

SCIENTIFIC REPORTS



OPEN

Association of three SNPs in *TOX3* and breast cancer risk: Evidence from 97275 cases and 128686 controls

Received: 11 December 2014

Accepted: 09 July 2015

Published: 04 August 2015

Li Zhang & Xinghua Long

The associations of SNPs in *TOX3* gene with breast cancer risk were investigated by some Genome-wide association studies and epidemiological studies, but the study results were contradictory. To derive a more precise estimate of the associations, we conducted a meta-analysis. ORs with 95% CI were used to assess the strength of association between *TOX3* polymorphisms and breast cancer risk in fixed or random effect model. A total of 37 publications with 97275 cases and 128686 controls were identified. We observed that the rs3803662 C>T, rs12443621 A>G and rs8051542 C>T were all correlated with increased risk of breast cancer. In the stratified analyses by ethnicity, significantly elevated risk was detected for all genetic models of the three SNPs in Caucasians. In Asian populations, there were significant associations of rs3803662 and rs8051542 with breast cancer risk. Whereas there was no evidence for statistical significant association between the three SNPs and breast cancer risk in Africans. Additionally, we observed different associations of rs3803662 with breast cancer risk based on different ER subtype and *BRCA1/BRCA2* mutation carriers. In conclusion, the meta-analysis suggested that three SNPs in *TOX3* were significantly associated with breast cancer risk in different populations.

Breast cancer is the most generally diagnosed cancer and the most common cause of cancer death for females all over the world, particularly in the economically developing countries¹. It is well known that breast cancer is a heterogeneous disease, not only in the aspect of various pathogenesis, but also in diversified clinical manifestation and outcome. Meanwhile, breast carcinoma is multifactorial disease, from a certain perspective, along with the combination of polygenic inheritance factor and environmental factor. Accompany with technological advances, more studies related with the genomic variation were conducted, in order to improve diagnosis and treatment for breast cancer patients. Mutations in some high and moderate penetrate genes, such as *BRCA1*, *BRCA2* and *ATM*, were verified to be connected with the increased risk of breast cancer^{2,3}. Nonetheless, these mutations constitute a part of the disease risk and it remains unclarified for the majority of genetic variations related with breast cancer susceptibility, particularly for low penetrate genes. It is noteworthy that genome-wide association studies (GWASs) about hundreds of single nucleotide polymorphisms (SNPs) provide strong evidences in elaborating the associations between low penetrate genes and breast cancer risk.

The *TOX3* gene, formerly known as trinucleotide repeat containing 9 (*TNRC9*), is located in the chromosome 16q 12 and has a tri-nucleotide repeat motive. The gene encoded a protein containing a putative high mobility group (HMG) box⁴, indicating that it might play a potential role in calcium dependent transcription as a transcription factor⁵. In the recent years, the associations between genetic variants in *TOX3* region and breast cancer susceptibility have been validated by GWASs and epidemiological studies in European, Asian and African American populations^{6–13}. The SNP rs3803662 is located in 8 kb

Zhongnan Hospital of Wuhan University, Wuhan, 430071, China. Correspondence and requests for materials should be addressed to X.L. (email: xlong888@yahoo.com)

upstream of *TOX3*, and the rs12443621 and rs8051542 are both lied in an linkage disequilibrium (LD) block containing the 5' end of *TOX3*⁶.

The *TOX3* rs3803662 was identified to exhibit association with breast cancer by GWASs^{6,7,10}, with ascertainment of the association in Hispanic and non-Hispanic white women by Slattery *et al.*¹¹. However, no significant association was found between rs3803662 and breast cancer risk in Asian and African ancestry^{8,13}. Analogously, there was no evidence for the association between rs12443621 or rs8051542 and increased risk of breast cancer in Chinese women^{14–16}. Whereas, Shan *et al.* reported that rs8051542 was significantly correlated with breast cancer risk in Tunisians¹⁷. Additionally, some studies found different relationships of three SNPs and breast cancer risk among different populations, which might result from different sample size or diverse allele frequencies and LD pattern among populations.

Meanwhile, the most recent meta-analysis related to the associations between the above-mentioned 3SNPs with breast cancer risk omitted some important studies¹⁸, and thus had limited statistical power to demonstrate the associations. Therefore, we performed an updated meta-analysis to aim to come up with the highest level of evidence for the associations between three SNPs in *TOX3* gene and breast cancer risk among diverse ancestry populations and distinct tumor subtypes stratified by estrogen receptor (ER) or *BRCA1/BRCA2*.

Materials and Methods

Literature search strategy. We carried out a comprehensive literature search from PubMed and EMBASE databases up to March 2015, using the following search terms “*TOX3*” or “*TNRC9*” and “polymorphism” or “genetic variant” or “rs3803662” or “rs12443621” or “rs8051542” and “breast cancer” or “breast carcinoma” or “breast tumor”. First, we retrieved all potentially relevant articles, whose abstracts contained information related to our research purpose. Second, the references from eligible studies were carefully checked for additional relevant literature. Finally, only the comprehensive or the most recent study was brought into this meta-analysis, in the case that the same study population was included in several different articles.

Selection criteria. Eligible studies had to fulfill the following criteria: (1) case-control studies or cohort studies evaluating the association between *TOX3* polymorphism (rs3803662, rs12443621 or rs8051542) and breast cancer risk; (2) odds ratio (OR) and 95% confidence interval (CI) or genotype data of rs3803662, rs12443621 or rs8051542 in breast cancer patients and cancer-free female to calculate OR and 95% CI; (3) studies were confined to human female groups; (4) articles in English.

Data extraction. A standard protocol was applied to extract data. For every eligible study, the following data were extracted: First author's surname, year of publication, country of origin, population ethnicity, genotyping method, the genotype counts in cases and control (TT, CT and CC genotypes for *TOX3* rs3803662; GG, AG and AA genotypes for rs12443621; TT, CT and CC genotypes for rs8051542) and P-value for the HWE in control groups. Two investigators independently extracted the above relative data with any disagreement resolved by discussion. If no consensus wasn't reached, another investigator joined in the discussion. And the final decision was made by the majority of the votes.

Statistical methods. The strength of associations between *TOX3* polymorphisms and breast carcinoma risk were estimated by OR with corresponding 95% CI. For all studies, we assessed the association under five different genetic models for calculating OR. Those were homozygote codominant model, heterozygote codominant model, dominant model, recessive model and allele model. Hardy-Weinberg equilibrium (HWE) was assessed by using χ^2 test to compare expected and actual genotype frequencies among controls of each study. Q-statistic was applied to investigate heterogeneity among studies. P-value greater than 0.1 for Q test suggested a lack of statistically significant heterogeneity, and the fixed-effect model (Mantel-Haenszel method)¹⁹ was used to calculate pooled ORs. Otherwise, heterogeneity was present and the random-effect model (DerSimonian-Laird method)²⁰ was more appropriate. In addition, the I^2 -test was employed to accurately measure the degree of heterogeneity. Furthermore, the I^2 -value less than 25% was equivalent to mild heterogeneity, and values between 25% and 50% was equivalent to moderate heterogeneity, whereas values greater than 50% was equivalent to large heterogeneity among studies. Potential publication bias was estimated by symmetry of funnel plot of OR versus the standard error of log (OR) and the visual symmetrical plot indicated that there was no publication bias among studies. Sensitivity analyses were conducted to assess the robustness of the results by eliminating each study in turn to show whether the individual data set influenced the pooled OR. Stratified analyses were conducted in terms of ethnicity, estrogen receptor (ER) status, *BRCA1* and *BRCA2* mutation. All statistical tests in this meta-analysis were two-tailed and P-value ≤ 0.05 was considered statistically significant unless otherwise noted. All statistical analyses were performed with Review Manager 5.2 software recommended by Cochrane Collaboration and Comprehensive Meta Analysis V2 software.

Result

Study Characteristics. Based on the above selection criteria, a total of 37 eligible studies were included in the pooled analyses, involving 97275 cases and 128686 controls for rs3803662 polymorphism^{7–17,21–46}. For rs12443621, 14 studies^{8,12,14–17,24,28,29,31,41,42,45,46} involved a total of 17750 cases and

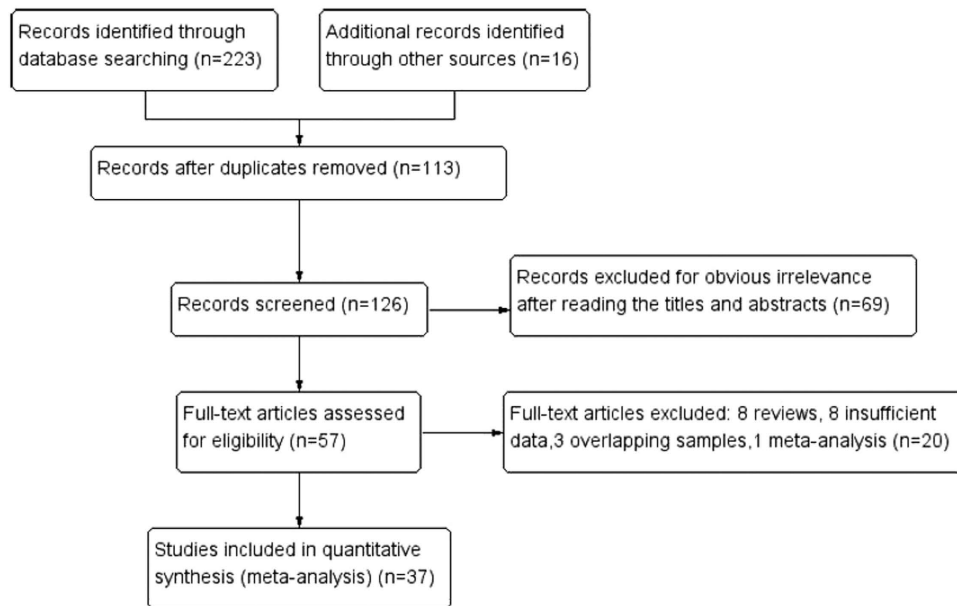


Figure 1. The flowchart of the study selection process.

19488 controls. Moreover, there were 13 studies^{8,12,15–17,28,29,31,39,41,42,45,46} with 20965 cases and 21580 controls for rs8051542. Of particular note was that it's smaller than 0.05 for the P-value of Hardy-Weinberg equilibrium in the controls of two studies, Campa *et al.* and Garcia-Closas *et al.*^{37,38}, but we still included the two studies after sensitivity analyses were done. Additionally, in three included studies, genotype frequencies were shown separately according to different ethnic groups^{7,31,39}. Therefore, the corresponding genotype counts in the study were separately considered for analyses. For rs3803662, five studies^{11,15,16,26,38} concerned with ER subtype of breast cancers and three studies^{26,30,35} related with BRCA1/2 mutation carriers were analysed as subgroups. The Fig. 1 expounded the study selection process. The Table 1 and 2 described the main features of these studies, especially for the genotype counts.

Meta-analysis results. The mixtures of adjusted and crude estimates were used to calculate pooled ORs. The available adjusted variables of included studies were listed in supplementary table 1. Owing to large heterogeneity among studies, we used random-effect model to calculate pooled ORs for the associations of rs3803662 and rs12443621 with breast cancer risk. In contrast, fix-effect model was applied to calculate pooled ORs for rs8051542. In aggregate, T-rs3803662 and T-rs8051542 were all statistically associated with increased risk of breast cancer in all genetic models. However, the association between G-rs12443621 and breast cancer risk was only observed in Caucasians under all genetic models. The pooled ORs and 95%CI for these associations in all genetic models were shown in detail in Table 3–5, respectively. Forest plots related to the association of rs3803662, rs12443621 and rs8051542 with breast cancer susceptibility in homozygote model were shown in Fig. 2, Fig. 3 and Fig. 4, respectively.

In the subgroup analysis by ethnicity, our results indicated statistically significant associations between the three SNPs and breast cancer susceptibility in Caucasians under all genetic models. Nevertheless, as for Asian populations, T-rs3803662 and T-rs8051542 were shown to be statistically significant correlated with increased risk of breast cancer in all genetic models. In addition, there was no evidence for the statistical significant associations between the three SNPs and increased risk of breast cancer in African population which were almost African-American in our study. The pool ORs and 95%CI for these stratified analyses were detailedly shown in Table 3–5 for all genetic modes.

When stratified by ER status for rs3803662, statistically significant increased risk was found in ER⁺ and ER⁻ tumor (Fig. 5 and Fig. 6). Moreover, a stronger association was identified in ER⁺ than ER⁻ subtype for breast cancer risk (Fig. 7). Additionally, our analysis demonstrated that there were significant relationships between elevated risk of breast cancer and BRCA1/2 mutation carriers for rs3803662 (Fig. 8 and Fig. 9). And the details about ORs and 95% CI under all genetic models were shown in Table 3.

Sensitivity analyses and publication bias. Sensitivity analyses were conducted to assess the robustness of the results by eliminating each study in turn and all the results were not essentially altered, suggesting that the results of our meta-analysis were statistically stable. Publication bias of the eligible literature was evaluated by funnel plots and the shapes of funnel plots for literature about association between three SNPs and breast cancer risk were mostly symmetrical, indicating that no publication bias was detected.

Author	Year	Country	Ethnicity	Genotyping method	Case			Control			P _{HWE}
					TT	CT	CC	TT	CT	CC	
Stacey ⁷	2007	Iceland	Caucasian	Illumina	469	1985	2100	1272	6913	9392	0.999
Stacey ⁷	2007	Iceland	AA	Illumina	95	211	116	130	222	95	0.990
Stacey ⁷	2007	Iceland	Asian	Illumina	155	275	122	147	278	132	0.980
Stacey ⁷	2007	Iceland	Mixed	Illumina	104	275	183	114	340	259	0.891
He ⁸	2014	China	Asian	Sequenom MassARRAY	271	280	72	270	278	72	0.973
Elematore ⁹	2014	Chile	Mixed	TaqMan	62	185	100	100	371	330	0.786
Low ¹⁰	2013	Japan	Asian	Illumina OmniExpress BeadChip	1801	2705	1016	1496	2739	1254	0.996
Slattery ¹¹	2011	USA	Caucasian	TaqMan	204	755	778	202	862	978	0.550
Udler ¹²	2010	UK	Caucasian	NA	194	942	1041	157	829	1273	0.167
Ruiz -Narváez ¹³	2010	USA	AA	Sequenom MassArray iPLEX	188	376	189	214	412	199	0.980
Jiang ¹⁴	2011	China	Asian	SNaPshot	233	212	48	232	224	54	0.995
Li ¹⁵	2009	China	Asian	PCR-LDR	118	141	32	123	128	40	0.470
Liang ¹⁶	2010	China	Asian	SNP stream highthroughput 12-plex	486	413	126	455	464	127	0.603
Shan ¹⁷	2012	Qatar	Caucasian	TaqMan	147	293	200	78	165	126	0.083
Long ²¹	2013	USA	AA	TaqMan/Sequenom	328	613	287	563	1027	469	0.988
Kim ²²	2012	Korea	Asian	Affymetrix/TaqMan	499	1939	1887	577	1967	1677	0.996
Huo ²³	2012	USA	African	Illumina GoldenGate	393	754	362	368	691	324	0.991
Chan ²⁴	2012	Singapore	Asian	TaqMan	541	499	134	629	656	181	0.622
Stevens ²⁵	2011	USA	Caucasian	iPLEX/Illumina	268	1252	1460	363	1960	2650	0.982
Mulligan ²⁶	2011	Canada	Mixed	TaqMan	585	2652	3109	426	2197	2899	0.730
Han ²⁷	2011	Korea	Asian	RT-PCR	1481	1435	369	1361	1617	516	0.317
Long ²⁸	2010	USA	Asian	Affymetrix/Sequenom MassARRAY	2934	2761	650	1603	1727	465	0.996
Long ²⁹	2010	USA	Caucasian	TaqMan	258	1172	1330	190	1028	1391	0.997
Latif ³⁰	2010	UK	Caucasian	TaqMan	84	395	422	19	137	217	0.660
Barnholtz-Sloan ³¹	2010	USA	AA	Illumina GoldenGate	196	378	166	182	333	142	0.654
Barnholtz-Sloan ³¹	2010	USA	Caucasian	Illumina GoldenGate	133	512	585	89	440	589	0.591
Gorodnova ³²	2010	Russia	Caucasian	RT-PCR	16	50	74	15	82	77	0.294
Hemminki ³³	2010	Germany	Caucasian	MALDITOF mass spectrometry	154	626	635	124	704	1002	0.982
Mcinerney ³⁴	2009	Ireland	Caucasian	KASPar SNP genotyping	82	382	486	58	396	532	0.161
Antoniou ³⁵	2008	UK	Caucasian	TaqMan/MALOL-TOF/iPLEX	497	2173	2422	382	1831	2244	0.756
Tapper ³⁶	2008	UK	Caucasian	iPLEX Service	76	371	452	196	1137	1647	0.990
Campa ³⁷	2011	Germany	Mixed	TaqMan	1071	3528	3706	1150	4724	5721	0.001
Garcia-Closas ³⁸	2008	USA	Mixed	TaqMan/MALDITOF	1848	7132	7759	2026	9705	13295	<0.0001
Barzan ³⁹	2013	Germany	Asian	Sequenom MassARRAY	482	413	89	961	990	255	0.999
Barzan ³⁹	2013	Germany	Caucasian	Sequenom MassARRAY	36	140	135	65	369	526	0.979
Rinella ⁴⁰	2013	USA	Caucasian	Affymetrix/KASPar	131	335	214	106	366	315	0.985
Zheng ⁴¹	2009	USA	AA	Sequenom MassARRAY	222	404	184	482	891	411	0.984
Butt ⁴²	2012	Sweden	Caucasian	Sequenom MassARRAY	64	278	353	95	512	780	0.380
Mizoo ⁴³	2013	Japan	Asian	TaqMan	160	230	74	142	227	91	0.987
Harlid ⁴⁴	2012	Sweden	Caucasian	MALDITOF mass spectrometry	330	1420	1794	352	1898	2768	0.280
Zheng ⁴⁵	2010	USA	Asian	Affymetrix	1401	1325	313	1286	1410	386	0.987
Tamimi ⁴⁶	2010	USA	Caucasian	Sequenom iPLEX/TaqMan	54	300	333	50	273	415	0.576

Table 1. Characteristics of studies for the association of *TOX3* rs3803662 with breast cancer risk included in the meta-analysis. AA: African Americans; NA: not available; HWE: Hardy-Weinberg equilibrium; PCR-LDR : Polymerase chain reaction–ligation detection reaction.

Discussion

The *TOX3* gene encoded a protein with an HMG box that is considered to be implicated in modification of DNA and chromatin structure⁴⁷. Moreover, increased expression of *TOX3* was relevant to bone

Author	rs12443621 Genotypes						P _{HWE}	rs8051542 Genotypes						P _{HWE}
	Cases			Controls				Cases			Controls			
	GG	AG	AA	GG	AG	AA		TT	CT	CC	TT	CT	CC	
He ⁸ 2014	110	304	209	115	304	201	0.809	25	199	399	18	175	427	0.989
Udler ¹² 2010	527	1111	546	497	1099	681	0.176	455	1089	611	425	1067	736	0.274
Jiang ¹⁴ 2011	170	239	84	162	251	97	0.990	—	—	—	—	—	—	—
Li ¹⁵ 2009	106	138	54	97	141	55	0.766	15	82	198	9	90	209	0.854
Liang ¹⁶ 2010	347	507	186	338	519	204	0.850	48	314	670	47	309	708	0.078
Shan ¹⁷ 2012	190	301	147	98	180	85	0.894	138	289	208	46	176	146	0.529
Chan ²⁴ 2012	404	573	198	532	669	262	0.419	—	—	—	—	—	—	—
Long ²⁸ 2010	546	1448	960	554	1469	974	0.998	246	1971	3941	118	1080	2460	0.968
Long ²⁹ 2010	286	573	286	274	571	297	0.989	336	788	463	279	709	451	0.991
Barnholtz-Sloan ³¹ 2010	164	370	208	165	329	164	0.999	87	342	313	59	304	295	0.121
Barnholtz-Sloan ³¹ 2010	337	580	313	242	574	302	0.319	257	587	386	201	559	358	0.501
Barzan ³⁹ 2013	—	—	—	—	—	—	—	43	327	614	72	651	1483	0.957
Barzan ³⁹ 2013	—	—	—	—	—	—	—	81	155	75	186	473	301	0.994
Zheng ⁴¹ 2009	189	405	216	423	891	470	0.986	80	349	381	170	762	852	0.984
Butt ⁴² 2012	165	338	195	275	657	451	0.203	149	338	192	272	637	443	0.119
Zheng ⁴⁵ 2010	1001	1486	552	1008	1509	565	0.995	118	961	1960	96	898	2088	0.963
Tamimi ⁴⁶ 2010	151	337	193	130	366	241	0.659	132	359	194	135	380	220	0.193

Table 2. Genotypes and P_{HWE} for TOX3 rs12443621 and rs8051542 polymorphisms included in the study. P_{HWE}: P value of Hardy-Weinberg equilibrium for control groups.

metastasis in breast cancer patients⁴⁸. Whereas, precise biological function of *TOX3* is undetermined. Some GWASs and epidemiological studies have identified the associations of *TOX3* polymorphisms with breast cancer susceptibility. However, study results were not consistent. Hence, in order to resolve the conflict, we performed this meta-analysis of the associations between the *TOX3* rs3803662, rs12443621 and rs8051542 polymorphism and breast cancer risk.

The three SNPs locate in the 5' end of *TOX3* gene and a hypothetical gene LOC643714 on 16q12, and the region is contained in a 133kb linkage disequilibrium (LD) block¹². Based on the International HapMap database, different LD patterns were observed between Asian and European ancestry. SNP rs3803662 was in moderate LD with rs12443621, with a Pearson's correlation coefficient (r^2) of 0.29 in the HapMap CEU population for European ancestry, but very weak LD was found between these two SNPs ($r^2 = 0.06$) in Chinese⁸. Similarly, there was very weak association ($r^2 = 0.08$) between rs3803662 and rs8051542 located 52 kb apart from each other in Chinese women⁴⁵. However, the two SNPs showed moderate association ($r^2 = 0.15$) with each other in European populations⁸. The substantial differences in genetic architecture among races, such as allele frequencies and LD structures, may partly account for our results which confirmed different association of the three SNPs with breast cancer risk in Caucasians, Asians and African-Americans. Rs3803662-T allele, rs12443621-G allele and rs8051542-T allele were statistically significantly associated with increased risk of breast cancer in Caucasians. Meanwhile, T-rs3803662 and T-rs8051542 were identified as risk factors of breast cancer in Asian populations. However, there was no evidence to prove that the three SNPs in African-Americans and G-rs12443621 in Asians were implicated in the breast tumor susceptibility, which was in line with the previous studies^{41,45}. It's worth mentioning that our study showed that T-rs3803662 and G-rs12443621 were protective factors in African-Americans in spite of no statistical significance.

In general, our study proved that the T-rs3803662 and T-rs8051542 in *TOX3* were correlated with elevated breast cancer risk in all genetic models. It is notable that a previous meta-analysis directed by Chen *et al.*¹⁸ has showed that rs3803662 polymorphism was significantly associated with breast cancer risk, but no significant associations were observed for the rs12443621 and rs8051542. In addition, it only included eight case-control studies without stratified analyses. Compared with the previous meta-analysis, our study had more powerful and detailed analyses to prove our results. First and most obviously, more eligible literature and larger sample size were included. Second, the associations between breast cancer risk and rs3803662 polymorphism were considered with respect to ER status and BRCA1/2 mutation carriers. Third, stratified analyses were performed based on Caucasians, Asians and Africans, which was in favor of a more comprehensive understanding the associations in diverse populations. Finally and most importantly, we used mixture of adjusted and crude ORs rather than unadjusted estimates to calculate the pooled ORs. Meanwhile, the original genotype counts of eligible studies were also used to calculate the crude ORs. Supplementary table 2 showed the pooled ORs of the associations between the

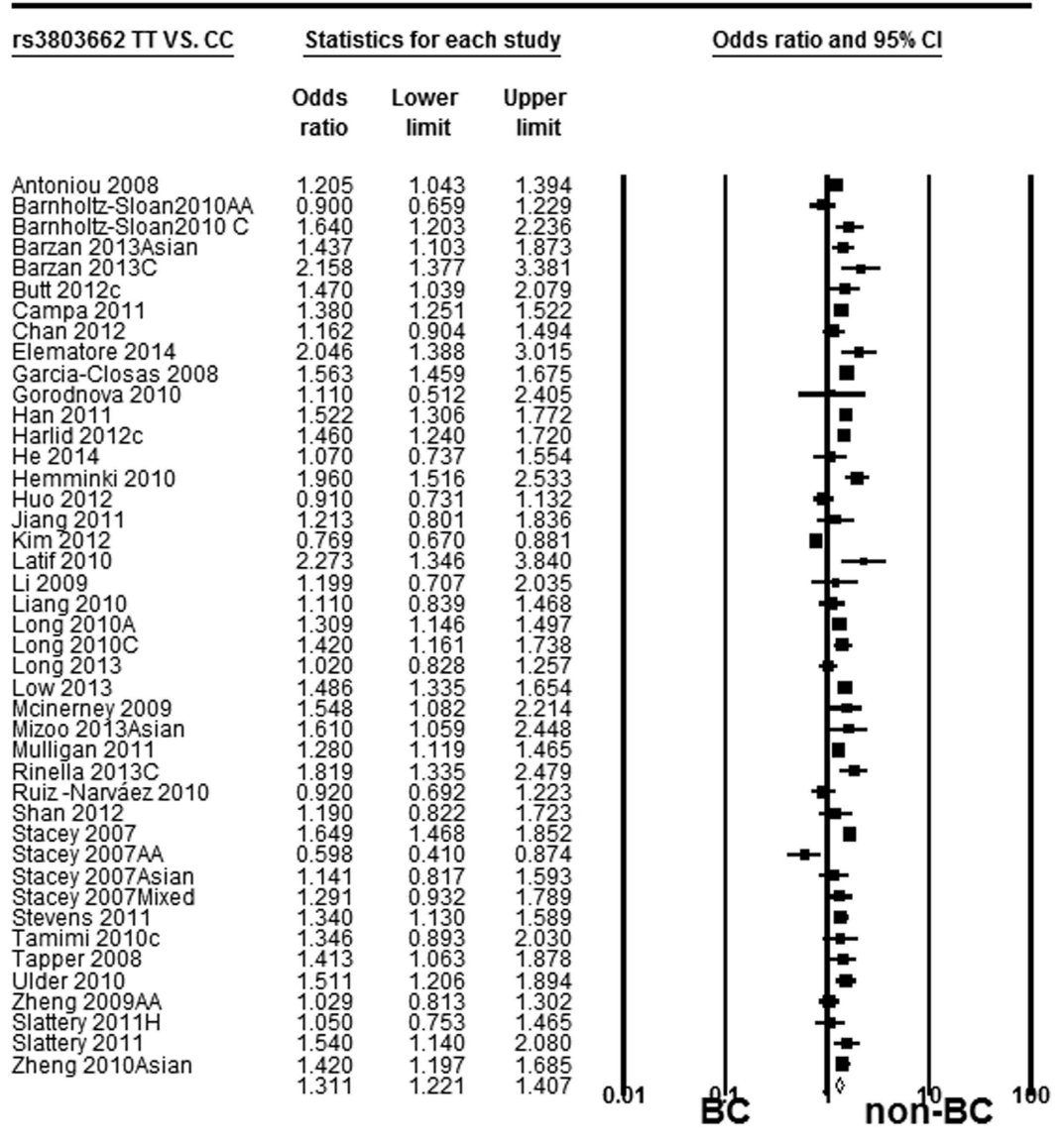


Figure 2. Forest plot of *TOX3* rs3803662 polymorphism and breast cancer risk. Random-effect model was used for the analysis (homozygote codominant model TT vs. CC). The squares and horizontal lines correspond to the specific OR and 95% CI for every study. The area of the squares reflects the study specific weight. The diamond stands for the pooled OR and 95% CI.

3SNPs and breast cancer risk by using crude estimates. And there was no significant difference among the two results of pooled ORs based on different estimates, except for rs12443621. The crude ORs were incorporated to result in marginally association of rs12443621 with breast cancer risk under homozygote, dominant and allele genetic mode, but no association was found by using mixture ORs. That was probably because that adjusted estimates could yield more accurate results. Nevertheless, the two ways both demonstrated the relationship between rs12443621 and elevated risk of breast cancer in Caucasians.

To date, more attention has been paid to the heterogeneity of associations between common genetic variants and breast cancer subtypes. The two large-scale studies^{37,38} and our result identified that rs3803662 polymorphism was associated with both ER⁺ and ER⁻ subtype of breast cancer, in spite of the slightly weaker association for ER⁻ breast cancer. Additionally, our study demonstrated that T-rs3803662 was statistically significant associated with increased risk in ER⁺ breast cancer compared with ER⁻ subtype, which was accordance with the researches done by Stacey *et al.* and Broeks *et al.*^{7,49}. Intriguingly, T-rs8051542 allele and rs12443621 AG/GG genotypes, rather than rs3803662, were significantly associated with elevated risk of ER⁺ breast cancer in Chinese women^{8,16}. By contrast, the significant associations of rs8051542 and rs12443621 were observed with luminal A (ER/PR⁺, Her2⁻) and Her2⁺/ER⁻ breast cancer only among whites, respectively⁵⁰. Furthermore, the association was strongly confirmed between

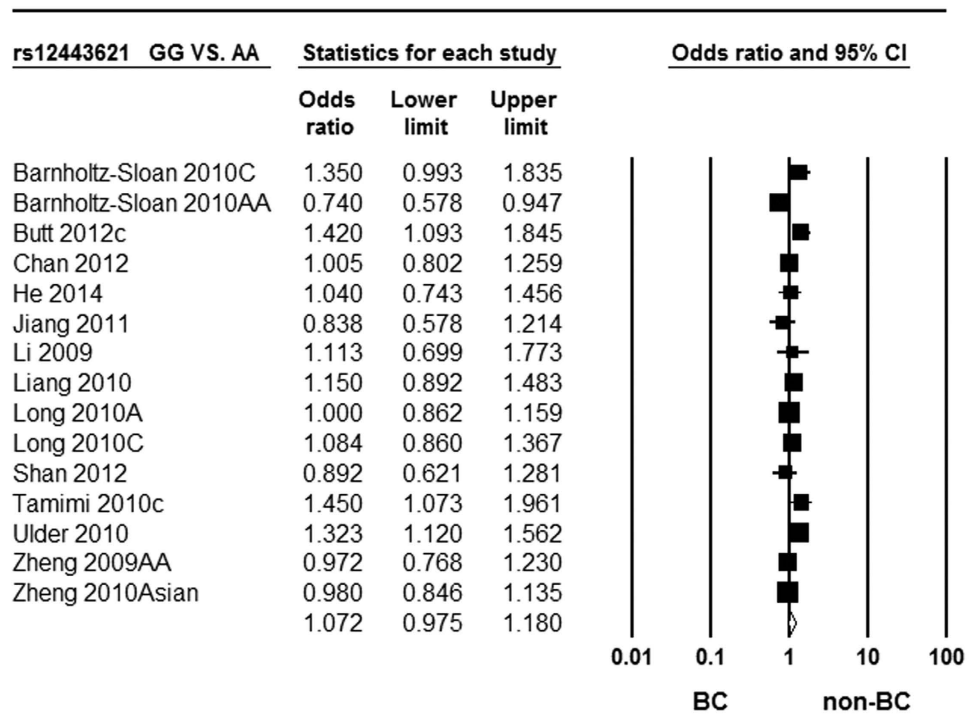


Figure 3. Forest plot of *TOX3* rs12443621 polymorphism and breast cancer risk. Random-effect model was used for the analysis (homozygote codominant model GG vs. AA).

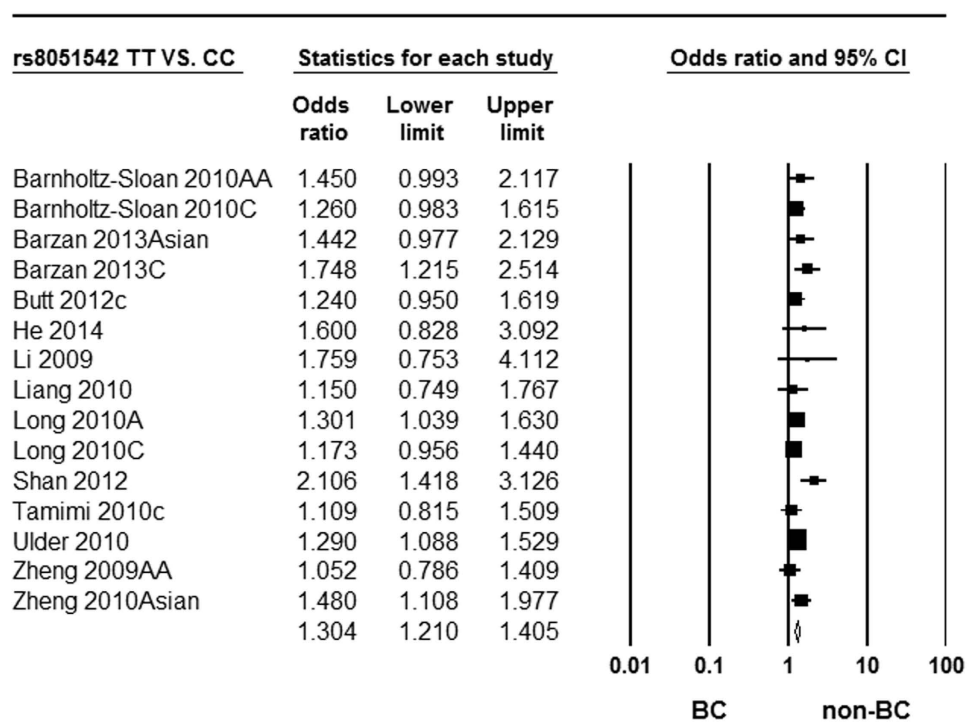


Figure 4. Forest plot of *TOX3* rs8051542 polymorphism and breast cancer risk. Fixed-effect model was used for the analysis (homozygote codominant model TT vs. CC).

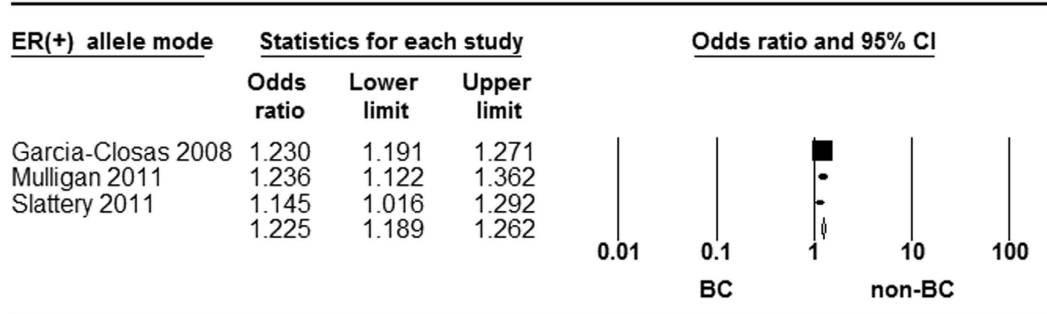


Figure 5. Forest plot of *TOX3* rs3803662 polymorphism and breast cancer risk stratified by ER (+). Fixed-effect model was used for the analysis (allele model T vs. C).

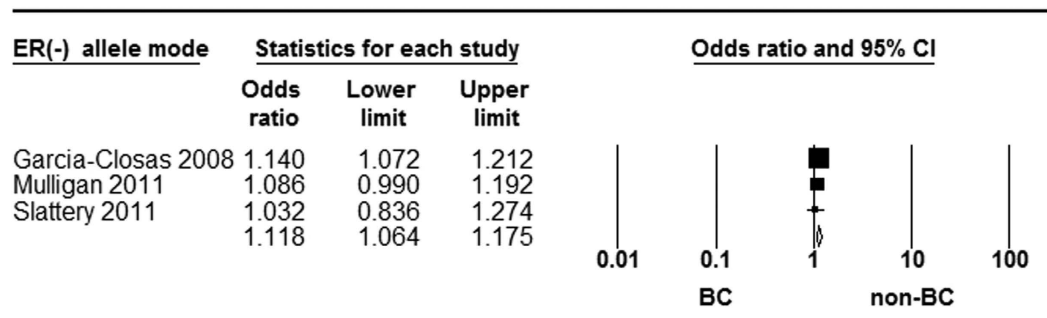


Figure 6. Forest plot of *TOX3* rs3803662 polymorphism and breast cancer risk stratified by ER (-). Fixed-effect model was used for the analysis (allele model T vs. C).

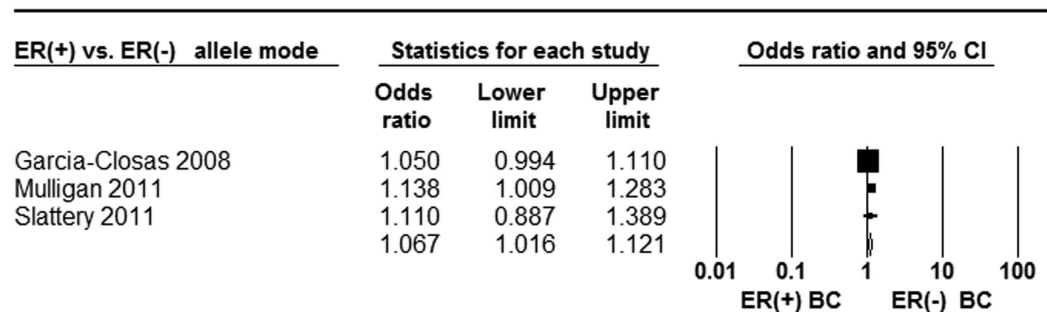


Figure 7. Forest plot of *TOX3* rs3803662 polymorphism and breast cancer risk in ER+ subtype compared with ER- tumors. Fixed-effect model was used for the analysis (allele model T vs. C).

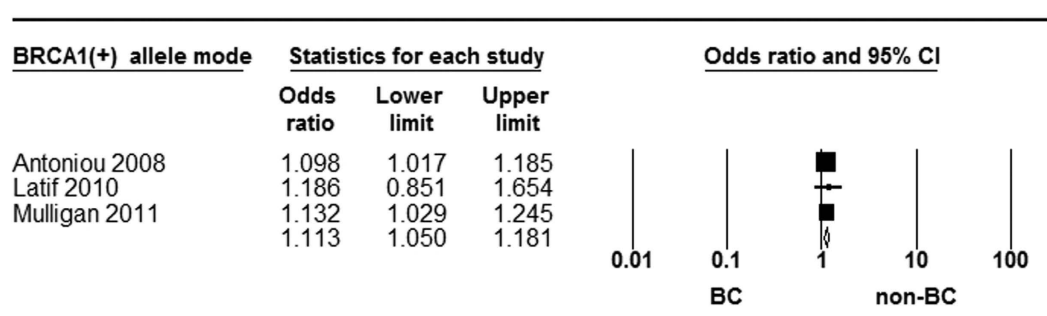


Figure 8. Forest plot of *TOX3* rs3803662 polymorphism and breast cancer risk stratified by BRCA1 mutation. Fixed-effect model was used for the analysis (allele contrast model T vs. C).

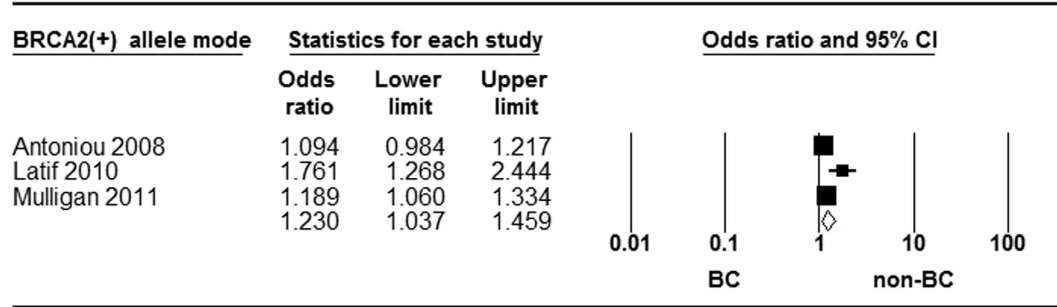


Figure 9. Forest plot of TOX3 rs3803662 polymorphism and breast cancer risk stratified by BRCA2 mutation. Fix -effect model was used for the analysis (allele contrast model T vs. C).

Variables	N	TT versus CC		CT versus CC		TT+CT versus CC		TT versus CT+CC		T versus C	
		OR (95% CI)	P _H	OR (95% CI)	P _H	OR (95% CI)	P _H	OR (95% CI)	P _H	OR (95% CI)	P _H
Total	43	1.311 (1.221–1.407)	<0.001	1.151 (1.103–1.201)	<0.001	1.160 (1.088–1.237)	<0.001	1.200 (1.140–1.263)	<0.001	1.145 (1.106–1.186)	<0.001
Ethnicity											
Asian	13	1.245 (1.076–1.440)	<0.001	1.121 (1.013–1.241)	<0.001	1.162 (1.030–1.311)	<0.001	1.133 (1.049–1.223)	<0.001	1.112 (1.034–1.195)	<0.001
Caucasian	19	1.483 (1.371–1.604)	0.029	1.212 (1.153–1.275)	0.005	1.259 (1.194–1.327)	<0.001	1.347 (1.280–1.418)	0.332	1.220 (1.171–1.271)	<0.001
African	6	0.929 (0.837–1.032)	0.240	0.961 (0.841–1.099)	0.044	0.989 (0.877–1.116)	0.048	0.952 (0.878–1.032)	0.52	0.962 (0.914–1.012)	0.296
Mixed	5	1.440 (1.288–1.611)	0.017	1.200 (1.109–1.298)	0.004	1.084 (0.927–1.268)	<0.001	1.358 (1.294–1.425)	0.265	1.202 (1.126–1.282)	<0.001
ER (+)	4	1.493 (1.391–1.603)	0.696	1.241 (1.188–1.297)	0.468	1.312 (1.262–1.364)	0.284	1.410 (1.321–1.505)	0.594	1.225 (1.189–1.262)	0.524
ER (-)	4	1.104 (0.885–1.376)	0.073	1.134 (1.066–1.206)	0.491	1.135 (1.008–1.278)	0.079	1.282 (1.163–1.414)	0.327	1.118 (1.064–1.175)	0.519
ER (+) vs. ER (-)	5	1.382 (0.998–1.915)	0.016	1.073 (1.002–1.149)	0.582	1.086 (1.019–1.157)	0.668	1.093 (0.986–1.212)	0.883	1.067 (1.016–1.121)	0.468
BRCA1	3	1.249 (1.087–1.436)	0.811	1.107 (1.022–1.198)	0.933	1.130 (1.048–1.219)	0.888	1.194 (1.044–1.365)	0.833	1.113 (1.050–1.181)	0.822
BRCA2	3	1.102 (0.921–1.319)	0.911	1.276 (1.018–1.599)	0.030	1.310 (1.039–1.650)	0.017	1.207 (1.018–1.432)	0.458	1.230 (1.037–1.459)	0.023

Table 3. Stratified analysis of TOX3 rs3803662 polymorphism on breast cancer. N: Numbers of data sets; P_H: P-value of Q-test for heterogeneity test; P_H < 0.1 indicates that there is heterogeneity and random-effect model is used to calculate pooled OR_s and 95% CI. Otherwise, fixed-effect model is used.

Variables	N	GG versus AA		AG versus AA		GG+AG versus AA		GG versus AG+AA		G versus A	
		OR (95% CI)	P _H	OR (95% CI)	P _H	OR (95% CI)	P _H	OR (95% CI)	P _H	OR (95% CI)	P _H
Total	15	1.072 (0.975–1.180)	0.005	1.033 (0.958–1.113)	0.020	1.049 (0.983–1.121)	0.047	1.061 (0.996–1.131)	0.069	1.033 (0.985–1.083)	0.003
Ethnicity											
Asian	7	1.006 (0.927–1.092)	0.885	1.002 (0.936–1.073)	0.663	1.006 (0.941–1.075)	0.696	1.007 (0.945–1.072)	0.802	1.000 (0.956–1.046)	0.843
Caucasian	6	1.264 (1.143–1.398)	0.201	1.156 (1.062–1.258)	0.281	1.163 (1.078–1.256)	0.147	1.181 (1.089–1.280)	0.273	1.125 (1.070–1.183)	0.190
African	2	0.854 (0.720–1.012)	0.117	0.863 (0.658–1.132)	0.062	0.931 (0.803–1.079)	0.355	0.926 (0.794–1.080)	0.371	0.928 (0.853–1.009)	0.145

Table 4. Stratified analysis of TOX3 rs12443621 polymorphism on breast cancer. N: Numbers of data sets; P_H: P-value of Q-test for heterogeneity test; P_H < 0.1 indicates that there is heterogeneity and random-effect model is used to calculate pooled OR_s and 95% CI. Otherwise, fixed-effect model is used.

rs3803662 and triple-negative tumors^{25,49}. Taken together, these studies indicated that there were some-how connections between the three SNPs in TOX3 gene and pathological subtype of breast tumor among different populations. And in another aspect, these studies provided further support for the hypothesis that different subtypes stem from diverse etiological pathways. Additionally, rs3803662 SNP in BRCA1 and BRCA2 mutation carriers was significantly associated with the increased risk of breast cancer in our analysis, which was in consistent with previous studies^{26,35}. While Latif *et al.* confirmed that the genetic variant was only associated with breast tumor in BRCA2 mutation carriers³⁰. Therefore, it's necessary to further elucidate the relevance of rs3803662 to breast cancer risk with BRCA1 and BRCA2 mutation.

Variables	N	TT versus CC		CT versus CC		TT + CT versus CC		TT versus CT + CC		T versus C	
		OR (95% CI)	P _H	OR (95% CI)	P _H	OR (95% CI)	P _H	OR (95% CI)	P _H	OR (95% CI)	P _H
Total	15	1.304 (1.210–1.405)	0.378	1.125 (1.076–1.175)	0.647	1.159 (1.112–1.208)	0.689	1.198 (1.121–1.280)	0.420	1.135 (1.099–1.171)	0.388
Ethnicity											
Asian	6	1.370 (1.185–1.584)	0.886	1.141 (1.075–1.211)	0.812	1.164 (1.101–1.231)	0.917	1.200 (1.041–1.383)	0.708	1.148 (1.093–1.206)	0.786
Caucasian	7	1.29 (1.181–1.425)	0.107	1.128 (1.046–1.215)	0.281	1.199 (1.116–1.288)	0.557	1.226 (1.082–1.389)	0.079	1.138 (1.086–1.191)	0.137
African	2	1.186 (0.941–1.494)	0.189	1.030 (0.896–1.183)	0.917	1.030 (0.913–1.161)	0.996	1.145 (0.970–1.351)	0.407	1.066 (0.964–1.179)	0.318

Table 5. Stratified analysis of TOX3 rs8051542 polymorphism on breast cancer. N: Numbers of data sets; P_H: P-value of Q-test for heterogeneity test; P_H < 0.1 indicates that there is heterogeneity and random-effect model is used to calculate pooled OR_s and 95% CI. Otherwise, fixed-effect model is used.

Despite the advantage of large sample size and stratified analyses, the meta-analysis had several limitations that should be taken into account. First, there was extreme heterogeneity for the outcomes of the association between rs3803662 polymorphism and breast cancer risk. Although we reduced the degree of heterogeneity by stratified analyses based on ethnicity, other sources of heterogeneity were not verified, such as different genotyping methods or tumor types. Second, the sample size of African populations (5462 cases and 7155 controls) was relatively small. Therefore there was insufficient statistical power to demonstrate the associations between the 3SNPs and breast cancer risk in Africans. Third, the criterion of control groups was not uniformly defined. The design of eligible studies was based on population or/and hospital patients, thus there were potential risks of breast cancer in control groups. Fourth, the mixtures of crude and adjusted publish estimates, rather than incorporation of adjusted ORs, were used in the meta-analysis. Because of the lack of some individual data, we were unable to adjust effect size with possible confounders related with lifestyle risk factors, such as age, obesity, smoking, alcohol consumption and menopausal status. Furthermore, we were unable to examine the interaction between genetic variables and environment. In recent years, some studies for gene-environment interactions showed that relative risks of breast cancer correlated with low-penetrance susceptibility variants (including rs3803662) didn't vary significantly with established environmental risk factors, such as reproductive history, menopausal status and body mass index^{51–53}. Nevertheless, more and more researches have elaborated combined effect of low-penetrance susceptibility loci with breast cancer risk. And the obviously elevated risk stemming from combining many low-penetrant risk alleles supports the polygenic inheritance model of breast cancer⁴⁴. Finally, owing to merely include English articles, there might be language bias on some level. Additionally, positive reports are tended to be published, which might make certain bias.

In conclusion, this meta-analysis indicated that there were different associations between the 3SNPs in TOX3 gene and breast cancer risk in different ethnic groups or subtype tumor. The 3SNPs were associated with the increased risk of breast cancer in Caucasians, while weren't correlated in Africans. Additionally, rs3803662 and rs8051542 were risk factors for breast cancer in Asians. Furthermore, there were stronger associations between rs3803662 polymorphism and breast cancer risk in ER⁺ subtype than ER⁻ tumors. Increased risk of breast cancer associated with rs3803662 was confirmed in BRCA1/BRCA2 mutation carriers. However, studies with larger sample size, which use uniform genotyping methods and criterion of control groups, have sufficiently corresponding individual data and consider the interactions of gene-gene and gene-environment will be needed to verify our results for TOX3 rs3803662, rs12443621 and rs8051542 as predisposition markers to breast cancer in clinical application.

References

- Jemal, A. *et al.* Global cancer statistics. *CA Cancer J Clin* **61**, 69–90 (2011).
- Gabai-Kapara, E. *et al.* Population-based screening for breast and ovarian cancer risk due to BRCA1 and BRCA2. *Proc Natl Acad Sci USA* **111**, 14205–14210 (2014).
- Tavtigian, S. V. *et al.* Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. *Am J Hum Genet* **85**, 427–446 (2009).
- O'Flaherty, E. & Kaye, J. TOX defines a conserved subfamily of HMG-box proteins. *BMC Genomics* **4**, 13–22 (2003).
- Yuan, S. H., Qiu, Z. & Ghosh, A. TOX3 regulates calcium-dependent transcription in neurons. *Proc Natl Acad Sci USA* **106**, 2909–2914 (2009).
- Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
- Stacey, S. N. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865–869 (2007).
- He, X., Yao, G., Li, F., Li, M. & Yang, X. Risk-association of five SNPs in TOX3/LOC643714 with breast cancer in southern China. *Int J Mol Sci* **15**, 2130–2141 (2014).
- Elematore, I. *et al.* Association of genetic variants at TOX3, 2q35 and 8q24 with the risk of familial and early-onset breast cancer in a South-American population. *Mol Biol Rep* **41**, 3715–3722 (2014).
- Low, S. K. *et al.* Genome-wide association study of breast cancer in the Japanese population. *PLoS One* **8**, e76463 (2013).
- Slattery, M. L. *et al.* Replication of five GWAS-identified loci and breast cancer risk among Hispanic and non-Hispanic white women living in the Southwestern United States. *Breast Cancer Res Treat* **129**, 531–539 (2011).
- Udler, M. S. *et al.* Fine scale mapping of the breast cancer 16q12 locus. *Hum Mol Genet* **19**, 2507–2515 (2010).

13. Ruiz-Narváez, E. A. *et al.* Polymorphisms in the *TOX3/LOC643714* locus and risk of breast cancer in African-American women. *Cancer Epidemiol Biomarkers Prev* **19**, 1320–1327 (2010).
14. Jiang, Y. *et al.* Risk of genome-wide association study newly identified genetic variants for breast cancer in Chinese women of Heilongjiang Province. *Breast Cancer Res Treat* **128**, 251–257 (2011).
15. Li, L. *et al.* TNRC9/LOC643714 polymorphisms are not associated with breast cancer risk in Chinese women. *Eur J Cancer Prev* **18**, 285–290 (2009).
16. Liang, J. *et al.* Genetic variants in trinucleotide repeat containing 9 (TNRC9) are associated with risk of estrogen receptor positive breast cancer in a Chinese population. *Breast Cancer Res Treat* **124**, 237–241 (2010).
17. Shan, J. *et al.* Genome-Wide Association Studies (GWAS) breast cancer susceptibility loci in Arabs: susceptibility and prognostic in Tunisians. *Breast Cancer Res Treat* **135**, 715–724 (2012).
18. Chen, M. B. *et al.* Association between polymorphisms of trinucleotide repeat containing 9 gene and breast cancer risk: evidence from 62,005 subjects. *Breast Cancer Res Treat* **126**, 177–183 (2011).
19. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* **22**, 719–748 (1959).
20. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin Trials* **7**, 177–188 (1986).
21. Long, J. *et al.* Evaluating genome-wide association study-identified breast cancer risk variants in African-American women. *PLoS One* **8**, e58350 (2013).
22. Kim, H. C. *et al.* A genome-wide association study identifies a breast cancer risk variant in *ERBB4* at 2q34: results from the Seoul Breast Cancer Study. *Breast Cancer Res* **14**, R56 (2012).
23. Huo, D. *et al.* Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis* **33**, 835–840 (2012).
24. Chan, M. *et al.* Association of common genetic variants with breast cancer risk and clinicopathological characteristics in a Chinese population. *Breast Cancer Res Treat* **136**, 209–220 (2012).
25. Stevens, K. N. *et al.* Common breast cancer susceptibility loci are associated with triple-negative breast cancer. *Cancer Res* **71**, 6240–6249 (2011).
26. Mulligan, A. M. *et al.* Common breast cancer susceptibility alleles are associated with tumour subtypes in *BRCA1* and *BRCA2* mutation carriers: results from the Consortium of Investigators of Modifiers of *BRCA1/2*. *Breast Cancer Res* **13**, R110 (2011).
27. Han, W. *et al.* Common genetic variants associated with breast cancer in Korean women and differential susceptibility according to intrinsic subtype. *Cancer Epidemiol Biomarkers Prev* **20**, 793–798 (2011).
28. Long, J. *et al.* Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol Biomarkers Prev* **19**, 2357–2365 (2010).
29. Long, J. *et al.* Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet* **6**, e1001002 (2010).
30. Latif, A. *et al.* Breast cancer susceptibility variants alter risks in familial disease. *J Med Genet* **47**, 126–131 (2010).
31. Barnholtz-Sloan, J. S. *et al.* *FGFR2* and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. *Carcinogenesis* **31**, 1417–1423 (2010).
32. Gorodnova, T. V. *et al.* Distribution of *FGFR2*, *TNRC9*, *MAP3K1*, *LSP1*, and 8q24 alleles in genetically enriched breast cancer patients versus elderly tumor-free women. *Cancer Genet Cytoenet* **199**, 69–72 (2010).
33. Hemminki, K. *et al.* Low-risk variants *FGFR2*, *TNRC9* and *LSP1* in German familial breast cancer patients. *In t J Cancer* **126**, 2858–2862 (2010).
34. Mcinerney, N. *et al.* Low penetrance breast cancer predisposition SNPs are site specific. *Breast Cancer Res Treat* **117**, 151–159 (2009).
35. Antoniou, A. C. *et al.* Common breast cancer-predisposition alleles are associated with breast cancer risk in *BRCA1* and *BRCA2* mutation carriers. *Am J Hum Genet* **82**, 937–948 (2008).
36. Tapper, W. *et al.* The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. *Breast cancer Res* **10**, R108 (2008).
37. Campa, D. *et al.* Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *J Natl Cancer Inst* **103**, 1252–1263 (2011).
38. Garcia-Closas, M. *et al.* Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* **4**, e1000054 (2008).
39. Barzan, D. *et al.* Comparison of genetic variation of breast cancer susceptibility genes in Chinese and German populations. *Eur J Hum Genet* **21**, 1286–1292 (2013).
40. Rinella, E. S. *et al.* Genetic variants associated with breast cancer risk for Ashkenazi Jewish women with strong family histories but no identifiable *BRCA1/2* mutation. *Hum Genet* **132**, 523–536 (2013).
41. Zheng, W. *et al.* Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev* **18**, 2761–2764 (2009).
42. Butt, S. *et al.* Genetic predisposition, parity, age at first childbirth and risk for breast cancer. *BMC Res Notes* **5**, 414 (2012).
43. Mizoo, T. *et al.* Effects of lifestyle and single nucleotide polymorphisms on breast cancer risk: a case-control study in Japanese women. *BMC Cancer* **13**, 565 (2013).
44. Harlid, S. *et al.* Combined effect of low-penetrant SNPs on breast cancer risk. *Br J Cancer* **106**, 389–396 (2012).
45. Zheng, W. *et al.* Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. *J Natl Cancer Inst* **102**, 972–981 (2010).
46. Tamimi, R. M. *et al.* Birth weight, breast cancer susceptibility loci, and breast cancer risk. *Cancer Causes Control* **21**, 689–696 (2010).
47. O'Flaherty, E. & Kaye, J. *TOX* defines a conserved subfamily of HMG-box proteins. *BMC Genomics* **4**, 13 (2003).
48. Smid, M. *et al.* Genes associated with breast cancer metastatic to bone. *J Clin Oncol* **24**, 2261–2267 (2006).
49. Broeks, A. *et al.* Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum Mol Genet* **20**, 3289–3303 (2011).
50. O'Brien, K. M. *et al.* Breast cancer subtypes and previously established genetic risk factors: a bayesian approach. *Cancer Epidemiol Biomarkers Prev* **23**, 84–97 (2014).
51. Milne, R. L. *et al.* Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the breast cancer association consortium: a combined case-control study. *Breast Cancer Res* **12**, R110 (2010).
52. Travis, R. C. *et al.* Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. *Lancet* **375**, 2143–2151 (2010).
53. Warren Andersen, S. *et al.* The associations between a polygenic score, reproductive and menstrual risk factors and breast cancer risk. *Breast Cancer Res Treat* **140**, 427–434 (2013).

Acknowledgements

The Project-sponsored by SRF for ROCS, SEM. This work was supported by grants from National Natural Science Foundation of China (30873044 and 81272372).

Author Contributions

L.Z. and X.L. conceived and designed the experiments; L.Z. analyzed the data, L.Z. and X.L. wrote the manuscript. All authors reviewed and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhang, L. and Long, X. Association of three SNPs in *TOX3* and breast cancer risk: Evidence from 97275 cases and 128686 controls. *Sci. Rep.* **5**, 12773; doi: 10.1038/srep12773 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>