Breast Cancer Res Treat
DOI 10.1007/s10549-011-1663-3

PRECLINICAL STUDY

Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and mRNA expression of the nearest genes

Muhammad Riaz · Els M. J. J. Berns · Anieta M. Sieuwerts · Kirsten Ruigrok-Ritstier · Vanja de Weerd · Arwin Groenewoud · Andrè G. Uitterlinden · Maxime P. Look · Jan G. M. Klijn · Stefan Sleijfer · John A. Foekens · John W. M. Martens

Received: 6 April 2011 / Accepted: 24 June 2011

Abstract To understand the biology of low-risk breast cancer alleles, and to investigate whether these loci also contribute to disease progression that was once established, we examined the association of SNPs tagging the low-risk breast cancer loci in or near *FGFR2*, *LSP1*, *MAP3K1, H19*, *TOX3*, *POU5F1P1*, *MYC*, and 2q35, with clinical, pathological characteristics, prognosis, and mRNA expression of the nearest genes. Tumor DNA samples of 2,480 breast cancer patients were available. Out of this cohort, 1,290 patients with lymph-node negative disease who did not receive adjuvant systemic therapy, the SNP status was associated with metastasis-free survival (MFS). In 1,401 patients, the mRNA expression levels of *FGFR2*, *LSP1*, *MAP3K1, H19*, *TOX3*, *POU5F1P1,* and *MYC* were determined and correlated with SNP genotypes. The SNP rs2981582 in *FGFR2* was significantly associated with positive ER and PgR status ($P < 0.001$ and $P = 0.003$, respectively). No other significant associations with patient or tumor characteristics were observed. Only rs2107425 near *H19* was significantly associated with shorter MFS in uni- and multi-variate analysis (HR: 1.53, CI: 1.12–2.08, $P = 0.006$ and HR: 1.59, CI: 1.16–2.20, $P = 0.004$, respectively), with the more aggressive minor allele displaying a recessive trait. The minor allele of SNP rs3803662 located near the *TOX3* gene was associated with lower mRNA expression of this gene. In conclusion, except for the association of rs13283662 with *TOX3* gene expression indicating a tumor suppressor role of *TOX3*, our findings suggest that breast cancer low-risk loci generally do not affect expression of the nearest gene in breast tumor tissue. Also the prognosis of patients is largely not affected by low-risk breast cancer loci except for the SNP near *H19*. How, this SNP affects prognosis warrants further study as it does not operate through altering *H19* mRNA expression.

Keywords Breast cancer · Low-risk loci · Prognosis · mRNA Expression

Abbreviations

MFS	Metastasis-free survival
GWAS	Genome-wide association studies
SNPs	Single nucleotide polymorphisms
ER	Estrogen receptor
PgR	Progesterone receptor
Ct	Cycle of threshold
HWE	Hardy–Weinberg equilibrium
LD	Linkage disequilibrium
MAF	Minor allele frequency
HR	Hazard ratio
CI	Confidence interval

Electronic supplementary material The online version of this article (doi:10.1007/s10549-011-1663-3) contains supplementary material, which is available to authorized users.

M. Riaz · E. M. J. J. Berns · A. M. Sieuwerts · K. Ruigrok-Ritstier · V. de Weerd · A. Groenewoud · M. P. Look · J. G. M. Klijn · S. Sleijfer · J. A. Foekens · J. W. M. Martens (✉)
Department of Medical Oncology, Josephine Nefkens Institute and Daniel den Hoed Cancer Center, Erasmus MC, Building BE, Room 400, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands
e-mail: j.martens@erasmusmc.nl

J. A. Foekens · J. W. M. Martens
Cancer Genomic Center, Erasmus MC, Rotterdam, The Netherlands

A. G. Uitterlinden
Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands

Published online: 12 July 2011

Introduction

Genome-wide association studies (GWAS) have recently identified various single nucleotide polymorphisms (SNPs) tagging several breast cancer predisposing loci [1–3]. The identified SNPs are: rs2981282 located within intron 2 of the fibroblast growth factor receptor 2 (*FGFR2)*; rs3817198 in the lymphocyte-specific protein 1 (*LSP1*)) gene; rs889312 located close to the mitogen-activated protein kinase (*MAP3K1)* gene; rs3803662, rs12443621, and rs8051542 located near the trinucleotide repeat—containing 9 (*TNRC9*)) gene, also known as *TOX3*; rs2107425 located close to the imprinted *H19* gene and rs13387042, and rs13281615 in the 2q35 and 8q24 regions, respectively, which lack nearby annotated genes. The most nearest genes to rs13281615 in the 8q24 region are *POU5F1P1*, a transcription factor with a POU homeodomain, and *MYC*, a proto-oncogene. Although several independent GWAS have replicated the association of these SNPs with increased cancer risk [3–6], the mechanisms by which these loci in the target tissue may exert their effect are poorly understood.

In analogy to high-risk and moderate-risk breast cancer genes, it was speculated that the identified SNPs tag causal variants in the coding region of the nearby genes [1]. However, extensive sequencing efforts of the nearest genes have not identified such variants in the SNP-associated haplotype blocks [1]. This suggests that an alternative disease mechanism in which low-risk loci might involve expression modulation of nearby located genes, and hence, predispose to breast cancer. For example, Meyer et al. [7] observed an association of risk alleles in *FGFR2* with increased *FGFR2* mRNA expression in 63 invasive breast cancer samples, and Sun et al. [8] reported an association of the normal allele with higher *FGFR2* mRNA expression in 81 normal breast tissue samples. Besides that these results appear contradictory, both studies lacked substantial power to test, conclusively, the hypothesis of expression modulation of the nearest genes by risk loci.

To further test the hypothesis, that in breast cancer these SNPs may operate through expression modulation of the nearest genes, we measured in a large subset of 1,401 primary breast cancer samples, mRNA transcript levels of the most plausible genes located nearest to the respective SNPs. In addition, we investigated whether these low-risk breast cancer loci not only related to onset of breast cancer, as previously reported, but also contributed to its progression once established. Therefore, in the present study we have studied the association of SNP genotypes of seven low-risk breast cancer loci with patient and clinico-pathological tumor characteristics, as well as with tumor aggressiveness.

Materials and methods

Patients and tumor characteristics

This retrospective study was approved by the medical ethical committee of the Erasmus Medical Centre Rotterdam, the Netherlands (MEC 02.953). The study was performed in accordance with the Code of Conduct of the Federation of Medical Scientific Societies in the Netherlands (http://www.federa.org/?s=1&m=99), and consent was not required, and, wherever possible, it has been reported in line with the Reporting Recommendations for Tumor Marker Prognostic Studies guidelines [9]. Frozen primary tumor samples from patients who entered the clinic between 1978 and 2004, and from whom detailed clinical follow-up was available, were used. The median follow-up of patients alive was 106 (range 3–315) months and mean age was 55.9 (range 22–88) years.

Tumor estrogen receptor (ER) and progesterone receptor (PgR) analyses were performed by routine ligand binding assay or enzyme-linked immunoassay [10]. The cutoff classied the tumors as ER and PgR positive was 10 fmol/mg cytosolic protein. Patients with missing values for ER and PgR were indicated as an unknown category in Table 1. Details of patients and tumor characteristics are presented in Table 1. Transcript level expression of nearby located candidate genes and *HER2* were determined in a subset of 1,401 tumors. Details of patients and tumor characteristics of this subgroup are presented in Supplementary Table S1.

SNP genotyping

Genomic DNA of 680 tumor homogenates has been isolated as described previously [11] and the genomic DNA of 1,800 tissue sections were isolated from two to ten 30 μm cryostat sections (5–20 mg) with the NucleoSpin® Tissue kit (Macherey–Nagel; Bioké, Leiden, The Netherlands) according to the protocol provided by the manufacturer. The quantity and quality of the isolated DNA were established by ultraviolet spectroscopy, and by examination of the product size after agarose gel electrophoresis, and by the ability of the sample to be linearly amplified by real-time PCR in a serial dilution with a set of primers located in an intron of the hydroxymethylbilane synthase on chromosome 11, and thymidine kinase on chromosome 17 [12]. Due to low sensitivity of the assay, SNP rs13281615 in the 8q24 region was genotyped in whole genome amplified DNA of 1,705 tumors as described before [12]. Briefly, 10 ng aliquots of genomic DNA of 1,705 tumors were amplified with the GenomiPhi V2 DNA amplification kit (GE Healthcare, Piscataway, NJ, USA) according to the protocol provided by the manufacturer [12], typically

Table 1 Associations of seven SNPs with patient and tumor characteristics ($n = 2{,}480$)

Characteristics	Subgroups	SNPs:		rs2981582			rs3817198			rs889312			rs2107425			rs3803662			rs13281615			rs13387042		
		Nearest genes:		*FGFR2*			*LSP1*			*MAP3K1*			*H19*			*TOX3*			8q24			2q35		
		Patients with genotype data (missing values):		$n = 2{,}410$ (70)			$n = 2{,}445$ (35)			$n = 2{,}460$ (20)			$n = 2{,}430$ (50)			$n = 2{,}379$ (101)			$n = 1{,}705$ (775)			$n = 2{,}362$ (118)		
		SNP genotypes:		CC	CT	TT	TT	TC	CC	AA	AC	CC	GG	GA	AA	GG	GA	AA	AA	AG	GG	GG	GA	AA
		Patients:																						
		n	%	33[b]	49	18	49	42	9	48	43	9	48	43	9	51	40	9	35	47	18	28	48	23
Age (years)	<40	305	12	34	48	18	49	43	8	43	45	12	44	47	9	52	38	10	39	43	18	26	49	25
	41–55	971	39	34	48	18	49	42	9	49	42	9	51	41	8	49	41	10	34	48	18	30	48	23
	56–70	799	32	31	50	19	50	42	9	48	44	8	46	45	9	51	40	9	35	47	19	29	47	25
	>70	405	16	34	50	16	47	43	9	52	39	9	50	39	11	56	37	7	36	46	18	26	52	21
				n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Menopausal status	Pre	1,078	43	34	49	18	48	42	9	47	43	10	50	42	8	49	41	9	35	48	17	28	49	24
	Post	1,402	57	33	49	18	49	42	9	50	42	8	48	43	9	53	39	9	35	46	19	29	48	23
				n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Tumor size	pT1	910	37	32	50	18	47	44	9	48	43	9	48	43	10	51	41	9	34	48	17	28	48	24
	pT2	1,293	52	34	48	18	50	41	9	49	43	8	49	43	8	51	40	10	36	45	19	29	48	24
	pT3	277	11	34	48	18	49	42	8	50	40	10	46	43	11	56	36	7	34	47	19	27	52	22
				n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Lymph nodes involved	0	1,290	52	33	49	18	49	42	8	48	42	9	50	42	8	51	39	10	35	46	19	28	47	24
	1–3	550	22	32	50	18	46	43	10	49	42	9	47	44	9	51	41	8	36	45	18	28	50	22
	>3	640	26	33	48	19	49	42	9	48	44	8	47	43	10	53	39	8	34	50	16	29	48	23
				n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Tumor grade	Poor	1,352	55	35	49	16	49	42	9	49	43	8	48	43	9	53	38	9	35	47	18	29	47	24
	Good/moderate	435	18	34	48	18	45	45	10	44	46	11	49	42	8	47	43	10	36	47	17	27	49	23
	Unknown	693	28	29	50	22	50	41	9	50	41	9	48	43	9	51	41	8	35	45	20	28	50	22
				n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		
Tumor histology	IDC	1,341	54	33	49	18	48	43	10	47	44	8	51	41	8	50	42	8	35	46	19	28	48	24
	ILC	190	8	26	54	20	47	45	8	52	39	9	48	42	10	51	40	9	39	43	18	39	46	15
	Muc	57	2	38	48	14	42	46	12	50	36	14	44	46	11	50	43	7	23	44	33	26	50	24
	Medul	57	2	34	55	11	60	33	7	54	38	9	44	55	2	48	44	7	32	44	24	30	49	21
	DCIS + IDC	315	13	32	50	18	51	41	8	48	41	11	44	44	12	51	38	11	32	54	14	28	46	26
	Unknown	520	21	35	46	19	50	42	8	49	42	9	47	45	9	56	33	11	38	45	17	27	51	22
				n.s.			n.s.			n.s.			n.s.			n.s			n.s.			n.s.		

Table 1 continued

Characteristics	Subgroups	SNPs:		rs2981582			rs3817198			rs889312			rs2107425			rs3803662			rs13281615			rs13387042		
		Nearest genes:		*FGFR2*			*LSP1*			*MAP3K1*			*H19*			*TOX3*			8q24			2q35		
		Patients with genotype data (missing values):		n = 2,410 (70)			n = 2,445 (35)			n = 2,460 (20)			n = 2,430 (50)			n = 2,379 (101)			n = 1,705 (775)			n = 2,362 (118)		
		SNP genotypes:		CC	CT	TT	TT	TC	CC	AA	AC	CC	GG	GA	AA	GG	GA	AA	AA	AG	GG	GG	GA	AA
		Patients:																						
		n	%	33[b]	49	18	49	42	9	48	43	9	48	43	9	51	40	9	35	47	18	28	48	23
Estrogen receptor	Negative	655	26	41	44	15	53	38	9	52	41	7	48	43	9	53	38	9	36	46	18	27	47	26
	Positive	1,809	73	30	51	19	47	44	9	47	43	10	49	43	9	51	40	9	35	47	18	29	49	22
	Unkown	16	1	44	44	13	25	56	19	50	50	0	56	44	0	53	40	7	44	33	22	38	25	38
				$P < 0.001$			**$P = 0.03$**			**$P = 0.03$**			n.s.			n.s.			n.s.			n.s.		
Progesterone receptor	Negative	782	32	38	46	16	52	39	9	50	43	7	49	41	10	52	41	8	36	44	20	28	46	26
	Positive	1,555	63	31	50	19	47	44	9	48	43	10	49	43	8	52	39	10	36	47	17	29	49	22
	Unknown	143	6	31	52	17	52	36	11	48	45	8	43	47	9	47	44	9	29	56	15	25	55	20
				$P = 0.003$			**$P = 0.05$**			n.s.			n.s.			n.s.			n.s.			n.s.		
HER2 status (mRNA)[a]	Negative	1,146	82	34	49	18	50	41	9	49	42	9	48	43	9	50	41	9	35	46	19	27	49	24
	Positive	234	17	32	49	19	53	39	8	43	50	7	49	42	9	55	38	7	34	48	18	31	45	24
	Unknown	20	1	40	55	5	45	45	10	60	25	15	30	45	25	47	42	11	40	40	20	30	50	20
				n.s.			n.s.			n.s.			n.s.			n.s.			n.s.			n.s.		

[a] *HER2* mRNA data is available only for 1,401 tumors, n.s. non-significant

[b] Percentage of patients per SNP genotype

All *P* values were derived from Pearson's chi-squared statistics. Minor variations in percentages of the patients among SNPs occur due to variation in missing values per SNP

yielding 4 μg amplified genomic DNA with 20 kb band still visible on gel. Genotyping was performed by a fluorescent 5′ exonuclease assay (TaqMan) using the ABI PRISM 7900HT Sequence Detection Sequence (PE Biosystems) according to the manufacturer's instructions. Primers and probes for SNPs rs2981582 (C_2917302_10), rs3817198 (C_27493923_10), rs889312 (C_8886795_10), rs3803662 (C_25968567_10), rs12443621 (C_30765260_10), rs8051542 (C_1230104_10), rs13387042 (C_32048042_10), and rs13281615 (C_1332250_10) were purchased from Applied Biosystems as Made to Order, and for rs2107425 (forward: ATA ATG CCC GAC CTG AAG ATC TG, reverse: GCG TCG CAG GGT TCA C, VIC-probe: CAC TCA TGG GAG CCG,and FAM-probe: ACA CTC ATA GGA GCC G) as Assay on Demand. All reactions were carried out with no template as negative controls. Duplicate samples ($n = 352$) extracted with two different DNA isolation procedures as mentioned above were analyzed to assess concordance and quality of genotyping. High concordance was found between the SNPs mentioned indicating that both procedures were equivalent. Due to sensitivity issue of the assays, the SNP genotypes of some patients could not be determined and were regarded as missing values for the corresponding SNPs. The missing values for each of the analyzed SNPs are also presented in Table 1.

RNA isolation and quantitative RT-PCR

RNA isolation, cDNA synthesis, quantification of specific mRNA species, and quality control checks were done as described in detail elsewhere [13]. In brief, total RNA was extracted with RNA-Bee (Campro, Veenendaal, the Netherlands) according to the manufacturer's instructions and 2–5 μg of total RNA sample aliquots were reverse transcribed with oligo(dT)12–18 and random hexamer primers using the RevertAid H Minus First Strand cDNA Synthesis Kit from Fermentas (St. Leon-Rot, Germany) or the Superscript II RNase H-kit from Invitrogen (Breda, the Netherlands), respectively, according to the protocols provided by the manufacturers. Before the PCR analysis, the resulting cDNA samples were treated with RNase H- (Ambion, Huntingdon, United Kingdom). The quantity and quality of the isolated RNA were established by UV spectroscopy, by examination of rRNA bands after agarose gel electrophoresis, and by the ability of the sample to be linearly amplified in a serial dilution with our housekeeping gene set (see next section for further details). Samples of total RNA that did not show both the 18S and 28S bands (6%) or 15 μg reverse-transcribed total RNA that was not amplifiable within 26 cycles at our fixed threshold value of 0.02 with our housekeeping set were excluded. Real-time quantitative PCR was done in a Mx3000P Real-time PCR System (Agilent, Amsterdam, The Netherlands) according to the recommended protocol. Commercially available Taqman Gene Expression Assays (Applied Biosystems) were used for *FGFR2* (Hs00240792_m1), *MAP3K1* (Hs00394890_m1), *TOX3* (Hs00300355_m1), *LSP1* (Hs00158885_m1), *H19* (Hs00399294_g1), *POU5F1/POU5F1P3/POU5F1P1* (Hs00999634_gH), and *MYC* (Hs00905030_m1). Forty rounds of amplification were performed according to the supplier's protocol, and at the end of the amplification fluorescent signals of the TaqMan probes were used to generate cycle threshold (Ct) values from which mRNA expression levels were calculated. To enable comparison of the levels of specific mRNAs in different samples, they were evaluated relative to the average expression levels of the three reference genes, porphobilinogen deaminase *(PBGD)*, hypoxanthine-guanine phosphoribosyltransferase *(HPRT)*, and β-2-microglobulin *(β2M)*. Expression levels of the target genes relative to the average expression levels of three reference genes were quantified as follows: mRNA target $= 2^{(\text{mean Ct reference} - \text{mean Ct target})}$. The primer sequences for *ER-α*, *PgR*, and *ERBB2/HER2* for the three reference genes, *PBGD*, *HPRT*, and *β2M*, have all been previously described [13, 14].

Statistical analyses

Pearson's χ^2 analysis was used to test for independence of the alleles [Hardy–Weinberg equilibrium (HWE)], the allelic distribution in breast tumors, to compare genotype frequencies between the three groups, and to check the relation of SNPs with patients, and tumor characteristics. The hazard ratios (HRs) and their 95% confidence interval (95% CI) for the SNP alleles and traditional prognostic factors were determined with Cox proportional hazards models for both univariate and multivariate metastasis-free survival (MFS) analyses. MFS was defined as the time from diagnosis until the occurrence of a distant metastasis. Differences between HR for SNP carrier versus non-carriers were tested using the likelihood ratio test associated with the Cox regression analysis. The assumption of proportional hazards was investigated using a test based on the Schoenfeld residuals and was not violated for SNP rs2107425. In the multivariate model, the contribution of the SNP alleles was adjusted for the classical prognostic factors; age, menopausal status, nodal status, tumor size, differentiation grade, and receptor status.

MFS probabilities were calculated using the actuarial method of Kaplan–Meier and a log–rank test was used to test for differences between the survival curves.

The association of SNPs and mRNA expression was determined using the Kruskal–Wallis test. A Bonferroni correction was applied to address the problem of multiple comparisons. All computations were carried out using the STATA statistical package, version 11.1 (Stata Corp.,

College Station, TX). A P value ≤ 0.05 was considered statistically significant. All P values are two-sided and relate to all data during the total period of follow-up.

Results

Frequencies of low-risk alleles in tumor samples and association with patient and tumor characteristics

We genotyped nine SNPs tagging linkage disequilibrium (LD) blocks containing seven loci, which have been reported to be associated with increased breast cancer risk [1, 2, 15] in target tissue of breast cancer tumors from a cohort of 2,480 patients. Genotype determination and test for HWE revealed that all nine genotyped SNPs in this cohort were in HWE ($P \geq 0.08$ for all SNPs). The calculated minor allele frequencies (MAF) of the SNPs in our cohort are given in Table 2. To check whether risk alleles showed an association with specific clinico-pathological characteristics of the patients and tumors, we investigated the correlation of the SNPs with patient and tumor characteristics (Table 1). The risk alleles of the SNP rs2981582 in *FGFR2* were significantly associated with ER- and PgR-positive tumors ($P < 0.001$ and $P = 0.003$, respectively). None of the other SNPs showed a strong association with either patient or tumor characteristics. After Bonferroni correction, only the association of *FGFR2* allele with ER and PgR level remained significant.

Association of low-risk alleles with prognosis of primary breast cancer

One of the aims of this study was to study a possible prognostic value of the low-risk SNPs. As prognosis is confounded by factors such as adjuvant treatment, we limited our analysis to 1,290 patients with lymph node-negative disease who did not receive any adjuvant systemic therapy. MFS was used as the primary end-point in this analysis. Compared with wild-type allele, patients carrying homozygous risk alleles AA of rs2107425 in *H19* were associated with a poor MFS in both univariate (HR: 1.53, CI: 1.12–2.08, $P = 0.006$), and multivariate analysis (HR: 1.59, CI: 1.16–2.20, $P = 0.004$) corrected for the base model that included age, menopausal status, pathological tumor size and grade, and ER and PgR protein levels (Supplementary Table S2). Kaplan–Meier survival curves were used to visualize the significant worse MFS (Logrank test $P = 0.006$) of patients homozygous for the risk allele of rs2107425 (Fig. 1). None of the other SNPs were related with the length of MFS in this cohort of 1,290 untreated lymph node-negative patients.

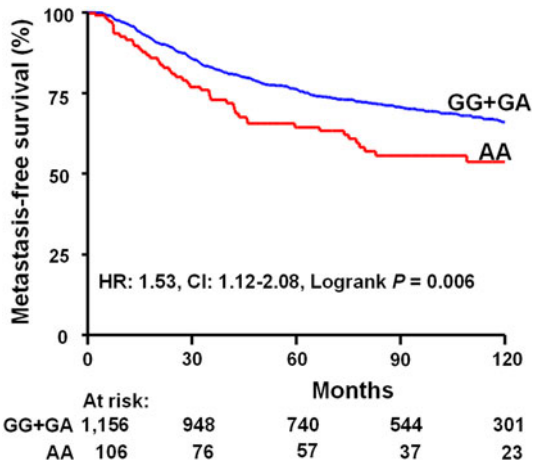

Fig. 1 Kaplan–Meier analysis of metastasis-free survival for rs2107425 SNP near *H19* in lymph node-negative patients who did not receive any adjuvant systemic therapy ($n = 1{,}290$). The *number* of patients at risk on various time intervals is indicated

Table 2 Minor allele frequencies of low-risk breast cancer loci in the target tissues

SNP	Locus	Nearest gene	MAF (European population)	MAF (target tissue)
rs2881582	10q26	*FGFR2*	0.38	0.42
rs3817198	11p15	*LSP1*	0.30	0.30
rs889312	5q11	*MAP3K1*	0.28	0.30
rs2107425	11p15	*H19*	0.31	0.30
rs3803662	16q12	*TOX3*	0.25	0.29
rs12443621	16q12	*TOX3*	0.46	0.46
rs8051542	16q12	*TOX3*	0.44	0.45
rs13281615	8q24	*POU5F1P1/MYC*	0.40	0.58
rs13387042	2q35	Not analyzed	0.52	0.50

Correlation of SNP genotype with expression changes of the nearest genes

The second aim of the study was to determine whether the risk alleles were associated with altered expression of the nearest gene. Therefore, we analyzed mRNA expression of the most plausible genes located close to the disease associated SNPs in a subset of 1,401 patients for which RNA was available (Fig. 2; clinical characteristics of this cohort are provided in Supplementary Table S1). For three SNPs near the *TOX3* gene (rs3803662, rs12443621, and rs8052542), which all have been associated with cancer risk, we found significant associations of SNPs rs3803662 and rs12443621 with a lower expression of *TOX3* mRNA ($P = 0.0019$ and $P = 0.029$, respectively; Kruskal–Wallis test). The lower *TOX3* mRNA expression was dose-dependent and decreased with the number of risk alleles ($P = 0.02$ for trend). However, rs8052542 near *TOX3*,

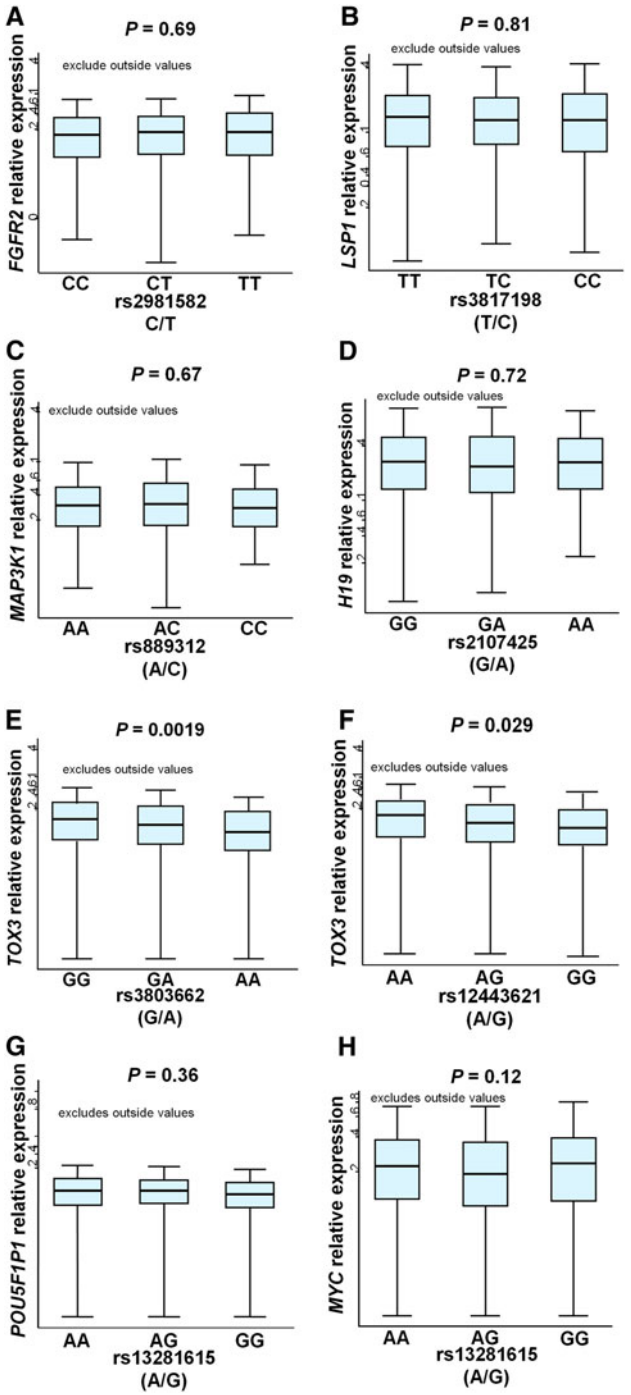

Fig. 2 Correlation of mRNA expression level of genes with genotype of their respective SNPs in 1,401 primary breast tumors: **a** *FGFR2,* **b** *LSP1*, **c** *MAP3K1*, **d** *H19*, **e** *TOX3,* **f** *TOX3,* **g** *POU5F1P1*, and **h** *MYC*. Kruskal–Wallis test was used to test for significant differences between the respective three genotype groups. Allele changes go from major allele to minor allele (major/minor). The rs3803662 and rs12443621 SNPs are located on the same LD block previously shown to be associated with breast cancer risk and these SNPs are correlated with each other ($r = 0.65$)

which has the weakest association with breast cancer risk, did not show an association with *TOX3* mRNA expression. Correlation analysis of these three SNPs near *TOX3* indicated that the highest risk-associated SNP rs3803662 showed a stronger correlation with rs12443621 than with rs8051542 (Pearson correlation coefficient $r = 0.65$ and $r = 0.20$, respectively). For none of the other loci investigated, did we observe a significant correlation of tumor mRNA expression of the analyzed nearby gene(s) with the tumor genotype of the corresponding SNP (Fig. 2). This includes the risk conferring SNP rs2981582 previously connected to expression modulation of *FGFR2* gene [7]. Also the reported physical interactions in breast cancer for rs13281615 in 8q24 with the *MYC* gene [16] did not translate into an altered *MYC* mRNA expression in primary tumors of our cohort (Fig. 2). Furthermore, the genotype of this SNP in the subset of breast tumors for which amplification status was available ($n = 163$) did not correlate with *MYC* amplification (Supplementary Table S3).

Discussion

Several large case–control studies have discovered various SNPs being associated with breast cancer risk [1, 17, 18]. Among these newly discovered SNPs, the associations of nine SNPs tagging seven low-risk breast cancer loci have been independently replicated in many populations of different ethnic backgrounds [3–6]. Following their discovery, some of these loci were found to be associated with important clinical and pathological characteristics of the resulting invasive breast cancer. Rs2981582 in *FGFR2* and rs13282615 in 8q24 have been shown to be related to ER-positive and PgR-positive tumors, whereas rs12383662 in *TOX3* was associated with ER-negative tumors [19]. Rs13387042 on 2q35 has been shown to be associated with both ER-positive and ER-negative breast cancer [20]. Consistent with previous findings, we also found rs2981582 in *FGFR2* to be associated with ER-positive and PgR-positive tumors in our cohort. The other associations with hormone receptor status could not be confirmed.

Additionally, the present retrospective study primarily aimed to address two important questions with regard to these risk alleles: firstly, do these alleles also affect the prognosis of the patients and, secondly, do these alleles affect the expression of adjacent genes?

To address the first question, whether these SNPs were associated with the natural course of the disease, i.e., tumor aggressiveness, we focused in this retrospective study on the length of MFS in patients with lymph node-negative disease who did not receive any systemic adjuvant treatment. As adjuvant systemic therapy has a confounding effect on the development of metastases, a population like this one is the most optimal to address such question. Of all the risk alleles investigated, only homozygosity for risk alleles of rs2107425 (AA) in the imprinted gene *H19* was significantly

associated with poor prognosis in both univariate and multivariate analysis. To our knowledge, this is the first report showing the pure prognostic value of one of these low-risk loci [1]. However, the underlying biological mechanism accounting for this association is not yet evident. Since this SNP near *H19* does not affect *H19* mRNA expression, it is unlikely that this risk allele and prognostic SNP operates through altering *H19* mRNA expression. This locus is located within the well-known imprinted *IGF2*/*H19* locus on chromosome 11, which contains various imprinted transcripts [22, 23]. Loss of imprinting and/or altered expression of other transcripts in this locus may be causal. Lack of an association of other loci with prognosis could be related to lack of power of our study, but may also indicate that most risk alleles only relate to onset of breast cancer and not to its progression once established. One limitation of our study is that SNP genotype analysis was performed on tumor and not on germline DNA, which is not available for this cohort, and that somatic acquired genetic changes in tumor tissues might affect SNP genotype analysis. To address this, we analyzed a subset of the cohort from which we also had copy number variation data available [21]. In this subgroup, we found no evidence that copy number variation in the regions containing risk-associated SNPs significantly affected SNP genotype frequency. Therefore, we have no evidence that genetic changes in tumor samples of our cohort have skewed the presented SNP genotype data.

With regard to the second hypothesis that whether these loci modulate the expression of the nearest genes, we measured the transcript expression levels of genes located nearby the risk loci in 1,401 primary breast tumors. For transcript expression analysis, we only focused on the most plausible, non-hypothetical genes in or nearest to the haplotype block containing the risk-associated loci. The rs2981582 SNP located in the intronic region of the *FGFR2* gene was not related to differences in *FGFR2* mRNA expression. This is puzzling as the *FGFR2* gene has been shown to be amplified and over-expressed in 5–10% of breast cancer patients [24], and because the risk allele itself alters the binding affinity for transcription factors of Oct-1/Runx2 and C/EBP*β,* which has been causally linked with increased expression of *FGFR2* in breast cancer cell lines [7]. However, such findings were not seen by Sun et al. [8], who reported an association of the normal allele with higher *FGFR2* expression in 81 samples. Of note, Sun et al. [8] analyzed normal breast tissue and not tumor tissue. All together, it is unlikely that this SNP acts through altering *FGFR2* mRNA level in breast tumor tissue.

We did not reveal an association of *MYC* transcript expression or *MYC* amplification with SNP genotype rs13281615 located in 8q24. Rs13281615 has no annotated gene in its vicinity. Nevertheless, a recent study showed a tissue-specific physical interaction of enhancers in the 8q24 region with *MYC* expression [16]. However, our analysis does not provide further support for a link between *MYC* mRNA expression and this risk allele. We also measured transcript levels of the class 5 homeobox transcription factor *POU5F1P1*, the nearest gene to this locus, which was recently shown to encode a functional nuclear transcriptional activator [25], and to be over-expressed in prostatic carcinoma [26]. With our assay we could not establish a significant correlation between *POU5F1P1* mRNA expression and the rs13281615 genotype in 1,401 primary breast tumors. It should however be noted that our assay as recently indicated by Applied Biosystems also detects the pseudogene *POU5F1P3* located on chromosome 12, and *POU5F1*, an intronless gene located on chromosome 6 that can encode a protein highly similar to POU class 5 homeobox 1.

For all other loci, we also did not observe any correlation of mRNA expression of the nearest gene with the genotype of the corresponding SNPs (Fig. 2), except for the risk loci (rs12443621 and rs3803662) on 16q12**.** These SNPs reside in the LD block containing the 5′ end of the *TOX3* gene and a hypothetical gene, *LOC643714*. A recent fine mapping of the LD block has further narrowed this risk-associated locus down to a region of 133 kb spanning the entire coding part of *LOC643714* and a small intergenic region. Moreover, this region exhibits evolutionary conservation and open chromatin conformation suggesting a regulatory function [27]. Our study shows that the alleles in this locus that have strongest association with risk (rs3803662 and rs12443621) were significantly related with a lower expression of *TOX3* mRNA. The lower expression of the *TOX3* gene, which encodes a high mobility group box nuclear protein involved in calcium dependent transcription and structural modification of chromatin, in tumors with risk alleles of rs3803662, suggests that *TOX3* might act as a tumor suppressor gene. It is important to note that although the risk alleles of rs3803662 are significantly associated with lower expression of *TOX3,* yet another recent study showed an association of the rs3803662 SNP with an increased expression of the *RBL2* gene, which is also located nearby this risk allele [27]. Although this study including only 77 breast tumors is underpowered compared with our analysis, it might however, be possible that this locus differentially regulates more than one distant gene in *cis* or in *trans*.

In summary, this study for the first time estimates the pure prognostic value of SNPs that are associated with low breast cancer risk. We have found that homozygous risk alleles of rs2107425 SNP were associated with an aggressive course of the disease. Secondly, we associated these low risk SNPs with expression of nearby genes. Based on our findings that the risk alleles of rs3803662 and rs12443621 near the *TOX3* gene were associated with a lower expression of *TOX3*

mRNA, we hypothesize a tumor suppressor role of this gene. However, functional studies to determine reliably the role of the *TOX3* gene in both normal and malignant settings are required. Finally, the breast cancer risk posed by rs2981582 SNP in *FGFR2* and rs13181615 in the 8q24 region may not operate through the expression modulation of *FGFR2* and *MYC* genes, respectively.

Acknowledgments The authors are especially grateful to Pascal Arp, Iris van Staveren, Marion Meijer-van Gelder, and Mieke Timmermans for their contribution and technical support. The authors thank also the surgeons, pathologists, and internists of St. Clara Hospital, Ikazia Hospital, St. Franciscus Gasthuis at Rotterdam, and Ruwaard van Putten Hospital at Spijkenisse for the supply of tumor tissues and/or their assistance in the collection of clinical follow-up data. This study was in part financed by the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research. MR was funded by the Higher Education Commission government of Pakistan.

Conflict of interest All authors declared that they have no competing interests.

References

1. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447:1087–1093
2. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J et al (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 39:865–869
3. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A et al (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet 42:504–507
4. McInerney N, Colleran G, Rowan A, Walther A, Barclay E et al (2009) Low penetrance breast cancer predisposition SNPs are site specific. Breast Cancer Res Treat 117:151–159
5. Hemminki K, Muller-Myhsok B, Lichtner P, Engel C, Chen B et al (2010) Low-risk variants FGFR2, TNRC9 and LSP1 in German familial breast cancer patients. Int J Cancer 126: 2858–2862
6. Liang J, Chen P, Hu Z, Shen H, Wang F et al (2010) 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
7. Meyer KB, Maia AT, O'Reilly M, Teschendorff AE, Chin SF et al (2008) Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer. PLoS Biol 6:e108
8. Sun C, Olopade OI, Di Rienzo A (2010) rs2981582 is associated with FGFR2 expression in normal breast. Cancer Genet Cytogenet 197:193–194
9. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M et al (2006) REporting recommendations for tumor MARKer prognostic studies (REMARK). Breast Cancer Res Treat 100:229–235
10. Foekens JA, Portengen H, van Putten WL, Peters HA, Krijnen HL et al (1989) Prognostic value of estrogen and progesterone receptors measured by enzyme immunoassays in human breast tumor cytosols. Cancer Res 49:5823–5828
11. Berns EM, Klijn JG, van Staveren IL, Portengen H, Noordegraaf E et al (1992) Prevalence of amplification of the oncogenes c-myc, HER2/neu, and int-2 in one thousand human breast tumours: correlation with steroid receptors. Eur J Cancer 28: 697–700
12. Hsieh SM, Look MP, Sieuwerts AM, Foekens JA, Hunter KW (2009) Distinct inherited metastasis susceptibility exists for different breast cancer subtypes: a prognosis study. Breast Cancer Res 11:R75
13. Sieuwerts AM, Meijer-van Gelder ME, Timmermans M, Trapman AM, Garcia RR et al (2005) How ADAM-9 and ADAM-11 differentially from estrogen receptor predict response to tamoxifen treatment in patients with recurrent breast cancer: a retrospective study. Clin Cancer Res 11:7311–7321
14. van Agthoven T, Sieuwerts AM, Meijer-van Gelder ME, Look MP, Smid M et al (2009) Relevance of breast cancer antiestrogen resistance genes in human breast cancer progression and tamoxifen resistance. J Clin Oncol 27:542–549
15. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M (2010) Genetic susceptibility to breast cancer. Mol Oncol 4:174–191
16. Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L et al (2010) 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc Natl Acad Sci USA 107:9742–9746
17. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A et al (2008) Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci USA 105:4340–4345
18. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S et al (2009) A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 41:579–584
19. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J et al (2008) Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet 4:e1000054
20. Milne RL, Benitez J, Nevanlinna H, Heikkinen T, Aittomaki K et al (2009) Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042. J Natl Cancer Inst 101:1012–1018
21. Zhang Y, Martens JW, Yu JX, Jiang J, Sieuwerts AM et al (2009) Copy number alterations that predict metastatic capability of human breast cancer. Cancer Res 69:3795–3801
22. Tilghman SM (1999) The sins of the fathers and mothers: genomic imprinting in mammalian development. Cell 96:185–193
23. Tycko B, Efstratiadis A (2002) Genomic imprinting: piece of cake. Nature 417:913–914
24. Adnane J, Gaudray P, Dionne CA, Crumley G, Jaye M et al (1991) BEK and FLG, two receptors to members of the FGF family, are amplified in subsets of human breast cancers. Oncogene 6:659–663
25. Panagopoulos I, Moller E, Collin A, Mertens F (2008) The POU5F1P1 pseudogene encodes a putative protein similar to POU5F1 isoform 1. Oncol Rep 20:1029–1033
26. Kastler S, Honold L, Luedeke M, Kuefer R, Moller P et al (2010) POU5F1P1, a putative cancer susceptibility gene, is overexpressed in prostatic carcinoma. Prostate 70:666–674
27. Udler MS, Ahmed S, Healey CS, Meyer K, Struewing J et al (2010) Fine scale mapping of the breast cancer 16q12 locus. Hum Mol Genet 19:2507–2515