

# A functional variant rs4442975 modulating FOXA1-binding affinity does not influence the risk or progression of breast cancer in Chinese Han population

Yi Zhang<sup>1,2</sup>, Yan Li<sup>2</sup>, Shaojie Jiang<sup>3</sup>, Wei Chen<sup>1,4</sup>, Jiao Lou<sup>1</sup>, Juntao Ke<sup>1</sup>, Jiaoyuan Li<sup>1</sup>, Ying Zhu<sup>1</sup>, Yajie Gong<sup>1</sup>, Yang Yang<sup>1</sup>, Jianbo Tian<sup>1</sup>, Xiating Peng<sup>1</sup>, Danyi Zou<sup>1</sup>, Jing Gong<sup>1</sup>, Jiang Chang<sup>1</sup>, Xiaoping Miao<sup>1</sup>, Rong Zhong<sup>1</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics and State Key Laboratory of Environment Health (Incubation), Ministry of Education Key Laboratory of Environment & Health, Ministry of Environmental Protection Key Laboratory of Environment and Health (Wuhan), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, PR China

<sup>2</sup>Department of Preventive Medicine, School of Public Health, Zunyi Medical University, Zunyi, Guizhou, China

<sup>3</sup>Department of Gastroenterology, Jing Zhou Central Hospital, Jingzhou, China

<sup>4</sup>Zhuhai Center for Chronic Disease Control, Zhuhai, China

**Correspondence to:** Rong Zhong, **email:** rongzhong91@gmail.com

Xiaoping Miao, **email:** miaoxp@mail.hust.edu.cn

**Keywords:** breast cancer, FOXA1, genetic variants, susceptibility, case-control study

**Received:** August 20, 2016

**Accepted:** October 19, 2016

**Published:** November 07, 2016

## ABSTRACT

The DNA-binding protein FOXA1 has been shown to regulate nearly all estrogen receptor-chromatin interactions, thereby influencing target gene expression levels in breast cancer (BC) cells. Recently, the rs4442975 T-allele, which disrupts the recruitment of FOXA1 and interacts with the *IGFBP5* promoter, was associated to BC susceptibility in a European population. We conducted a hospital-based case-control study that included 1227 cases and 1285 controls to explore the potential association between rs4442975 and BC risk in Chinese Han population, and the effect of this SNP on BC progression was also observed in cases. No significant associations between rs4442975 and BC risk were observed under any genetic models, with an odds ratio of 0.96 (95% confidence interval = 0.81-1.15) under the additive model. When stratified based on estrogen or progesterone receptor expression, smoking or drinking habits, or menopausal status, similar negative associations were observed for all subgroups. No significant associations were observed between rs4442975 and traditional progression factors such as tumor size, nodal status, distant metastasis, or TNM staging. These results reveal that rs4442975 may not confer a risk of BC occurrence or progression in the Chinese Han population, which indicates a distinct association related to genetic heterogeneity across ethnic populations.

## INTRODUCTION

Breast cancer (BC) remains a global public health issue and the most common cancer in females worldwide. An estimated 1.67 million newly diagnosed BC cases and 521,907 deaths occurred in 2012 according to the database of the International Agency for Research on Cancer (IARC) [1], accounting for 25.2% of total cancer cases and 14.7% of cancer deaths among females. In China, the incidence of breast cancer has continuously increased over the past few decades. It was estimated that the morbidity

and mortality of BC in Chinese females in 2015 were 26.86/100,000 and 6.95/100,000, respectively [2]. In order to effectively reduce the morbidity and mortality, it is urgent to clarify the mechanism of BC pathogenesis.

Genome-wide association studies (GWASs) have revealed multiple loci and single nucleotide polymorphisms (SNPs) associated with BC risk [3, 4]. The function of the majority of BC risk-associated SNPs, located in non-coding regions, remains unclear, complicating the search for potential mechanisms underlying those associations [5, 6]. Regulatory mechanisms of risk-associated SNPs identified

by GWAS have, to date, been gradually discovered. As a consequence, fine mapping of the non-coding variants associated with BC risk has been the main point of systemic post-GWAS functional characterization. Remarkably, BC risk-associated SNPs identified by GWAS were enriched in the binding sites for the transcription factors FOXA1 and estrogen receptor 1 (ESR1) [7], which suggested FOXA1 contributes to BC development by regulatory mechanisms related to estrogen receptor (ER) signaling [8, 9].

FOXA1 is a pioneer factor that directly binds and modulates compacted chromatin [10], which in turn facilitates or inhibits the recruitment of other transcription factors, particularly steroid hormone receptors including ER [11–13]. Direct evidence suggests that FOXA1 regulates almost all estrogen receptor-chromatin interactions, thus influencing target gene expression levels in breast cancer cells [14, 15]. Silencing of FOXA1 leads to the global inhibition of ER binding and transcriptional activity [15, 16]. Previous studies showed that the deregulation of FOXA1 played a potentially critical role in the carcinogenesis of breast cancer by disturbing ER binding [15]. Inactivating mutations of FOXA1 have been observed in breast cancer by exome sequencing study [17]. Additionally, differential ER binding is associated with the prognosis of BC patients [18], and co-recruitment of FOXA1 appears to be related to reprogramming of ER binding. Several reports showed that FOXA1 expression levels significantly correlated with better breast cancer-specific survival [19–24]. Therefore, it has been proposed that genetic variants in FOXA1 binding sites might affect ER binding to FOXA by modulating the affinity of chromatin, thereby regulating target gene expression levels and eventually contributing to the risk of breast cancer initiation and progression.

A recent GWAS study [25] from Ghousaini M et al. demonstrated that the variant rs4442975, which is strongly correlated with rs13387042 [26, 27] ( $r^2 > 0.8$ ), and disrupts the recruitment of FOXA1 [25, 28], was associated with elevated BC risk in Europeans due to a resulting reduction in *IGFBP5* expression. However, the association between rs4442975 and BC risk in Chinese Han population has yet to be reported.

Given the genetic heterogeneity across populations, cell types, and tissue of breast cancer, we carried out a hospital-based case-control study in a Chinese Han population to explore the role of rs4442975 in susceptibility to BC, and the effect of this SNP on BC progression was also observed in cases.

## RESULTS

### Subject characteristics

The characteristics of study participants are presented in Table 1. The controls and cases were well matched in the distribution of age and menopausal status, with  $P$ -values of 0.483 ( $t=0.701$ ) and 0.619 ( $\chi^2=0.247$ ),

respectively. Additionally, there were no differences in drinking or smoking habits between control and case group. Among BC patients, 760 (61.94%) cases were ER positive and 463 (37.73%) were ER negative, 4 (0.33%) cases were ER unknown. 683 (55.66%) and 537 (43.77 %) cases were classified as progesterone receptor (PR) positive and PR negative, respectively. Additionally, there were 615 (50.12%) cases with lymph node metastasis.

### The association between rs4442975 and BC risk

The genotypes of rs4442975 in controls conformed to the Hardy-Weinberg equilibrium ( $P = 0.829$ ), and the minor allele frequency was consistent with the 1000 Genomes Project. The genotype distribution of rs4442975 and its associations with BC risk are presented in Table 2. Compared with individuals carrying the TT genotype, there was no increased risk of BC for individuals with TG and GG genotypes, with an OR of 0.98 (95% CI: 0.81-1.19) after adjusting for drinking, smoking, menopausal status, and age. Meanwhile, no significant association between rs4442975 and BC risk was found under the heterozygote or homozygote models (OR=0.96, 95% CI: 0.78-1.17 and OR=0.97, 95% CI: 0.50-1.88 respectively). A similar negative result was also observed in the additive model.

Stratified analysis was performed according to ER and PR expression, smoking and drinking habits, and menopausal status. No significant association between rs4442975 and BC risk was found in ER+, ER-, PR+, PR-, premenopausal, postmenopausal, non-smoker, or nondrinker subgroups (Supplementary Table S1, Supplementary Table S2, and Supplementary Table S3).

### The association between rs4442975 and progression of BC

We further evaluated the association between rs4442975 and clinicopathological features such as tumor size, lymph node involvement, distant metastasis, and TNM stage (Table 3), which are typical indicators of BC progression. No associations were found between the SNP and any of the studied prognostic factors under the additive model: tumor size ( $P = 0.823$ ), lymph node involvement ( $P = 0.933$ ), distant metastasis ( $P = 0.154$ ), and advanced TNM stage ( $P = 0.765$ ). A similar negative result was observed in the dominant model.

## DISCUSSION

In the present study, we conducted a hospital-based case-control study in a Chinese Han population to explore the potential association between the FOXA1-binding functional variant rs4442975 and BC risk. Our results demonstrated that rs4442975 was not associated with

**Table 1: Characteristics of the participants in the case-control study**

| Variables                    | Case (1227) No(%) | Control (1285) No(%) | P                  |
|------------------------------|-------------------|----------------------|--------------------|
| Age (Mean ± SD)              | 48.91±9.69        | 48.64±9.50           | 0.483 <sup>a</sup> |
| <b>Smoking</b>               |                   |                      | 0.581 <sup>b</sup> |
| Yes                          | 9(0.73)           | 12(0.90)             |                    |
| No                           | 1218(99.27)       | 1273(99.10)          |                    |
| <b>Alcohol use</b>           |                   |                      | 0.759 <sup>b</sup> |
| Yes                          | 20(1.63)          | 19(1.50)             |                    |
| No                           | 1207(98.37)       | 1266(98.50)          |                    |
| <b>Menopausal status</b>     |                   |                      | 0.619 <sup>b</sup> |
| Premenopausal                | 690(56.23)        | 743(57.80)           |                    |
| Postmenopausal               | 524(42.71)        | 542(42.20)           |                    |
| <b>Estrogen receptor</b>     |                   |                      |                    |
| Positive                     | 760(61.94)        |                      |                    |
| Negative                     | 463(37.73)        |                      |                    |
| <b>Progesterone receptor</b> |                   |                      |                    |
| Positive                     | 683(55.66)        |                      |                    |
| Negative                     | 537(43.77)        |                      |                    |
| <b>Tumor size</b>            |                   |                      |                    |
| ≤ 5 cm                       | 859(70.01)        |                      |                    |
| > 5 cm                       | 95(7.74)          |                      |                    |
| <b>Lymph node metastasis</b> |                   |                      |                    |
| Yes                          | 615(50.12)        |                      |                    |
| No                           | 505(41.16)        |                      |                    |
| <b>Distant metastasis</b>    |                   |                      |                    |
| Yes                          | 96(7.82)          |                      |                    |
| No                           | 1052(85.74)       |                      |                    |
| <b>TNM stages</b>            |                   |                      |                    |
| TNM I-II                     | 775(63.16)        |                      |                    |
| TNM III-IV                   | 304(24.78)        |                      |                    |

<sup>a</sup>P value was calculated by *t* test; <sup>b</sup>P value was calculated by  $\chi^2$  test.

**Table 2: The associations between rs4442975 and BC risk in Chinese population**

| Variables        | Case(1227)<br>No. (%) | Control(1285)<br>No. (%) | MAF in Control | MAF in<br>CHB <sup>a</sup> | OR(95%CI); P <sup>b</sup> |
|------------------|-----------------------|--------------------------|----------------|----------------------------|---------------------------|
| <b>rs4442975</b> | 1201(97.88)           | 1270(98.83)              | 0.119          | 0.112                      |                           |
| TT               | 942(78.43)            | 985(77.56)               |                |                            | 1.00                      |
| TG               | 242(20.15)            | 266(20.94)               |                |                            | 0.96(0.78-1.17); 0.662    |
| GG               | 17(1.42)              | 19(1.50)                 |                |                            | 0.97(0.50-1.88); 0.928    |
| Dominant         |                       |                          |                |                            | 0.98(0.81-1.19); 0.812    |
| Additive         |                       |                          |                |                            | 0.96(0.81-1.15); 0.678    |

<sup>a</sup>MAF was downloaded from 1000 Genome Project data among Han Chinese in Beijing individuals;

<sup>b</sup>ORs and 95% CIs were calculated by unconditional logistic regression after adjusting for age, smoking, alcohol use, and menopausal status.

**Table 3: Association of rs4442975 genotypes with clinicopathologic parameters**

| Parameter                    | Parameter category No. (%) |                      | OR(95%CI); Pa          |
|------------------------------|----------------------------|----------------------|------------------------|
| <b>Tumor Sizes</b>           | <b>≤ 5 cm(859)</b>         | <b>&gt; 5 cm(95)</b> |                        |
| <b>rs4442975</b>             | 848(98.72)                 | 92(96.84)            |                        |
| TT                           | 673(79.36)                 | 71(77.17)            | 1.00                   |
| TG                           | 162(19.11)                 | 21(22.83)            | 1.31(0.77-2.21); 0.320 |
| GG                           | 13(1.53)                   | 0(0)                 | —                      |
| Dominant                     |                            |                      | 1.15(0.68-1.94); 0.601 |
| Additive                     |                            |                      | 1.06(0.66-1.70); 0.823 |
| <b>Lymph node metastasis</b> | <b>Negative(505)</b>       | <b>Positive(615)</b> |                        |
| <b>rs4442975</b>             | 493(97.62)                 | 608(98.86)           |                        |
| TT                           | 390(79.11)                 | 471(77.47)           | 1.00                   |
| TG                           | 93(18.86)                  | 130(21.38)           | 1.14(0.83-1.55); 0.423 |
| GG                           | 10(2.03)                   | 7(1.15)              | 0.56(0.21-1.50); 0.251 |
| Dominant                     |                            |                      | 1.07(0.80-1.45); 0.642 |
| Additive                     |                            |                      | 1.01(0.77-1.32); 0.933 |
| <b>Distant metastasis</b>    | <b>Negative(1052)</b>      | <b>Positive(96)</b>  |                        |
| <b>rs4442975</b>             | 1032(98.10)                | 92(95.83)            |                        |
| TT                           | 806(78.10)                 | 76(82.61)            | 1.00                   |
| TG                           | 209(20.25)                 | 16(17.39)            | 0.78(0.44-1.39); 0.400 |
| GG                           | 17(1.65)                   | 0(0)                 | —                      |
| Dominant                     |                            |                      | 0.69(0.39-1.24); 0.214 |
| Additive                     |                            |                      | 0.68(0.39-1.16); 0.154 |
| <b>TNM stages</b>            | <b>I-II(775)</b>           | <b>III-IV(304)</b>   |                        |
| <b>rs4442975</b>             | 758(97.81)                 | 297(97.70)           |                        |
| TT                           | 601(79.29)                 | 232(78.11)           | 1.00                   |
| TG                           | 143(18.86)                 | 64(21.55)            | 1.15(0.82-1.62); 0.419 |
| GG                           | 14(1.85)                   | 1(0.34)              | 0.16(0.02-1.24); 0.080 |
| Dominant                     |                            |                      | 1.08(0.77-1.51); 0.658 |
| Additive                     |                            |                      | 0.95(0.70-1.30); 0.765 |

\*P values were calculated using unconditional logistic regression after adjusting for age, smoking, alcohol use, menopausal status, ER and PR.

the risk of BC incidence or progression among Chinese population, and negative results were also observed in all of the subgroups stratified by ER, PR, smoking, drinking, and menopausal status.

Rs4442975 is located near a putative regulatory element and interacts with the *IGFBP5* promoter. It has been reported that the cancer-protective T allele of rs4442975 creates a stronger interaction with *IGFBP5* than the G allele, resulting in increased *IGFBP5* expression. Additionally, rs4442975 resides in the FOXA1 binding site with the T allele leading to increased FOXA1 binding [25, 28]. FOXA1 is a pioneer factor and plays an important role in BC risk by interrupting the FOXA1-binding ability which then influences the ER binding and transcription of its target genes [29]. However, no association between

rs4442975 and risk of BC incidence or progression was found in this Chinese population.

We propose the following explanation for the conflicting results between different ethnic groups. The differential role of rs4442975 in BC risk between European and Chinese women may be partly due to different genetic backgrounds among populations: MAF (G) was 0.112 in CHB versus 0.509 in EUR. Differences in linkage disequilibrium (LD) patterns across populations may result in the heterogeneity of effect size. So, the variant observed to associate with different effect sizes among different populations may not be the actual causal variant, whereas a variant with similar effect sizes can be causal. Rs4442975 was strongly correlated with rs13387042 ( $r^2=0.93$ ), which is located in 2q35 and

first identified to contribute to both ER positive and ER negative BC susceptibility in an Icelandic GWAS and then confirmed by a large European replication study [26, 27]. Similarly, rs13387042 was located in a gene desert region with unknown biological function. According to the “multiple enhancers variant” hypothesis [30], we cannot exclude situations where another SNP, yet to be detected, may be in the same LD block of rs4442975 and influence FOXA1 binding. Additionally, breast cancer is highly heterogeneous, both at the tissue and cell level, and risk-associated SNPs may act only on one specific cell type.

The study still had some limitations. Selection bias may exist due to the hospital-based case-control study design. The sample size was not very large to detect the modest effect of the SNP on the risk of BC. Unfortunately we do not have 5-years survival rate, so we could not estimate the effect of rs4442975 on the prognosis.

In conclusion, our study indicates that the variant rs4442975 did not confer increased risk of BC incidence or progression in this population of Chinese Han women. Furthermore, our results emphasize the distinct mechanisms behind genetic heterogeneity across populations, cell types and tissue heterogeneity of breast cancer. Although well designed and larger population-based case-control studies are required to confirm the role of this polymorphism in BC risk across different populations, it is critical to clarify whether the significant association between rs4442975 and BC risk in the European population is a result of other genetically linked causal variants.

## MATERIALS AND METHODS

### Subjects for the case-control study

A hospital-based case-control study of 1227 breast cancer patients and 1285 control subjects was performed to comprehensively examine the association between rs4442975 and the risk of BC in the Chinese population. All subjects in this study are female and unrelated ethnic Han Chinese. There were no age or histology restrictions. Patients with newly diagnosed, histopathologically confirmed, and previously untreated (by radiotherapy or chemotherapy) breast cancer were consecutively recruited between June 2010 and June 2014 at Union Hospital of Huazhong University of Science and Technology (HUST) of Wuhan, Central China. Patients with metastasized cancer from other organs or a history of other cancers were not included in the case group. Control subjects were randomly chosen from a pool of cancer-free people who participated in health examination during the same period at the same hospital. Controls were frequency-matched to the cases by age ( $\pm 5$  years). Tumor stage was evaluated according to the 2002 American Joint Committee on Cancer staging system. At recruitment, written informed

consent was obtained from each participant, and then personal characteristics and peripheral venous blood sample (2 ml) were collected from each participant. The definitions of smokers and drinkers have been described previously [31, 32]. Briefly, subjects who had smoked less than 1 cigarette per day and  $< 1$  year or never smoked were classified as non-smokers; otherwise, they were defined as smokers. Subjects who drank more than twice a week and continued for at least 1 year were defined as drinkers, while others were considered as nondrinkers. The study was approved by the institutional review boards of Tongji Medical College of Huazhong University of Science and Technology.

### Genotyping

Genomic DNA was isolated from the 2ml peripheral venous blood samples of all participants using Relax Gene Blood DNA System DP319-02 (Tiangen, Beijing, China) in accordance with the manufacturer’s recommended protocol. The genetic variant rs4442975 was genotyped by TaqMan SNP Genotyping Assay using the 7900HT Fast Real-Time PCR System (Applied Bio-systems, Foster City, CA). Genotyping was performed without knowledge of case/control status of the subjects. Approximately 5% of the random samples from cases and controls were genotyped twice, and the results were in 100 % concordance. All methods were carried out in accordance with the approved guidelines.

### Statistical analysis

The statistical power to detect the association between BC and the SNP was calculated by Power v3.0 [33–35]. The distribution difference of demographic characteristics was evaluated by the *t*-test or chi-square test between case and control group. The Hardy-Weinberg equilibrium (HWE) for the genotype frequencies was tested by the goodness-of-fit chi-square test in controls. The odds ratios (ORs) and their 95% confidence intervals (CIs) were calculated by unconditional multivariate logistic regression analysis to estimate the association between genotypes and BC risk after adjusted by smoking, drinking, menopause status, and age.

For rs4442975 with minor allele frequencies (MAF) of 0.119 in controls, the statistical power for our sample size to detect an OR of 1.50 is 0.939. All *P* values were two-sided and  $P < 0.05$  was considered statistically significant. All statistical analyses were performed using the SPSS software (V18.0).

## ACKNOWLEDGMENTS

The authors wish to thank all the study participants, research staff, and students who participated in this work, especially blood sample donors.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

## FINANCIAL SUPPORT

This work was supported by the National Science Foundation of China (NSFC-81402746) to R.Z., and National Key Research and Development Plan Program (2016YFC1302702, 2016YFC1302703).

## REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015; 65:87-108.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ, He J. Cancer statistics in China, 2015. *CA Cancer J Clin.* 2016; 66:115-132.
3. Michailidou K, Beesley J, Lindstrom S, Canisius S, Dennis J, Lush MJ, Maranian MJ, Bolla MK, Wang Q, Shah M, Perkins BJ, Czene K, Eriksson M, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nature genetics.* 2015; 47:373-380.
4. Fachal L, Dunning AM. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. *Curr Opin Genet Dev.* 2015; 30:32-41.
5. Manolio TA. Bringing genome-wide association findings into clinical use. *Nat Rev Genet.* 2013; 14:549-558.
6. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A.* 2009; 106:9362-9367.
7. Cowper-Salari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoutte J, Moore JH, Lupien M. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nature genetics.* 2012; 44:1191-1198.
8. Lam EW, Brosens JJ, Gomes AR, Koo CY. Forkhead box proteins: tuning forks for transcriptional harmony. *Nat Rev Cancer.* 2013; 13:482-495.
9. Long J, Cai Q, Shu XO, Qu S, Li C, Zheng Y, Gu K, Wang W, Xiang YB, Cheng J, Chen K, Zhang L, Zheng H, et al. Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet.* 2010; 6:e1001002.
10. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell.* 2002; 9:279-289.
11. Lupien M, Eeckhoutte J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell.* 2008; 132:958-970.
12. Zhang L, Rubins NE, Ahima RS, Greenbaum LE, Kaestner KH. Foxa2 integrates the transcriptional response of the hepatocyte to fasting. *Cell Metab.* 2005; 2:141-148.
13. Zhang Y, Zhang D, Li Q, Liang J, Sun L, Yi X, Chen Z, Yan R, Xie G, Li W, Liu S, Xu B, Li L, et al. Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering. *Nature genetics.* 2016.
14. Laganier J, Deblois G, Lefebvre C, Bataille AR, Robert F, Giguere V. From the Cover: Location analysis of estrogen receptor alpha target promoters reveals that FOXA1 defines a domain of the estrogen response. *Proc Natl Acad Sci U S A.* 2005; 102:11651-11656.
15. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoutte J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell.* 2005; 122:33-43.
16. Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature genetics.* 2011; 43:27-33.
17. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012; 490:61-70.
18. Watters RJ, Benos PV, Oesterreich S. To bind or not to bind—FoxA1 determines estrogen receptor action in breast cancer progression. *Breast cancer research.* 2012; 14:312.
19. Mehta RJ, Jain RK, Leung S, Choo J, Nielsen T, Huntsman D, Nakshatri H, Badve S. FOXA1 is an independent prognostic marker for ER-positive breast cancer. *Breast cancer research and treatment.* 2012; 131:881-890.
20. Albergaria A, Paredes J, Sousa B, Milanezi F, Carneiro V, Bastos J, Costa S, Vieira D, Lopes N, Lam EW, Lunet N, Schmitt F. Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. *Breast cancer research.* 2009; 11:R40.
21. Xu C, Wei Q, Guo J, Zhou JC, Mei J, Jiang ZN, Shen JG, Wang LB. FOXA1 Expression Significantly Predicts Response to Chemotherapy in Estrogen Receptor-Positive Breast Cancer Patients. *Annals of surgical oncology.* 2015; 22:2034-2039.
22. Horimoto Y, Arakawa A, Harada-Shoji N, Sonoue H, Yoshida Y, Himuro T, Igari F, Tokuda E, Mamat O, Tanabe M, Hino O, Saito M. Low FOXA1 expression predicts good response to neo-adjuvant chemotherapy resulting in good outcomes for luminal HER2-negative breast cancer cases. *British journal of cancer.* 2015; 112:345-351.
23. Habashy HO, Powe DG, Rakha EA, Ball G, Paish C, Gee J, Nicholson RI, Ellis IO. Forkhead-box A1 (FOXA1) expression in breast cancer and its prognostic significance. *European journal of cancer.* 2008; 44:1541-1551.

24. Badve S, Turbin D, Thorat MA, Morimiya A, Nielsen TO, Perou CM, Dunn S, Huntsman DG, Nakshatri H. FOXA1 expression in breast cancer—correlation with luminal subtype A and survival. *Clinical cancer research*. 2007; 13:4415-4421.
25. Ghoussaini M, Edwards SL, Michailidou K, Nord S, Cowper-Sal Lari R, Desai K, Kar S, Hillman KM, Kaufmann S, Glubb DM, Beesley J, Dennis J, Bolla MK, et al. Evidence that breast cancer risk at the 2q35 locus is mediated through IGFBP5 regulation. *Nat Commun*. 2014; 4:4999.
26. Milne RL, Benitez J, Nevanlinna H, Heikkinen T, Aittomaki K, Blomqvist C, Arias JI, Zamora MP, Burwinkel B, Bartram CR, Meindl A, Schmutzler RK, Cox A, et al. Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042. *J Natl Cancer Inst*. 2009; 101:1012-1018.
27. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nature genetics*. 2007; 39:865-869.
28. Dryden NH, Broome LR, Dudbridge F, Johnson N, Orr N, Schoenfelder S, Nagano T, Andrews S, Wingett S, Kozarewa I, Assiotis I, Fenwick K, Maguire SL, et al. Unbiased analysis of potential targets of breast cancer susceptibility loci by Capture Hi-C. *Genome Res*. 2014; 24:1854-1868.
29. Jozwik KM, Carroll JS. Pioneer factors in hormone-dependent cancers. *Nat Rev Cancer*. 2012; 12:381-385.
30. Corradin O, Saiakhova A, Akhtar-Zaidi B, Myeroff L, Willis J, Cowper-Sal Lari R, Lupien M, Markowitz S, Scacheri PC. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res*. 2014; 24:1-13.
31. Gong J, Tian J, Lou J, Ke J, Li L, Li J, Yang Y, Gong Y, Zhu Y, Zhang Y, Zhong R, Chang J, Miao X. A functional polymorphism in lnc-LAMC2-1:1 confers risk of colorectal cancer by affecting miRNA binding. *Carcinogenesis*. 2016; 37:443-451.
32. Zhong R, Liu L, Zou L, Sheng W, Zhu B, Xiang H, Chen W, Chen J, Rui R, Zheng X, Yin J, Duan S, Yang B, et al. Genetic variations in the TGFbetasignaling pathway smoking and risk of colorectal cancer in a Chinese population. *Carcinogenesis*. 2013; 34:936-942.
33. Garcia-Closas M, Lubin JH. Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches. *American journal of epidemiology*. 1999; 149:689-692.
34. Foppa I, Spiegelman D. Power and sample size calculations for case-control studies of gene-environment interactions with a polytomous exposure variable. *American journal of epidemiology*. 1997; 146:596-604.
35. Lubin JH, Gail MH. On power and sample size for studying features of the relative odds of disease. *American journal of epidemiology*. 1990; 131:552-566.