A genome-wide association study identifies WT1 variant with better response to 5-fluorouracil, pirarubicin and cyclophosphamide neoadjuvant chemotherapy in breast cancer patients

Lina Wu²,*, Lu Yao¹,* , Hong Zhang² , Tao Ouyang¹ , Jinfeng Li¹ , Tianfeng Wang¹ , Zhaoqing Fan¹ , Tie Fan¹ , Benyao Lin¹ , C. Cameron Yin³ , Yuntao Xie¹

¹Breast Center, Key Laboratory of Carcinogenesis and Translational Research, Ministry of Education, Beijing Cancer Hospital & Institute, Peking University Cancer Hospital, Beijing, P. R. China
²Central Laboratory, Key Laboratory of Carcinogenesis and Translational Research, Ministry of Education, Beijing Cancer Hospital & Institute, Peking University Cancer Hospital, Beijing, P. R. China
³Department of Hematopathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

*These authors have contributed equally to this work

Correspondence to: Yuntao Xie, e-mail: zlxyt2@bjmu.edu.cn

Keywords: WT1, rs1799937, breast cancer, anthracycline, neoadjuvant chemotherapy

ABSTRACT

Breast cancer is believed to result from the interplay of genetic and non-genetic risk factors, and individual genetic variation may influence the efficacy of chemotherapy. Here we conducted a genome-wide association study to identify single nucleotide polymorphisms (SNPs) associated with response to anthracycline- and taxane-based neoadjuvant chemotherapy in breast cancer patients. In the discovery stage, we divided 92 patients who received anthracycline-based neoadjuvant chemotherapy into 2 groups according to pathologic response and performed a genome-wide study using Affymetrix SNP6.0 genechip. Of 389,795 SNPs associated with pathologic complete response (pCR), we identified 2 SNPs, rs6044100 and rs1799937, that were significantly associated with pCR after neoadjuvant chemotherapy. In the validation stage, genotype analysis of samples from an independent cohort of 401 patients who received anthracycline-based neoadjuvant regimens and 467 patients who received taxane-based regimens was performed using sequencing analysis. We found that only SNP rs1799937, located in the WT1 gene, was associated with pCR after anthracycline-based neoadjuvant therapy (AA vs GG; odds ratio [OR], 2.81; 95% confidence interval [CI], 1.13–6.98; P < 0.05) but not after taxane-based neoadjuvant therapy (AA vs GG; OR, 0.85; 95% CI, 0.36–2.04; P = 0.72). These results suggest that WT1 may be a potential target of anthracycline-based neoadjuvant therapy for breast cancer.

INTRODUCTION

Breast cancer is the most common malignant neoplasm among women in the world. It is believed to result from a complex interplay of genetic and non-genetic risk factors. Linkage and family-based studies have shown that germline mutations of high- and moderate-penetrance genes such as BRCA1/2, PTEN, ATM, CHEK2, TP53, PALB2, and BRIP1 account for 20%–25% of the genetic component of breast cancer [1]. Genetic association studies, such as genome-wide association studies (GWASs), have identified low-penetrance common variants (predominantly single nucleotide polymorphisms [SNPs]) associated with breast cancer risk. Since the first 3 GWASs that identified SNPs associated with breast cancer susceptibility were conducted in early 2007 [2–4], over 80 loci associated with increased breast cancer risk have been found [5–16]. These studies have suggested that high-
and moderate-penetrance genes and GWAS-identified low-penetrance common variants together may explain about 50% of the familial risk of breast cancer [17].

Accumulating evidence has shown that certain genetic variations are associated with response to chemotherapy in patients with lung cancer [18, 19]. However, genetic variants associated with tumor response to specific chemotherapy regimens in breast cancer have rarely been studied [20–22]. Neoadjuvant chemotherapy is increasingly used for operable primary breast cancer because it yields survival rate at least equal to those of adjuvant chemotherapy [23–25]. Some of the most effective cytotoxic agents against breast cancer are the anthracyclines, which have formed the backbone of most adjuvant and neoadjuvant regimens for more than 2 decades [23, 26–29]. Another class of effective cytotoxic agents, taxanes (such as paclitaxel and docetaxel), are also used as standard therapy for breast cancer, as monochemotherapy, in combination therapy, or in sequential therapy [30–32]. As neoadjuvant therapy, taxanes have produced dramatic increases in response rates [33–35]. Still, only a minority of patients achieve pathologic complete response (pCR) after receiving either anthracycline- or taxane-based neoadjuvant regimens. Given the lack of a preferred neoadjuvant treatment regimen, identifying which patients are most likely to respond to a specific regimen could significantly improve breast cancer treatment outcomes and facilitate the development of targeted therapies. Since GWASs have been successful in identifying low-penetrance common variants for breast cancer predisposition, several studies have investigated the potential roles of GWAS-identified SNPs in the choice of neoadjuvant therapy and evaluation of patient prognosis [18, 19]. In the current GWAS, we aimed to identify associations between SNPs associated with breast cancer susceptibility and response to anthracycline- and taxane-based neoadjuvant chemotherapy in patients with breast cancer.

RESULTS

Patient characteristics

The demographic and clinicopathologic characteristics of the 960 breast cancer patients who received neoadjuvant chemotherapy and were included in the discovery or validation cohort are shown in Table 1. In the discovery stage, the pCR rate was 32.6%. In the validation stage, the pCR rate was 15.5% in patients who received anthracycline-based neoadjuvant therapy and 17.1% in those who received a taxane-based regimen.

Identification of SNPs

The discovery cohort included 92 patients who each received 4 cycles of the anthracycline-based CTF neoadjuvant regimen. pCR was used to differentiate cases from controls. All individuals had < 10% of missing genotypes, and all were genotyped as females based on X-chromosome genotypes, and therefore all were included. GWAS was performed using PLINK on genotype data from 30 patients with pCR and 62 patients with non-pCR. Of 389,795 SNPs included in the analysis, 351 SNPs showed significant associations with pCR. We then employed 3 additional criteria to select SNPs for inclusion in the validation stage. First, the 351 SNPs that showed significant association with pCR were confirmed using a genome-wide significance at \( P < 1 \times 10^{-3} \) (Figure 1). Second, the data set was narrowed down to 32 SNPs that are known to be related to genes that correlate with cancer using a search of gene annotation databases and published literature (ORs ranging from 2.89 [95% CI, 1.53–5.48] to 6.14 [95% CI, 2.07–18.20]) (Table 2). Third, using unadjusted univariate Cox proportional hazard models \( (P < 0.05) \), we ended up with only 2 SNPs with a pCR rate that was significantly higher for both homozygotes (Table 3).

One of these two SNPs, rs1799937, is located in an intron of the WT1 gene at 11p13; the A allele had an OR of 3.33 (95% CI, 1.70–6.51) compared with the G allele. The other SNP, rs6044100, is located in an intron of the KIF16B gene at 20p11, and the T allele had an OR of 4.05 (95% CI, 1.98–8.29) compared with the C allele. We found significantly higher pCR rates in patients with the rs1799937 AA genotype (60.0%) and the rs6044100 TT genotype (71.4%) than in patients with the rs1799937 GG genotype (16.7%) and the rs6044100 CC genotype (16.4%) (Table 3).
Table 1: Patient characteristics

| Characteristic                              | No. of patients | Discovery cohort (N = 92) | Validation cohort | P value |
|---------------------------------------------|-----------------|---------------------------|-------------------|---------|
|                                             |                 | Anthracyline (N = 401)    | Taxane (N = 467)  |         |
|                                             | N   | %   | N   | %   | N   | %   |         |
| Age (years)                                 |     |     |     |     |     |     |         |
| ≤ 50                                        | 571 | 59.5| 65  | 70.7| 250 | 62.3| 256     | 54.8   | 0.006* |
| > 50                                        | 389 | 40.5| 27  | 29.3| 151 | 37.7| 211     | 45.2   |         |
| Tumor size                                  |     |     |     |     |     |     |         |        |
| ≤ 2cm                                       | 381 | 40.0| 22  | 23.9| 147 | 37.3| 212     | 45.5   | < 0.001*|
| > 2cm                                       | 571 | 60.0| 70  | 76.1| 247 | 62.7| 254     | 54.5   |         |
| Unknown                                     | 8   | 0   | 7   | 1   |     |     |         |        |
| Tumor grade                                 |     |     |     |     |     |     |         |
| I                                           | 95  | 10.3| 6   | 6.7 | 45  | 11.6| 44      | 9.8    | 0.21   |
| II                                          | 691 | 74.7| 68  | 76.4| 277 | 71.2| 346     | 77.4   |         |
| III                                         | 139 | 15.0| 15  | 16.9| 67  | 17.2| 57      | 12.8   |         |
| Unknown                                     | 35  | 3   | 12  | 2   |     |     |         |        |
| Lymph node involvement                      |     |     |     |     |     |     |         |
| Negative                                    | 525 | 55.4| 63  | 69.2| 230 | 57.6| 232     | 50.7   | 0.002* |
| Positive                                    | 423 | 44.6| 28  | 30.8| 169 | 42.4| 226     | 49.3   |         |
| Unknown                                     | 12  | 1   | 2   | 2   |     |     |         |        |
| Estrogen receptor status                    |     |     |     |     |     |     |         |
| Negative                                    | 377 | 39.4| 44  | 48.4| 155 | 38.8| 178     | 38.2   | 0.18   |
| Positive                                    | 579 | 60.6| 47  | 51.6| 244 | 61.2| 288     | 61.8   |         |
| Unknown                                     | 4   | 1   | 2   | 1   |     |     |         |        |
| Progesterone receptor status                |     |     |     |     |     |     |         |
| Negative                                    | 479 | 50.2| 51  | 56.7| 206 | 51.6| 222     | 47.7   | 0.23   |
| Positive                                    | 475 | 49.8| 39  | 43.3| 193 | 48.4| 243     | 52.3   |         |
| Unknown                                     | 6   | 2   | 2   | 2   |     |     |         |        |
| HER2 status                                 |     |     |     |     |     |     |         |
| Negative                                    | 694 | 72.4| 56  | 61.5| 296 | 73.8| 342     | 73.4   | 0.05   |
| Positive                                    | 264 | 27.6| 35  | 38.5| 105 | 26.2| 124     | 26.6   |         |
| Unknown                                     | 2   | 1   | 0   | 1   |     |     |         |        |
| Surgery type                                |     |     |     |     |     |     |         |
| BCS                                         | 406 | 42.3| 51  | 55.4| 171 | 42.6| 184     | 39.4   | 0.017* |
| Mastectomy                                  | 554 | 57.7| 41  | 44.6| 230 | 57.4| 283     | 60.6   |         |
| Chemotherapy cycles                         |     |     |     |     |     |     |         |
| ≥ 4                                         | 897 | 93.4| 92  | 100.0| 361 | 90.0| 444     | 95.1   | < 0.001*|
| < 4                                         | 63  | 6.6 | 0   | 0   | 40  | 10.0| 23      | 4.9    |         |

(Continued)
Association of rs1799937 genotype and response to anthracycline-based regimens

Having confirming that the SNPs rs1799937 AA genotype and rs6044100 TT genotype are associated with pCR, we investigated the potential association between these two genetic variants and pCR after anthracycline-based neoadjuvant chemotherapy. We attempted to validate rs1799937 and rs6044100 in an independent cohort of 401 patients who received CTF regimens by sequencing analysis. The results showed that patients with the rs1799937 AA genotype benefited from anthracycline-based neoadjuvant chemotherapy more than patients with the GG genotype; the pCR rates of the AA versus GG genotypes were 26.7% and 11.5%, respectively ($P = 0.035$). Logistic regression models showed an odds ratio for patients with the rs1799937 AA genotype compared with those with the GG genotype of 2.81 (95% CI, 1.13–6.98; $P = 0.026$, Table 4). However, no association was identified between rs6044100 and pCR; the pCR rates after CTF therapy for patients with the TT genotype versus those with the CC genotype were 18.8% and 15.2%, respectively ($P = 0.86$).

Association between rs1799937 genotype and survival

Among the 493 patients who received CTF regimen, including 92 from the discovery stage and 401 from the validation stage, patients with the rs1799937 AA genotype had a slightly higher but not statistically significant 5-year recurrence-free survival (RFS) rate (89.3%; 95% CI, 79.5%–99.1%; AA vs GA+GG; $P = 0.35$) and 5-year distant recurrence-free survival (DRFS) rate (92.2%; 95% CI, 83.8%-100.1%; AA vs GA+GG; $P = 0.31$) than those with the rs1799937 GG genotype (RFS, 84.0%; 95% CI, 79.1%-88.9%; DRFS, 87.4%; 95% CI, 82.9%-91.9%; Figures 2a and 2b). However, no difference in RFS and DRFS was detected between rs1799937 AA versus GG genotype in the 467 breast cancer patients treated with taxane-based neoadjuvant chemotherapy (Figures 2c and 2d).

**DISCUSSION**

We performed a GWAS to assess genetic variants that may be associated with pCR in patients receiving neoadjuvant chemotherapy for breast cancer. We identified 1 SNP, rs1799937 A > G, as a genetic variant associated with response to anthracycline-based CTF neoadjuvant chemotherapy. Patients with the rs1799937 AA genotype showed significantly higher rates of pCR and better prognosis after treatment with CTF than patients with the GG genotype. However, we found that the AA allele of rs1799937 was not associated with sensitivity to taxane-based neoadjuvant chemotherapy regimens.

The rs1799937 A > G SNP is located in an intron region of the **WT1** gene at 11p13. **WT1** encodes a transcription factor that contains 4 zinc-finger motifs.
Table 2: SNPs significantly associated with breast cancer in patients who received anthracycline-based CTF chemotherapy regimens in discovery cohort

| SNP         | Minor allele | Chr. | Position | Gene   | Location | pCR | non-pCR | OR (95%CI)  | P value          |
|-------------|--------------|------|----------|--------|----------|-----|---------|-------------|-----------------|
| rs6135795   | G            | 20   | 16534948 | KIF16B | Intron   | 0.62| 0.28   | 4.07 (2.11–7.83) | 1.61 × 10⁻⁵      |
| rs4748316   | T            | 10   | 16779881 | RSU1   | Intron   | 0.48| 0.18   | 4.17 (2.10–8.27) | 2.55 × 10⁻⁵      |
| rs6686072   | C            | 1    | 57936193 | DAB1   | Intron   | 0.40| 0.13   | 4.33 (2.07–9.06) | 4.98 × 10⁻⁵      |
| rs687660    | A            | 11   | 7011532  | PPFIA1 | Upstream | 0.32| 0.08   | 5.10 (2.19–11.87) | 5.97 × 10⁻⁵      |
| rs3102072   | A            | 1    | 20964066 | PINK1  | Intron   | 0.62| 0.31   | 3.61 (1.89–6.90)  | 7.39 × 10⁻⁵      |
| rs6044100   | T            | 20   | 16524972 | KIF16B | Intron   | 0.42| 0.15   | 4.05 (1.98–8.29)  | 7.65 × 10⁻⁵      |
| rs9806453   | G            | 15   | 60943845 | RORA   | Intron   | 0.67| 0.37   | 3.46 (1.80–6.64)  | 1.43 × 10⁻⁵      |
| rs6075070   | G            | 20   | 16536424 | KIF16B | Intron   | 0.42| 0.16   | 3.80 (1.87–7.72)  | 1.44 × 10⁻⁵      |
| rs4811431   | T            | 11   | 70203569 | TSHZ2  | Intron   | 0.28| 0.08   | 4.88 (2.02–11.77) | 1.48 × 10⁻⁵      |
| rs6686072   | C            | 1    | 57936193 | DAB1   | Intron   | 0.40| 0.13   | 4.33 (2.07–9.06)  | 4.98 × 10⁻⁵      |
| rs687660    | A            | 11   | 7011532  | PPFIA1 | Upstream | 0.32| 0.08   | 5.10 (2.19–11.87) | 5.97 × 10⁻⁵      |
| rs3102072   | A            | 1    | 20964066 | PINK1  | Intron   | 0.62| 0.31   | 3.61 (1.89–6.90)  | 7.39 × 10⁻⁵      |
| rs6044100   | T            | 20   | 16524972 | KIF16B | Intron   | 0.42| 0.15   | 4.05 (1.98–8.29)  | 7.65 × 10⁻⁵      |
| rs9806453   | G            | 15   | 60943845 | RORA   | Intron   | 0.67| 0.37   | 3.46 (1.80–6.64)  | 1.43 × 10⁻⁵      |
| rs6075070   | G            | 20   | 16536424 | KIF16B | Intron   | 0.42| 0.16   | 3.80 (1.87–7.72)  | 1.44 × 10⁻⁵      |
| rs4811431   | T            | 11   | 70203569 | TSHZ2  | Intron   | 0.28| 0.08   | 4.88 (2.02–11.77) | 1.48 × 10⁻⁵      |
| rs6686072   | C            | 1    | 57936193 | DAB1   | Intron   | 0.40| 0.13   | 4.33 (2.07–9.06)  | 4.98 × 10⁻⁵      |
| rs687660    | A            | 11   | 7011532  | PPFIA1 | Upstream | 0.32| 0.08   | 5.10 (2.19–11.87) | 5.97 × 10⁻⁵      |
| rs3102072   | A            | 1    | 20964066 | PINK1  | Intron   | 0.62| 0.31   | 3.61 (1.89–6.90)  | 7.39 × 10⁻⁵      |

SNP, single nucleotide polymorphism; CTF, 5-fluouracil, pirarubicin, and cyclophosphamide; Chr., chromosome; MAF, minor allele frequency; pCR, pathologic complete response; OR, odds ratio; CI, confidence interval

*A P value < 0.05 was considered statistically significant.
The two SNPs examined in the validation phase were marked by bold.
Table 3: SNPs for which homozygosity was significantly associated with pCR in the discovery cohort

| SNP     | Genotype | No. (N = 92) | %   | Pathologic response | P value       |
|---------|----------|--------------|-----|---------------------|--------------|
|         |          |              |     | non-pCR (N = 62)    | pCR (N = 30) |
|         |          |              |     | No. | %   | No. | %   |              |
| rs1799937 | A A     | 10           | 10.9 | 4   | 40.0 | 6   | 60.0 | 0.002*       |
|          | G A     | 34           | 37.0 | 18  | 52.9 | 16  | 47.1 |              |
|          | G G     | 48           | 52.2 | 40  | 83.3 | 8   | 16.7 |              |
| rs6044100 | C C     | 55           | 59.8 | 46  | 83.6 | 9   | 16.4 | < 0.001*     |
|          | T C     | 30           | 32.6 | 14  | 46.7 | 16  | 53.3 |              |
|          | T T     | 7            | 7.6  | 2   | 28.6 | 5   | 71.4 |              |

SNP, single nucleotide polymorphism; pCR, pathologic complete response
*A P value < 0.05 was considered statistically significant.*AA vs GG in rs11799937; *CC vs TT in rs6044100.

Table 4: Associations between SNP genotypes and pCR in 401 validation cohort patients who underwent anthracycline-based (CTF) chemotherapy. Two SNPs were examined

| SNP     | Genotype | No. (N = 401) | %   | Pathologic response | P value | OR (95%CI) | P value |
|---------|----------|--------------|-----|---------------------|---------|------------|---------|
|         |          |              |     | non-pCR (N = 339)   | pCR (N = 62) |
|         |          |              |     | No. | %   | No. | %   |               |
| rs1799937 | A A     | 30           | 7.5  | 22  | 73.3 | 8   | 26.7 | 0.035*       |
|          | G A     | 153          | 38.2 | 124 | 81   | 29  | 19   | 1.81(1.01–3.23) | 0.046*   |
|          | G G     | 218          | 54.4 | 193 | 88.5 | 25  | 11.5 |               |
| rs6044100 | C C     | 224          | 57.3 | 190 | 84.8 | 34  | 15.2 | 0.86       |
|          | T C     | 135          | 34.5 | 113 | 83.7 | 22  | 16.3 | 0.84(0.31–2.29) | 0.74*   |
|          | T T     | 32           | 8.2  | 26  | 81.2 | 6   | 18.8 |               |
|          | Unknown | 10           | 10   | 0   |      |      |      |               |

SNP, single nucleotide polymorphism; pCR, pathologic complete response; CTF, 5-fluorouracil, pirarubicin, and cyclophosphamide; OR, odds ratio; CI, confidence interval
*A P value of < 0.05 was considered statistically significant.*AA vs GG; *AG vs GG; *CC vs TT; *CT vs TT.

Table 5: Associations between SNP genotypes and pCR in 467 validation cohort patients who underwent taxane-based chemotherapy

| SNP     | Genotype | No. (N = 467) | %   | Pathologic response | P value | OR (95%CI) | P value |
|---------|----------|--------------|-----|---------------------|---------|------------|---------|
|         |          |              |     | non-pCR (N = 387)   | pCR (N = 80) |
|         |          |              |     | No. | %   | No. | %   |                  |
| rs1799937 | A A     | 41           | 8.8  | 34  | 82.9 | 7   | 17.1 | 0.34       |
|          | G A     | 185          | 39.6 | 159 | 85.9 | 26  | 14.1 | 0.85(0.36–2.04) | 0.72*   |
|          | G G     | 241          | 51.6 | 194 | 80.5 | 47  | 19.5 |                  |

pCR, pathologic complete response; SNP, single nucleotide polymorphism; OR, odds ratio; CI, confidence interval
*A P value of < 0.05 was considered statistically significant.*AA vs GG; *GA vs GG.
Figure 2: Kaplan-Meier estimates of recurrence-free survival (RFS) and distant recurrence-free survival (DRFS) by rs1799937 genotype and neoadjuvant chemotherapy regimen. Panel A, 10-year RFS by rs1799937 genotype in patients receiving anthracycline-based regimen. Panel B, 10-year DRFS by rs1799937 genotype in patients receiving anthracycline-based regimen. Panel C, 10-year RFS by rs1799937 genotype in patients receiving taxane-based regimen. Panel D, 10-year DRFS by rs1799937 genotype in patients receiving taxane-based regimen.
at the C-terminus and plays an important role in organ development and cell survival [36]. In addition to regulating nuclear transcription, WT1 is also involved in RNA splicing and metabolism [37, 38]. WT1 was first identified as a tumor suppressor gene in Wilms’ tumor. However, accumulating evidence has shown that overexpression of WT1 exerts an oncogenic effect in other tumors [36, 39, 40] and is linked to poor prognoses in leukemia and breast cancer [41, 42].

Previous studies have also shown that WT1 gene expression is related to sensitivity to chemotherapy. Perugorria et al. have reported that WT1 knockdown markedly sensitized hepatocellular carcinoma cell lines to doxorubicin-induced apoptosis [43]. Interestingly, the effect of WT1 expression on apoptosis of hepatocellular carcinoma cells seems to be specific to doxorubicin, and not to other chemotherapeutic agents. Doxorubicin, an anthracycline, is a potent inhibitor of Top2 isozymes and exerts its genotoxic effect by impeding DNA duplication [44–36]. We speculate that rs1799937 may function by altering WT1 gene expression, thus rendering cancer cells in patients with the AA allele more sensitive to anthracyline-based chemotherapy.

In summary, our results indicate that anthracycline-based neoadjuvant chemotherapy may benefit breast cancer patients with the rs1799937 AA allele more than those with the GG allele. Therefore, further investigation of WT1 as a potential target of chemotherapy for breast cancer is warranted.

MATERIALS AND METHODS

Patient selection

The GWAS included 960 patients with breast cancer who were treated at Peking University Cancer Hospital between October 2003 and November 2011. Written informed consent was obtained from all patients, and this study was approved by the Research and Ethical Committee of Peking University Cancer Hospital.

The discovery stage included 92 patients from a pool of 493 patients who had received 4 cycles of an anthracycline-based neoadjuvant regimen consisting of cyclophosphamide, pirarubicin (THPADM), 5-fluorouracil (CTF). This regimen included pirarubicin 35 mg/m² i.v. on day 1 and day 8, cyclophosphamide 500 mg/m² i.v. on day 1 and day 8, and 5-fluorouracil 200 mg/m²/day i.v. continuous infusion for 28 days, every 4 weeks. All the 92 patients received an identical CTF regimen with the same dosage and same number of cycles. There was no bias in terms of any patient characteristics. Using pCR, defined as the absence of invasive breast tumor cells in the breast and axillary lymph nodes after completion of neoadjuvant chemotherapy, the patients were divided into 2 pathologic response groups with 30 in the pCR group and 62 in the non-pCR group. In the validation stage, 401 patients who had received 2–6 cycles of CTF and 467 patients who had received a taxane-based regimen using paclitaxel were included.

Genotyping methods and quality control

In the discovery stage, we isolated DNA from peripheral blood lymphocytes using phenol-chloroform extraction and ultrapurification. Genotyping was conducted using Affymetrix Genome-Wide Human SNP Array 6.0 chips (Santa Clara, CA). We performed systematic quality control on the raw genotyping data to filter out unqualified samples and SNPs. As quality control, we excluded DNA samples that failed to be genotyped as female based on X-chromosome genotypes. SNPs were excluded if (1) they were not mapped on autosomal chromosomes; (2) they had a call rate < 95%; (3) they had minor allele frequency < 0.01; (4) they deviated from the Hardy-Weinberg equilibrium at P < 1e-05; or (5) their genotyping clusters had poor resolution. These led to a total of 389,795 SNPs to be analyzed using a GWAS toolset PLINK [47].

The genotypes of the patients with and without pCR were compared. We selected SNPs that were (1) associated with pCR in the discovery stage at P < 1e-03; (2) located in genes that have been correlated with cancer in gene annotation databases and published literature; and (3) associated with pCR status in the discovery stage at P < 0.05 using unadjusted univariate Cox proportional hazard models. Finally, 2 SNPs, rs1799937 and rs6044100, were identified.

For the validation phase, genomic DNA samples were extracted from peripheral blood lymphocytes of an additional 868 patients, including 401 who received anthracyclin-based regimens and 467 who received taxane-based regimens. Genotyping of the samples was conducted using a polymerase chain reaction/sequencing assay (ABI 3730 system, Applied Biosystems, Foster City, CA).

Statistical analysis

Association between SNPs and response to neoadjuvant chemotherapy in patients with breast cancer was assessed using logistic regression models and reported as odds ratios (ORs) and 95% confidence intervals (CIs). ORs were estimated for heterozygotes and homozygotes for the variant allele compared with homozygotes for the common allele. A logistic regression model was applied to determine whether a factor was an independent predictor of pCR in a multivariate analysis. Survival analysis was performed using the Kaplan-Meier curves. All statistical tests were two-sided. All statistical analyses were performed using SPSS 16.0 software (SPSS Inc., Chicago, IL).
Abbreviations

Abbreviations: GWAS, genome-wide association study; SNP, single nucleotide polymorphism; pCR, pathologic complete response; CTF, 5-fluorouracil, pirarubicin, and cyclophosphamide; OR, odds ratio; CI, confidence interval; RFS, recurrence-free survival; DRFS, distant recurrence-free survival.

ACKNOWLEDGMENTS

This study was supported by the National Key Technology Research and Development Program of the Ministry of Science and Technology of China (No. 2014BAI09B08); the 973 project 2013CB911004; the 985-III project; and grants from the National Natural Science Foundation of China (No.81202106, No.81302330, No.30973436 and No.81071629). We thank Amy L. Ninetto from Scientific Publication at The University of Texas MD Anderson Cancer Center for editorial review of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Campeau PM, Foulkes WD, Tischkowitz MD. Hereditary breast cancer: new genetic developments, new therapeutic avenues. Hum Genet. 2008; 124:31–42.
2. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447:1087–1093.
3. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet. 2007; 39:870–874.
4. 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. Nat Genet. 2007; 39:865–869.
5. Low SK, Takahashi A, Ashikawa K, Inazawa J, Miki Y, Kubo M, Nakamura Y, Katagiri T. Genome-wide association study of breast cancer in the Japanese population. PLoS One. 2013; 8:e76463.
6. Chung S, Low SK, Zembutsu H, Takahashi A, Kubo M, Sasa M, Nakamura Y. A genome-wide association study of chemotherapy-induced alopecia in breast cancer patients. Breast Cancer Res. 2013; 15:R81.
7. Sapkota Y, Yasui Y, Lai R, Sridharan M, Robson PJ, Cass CE, Mackey JR, Damaraju S. Identification of a breast cancer susceptibility locus at 4q31.22 using a genome-wide association study paradigm. PLoS One. 2013; 8:e62550.
8. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, Orr N, Rhie SK, Riboli E, Feigelson HS, Le Marchand L, Buring JE, Eccles D, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet. 2013; 45:392–398. 398:e391-392.
9. Kim HC, Lee JY, Sung H, Choi JY, Park SK, Lee KM, Kim YJ, Go MJ, Li L, Cho YS, Park M, Kim DJ, Oh JH, 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. 2012; 14:R56.
10. Ghoussaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, Dennis J, Wang Q, Humphreys MK, Luccioni C, Baynes C, Conroy D, Maranian M, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. Nat Genet. 2012; 44:312–318.
11. Cai Q, Long J, Lu W, Qu S, Wen W, Kang D, Lee JY, Chen K, Shen H, Shen CY, Sung H, Matsuo K, Haiman CA, et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. Hum Mol Genet. 2011; 20:4991–4999.
12. Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, Healey S, Morrison J, Kartsonaki C, Lesnick T, Ghoussaini M, Barrowdale D, Peock S, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. Nat Genet. 2010; 42:885–892.
13. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, Tapper W, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet. 2010; 42:504–507.
14. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet. 2009; 41:579–584.
15. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, Eccles D, Evans DG, Fletcher O, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet. 2009; 41:585–590.
16. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL, Gu K, Fair AM, Cai Q, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet. 2009; 41:324–328.
17. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen...
18. Wang Z, Xu B, Lin D, Tan W, Leaw S, Hong X, Hu X. XRCC1 polymorphisms and severe toxicity in lung cancer patients treated with cisplatin-based chemotherapy in Chinese population. Lung Cancer. 2008; 62:99–104.

19. Yu D, Zhang X, Liu J, Yuan P, Tan W, Guo Y, Sun T, Zhao D, Yang M, Xu B, Lin D. Characterization of functional excision repair cross-complementation group 1 variants and their association with lung cancer risk and prognosis. Clin Cancer Res. 2008; 14:2878–2886.

20. Marme F, Werft W, Benner A, Burwinkel B, Sinn P, Sohn C, Lichter P, Hahn M, Schneeweiss A. FGFR4 Arg388 genotype is associated with pathological complete response to neoadjuvant chemotherapy for primary breast cancer. Ann Oncol. 2010; 21:1636–1642.

21. Zhou X. CD24 polymorphisms cannot predict pathologic complete response to anthracycline- and taxane-based neoadjuvant chemotherapy in breast cancer. Clin Breast Cancer. 2014; 14:e33–40.

22. Fagerholm R, Schmidt MK, Khan S, Rafiq S, Tapper W, Aittomaki K, Greco D, Heikkinen T, Muranen TA, Fasching PA, Janni W, Weinshilboum R, Loehberg CR, et al. The SNP rs6500843 in 16p13.3 is associated with survival specifically among chemotherapy-treated breast cancer patients. Oncotarget. 2015; 6:7390–7407. doi: 10.18632/oncotarget.3506.

23. Rastogi P, Anderson SJ, Bear HD, Geyer CE, Kahlenberg MS, Robidoux A, Margolese RG, Hoehn JL, Vogel VG, Dakhil SR, Tamkun D, King KM, Pajon ER, et al. Preoperative chemotherapy: updates of National Surgical Adjuvant Breast and Bowel Project Protocols B-18 and B-27. J Clin Oncol. 2008; 26:778–785.

24. Fisher B, Bryant J, Wolmark N, Mamounas E, Brown A, Fisher ER, Wickerham DL, Begovic M, DeCillis A, Robidoux A, Margolese RG, Cruz AB Jr., Hoehn JL, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol. 1998; 16:2672–2685.

25. van der Hage JA, van de Velde CJ, Julien JP, Tubiana-Hulin M, Vandervelden C, Duchateau L. Preoperative chemotherapy in primary operable breast cancer: results from the European Organization for Research and Treatment of Cancer trial 10902. J Clin Oncol. 2001; 19:4224–4237.

26. Untch M, Mobus V, Kuhn W, Muck BR, Thomssen C, Bauerfeind I, Harbeck N, Werner C, Lebeau A, Schneeweiss A, Kahler S, von Koch F, Petry KU, et al. Intensive dose-dense compared with conventionally scheduled preoperative chemotherapy for high-risk primary breast cancer. J Clin Oncol. 2009; 27:2938–2945.

27. Crozier JA, Swaika A, Moreno-Aspitia A. Adjuvant chemotherapy in breast cancer: To use or not to use, the anthracyclines. World J Clin Oncol. 2014; 5:529–538.
41. Inoue K, Sugiyama H, Ogawa H, Nakagawa M, Yamagami T, Miwa H, Kita K, Hiraoka A, Masaoka T, Nasu K, et al. WT1 as a new prognostic factor and a new marker for the detection of minimal residual disease in acute leukemia. Blood. 1994; 84:3071–3079.
42. Miyoshi Y, Ando A, Egawa C, Taguchi T, Tamaki Y, Tamaki H, Sugiyama H, Noguchi S. High expression of Wilms’ tumor suppressor gene predicts poor prognosis in breast cancer patients. Clin Cancer Res. 2002; 8:1167–1171.
43. Perugorria MJ, Castillo J, Latasa MU, Goni S, Segura V, Sangro B, Prieto J, Avila MA, Berasain C. Wilms’ tumor 1 gene expression in hepatocellular carcinoma promotes cell dedifferentiation and resistance to chemotherapy. Cancer Res. 2009; 69:1358–1367.
44. Gewirtz DA. A critical evaluation of the mechanisms of action proposed for the antitumor effects of the anthracycline antibiotics adriamycin and daunorubicin. Biochem Pharmacol. 1999; 57:727–741.
45. Harper JW, Elledge SJ. The DNA damage response: ten years after. Mol Cell. 2007; 28:739–745.
46. Zhang A, Lyu YL, Lin CP, Zhou N, Azarova AM, Wood LM, Liu LF. A protease pathway for the repair of topoisomerase II-DNA covalent complexes. J Biol Chem. 2006; 281:35997–36003.
47. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575.