npg

ARTICLE

C6ORF97-ESR1 breast cancer susceptibility locus: influence on progression and survival in breast cancer patients

Mutsuko Yamamoto-Ibusuki[1], Yutaka Yamamoto[1,2], Saori Fujiwara[1], Aiko Sueta[1,2], Satoko Yamamoto[1], Mitsuhiro Hayashi[1], Mai Tomiguchi[1], Takashi Takeshita[1] and Hirotaka Iwase*,[1]

Genome-wide association studies have identified a single-nucleotide polymorphism (SNP) to be associated with an increased risk of breast cancer. The biology of one of the susceptibility locus C6ORF-ESR1 and whether it also contributes to progression of established disease has not yet been ascertained. We examined the association of rs2046210 and its six linkage disequilibrium SNPs with clinicopathological characteristics, prognosis, and gene expression levels of ESR1 and the C6ORFs (C6ORF97:CCDC170, C6ORF211, C6ORF96:RMND1) in 344 breast cancer tissue samples and 253 corresponding samples of adjacent normal tissue. Tumor genotypes with homozygous risk alleles were more frequent than normal tissues. The tumor genotypes of rs2046210 and rs6929137 with homozygous risk alleles showed worse relapse-free survival (RFS, $P=0.038$ and $P=0.031$, respectively), whereas no notable associations were observed with either clinicopathological characteristics or expression of the peripheral genes. Higher C6ORF97 expression correlated with ER negativity ($P<0.0001$), highly proliferative characteristics ($P=0.0005$ for Ki67, $P<0.0001$ for nuclear grade) and worse RFS in the ER+/HER2− cohort ($P=0.013$), whereas the other two C6ORFs showed the inverse associations. Furthermore, C6ORF97 showed significant worse prognostic values especially in luminal B subtype in the publically available data sets. rs2046210 and the upstream gene C6ORF97 might have substantial roles not only in carcinogenesis but also in progression toward a more aggressive phenotype in breast cancer patients, which suggests that functional studies of this locus are imperative.

INTRODUCTION

ESR1, which encodes estrogen receptor α (ERα), has long been studied in correlation with breast carcinogenesis, development, and therapeutics. Elevated estrogen levels have been associated with increased breast cancer risk because estrogens can stimulate the proliferation of mammary epithelial tissue, mainly by interacting with the ER, and altering the expression of downstream genes.[1] It has been assumed for a long time that the action of estrogens in carcinogenesis is via the ERα signaling leading to the development of ERα-positive tumors.

Recently, Zheng et al[2] reported that the single-nucleotide polymorphism (SNP) rs2046210, which is located 29 kb upstream of the first untranslated region of the ESR1 gene, and also 180 kb upstream of the transcription start site of its first exon and 6 kb downstream of the C6ORF97/CCDC170 gene, increased breast cancer risk in a genome-wide association study in a large sample of the Chinese population. Further, the association was observed to be stronger for ERα-negative tumors than for ERα-positive tumors. Turnbull et al[3] confirmed the risk of breast cancer associated with rs3757318, in intron 7 of C6ORF97, especially in a western population. Also, in a Japanese population, Sueta et al[4] analyzed these two SNPs and showed that only rs2046210 was significantly associated with the risk of breast cancer. Cai et al[5] also reported that rs2046210 was associated with risk in a much larger Japanese population and found that the homogeneity of

SNPs between the Chinese and Japanese showed similar patterns. Their results showed that rs2046210 is more evident in ER-negative compared with ER-positive breast cancer in East Asian women including Japanese, but not in women of European ancestry.

Using ancestry-shift refinement mapping, Stacey et al[6] focused on the identification of a panel of linkage disequilibrium (LD) SNPs located in the region of C6ORF97-ESR1; rs12662670, rs6929137, rs3734805, rs9383589, rs12665607, rs2046210, and rs9397435 (from upstream to downstream) throughout Asian, European and African populations. Among them, they suggested that not rs2046210 but rs9397435 might be a pathogenic variant associated with common breast cancer risk among these three main ancestries. They tested whether the rs9397435 variant affected the levels of ESR1 or progesterone receptor (PGR) mRNAs in more than 1000 tumor samples from western breast cancer patients and showed that expression of the homozygous risk allele genotype corresponded to a four- to fivefold increase in the expression levels of ESR1 and PGR compared with the other genotypes. They speculated that the rs9397435 risk allele might act to increase expression of ESR1 and, as a consequence, increase PGR expression, which might affect breast cancer risk through modulating ER signaling or having some other impact on positive transcription of additional genes in that region. The variant genotype of SNP rs2046210 was also reported to be associated

[1]Department of Breast and Endocrine Surgery, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan; [2]Department of Molecular Targeting Therapy for Breast Cancer, Graduate School of Medical Sciences, Kumamoto University, Chuo-ku, Kumamoto, Japan
*Correspondence: Dr H Iwase, Department of Breast and Endocrine Surgery, Graduate School of Medical Sciences, Kumamoto University, 1-1-1 Honjo, Kumamoto, Kumamoto 860-8556, Japan. Tel: +81-96-373-5521; Fax: +81-96-373-5525; E-mail: hiwase@kumamoto-u.ac.jp
Received 22 March 2014; revised 31 August 2014; accepted 16 September 2014; published online 5 November 2014

with increased tumor ERα expression when assessed by immunohistochemistry.[7] On the other hand, Dunbier et al[8] demonstrated that ESR1 and its upstream genes C6ORF97 (CCDC170), C6ORF211, and C6ORF96 (RMND1) are co-regulated at this locus in breast cancer specimens treated with neoadjuvant endocrine therapy. They showed that a positive correlation between ESR1 and these ORFs was highly significant in their cDNA microarray data set, and in all of the publicly available data sets mainly of ERα positive tumors they examined. They also reported that C6ORF97 gene expression correlated negatively with the proliferation metagene, which is the reverse of C6ORF211, and higher quartile population of C6ORF97 gene expression, increased four- to fivefold, was predicted for improved disease-free survival in a tamoxifen-treated published data set, but not for C6ORF211. Overall, the precise functions of these genes have not yet been clearly elucidated.

If the SNPs or peripheral genes present at this locus are the mediators of changes in expression levels of ESR1, this would support the hypothesis that ER signaling is also important for the development of ER-negative breast cancer. Thus, we explored the relationship between these SNPs, ESR1, and upstream ORFs and also investigated their clinical significance for breast cancer patients to identify any putative relationships leading to breast carcinogenesis caused by this region.

SUBJECTS AND METHODS
Breast cancer patients and tissue samples
Three hundred and forty four breast tumors and 253 corresponding adjacent normal specimens from consecutive female patients with primary breast carcinoma, who had been treated at Kumamoto University Hospital between 2001 and 2008, were included in this study. The study was reported according to the Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) criteria.[9] All patients had undergone pretherapeutic biopsy or surgical treatment. Macrodissected samples (50 mg) were quickly frozen in liquid nitrogen and stored at $-80\,^{\circ}$C. All the treatment were decided by risk evaluation in accordance with the recommendations of the St Gallen international expert consensus on the primary therapy of early breast cancer.[10–12]

Genotyping of SNPs and gene expression analysis
Genomic DNA/RNA of tumor and DNA of adjacent normal specimens were extracted using the Allprep DNA/RNA Mini Kit (Qiagen, Tokyo, Japan) and Allprep DNA Mini Kit (Qiagen), respectively, following the manufacturer's protocol without any modification. DNA/RNA quantification was performed using NanoDrop 2000 spectrometer (NanoDrop Technologies, Wilmington, DE, USA) determined with the A260/A280 absorbance ratios.

We evaluated six SNPs of LD with rs2046210 among the C6ORF97-ESR1; namely rs12662670 (NC_000006.12:g.151597721T>G), rs6929137(NC_000006.12:g.151615542G>A), rs3734805 (NC_000006.12:g.151618215A>C), rs9383589 (NC_000006.12:g.151619125A>G), rs12665607 (NC_000006.12:g.151625494T>A), rs2046210 (NC_000006.12:g.151627231G>A), and rs9397435 (NC_000006.12:g.151630085A>G) for each subject. All the minor alleles are the risk (dominant) alleles for breast cancer susceptibility. Genotyping was carried out using TaqMan Single Nucleotide Polymorphism Genotyping Assays (Applied Biosystems Foster City, CA, USA) completely according to manufacturer's instruction with the ABI Prism 7900HT system (Applied Biosystems); namely assay ID: rs12662670 (C_32238365), rs6929137 (C_25642986), rs3734805 (PN4331349), rs9383589 (C_30434480), rs12665607 (C_32238349), rs2046210 (C_12034236), and rs9397435 (C_29586832). We used a X^2-test with one degree of freedom to verify that the allele distributions for each SNP were in Hardy–Weinberg equilibrium (HWE) and confirmed that these SNPs did not deviate from HWE in normal specimens. Data from this association study had been submitted to the ClinVar database at https://jira.ncbi.nlm.nih.gov/browse/MDI-3910.

Total RNA was isolated from 289 patients' tumor specimens. RNA qualification was performed using the Agilent 2100 Bioanalyzer (Expert Software version B.02.03) with RNA Nano LabChip Kits (Agilent Technologies, Stockport, UK). The specimens with RNA integrity number >6 were used for further step. Reverse transcription from 500 ng of RNA to cDNA was performed using PrimeScript RT Master Mix (Takara Bio Inc., Otsu, Japan) according to the manufacturer's instruction. Real-time quantitative PCR was performed using the relative quantification method with TaqMan chemistry with our in-house multiple reference genes[13] with the ABI Prism 7900HT system (Applied Biosystems): TaqMan gene expression assay for ESR1 (Hs01046815_m1), C6ORF97 (Hs01563345_m1), C6ORF211 (Hs00226188_m1), C6ORF96 (Hs00215537_m1), ACTB (Hs01060665_g1), PUM1 (Hs00982775_m1), and TAF-10 (Hs00359540_g1) (Applied Biosystems). All samples were amplified in a minimum of triplicates and a no-template-control was included in each reaction. Relative expression values of each gene per sample (the raw Ct data) were calculated by SDS 2.2 software (Applied Biosystems), defined as the point at which the fluorescence rises above the background fluorescence. Data Assist software (Applied Biosystems) was used to calculate relative gene expression by ΔC_{t} method normalized with the three reference genes.

Immunohistochemical analysis
Histological sections (4 μm) were deparaffinized and incubated for 10 min in methanol containing 0.3% hydrogen peroxide to block endogenous peroxidase. We used mouse monoclonal antibodies against ERα (SP1, Ventana Japan, Tokyo, Japan), PgR (1E2, Ventana Japan) and Ki67 (MIB1, Dako Japan, Tokyo, Japan), and a polyclonal antibody against Her2 (Dako Japan, 1:200); staining was carried out in the NexES IHC Immunostainer (Ventana Medical Systems, Tucson, AZ), in accordance with the manufacturer's instructions. ER and PgR status was evaluated based on the percentage of positively stained nuclei and each was considered positive when $\geq$1% of nuclei were stained.[14] Her2 immunostaining was evaluated using the same method as the HercepTest (Dako Japan, Tokyo, Japan); membranous staining was scored on a scale of 0 to 3+. Tumors with scores of $\geq$3 or with a $\geq$2.2-fold increase in Her2 gene amplification as determined by fluorescence in situ hybridization were considered to be positive for Her2 overexpression. Ki67 was scored as the percentage of nuclear-stained cells out of all cancer cells along the invasive front of the tumor in ×400 high-power fields; this gave the Ki67 labeling index.

Statistical analysis
Statistical analysis of the associations between different genotypes or gene expression status and clinicopathological factors was performed using the following tests; for the nonparametric variables, Wilcoxon (for univariable) and Kruskal–Wallis test (for multivariables). The X^2-test was adopted for categorical comparison. The Spearman rank correlation coefficients were used to assess the correlation between gene expression levels among C6ORF molecules and ESR1. Relapse-free survival (RFS) curves were calculated according to the Kaplan–Meier method and verified by the log-rank test. In the analysis of RFS, both local recurrences and distant metastases were considered as events. Among 38 recurrent cases, there were 23 cases of distant metastases and six local recurrences. Twenty patients died as a result of breast cancer. For Kaplan–Meier analysis using KMplot software, a database of public microarray data sets[15] (http://kmplot.com/analysis), the results from 3597 patients, was analyzed. Patients were divided according to the automatically generated best cutoff value, and luminal A, luminal B, endocrine-treated, or each subtype patients were included in the analysis. In all analyses, the statistical significance was defined as $P<0.05$. JMP software version 10.0.0 for Windows (SAS institute Japan, Tokyo, Japan) was used for all statistical analyses of our own cohort.

RESULTS
Tumor genotypes with homozygous risk alleles are more frequent than normal tissues
In Table 1, the minor allele frequency (MAF) in normal tissues was almost equal (0.22–0.25) compared with Japanese germline MAFs (0.23–0.33) identified in the HapMap project, and were slightly lower to those of Asian ancestry controls reported by Stacey et al (0.32–0.36).

Table 1 Comprehensive analysis of SNPs associated with breast cancer risk in normal and tumor tissue

		Japanese controls HapMap	Asian controls Stacey et al	Euorpean controls Stacey et al		Normal tissue genotypes				Tumor tissue genotypes		
SNPs ID	HGVS	MAF	MAF	MAF	N	Genotype	N	MAF	N	Genotype	N	MAF
rs12662670	NC_000006.12: g.151597721T>G	0.25	0.34	0.07	236	TT	140	0.22	343	TT	196	0.33
						TG	82			GT	118	
						GG	14			GG	29	
rs6929137	NC_000006.12: g.151615542G>A	0.29	0.35	0.32	253	GG	152	0.22	344	GG	176	0.34
						GA	86			GA	133	
						AA	15			AA	35	
rs3734805	NC_000006.12: g.151618215A>C	0.23	0.32	0.07	243	AA	133	0.24	340	AA	185	0.40
						AC	98			AC	105	
						CC	12			CC	50	
rs9383589	NC_000006.12: g.151619125A>G	0.24	0.32	0.07	228	AA	127	0.24	343	AA	188	0.38
						AG	86			AG	109	
						GG	15			GG	46	
rs12665607	NC_000006.12: g.151625494T>A	0.23	0.32	0.07	252	TT	131	0.25	344	AA	183	0.38
						TA	107			AT	115	
						AA	14			TT	46	
rs2046210	NC_000006.12: g.151627231G>A	0.33	0.36	0.34	245	GG	147	0.22	344	GG	176	0.33
						GA	82			GA	134	
						AA	16			AA	34	
rs9397435	NC_000006.12: g.151630085A>G	0.23	0.33	0.06	253	AA	139	0.24	344	AA	180	0.34
						AG	97			AG	128	
						GG	17			GG	36	

Abbreviations: HGV, human genome variation; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

The MAFs of tumor tissue were apparently higher than those of HapMap germline and normal tissue, where HWE collapsed in tumor genotypes: rs3734805 (X^2-test 24.2, $P<0.05$), rs938589 (X^2-test 18.7, $P<0.05$), and rs12665607 (X^2-test 14.5, $P<0.05$).

Tumor genotypes of rs2046210 and rs6929137 with homozygous risk alleles show worse prognoses

The SNP rs2046210 with homozygous risk alleles (AA: $n=32$) showed significantly lower survival ($P=0.038$; Figure 1a) compared with other genotypes, whereas tendency for survival correlation was observed in normal tissue ($P=0.10$; Figure 1b). In tumor-normal corresponding cases, genotype conversion to homozygous risk alleles from normal to tumor tissue were observed in 10 patients, which showed worse prognosis than the patients with no genotype conversion ($P=0.026$; Figure 1c). For rs6929137, there also were correlations with RFSs between the genotypes of tumor and normal tissues ($P=0.031$ in tumor tissue genotype; Figure 1d, $P=0.48$ in normal tissue genotype; Figure 1e). Genotype conversion to homozygous risk alleles were shown in 18 patients, which had marginal survival significance in RFS (Figure 1f) Other LD SNPs showed no correlation with prognoses (Supplementary Table 1). rs2046210 and rs6929137 showed no apparent correlation with almost all the factors including tumor gene expression levels of ESR1 or C6ORFs (Supplementary Table 2).

Higher C6ORF97 expression correlates with ER negativity, highly proliferative features and worse clinical outcome

C6ORF97 gene expression showed a significant inverse correlation with ESR1 gene expression levels (Spearman's $\rho=-0.22$, $P=0.00022$), and had a positive correlation with C6ORF211 (Spearman's $\rho=0.27$, $P<0.0001$) and C6ORF96 (Spearman's $\rho=0.36$, $P<0.0001$; Supplementary Table 3). C6ORF211 and C6ORF96 showed higher positive correlations with each other (Spearman's $\rho=0.89$, $P<0.0001$). C6ORF97 gene expression levels were higher in the group of patients with larger diameter tumors ($P=0.0041$), positive nodal status ($P=0.029$), higher nuclear grade ($P<0.001$), negative ER and PgR status ($P<0.001$), and positive HER2 status ($P=0.0058$), as well as higher Ki67 labeling index ($P=0.0005$; Table 2). Among the subtype groups, higher C6ORF97 gene expression was observed both in ER+HER2+ or ER-HER2+ and in ER−HER2− subgroups. Higher C6ORF211 gene expression levels were seen in the group of patients with lower nuclear grade ($P=0.0027$ for C6ORF211; $P=0.0094$ for C6ORF96), positive ER status ($P<0.0001$ for both C6ORF211 and C6ORF96), positive PgR status ($P<0.001$ for C6ORF211; $P=0.0037$ for C6ORF96), negative HER2 status ($P=P=0.0007$ for C6ORF211; $P=0.0014$ for C6ORF96), and no correlation with Ki67. These two ORFs' gene expression was high in the ER+HER2− subtype ($P<0.0001$ for both C6ORF211 and C6ORF96).

The prognostic relevancies of C6ORFs are shown in Figure 2 and Figure 3 (for RFS). A C6ORF97 gene expression level of 0.080 (high,

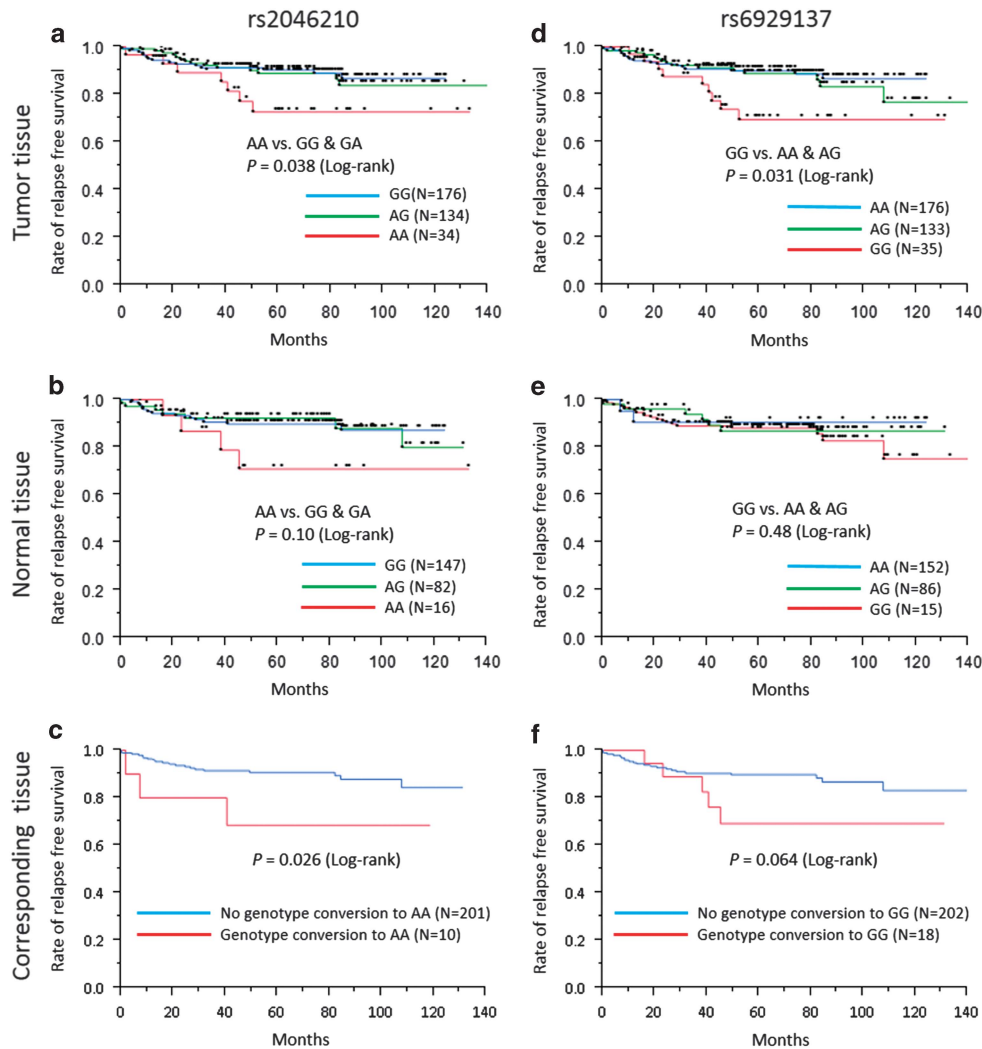**Figure 1** RFS according to the genotype of rs2046210 and rs6929137 in normal and tumor tissue. (**a**) rs2046210 in tumor tissue and RFS. Homozygous risk alleles (AA: $n=32$) showed significantly lower survival (log-rank correlation $P=0.038$) compared with other genotypes in tumor tissue. (**b**) rs2046210 in normal tissue and RFS. No significant survival correlation was observed (log-rank correlation $P=0.10$). (**c**) rs2046210 genotype change in tumor-normal corresponding tissue and RFS. Genotype conversion to risk alleles (AA) showed a significant survival correlation (log-rank correlation $P=0.026$). (**d**) rs6929137 in tumor tissue and RFS. rs6929137 with homozygous risk alleles (GG: $n=33$) showed a correlation with RFS (log-rank correlation $P=0.031$). (**e**) rs6929137 in normal tissue and RFS. No significant survival correlation was observed (log-rank correlation $P=0.48$). (**f**) rs2046210 genotype change in tumor-normal corresponding tissue and RFS. Genotype conversion to risk alleles (GG) showed marginal survival correlation (log-rank correlation $P=0.064$).

$n=37$; low, $n=252$), a C6ORF211 gene expression level of 0.72 (high, $n=171$, low, $n=118$), and a C6ORF96 gene expression level of 0.72 (high, $n=225$, low, $n=64$) were identified as providing the most significant association with RFS. In this context, higher C6ORF97 gene expression correlated with a significantly poorer prognosis both in RFS ($P<0.0001$; Figure 2a) and BCSS ($P<0.0001$; Supplementary Figure 1a) for the entire cohort. In contrast, higher gene expression levels of C6ORF211 and C6ORF96 tended to correlate with better prognosis ($P=0.0054$ for C6ORF211, $P=0.0084$ for C6ORF96; Figures 2b and c, respectively) As for BCSS, only C6ORF211 showed significant correlation with better survival ($P=0.0040$; Supplementary Figure 1d). Further, we studied the prognostic value of C6ORFs in the ER +/HER2 − subtype ($n=214$) of our cohort. Only higher C6ORF97 gene expression was significantly associated with both poor RFS ($P=0.013$; Figure 2b) and BCSS ($P=0.0045$; Supplementary Figure 1b), however not in other subtypes.

C6ORF97 has opposing prognosis within the ER+/HER2- subtype

To identify the detailed prognostic value of C6ORFs in the ER +/HER2 − population, we used a web-based analysis tool with the public data source of microarray platforms.and tested the RFS value affected by C6ORF gene expression in the category of luminal A and B subtype (Figure 3). In the untreated cohort, higher C6ORF97 gene expression correlated with better prognosis ($P=0.0045$; Figure 3a), otherwise the high C6ORF97 gene expression population seemed to show worse prognosis, which was not significant in the untreated luminal B cohort ($P=0.19$; Figure 3d). In the second cohort treated only with hormonal therapy, this tendency of inverse correlation among the two luminal subtypes was more clearly observed ($P=0.024$ for luminal A, $P=0.053$ for luminal B; Figure 3e vs d). This distinctive inverse correlation of prognosis was also observed (Figure 3b vs e). In C6ORF211, the higher expression of C6ORF211 was correlated with worse prognosis mainly in the luminal B cohort (Supplementary

Table 2 Relationship between gene expression levels of *C6ORF*s near *ESR1* and clinicopathological characteristics

Clinical parameters	No. of patients	C6ORF97			C6ORF211			C6ORF96		
		Median	(25%, 75%)	P-value	Median	(25%, 75%)	P-value	Median	(25%, 75%)	P-value
Menopause										
Pre –	78	0.022	(0.0081, 0.044)	0.23	0.12	(0.0051, 038)	0.19	0.15	(0.0051, 038)	0.34
Post –	210	0.019	(0.0097, 0.056)		0.097	(0.040, 0.28)		0.17	(0.040, 0.28)	
Invasive diameter (mm)										
0<, <20	129	0.015	(0.011, 0.053)	0.0041	0.11	(0.040, 0.32)	0.4	0.15	(0.070, 0.41)	0.89
≥20	160	0.025	(0.0073, 0.044)		0.1	(0.043, 0.28)		0.16	(0.065, 0.39)	
Nodal status										
–	181	0.018	(0.0082, 0.042)	0.029	0.11	(0.043, 0.31)	0.94	0.15	(0.068, 0.39)	0.78
+	108	0.025	(0.0092, 0.058)		0.1	(0.047, 0.28)		0.17	(0.066, 0.40)	
Clinical stage										
0	14	0.0073	(0.0042, 0.02)	0.001	0.073	(0.0044, 0.02)	0.027	0.096	(0.36, 0.41)	0.19
I	106	0.014	(0.0064, 0.032)		0.1	(0.042, 0.31)		0.15	(0.070, 0.41)	
II	155	0.026	(0.010, 0.055)		0.12	(0.048, 0.37)		0.18	(0.068, 0.42)	
III	28	0.029	(0.014, 0.071)		0.057	(0.021, 0.12)		0.1	(0.044, 0.24)	
Nuclear grade										
1	146	0.016	(0.0069, 0.03)	<0.0001	0.13	(0.051, 0.38)	0.0027	0.18	(0.093, 0.54)	0.0094
2	71	0.021	(0.0081, 0.049)		0.1	(0.050, 0.29)		0.15	(0.067, 0.41)	
3	70	0.042	(0.022, 0.073)		0.057	(0.019, 0.22)		0.092	(0.043, 0.30)	
ERα (IHC)										
–	64	0.038	(0.019, 0.074)	<0.0001	0.037	(0.016, 0.10)	<0.0001	0.075	(0.036, 0.19)	<0.0001
+	225	0.018	(0.0075, 0.039)		0.13	(0.058, 0.38)		0.18	(0.085, 0.47)	
PgR (IHC)										
–	98	0.031	(0.014, 0.068)	<0.0001	0.058	(0.021, 0.17)	<0.0001	0.11	(0.046, 0.29)	0.0037
+	191	0.017	(0.0069, 0.039)		0.14	(0.059, 0.38)		0.18	(0.084, 0.41)	
HER2 (IHC/FISH)										
–	249	0.019	(0.0081, 0.045)	0.0058	0.11	(0.050, 0.33)	0.0007	0.17	(0.072, 0.43)	0.0014
+	40	0.034	(0.013, 0.058)		0.043	(0.017, 0.12)		0.081	(0.044, 0.23)	
Ki67 (IHC LI)										
<0.20	121	0.014	(0.0068, 0.041)	0.0005	0.12	(0.054, 0.32)	0.063	0.18	(0.087, 0.42)	0.099
≥0.20	160	0.027	(0.012, 0.057)		0.083	(0.033, 0.25)		0.13	(0.058, 0.34)	
Tumor subtype										
ER+/HER2–	214	0.017	(0.0073, 0.0407)	<0.0001	0.15	(0.059, 0.39)	<0.0001	0.19	(0.094, 0.52)	<0.0001
ER+ or –/HER2+	38	0.035	(0.014, 0.060)		0.046	(0.018, 0.13)		0.078	(0.043, 0.21)	
ER-/HER2-	37	0.031	(0.017, 0.084)		0.043	(0.017, 0.071)		0.075	(0.033, 0.13)	

Abbreviations: ER, estrogen receptor; FISH, fluorescent *in situ* hybridization; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; LI, labeling index; PgR, progesterone receptor.

Figures 2b–e). In addition, *C6ORF96* gene expression showed a marked significant impact on RFS (Supplementary Figures 3a–f).

DISCUSSION

Our study is the first, comprehensive analysis to estimate the prognostic value of rs2046210, its LD SNPs, and surrounding *C6ORF*s-*ESR1* gene expression, which have been expected to associate with breast cancer risk. The genotypes are suggested to undergo conversion to risk alleles during the development from normal gland to tumor, and the tumor genotypes of rs2046210 and rs6929137 with homozygous risk alleles were found to correlate with aggressive course

of disease, with no relation to expression of the peripheral genes. On the other hand, higher *C6ORF97* expression correlates with ER negativity, highly proliferative features and worse clinical outcomes. *C6ORF211* and *C6ORF96* showed the opposite effects. We demonstrate that *C6ORF97* may have different roles between luminal A and B breast cancer, as revealed by its opposite prognostic value in each subtype.

Since Zheng *et al*[17] reported the novel relationship of the SNP rs2046210, many studies have been conducted to replicate this finding in diverse populations. A subset of these was summarized in the meta-analytic study[16] up to 14 evaluable populations by Zheng *et al*. In the
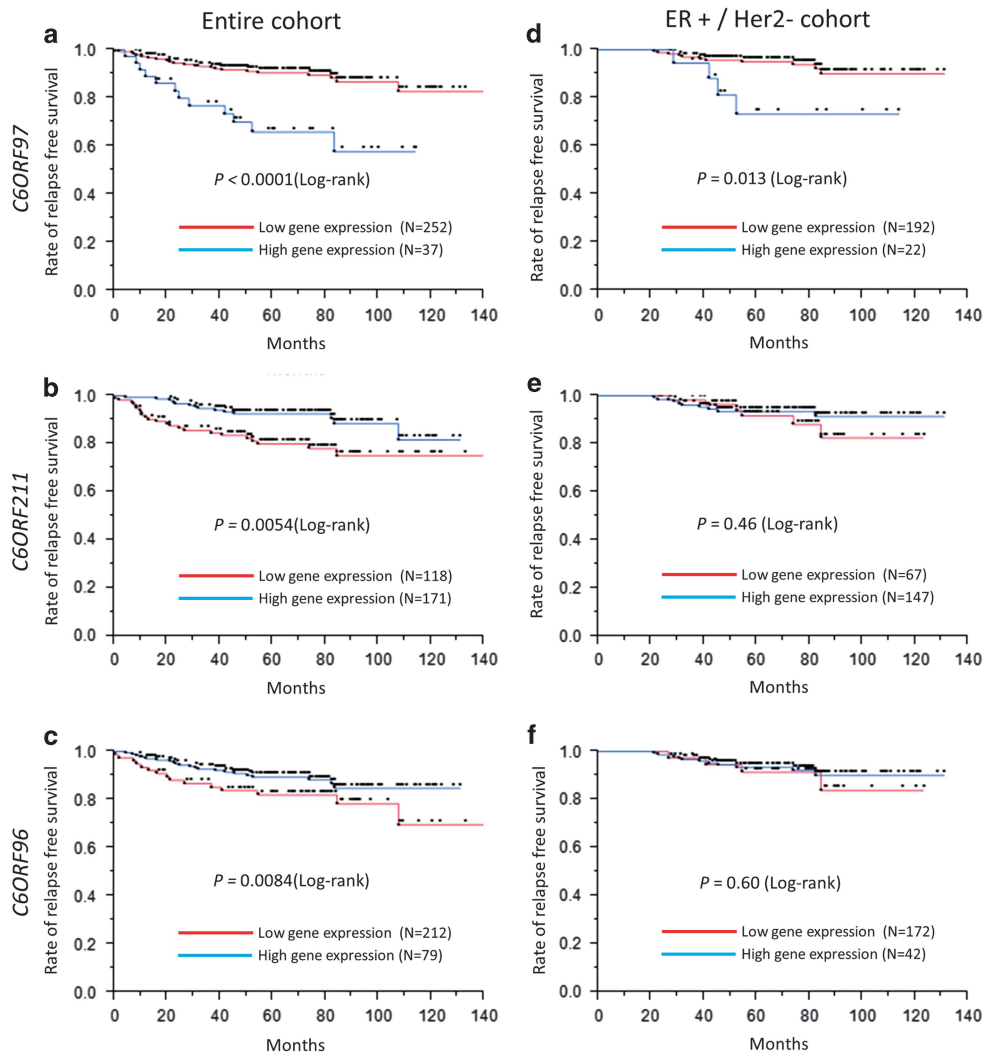
Figure 2 RFS according to tumor gene expression levels of *C6ORFs* among the entire cohort (**a**–**c**) and the ER+/HER2– cohort (**d**–**f**). (**a**) Patients with higher *C6ORF97* gene expression showed significantly poorer prognosis in RFS (log-rank correlation *P*<0.0001) for the entire cohort. Higher gene expression levels of *C6ORF211* (**b**) and *C6ORF96* (**c**) tended to correlate with better prognosis, whereas there were no significant findings in the ER+/HER2– cohort ((**e**) *C6ORF211*, (**f**) *C6ORF96*).

Japanese population, Sueta *et al* reported that rs2046210 had an OR of 1.44 (95%CI 1.18–1.74, by dominant model). Mizoo *et al*[18] also reported the odds ratio of 1.33 (95%CI 1.03–1.72 by dominant model) in their independent population. Mizoo *et al* recently showed that higher breast density (*P*=0.027) in the control cohort and lower body mass index (BMI) (*P*=0.0022) in the breast cancer cohort correlate with the risk allele of rs2046210, which tends to be associated with negative ER status of the tumor. In our cohort, rs9383589 in adjacent normal tissue showed a correlation with lower BMI (*P* trend = 0.02; Supplementary Table 2); and rs6929137 in adjacent normal tissue showed a significant correlation with lower ER positivity (GG *vs* AA and AG: median 60 *vs* 80%, *P*=0.013 by Wilcoxon test; data not shown). In addition, the genotype with homozygous risk alleles of rs2046210 in adjacent normal tissue showed significant correlation with high Ki67 LI (AA *vs* GG and GA: median 0.25 *vs* 0.41, *P*=0.025 by Wilcoxon test; data not shown). In other cohorts, Drury *et al*[7] reported that the variant germline genotype of rs2046210 showed quite a modest effect on higher ERα-histoscore in tumor samples in a British population. Horne *et al*[19] showed that rs2046210 risk allele significantly suppressed E-cadherin expression (*P*=0.002). In all cases,

it is suggested that rs2046210 and surrounding SNPs in the germline may have some effects on its tumor characteristics.

The changes in genotype from adjacent normal to tumor tissue revealed that the genotype with homozygous risk alleles of both rs2046210 and rs6929137 showed prognostic enhancement in RFS (Figure 1). The genotype changes to risk alleles including hetero-geneous and homogenous genotypes were observed in 7.2 and 4.3% respectively, which was not higher than other SNPs, either no significant correlation with patients' clinical characteristics. We speculate it may not be deniable that there is some functional effects forward carcinogenesis or tumor progression from risk allele, with no relation to the expression of surrounding genes (Supplementary Table 2). Even when the tumor genotype showed no obvious correlation trend with clinical characteristics, there may exist some tendency for higher Ki67 LI in relation to the genotype of homozygous risk alleles (*P*=0.045 for rs9383589, *P*=0.069 for rs3734805, *P*=0.12 for rs2046210 and rs12665601; Supplementary Table 2). Riaz *et al*[20] reported that there was no correlation between the breast cancer susceptibility SNP rs2107425 genotype near *H19* was significantly associated with shorter survival, but had no
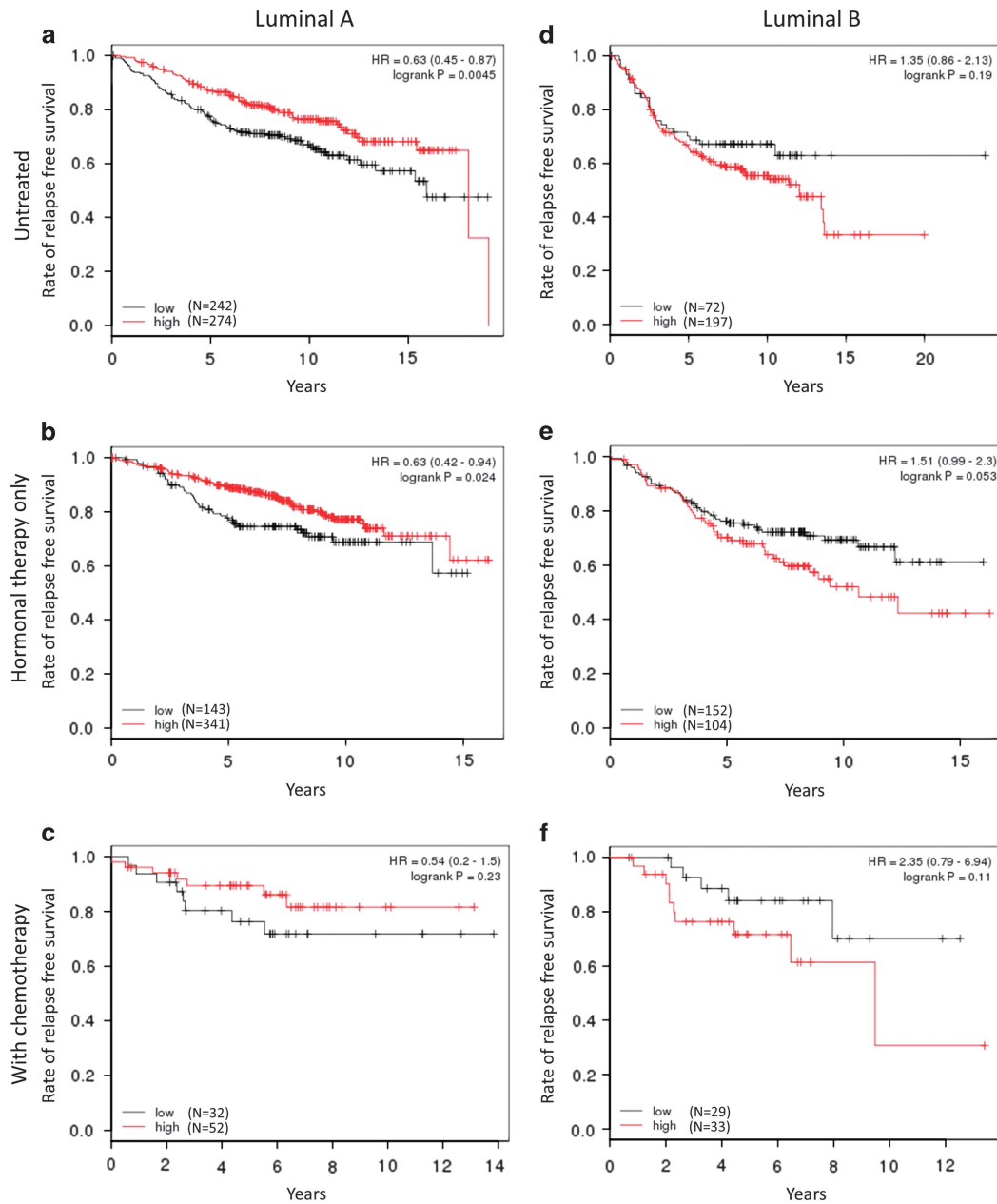
Figure 3 RFS according to the publically available gene expression data of *C6ORF97* among luminal A (**a**–**c**) and luminal B (**d**–**f**) subtypes, divided into three categories according to adjuvant therapies: untreated (**a** and **d**), treated with hormonal therapy alone (**b** and **e**), and treated with chemotherapy in addition to hormonal therapy (**c** and **f**). *C6ORF97* gene expression correlated with better prognosis in the luminal A subtype cohort (**a**–**c**). High *C6ORF97* gene expression seemed to show an inverse relationship, with worse prognosis in the luminal B subtype cohort (**d**–**f**).

relationship with the expression levels of *H19*. The underlying biological mechanism is not yet evident because these SNPs are located in non-coding regions of the genome. A large number of breast cancer susceptibility alleles have been detected, but to date none have been obviously associated with disease progression.[21] Cai et al[5] used MCF-7 *in vitro* and proved that rs6913578, which is in LD with rs2046210, may influence DNA binding protein interactions, but did not confirm interactions with any putative transcription factors involved in the regulation of nearby genes. In addition, there is no transcriptional factor binding site on rs2046210.[5] The mechanisms by which SNP rs2046210, or SNPs in LD with it, promote disease risk have remained elusive.

On the other hand, *ORF*s presented significant relationships with prognoses, most markedly in *C6ORF97* (Figure 2,Supplementary Figure 1). Dunbier et al[8] in the only previous report on *C6ORF*s, have described dissimilar findings that *C6ORF97* negatively correlated with proliferation in ER+ breast tumors and predicts better outcome in other adjuvant tamoxifen-treated public data sets. This was quite opposite to our results. Thus, we examined the motion of *C6ORF97* among the ER+ cohort in public database, and this revealed an inconsistent impact on prognosis between luminal A and luminal B regardless of the type of therapy. When limited to tamoxifen-treated patients, the results showed that high *C6ORF97* gene expression is related with good prognosis (Supplementary Figure 4a). However,

when divided into luminal A and luminal B, this association with prognosis was lost (Supplementary Figures 4b and c). In general, luminal B cancers are associated with early relapse following endocrine therapy and a prognosis that is partly similar to that of the aggressive basal subtype. Dunbier et al also reported that *C6ORF*s are co-expressed with *ESR1* in ER+ tumor specimens, which is supported by *in vitro* data that the upregulation of co-expression of these *ORF*s' was induced under estrogen deprivation in MCF-7 cells, generally subtyped in Luminal B. With regard to the distinctive feature of *C6ORF97* we observed in our cohort, this prognostic dissociation may be reasonable in relation to resistance against endocrine therapy generally shown in Luminal B breast cancer. There were a number of borderline significance (*P*-value around 0.05), thus the existence of accidental significance cannot be denied. If both our conflicting clinical data of *C6ORF97* and the prognostic dissociation between luminal A and B indicate potential functional differences of *C6ORF97-ESR1* lesion, it will be extremely interesting to further explore on rs2046210 and *C6ORF97* to reveal the mechanism of carcinogenesis and the connection to cancer progression, which might result in more complete risk assessment for relapse and thereby guide treatment choices.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Germain D. Estrogen carcinogenesis in breast cancer. *Endocrin Metab Clin* 2011; **40** 473–484.

2 Zheng W, Long J, Gao YT et al: Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 2009; **41**: 324–328.

3 Turnbull C, Ahmed S, Morrison J et al: Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* 2010; **42**: 504–507.

4 Sueta A, Ito H, Kawase T et al: A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. *Breast Cancer Res Treat* 2012; **132**: 711–721.

5 Cai Q, Wen W, Qu S et al: Replication and functional genomic analyses of the breast cancer susceptibility locus at 6q25.1 generalize its importance in women of chinese, Japanese, and European ancestry. *Cancer Res* 2011; **71**: 1344–1355.

6 Stacey SN, Sulem P, Zanon C et al: Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. *PLoS Genet* 2010; **6**: e1001029.

7 Drury S, Johnson N, Salter J et al: A breast cancer-associated SNP Adjacent to ESR1 correlated with oestrogen receptor-alpha level in invasive breast tumors. *PLoS Genet* 2011; **7**: e1001382.

8 Dunbier AK, Anderson H, Ghazoui Z et al: ESR1 is co-expressed with closely adjacent uncharacterised genes spanning a breast cancer susceptibility locus at 6q25.1. *PLoS Genet* 2011; **7**: e1001382.

9 McShane LM, Altman DG, Sauerbrei W et al: REporting recommendations for tumour MARKer prognostic studies (REMARK). *Eur J Cancer (Oxford, England: 1990)* 2005; **41**: 1690–1696.

10 Goldhirsch A, Wood WC, Gelber RD et al: Meeting highlights: updated international expert consensus on the primary therapy of early breast cancer. *J Clin Oncol* 2003; **21**: 3357–3365.

11 Goldhirsch A, Glick JH, Gelber RD et al: Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. *Ann Oncol* 2005; **16**: 1569–1583.

12 Goldhirsch A, Wood WC, Gelber RD et al: Progress and promise: highlights of the international expert consensus on the primary therapy of early breast cancer 2007. *Ann Oncol* 2007; **18**: 1133–1144.

13 Ibusuki M, Fu P, Yamamoto S et al: Establishment of a standardized gene-expression analysis system using formalin-fixed, paraffin-embedded, breast cancer specimens. *Breast cancer* 2013; **20**: 159–166.

14 Tapper W, Hammond V, Gerty S et al: The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. *Breast Cancer Res* 2008; **10**: R108.

15 Gyorffy B, Lanczky A, Eklund AC et al: An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat* 2010; **123**: 725–731.

16 Yang Z, Shen J, Cao Z, Wang B. Association between a novel polymorphism (rs2046210) of the 6q25.1 locus and breast cancer risk. *Breast Cancer Res Treat* 2013; **139**: 267–275.

17 Zheng W, Zhang B, Cai Q et al: Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. *Hum Mol Genet* 2013; **22**: 2539–2550.

18 Mizoo T, Taira N, Nishiyama K et al: Effects of lifestyle and single nucleotide polymorphisms on breast cancer risk: a case-control study in Japanese women. *BMC Cancer* 2013; **13**: 565.

19 Horne HN, Sherman ME, Garcia-Closas M et al: Breast cancer susceptibility risk associations and heterogeneity by E-cadherin tumor tissue expression. *Breast Cancer Res Treat* 2014; **143**: 181–187.

20 Riaz M, Berns EM, Sieuwerts AM et al: Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and mRNA expression of the nearest genes. *Breast Cancer Res Treat* 2012; **133**: 843–851.

21 Fasching PA, Pharoah PD, Cox A et al: The role of genetic breast cancer susceptibility variants as prognostic factors. *Hum Mol Genet* 2012; **21**: 3926–3939.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)