45-54)	49 (46-54)	0.9
Premenopausal	191 (52%)	316 (48%)	0.06
Perimenopausal	19 (5%)	28 (4%)	
Postmenopausal	159 (43%)	318 (48%)	
BMI, median (IQR)	25.7 (22.7-28.7)	25.4 (22.9-29.4)	0.3
TC, median RR (IQR)	2.4 (1.9-3.0)	2.1 (1.7-2.7)	<0.001

BMI: body mass index; ER: estrogen-receptor; IQR: inter-quartile range; RR: risk relative to general population; TC: Tyrer-Cuzick model.

Table 2: Genotype results for minor (m) and major (M) allele combinations, by trial arm and case status.

SNP	Status	Arm	n	mm	mM	MM	OR (95% CI)	P	Ingle OR (95% CI)
s8060157	Case	Placebo	220	49 (22%)	100 (45%)	71 (32%)	1.10 (0.87 - 1.40)	0.409	
	Control		394	72 (18%)	190 (48%)	132 (34%)			
	Case	Tamoxifen	147	26 (18%)	76 (52%)	45 (31%)	0.99 (0.73 - 1.34)	0.939	0.70 (0.60 - 0.81)
	Control		267	54 (20%)	121 (45%)	92 (34%)			
rs10030044	Case	Placebo	218	40 (18%)	103 (47%)	75 (34%)	1.01 (0.79 - 1.29)	0.929	
	Control		392	70 (18%)	190 (48%)	132 (34%)			
	Case	Tamoxifen	144	24 (17%)	66 (46%)	54 (38%)	1.00 (0.76 - 1.33)	0.991	1.42 (1.23 - 1.65)
	Control		265	41 (15%)	128 (48%)	96 (36%)			

OR, odds ratio; CI, confidence interval

Legends

Table 1: Baseline characteristics according to case control status.

Table 2: Genotype results for minor (m) and major (M) allele combinations, by trial arm and case status.

Figure 1: Consort diagram

Figure 2: P-values from conditional logistic regression likelihood ratio tests for the Oncoarray SNPs in *ZNF423* and in/near *CTSO*. Each point is a SNP with minor allele frequency greater than 1%. The expected distribution of P under the null is the diagonal.

Figure 1: CONSORT diagram.

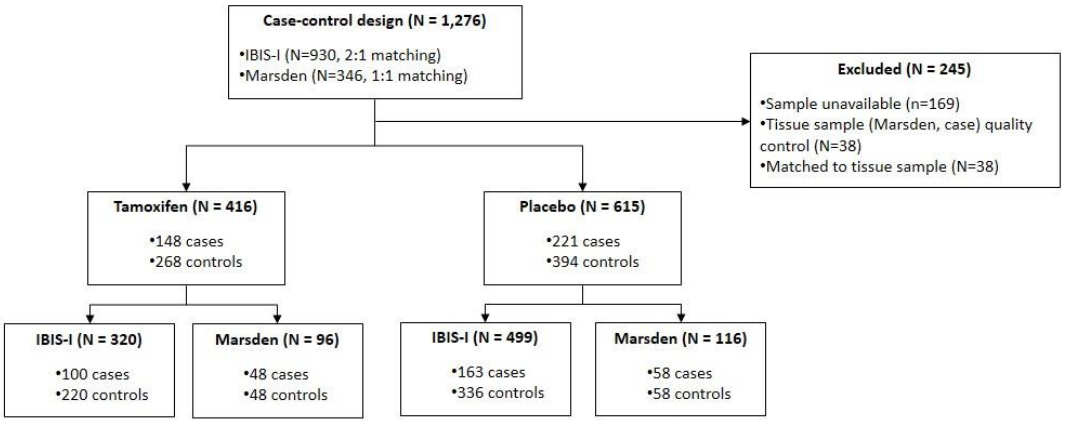
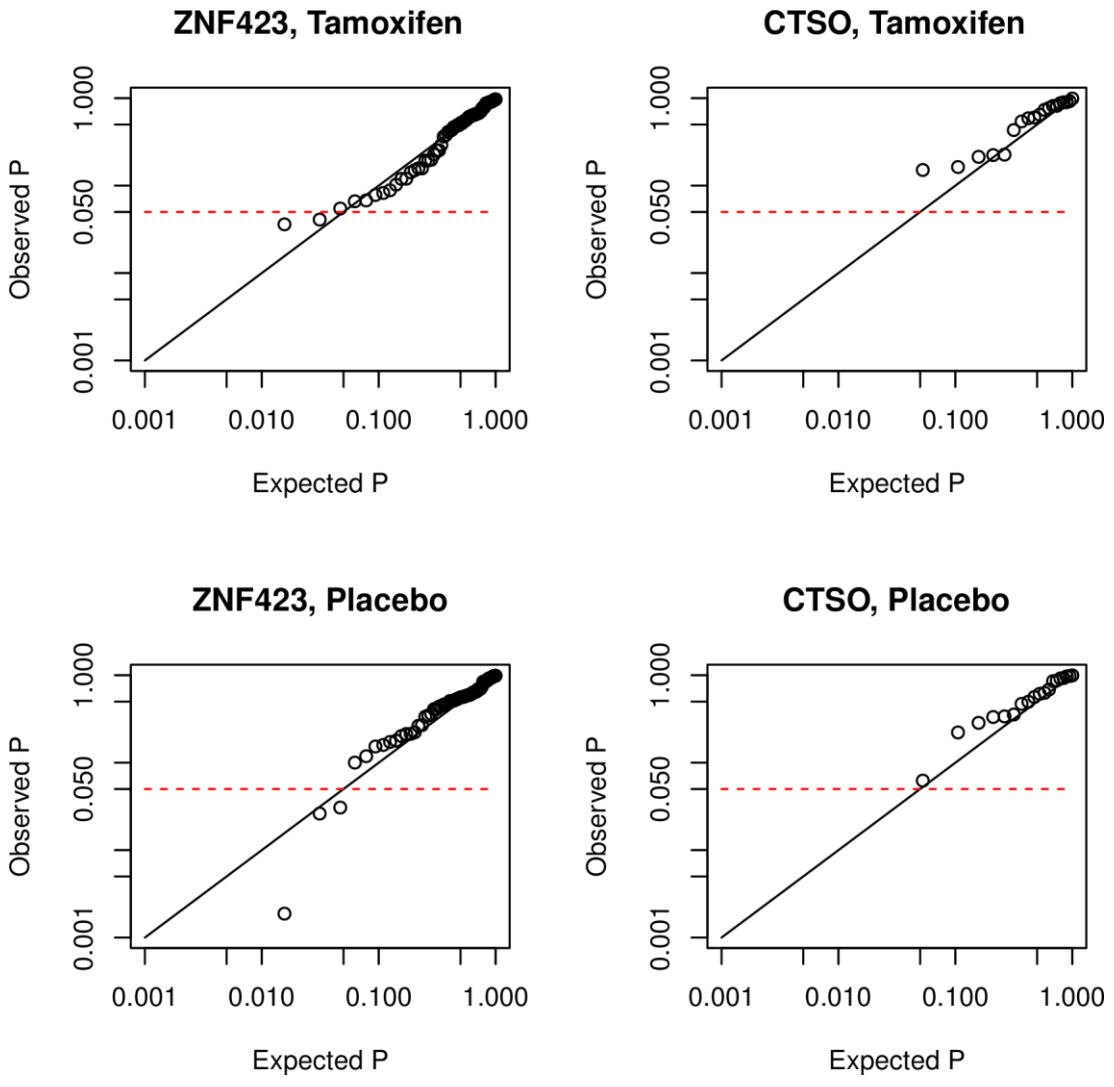


Figure 2: P-values from conditional logistic regression likelihood ratio tests for the Oncoarray SNPs in *ZNF423* and in/near *CTSO*. Each point is a SNP with minor allele frequency greater than 1%. The expected distribution of P under the null is the diagonal.



Supplementary Table S1: Number of failed SNPs on the OncoArray in ZNF423 and CTSO (from 104 in total), by arm and trial.

	Number of failed SNPs					
	0	1-5	6-10	11-20	21-30	31-46
All	936 (84.6%)	109 (9.8%)	22 (2.0%)	24 (2.2%)	9 (0.8%)	7 (0.6%)
Cases	326 (80.1%)	36 (8.8%)	15 (3.7%)	18 (4.4%)	8 (2.0%)	4 (1.0%)
Controls	610 (87.1%)	73 (10.4%)	7 (1.0%)	6 (0.9%)	1 (0.1%)	3 (0.4%)
IBIS-I	716 (87.4%)	82 (10.0%)	5 (0.6%)	7 (0.9%)	5 (0.6%)	4 (0.5%)
Cases	232 (88.2%)	24 (9.1%)	0 (0.0%)	1 (0.4%)	4 (1.5%)	2 (0.8%)
Controls	484 (87.1%)	58 (10.4%)	5 (0.9%)	6 (1.1%)	1 (0.2%)	2 (0.4%)
MARSDEN	220 (76.4%)	27 (9.4%)	17 (5.9%)	17 (5.9%)	4 (1.4%)	3 (1.0%)
Cases	94 (65.3%)	12 (8.3%)	15 (10.4%)	17 (11.8%)	4 (2.8%)	2 (1.4%)
(tissue)	0 (0.0%)	4 (10.3%)	15 (38.5%)	14 (35.9%)	4 (10.3%)	2 (5.1%)
(blood)	94 (89.5%)	8 (7.6%)	0 (0.0%)	3 (2.9%)	0 (0.0%)	0 (0.0%)
Controls	126 (87.5%)	15 (10.4%)	2 (1.4%)	0 (0.0%)	0 (0.0%)	1 (0.7%)

Supplementary Table S2: Baseline characteristics overall and for both trials individually.

Factor	Value	All (Case)	All (Control)	P	IBIS (Case)	IBIS (Control)	P	Marsden (Case)	Marsden (Control)	P
Number		369	662		263	556		106	106	
Arm	Placebo	221	394		163	336		58	58	
	Tamoxifen	148	268		100	220		48	48	
ER Status	Positive	273	489		198	414		75	75	
	Negative	65	118		45	98		20	20	
	Unknown	31	55		20	44		11	11	
Age	Median (IQR)	50 (45-54)	49 (46-54)	0.926	50 (46-54)	50 (46-54)	0.927	50 (44-54)	48 (44-54)	0.700
Menarche	<12	82 (23%)	156 (24%)	0.988	57 (22%)	133 (24%)	0.945	25 (24%)	23 (22%)	0.912
	12	68 (19%)	109 (17%)		52 (20%)	92 (17%)		16 (15%)	17 (16%)	
	>12	214 (59%)	392 (60%)		150 (58%)	326 (59%)		64 (61%)	66 (62%)	
	Unknown	5	5		4	5		1	0	
Parity	Parous	62 (17%)	86 (13%)	0.150	43 (16%)	72 (13%)	0.221	19 (18%)	14 (13%)	0.448
	Age first (IQR)	24.0 (21.0-27.0)	24.0 (21.0-27.0)	0.943	23.0 (21.0-27.0)	24.0 (21.0-27.0)	0.921	25.0 (21.0-28.5)	26.0 (22.0-28.0)	0.982
	Unknown	6	10		6	10		0	0	
Menopause	Pre	191 (52%)	316 (48%)	0.064	130 (49%)	265 (48%)	0.339	61 (58%)	51 (48%)	0.043
	Peri	19 (5%)	28 (4%)		12 (5%)	20 (4%)		7 (7%)	8 (8%)	
	Post	159 (43%)	318 (48%)		121 (46%)	271 (49%)		38 (36%)	47 (44%)	
BMI	Median (IQR)	25.7 (22.7-28.7)	25.4 (22.9-29.4)	0.303	26.4 (23.3-29.6)	25.8 (23.2-29.8)	0.441	23.4 (22.0-26.4)	23.4 (21.5-26.2)	0.318
	Unknown	11	8		5	7		6	1	
FDR	0	39 (11%)	61 (9%)	0.791	35 (13%)	55 (10%)	0.637	4 (4%)	6 (6%)	0.727
	1	268 (73%)	501 (76%)		186 (71%)	420 (76%)		82 (77%)	81 (76%)	
	>1	62 (17%)	100 (15%)		42 (16%)	81 (15%)		20 (19%)	19 (18%)	
TC Familial RR	Median (IQR)	2.2 (1.9-2.4)	2.2 (2.0-2.3)	0.094	2.2 (2.0-2.4)	2.2 (2.0-2.3)	0.157	2.2 (1.9-2.6)	2.2 (1.8-2.4)	0.364
TC Overall RR	Median (IQR)	2.4 (1.9-3.0)	2.1 (1.7-2.7)	<0.001	2.3 (1.9-3.0)	2.1 (1.7-2.6)	<0.001	2.5 (2.0-3.1)	2.3 (1.9-3.1)	0.127

BMI: body mass index; Cont: Control; ER: estrogen-receptor; FDR: number of affected first-degree relatives; IQR: inter-quartile range; RR: risk relative to general population; TC: Tyrer-Cuzick model.

Supplementary Table S3a: Genotype results for minor (m), major (M) and minor alleles by trial arm and case status; IBIS only.

SNP	Status	Arm	n	mm	mM	MM	OR (95% CI)	Ingle OR (95% CI)
rs8060157	Case	Placebo	162	40 (25%)	64 (40%)	58 (36%)	1.04 (0.80 - 1.34)	
	Control		336	63 (19%)	164 (49%)	109 (32%)		
	Case	Tamoxifen	99	19 (19%)	49 (49%)	31 (31%)	**1.07 (0.76 - 1.52)**	**0.70 (0.60 - 0.81)**
	Control		219	42 (19%)	99 (45%)	78 (36%)		
rs10030044	Case	Placebo	161	27 (17%)	77 (48%)	57 (35%)	0.96 (0.73 - 1.26)	
	Control		334	60 (18%)	164 (49%)	110 (33%)		
	Case	Tamoxifen	97	15 (15%)	48 (49%)	34 (35%)	**1.00 (0.71 - 1.41)**	**1.42 (1.23 - 1.65)**
	Control		217	34 (16%)	107 (49%)	76 (35%)		

Supplementary Table S3b: Genotype results for minor (m), major (M) and minor alleles by trial arm and case status; Marsden only.

SNP	Status	Arm	n	mm	mM	MM	OR (95% CI)	Ingle OR 95% CI)
rs8060157	Case	Placebo	58	9 (16%)	36 (62%)	13 (22%)	1.60 (0.84 - 3.02)	
	Control		58	9 (16%)	26 (45%)	23 (40%)		
	Case	Tamoxifen	48	7 (15%)	27 (56%)	14 (29%)	**0.78 (0.43 - 1.43)**	**0.70 (0.60 - 0.81)**
	Control		48	12 (25%)	22 (46%)	14 (29%)		
rs10030044	Case	Placebo	57	13 (23%)	26 (46%)	18 (32%)	1.27 (0.73 - 2.21)	
	Control		58	10 (17%)	26 (45%)	22 (38%)		
	Case	Tamoxifen	47	9 (19%)	18 (38%)	20 (43%)	**1.00 (0.61 - 1.64)**	**1.42 (1.23 - 1.65)**
	Control		48	7 (15%)	21 (44%)	20 (42%)		

Supplementary Table S4: Interaction subgroup tests for rs8060157 (*ZNF423*) and rs10030044 (near *CTSO*).

	Interaction group	ZNF423 OR (95% CI)	ZNF423 P-value	CTSO OR (95% CI)	CTSO P-value
Randomisation	Tamoxifen	0.86 (0.59-1.25)	0.433	0.88 (0.62-1.27)	0.498
Trial	Marsden	1.11 (0.72-1.71)	0.626	1.15 (0.78-1.70)	0.487
Age	Per year	1.01 (0.98-1.04)	0.370	0.99 (0.96-1.02)	0.372
Menopause	Post	0.91 (0.64-1.30)	0.611	0.74 (0.52-1.06)	0.099
BMI	> 25.4	0.77 (0.53-1.10)	0.152	1.09 (0.76-1.57)	0.647
Parity	No	0.54 (0.32-0.90)	0.016	0.70 (0.43-1.14)	0.154
Age at first birth	>30	0.73 (0.48-1.11)	0.136	0.73 (0.49-1.11)	0.139
Age at menarche	<13	0.65 (0.45-0.94)	0.021	0.84 (0.58-1.22)	0.364
TC family history (RR)	> 2.2	0.82 (0.57-1.17)	0.274	1.13 (0.78-1.62)	0.520
TC All (RR)	> 2.2	0.77 (0.54-1.10)	0.153	1.04 (0.72-1.49)	0.853

Supplementary Table S5: Genotype results for minor (m), major (M) and minor alleles by trial arm and case status including tissue based assessment from Marsden.

SNP	Status	Arm	n	mm	mM	MM	OR (95% CI)	Ingle OR (95% CI)
rs8060157	Case	Placebo	233	54 (23%)	107 (46%)	72 (31%)	1.14 (0.90 - 1.43)	
	Control		414	76 (18%)	202 (49%)	136 (33%)		
	Case	Tamoxifen	157	28 (18%)	83 (53%)	46 (29%)	**0.98 (0.73 - 1.32)**	**0.70 (0.60 - 0.81)**
	Control		285	59 (21%)	130 (46%)	96 (34%)		
rs10030044	Case	Placebo	238	44 (18%)	113 (47%)	81 (34%)	1.07 (0.85 - 1.36)	
	Control		412	72 (17%)	196 (48%)	144 (35%)		
	Case	Tamoxifen	160	26 (16%)	73 (46%)	61 (38%)	**0.95 (0.72 - 1.24)**	**1.42 (1.23 - 1.65)**
	Control		283	44 (16%)	140 (49%)	99 (35%)		