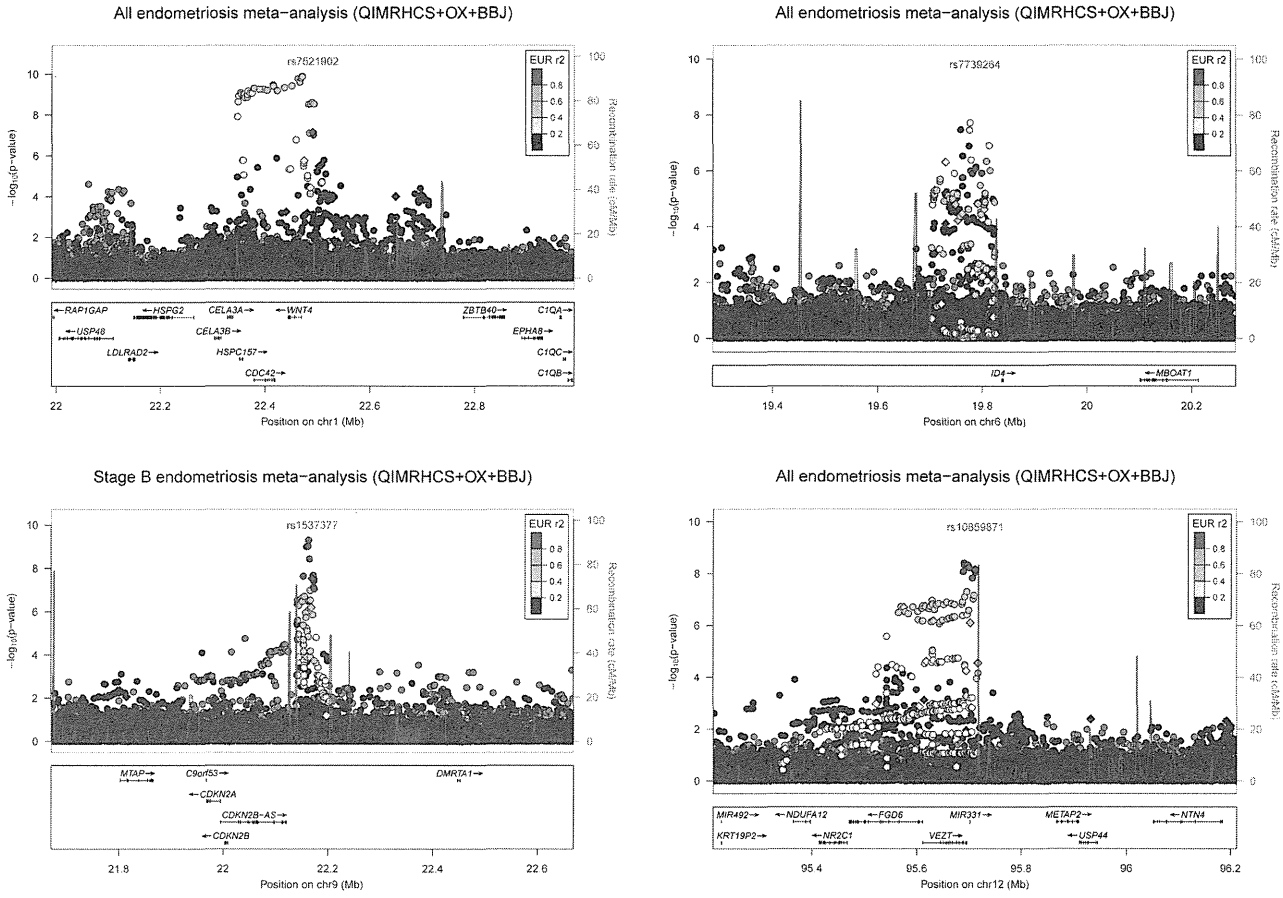

Figure 2.

Evidence for association with endometriosis from the QIMRHCS+OX+BBJ GWA meta-analysis across the 1p36.12 **(a)**, 6p22.3 **(b)**, 9p21.3 **(c)** and 12q22 **(d)** regions following imputation using the 1000 Genomes Project reference panel. Diamond and circle symbols represent genotyped and imputed SNPs, respectively. The most significant genotyped SNP is represented by a purple diamond. All other SNPs are color coded according to the strength of LD with the top genotyped SNP (as measured by r^2 in the European 1000 Genomes data).

Europe PMC Funders Author Manuscripts

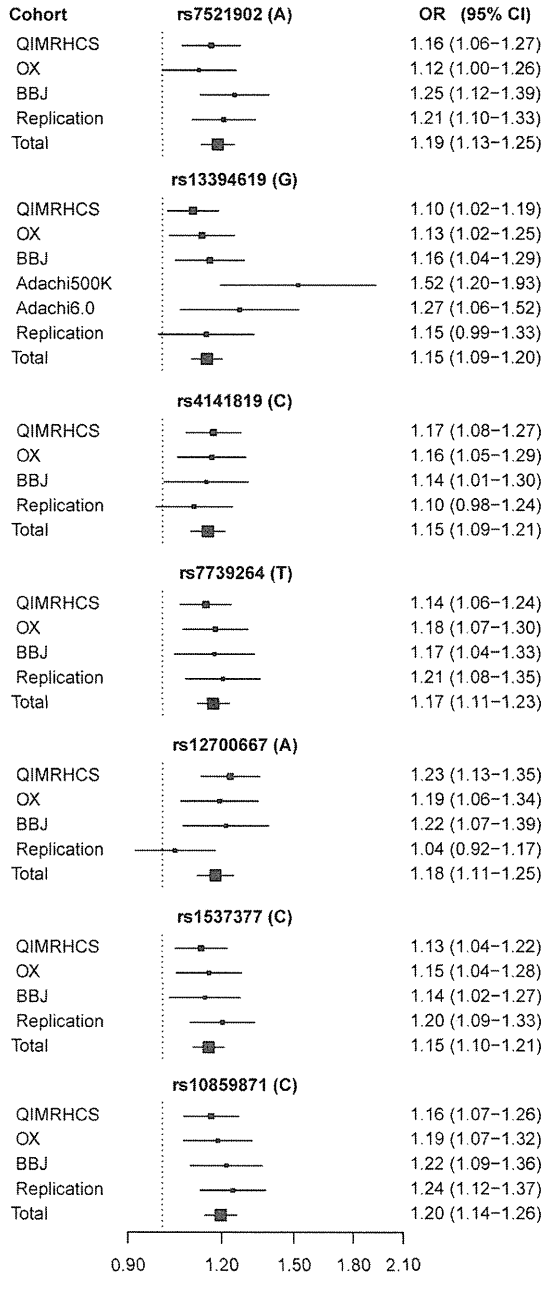

Figure 3.

Forest plots of risk allele effects for the seven genome-wide significant SNP loci in the individual and total endometriosis case-control cohorts.

Europe PMC Funders Author Manuscripts

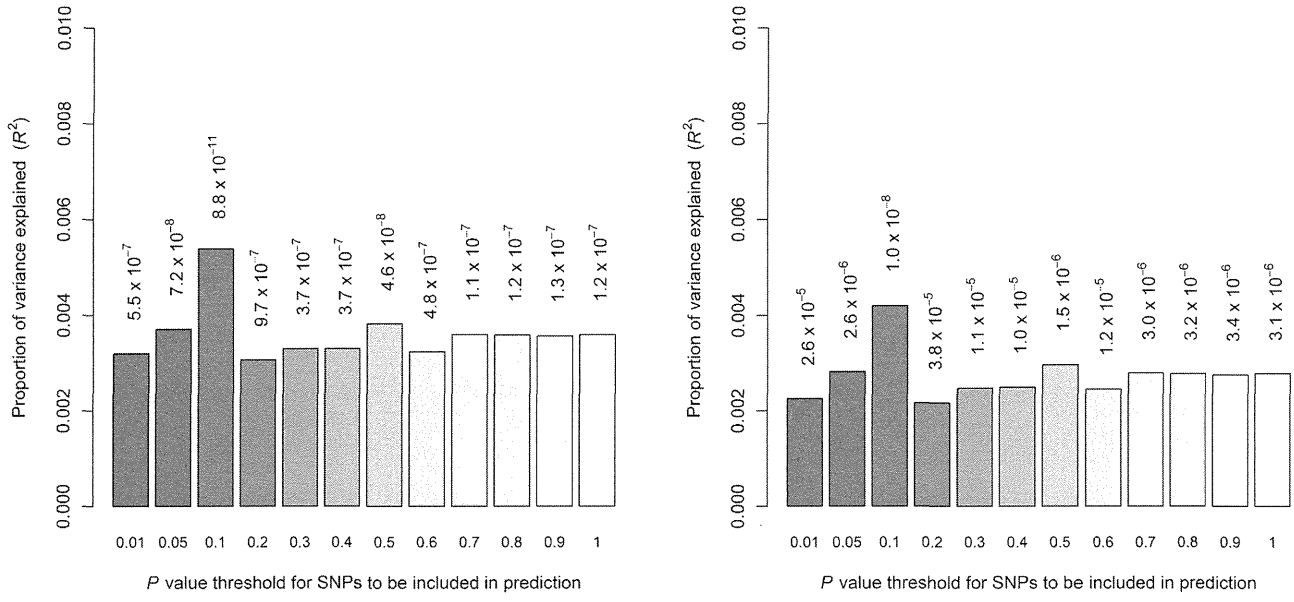

Figure 4.

Allele-specific score prediction for endometriosis, using the BBJ population as the discovery dataset and the QIMRHCS+OX population as the target dataset. The variance explained in the target dataset on the basis of allele-specific scores derived in the discovery dataset for twelve significance thresholds ($P < 0.01$, $P < 0.05$, $P < 0.1$, $P < 0.2$, $P < 0.3$, $P < 0.4$, $P < 0.5$, $P < 0.6$, $P < 0.7$, $P < 0.8$, $P < 0.9$, $P < 1.0$, plotted left to right). The y-axis indicates Nagelkerke's pseudo R^2 representing the proportion of variance explained. The number above each bar is the P value for the target dataset prediction analysis (i.e. R^2 significance). Prediction was performed using all GWA meta-analysis SNPs **(a)** and after excluding all SNPs within ±2500 kb of the seven implicated SNPs listed in Table 1 **(b)**. These figures show that the results were not driven by a few highly associated regions, indicating a substantial number of common variants underlie endometriosis risk.

Europe PMC Funders Author Manuscripts

Table 1

Summary of the endometriosis case-control cohorts

Cohort	Ancestry	No. of cases (stage B)	No. of controls
QIMRHCS GWA	European	2,262 (905)	2,924
OX GWA	European	919 (452)	5,151
BBJ GWA	Japanese	1,423	1,318
GWA meta-analysis		4,604	9,393
Replication	Japanese	1,044	4,017
Total		5,648	13,410

Europe PMC Funders Author Manuscripts

Table 2

Summary of the GWA and replication study results for the seven genome-wide significant loci

					GWA												
					QIMRHCS		OX		BBJ		Meta-analysis		Replication			Total	
Chr	SNP	Position	RA	OA	RAF_{case}	$RAF_{control}$	RAF_{case}	$RAF_{control}$	RAF_{case}	$RAF_{control}$	P_{all}	P_{stageB}	RAF_{case}	$RAF_{control}$	P	P_{all}	P_{stageB}
1	rs7521902	22490724	A	C	0.265	0.236	0.259	0.238	0.570	0.514	4.6×10^{-8}	2.3×10^{-9}	0.568	0.521	6.5×10^{-5}	3.2×10^{-11}	7.6×10^{-13}
2	rs13394619 *	11727507	G	A	0.538	0.514	0.551	0.521	0.485	0.449	6.1×10^{-8}	7.0×10^{-8}	0.489	0.455	3.5×10^{-2}	6.1×10^{-9}	6.7×10^{-9}
2	rs4141819	67864675	C	T	0.331	0.298	0.343	0.309	0.226	0.203	4.0×10^{-7}	6.5×10^{-8}	0.220	0.203	5.1×10^{-2}	8.5×10^{-8}	4.1×10^{-8}
6	rs7739264	19785588	T	C	0.545	0.512	0.556	0.515	0.772	0.742	1.3×10^{-7}	5.8×10^{-8}	0.778	0.744	6.9×10^{-4}	3.6×10^{-10}	2.1×10^{-10}
7	rs12700667	25901639	A	G	0.769	0.730	0.776	0.744	0.221	0.189	9.3×10^{-10}	3.8×10^{-11}	0.197	0.191	2.6×10^{-1}	3.6×10^{-9}	1.1×10^{-9}
9	rs1537377	22169700	C	T	0.424	0.395	0.436	0.401	0.410	0.379	2.5×10^{-6}	1.0×10^{-8}	0.402	0.359	1.3×10^{-4}	2.4×10^{-9}	5.8×10^{-12}
12	rs10859871	95711876	C	T	0.332	0.299	0.332	0.295	0.373	0.328	5.5×10^{-9}	3.7×10^{-7}	0.377	0.328	1.1×10^{-5}	5.1×10^{-13}	2.6×10^{-11}

Chr = Chromosome, Position = GRCh37 (hg19) bp position, RA = risk allele, OA = other allele, RAF = risk allele frequency.

* = GWA meta-analysis and total P values for rs13394619 include results published in Adachi et al. (2010), consisting of $P = 6.1 \times 10^{-4}$ ($RAF_{case} = 0.517$, $RAF_{control} = 0.414$) and $P = 1.0 \times 10\text{-}2$ ($RAF_{case} = 0.488$, $RAF_{control} = 0.429$) obtained in their 500K and 6.0 case-control cohorts, respectively. P_{all} includes all available endometriosis cases. P_{stageB} excludes unknown and minimal (rAFS I-II) endometriosis stage cases where detailed stage data was available.

Journal of Human Genetics (2012), 1–6

www.nature.com/jhg

ORIGINAL ARTICLE

A genome-wide association study identifies a genetic variant in the *SIAH2* locus associated with hormonal receptor-positive breast cancer in Japanese

Seham Elgazzar[1], Hitoshi Zembutsu[1], Atsushi Takahashi[2], Michiaki Kubo[3], Fuminori Aki[4], Koichi Hirata[5], Yuichi Takatsuka[6], Minoru Okazaki[7], Shozo Ohsumi[8], Takashi Yamakawa[9], Mitsunori Sasa[10], Toyomasa Katagiri[11], Yoshio Miki[12] and Yusuke Nakamura[1]

In Japan, breast cancer is the most common cancer among women and the second leading cause of cancer death among women worldwide. To identify genetic variants associated with the disease susceptibility, we performed a genome-wide association study (GWAS) using a total of 1086 Japanese female patients with hormonal receptor-positive (HRP) breast cancer and 1816 female controls. We selected 33 single-nucleotide polymorphisms (SNPs) with suggestive associations in GWAS (*P*-value of $<1 \times 10^{-4}$) as well as 4 SNPs that were previously implicated their association with breast cancer for further replication by an independent set of 1653 cases and 2797 controls. We identified significant association of the disease with a SNP rs6788895 ($P_{combined}$ of 9.43×10^{-8} with odds ratio (OR) of 1.22) in the *SIAH2* (*intron of seven in absentia homolog 2*) gene on chromosome 3q25.1 where the involvement in estrogen-dependent diseases was suggested. In addition, rs3750817 in intron 2 of the *fibroblast growth factor receptor 2* gene, which was reported to be associated with breast cancer susceptibility, was significantly replicated with $P_{combined}$ of 8.47×10^{-8} with OR = 1.22. Our results suggest a novel susceptibility locus on chromosome 3q25.1 for a HRP breast cancer.
Journal of Human Genetics advance online publication, 6 September 2012; doi:10.1038/jhg.2012.108

Keywords: breast cancer in the Japanese population; *FGFR2* gene; GWAS; hormonal receptor-positive breast cancer; *SIAH2* gene; 3q25.1 locus; 10q26 locus

INTRODUCTION

Nearly 70% of breast cancer is known to be hormone dependent, as estrogen and progesterone have key roles both in the development and progression of the disease.[1,2] The exposures to higher level and/or for longer period of estrogen such as early menarche, late menopause, late age at first pregnancy, nulliparity, postmenopausal obesity and high serum estrogen level in postmenopausal women is considered to be risk factors for breast cancer.[3–5] Furthermore, progestin, synthetic progesterone, was shown to markedly increase the risk of breast cancer in postmenopausal women when this hormonal therapy was provided for >10 years.[6] In Japan, breast cancer is the most common cancer among women and its incidence has been doubled in both pre- and postmenopausal women in the last 20 years, mainly as an estrogen receptor-positive subgroup.[7] Although hormone therapy and radiotherapy are effective, cancer cells often become resistant to these treatments; nearly half of estrogen receptor-positive breast cancer patients at an advanced stage suffer from recurrence[8–10] and only one-third of hormonal receptor-positive (HRP) patients with metastatic disease respond to radiotherapy.[11] Therefore, new therapeutic options for the disease are eagerly awaited.

The aim of this study is to identify the genetic factors susceptible to HRP breast cancer in the Japanese population and should facilitate the development of novel approaches to prevent and/or treat breast cancer.

MATERIALS AND METHODS

Samples

Characteristics of study subjects are shown in Table 1. Most of the breast cancer cases and all the controls in this study were registered in the BioBank Japan, which begun in 2003 with the goal of collecting DNA and serum

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; [2]Laboratory for Statistical Analysis, RIKEN Center for Genomic Medicine, Yokohama, Japan; [3]Laboratory for Genotyping Development, RIKEN Center for Genomic Medicine, Yokohama, Japan; [4]Department of Surgery, Itoh Surgery and Breast Clinic, Kochi, Japan; [5]First Department of Surgery, Sapporo Medical University, Sapporo, Japan; [6]Department of Breast Surgery, Kansai Rosai Hospital, Hyogo, Japan; [7]Department of Surgery, Sapporo Breast Surgical Clinic, Sapporo, Japan; [8]Department of Breast Oncology, Shikoku Cancer Center, Ehime, Japan; [9]Department of Surgery, Yamakawa Breast Clinic, Kochi, Japan; [10]Department of Surgery, Tokushima Breast Care Clinic, Tokushima, Japan; [11]Division of Genome Medicine, Institute for Genome Research, The University of Tokushima, Tokushima, Japan and [12]Genome Center, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan
Correspondence: Dr Y Nakamura, Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.
E-mail: yusuke@ims.u-tokyo.ac.jp
Received 3 April 2012; revised 5 August 2012; accepted 8 August 2012

Table 1 Characterization of samples used in hormonal receptor-positive breast cancer

	GWAS	*Replication*
Case		
Number of subjects	1086	1653
Mean age at interview (±s.d.)	66.7 (18.5)	60.7 (9.3)
Mean age of menarche	12.4	12.2
Mean age of menopause	48.3	47.9
Cases with DCIS	52	207
Cases with invasion	1034	1446
Body mass index prime	1.08	1.03
Platform	Illumina HumanHap 610K	Invader assay
Source	BioBank Japan Collaborative hospitals[a]	BioBank Japan Collaborative hospitals[a]
Control		
Number of subjects	1816	2797
Mean age at interview (±s.d.)	61.3 (12.6)	65.9 (13.2)
Body mass index prime	1.06	1.02
Platform	Illumina HumanHap 610K	Illumina HumanHap 610K
Source	BioBank Japan	BioBank Japan
Diseases in control[b]	MRC healthy volunteer Hepatitis B Keloid Drug eruption Pulmonary tuberculosis Peripheral artery disease Arrhythmias Stroke Myocardial infarction	Rheumatoid arthritis Amyotrophic lateral sclerosis Liver cirrhosis

Abbreviations: DCIS, ductal carcinoma in situ; GWAS, genome-wide association study.
[a]Tokushima Breast Care Clinic, Yamakawa Breast Clinic, Shikoku Cancer Center, and Itoh Surgery and Breast Clinic, Kansai Rosai Hospital, Sapporo Breast Surgical Clinic and Sapporo Medical University Hospital.
[b]The control groups from BioBank Japan consisted of female individuals without cancer also without any disease related to breast cancer.

samples, along with clinical information from 300 000 individuals who were diagnosed to have any of 47 different diseases from a collaborative network of 66 hospitals in Japan. All cases were diagnosed to have a HRP breast cancer by the following examinations: examination of breast tissue (biopsy or cytology), estrogen receptor and progesterone receptor positivities were evaluated by immunohistochemistry. For the genome-wide association study (GWAS) study, 1086 subjects with HRP breast cancer had been selected as cases (Table 1); 846 samples were collected from the BioBank Japan and the remaining 240 samples were collected from collaborative hospitals. Controls for the GWAS consisted of 1816 females including 231 healthy volunteers from the Midosuji Rotary Club, Osaka, Japan. In addition, we also used genome-wide screening data of 1585 female samples for 8 diseases registered in the BioBank Japan (Table 1). In the replication stage, 1547 cases were obtained from BioBank Japan and 105 cases from the collaborative hospitals. In all, 2797 female controls were registered in BioBank Japan and were genotyped in GWAS for other diseases (Table 1).

For re-sequencing analysis, we selected 2266 cases with HRP breast cancer from the BioBank Japan. We used 497 female controls with 4 diseases (hepatitis B, keloid, drug eruption and pulmonary tuberculosis) from the BioBank Japan as well as 231 healthy volunteers from the Midosuji Rotary Club, Osaka, Japan. All participating subjects provided written informed consent to participate in the study in accordance with the process approved by Ethical Committee at each of the Institute of Medical Science of the University of Tokyo and the Center for Genomic Medicine of RIKEN.

SNP genotyping

For the first stage, we genotyped 1086 female individuals with HRP breast cancer and 1816 female controls using the Illumina HumanHap 610 Genotyping BeadChip (Illumina, San Diego, CA, USA). We applied our single-nucleotide polymorphism (SNP) quality control standard (call rate of $\geqslant 0.99$ in both cases and controls, and Hardy–Weinberg equilibrium test of $P < 1.0 \times 10^{-6}$ in controls). A total of 453 627 SNPs on autosomal chromosomes and 10 525 SNPs on X chromosome passed the quality control filters and were further analyzed. All control samples for the replication stage were genotyped using the Illumina HumanHap 610 BeadChip (female samples of three diseases as controls). All cluster plots were checked by visual inspection by trained personnel, and SNPs with ambiguous calls were excluded. For cases in the replication study, we used the multiplex PCR-based Invader assay (Third Wave Technologies).[12] In addition, 22 variations resulted from re-sequencing analysis were selected and genotyped in 2266 cases and 728 female controls also using the multiplex PCR-based Invader assay (Third Wave Technologies, Madison, WI, USA).

Statistical analysis

Associations of SNPs were tested by employing the Cochran–Armitage trend test in both the GWA and replication stages. For the combined study, the simple combined method was applied. In the replication analyses, significance level was applied to be *P*-value of $< 1.35 \times 10^{-3}$ (calculated as 0.05/37) by Bonferroni correction. Odds ratios (ORs) and confidence intervals were calculated using the non-susceptible allele as a reference. Heterogeneity between the GWAS and replication sets was examined using the Breslow–Day test. The genomic inflation factor (λGC) was calculated from the median of the Cochran–Armitage trend test statistics. The quantile–quantile plot of the logarithms of the genome-wide *P*-values was generated by the 'snpMatrix' package in R program v2.10.0 (see URLs), and the Manhattan plot was generated using Haploview v4.1 (see URLs). Haplotype analysis was performed by the use of Haploview v4.1 by considering genotyped SNPs located within 500 kb upstream or downstream of the marker SNP. *In silico* prediction of functional consequences of SNP was done by the use of the SNP info web server (see URLs). (Haploview software was used to analyze linkage disequilibrium (LD) values, visualize haplotype.)

Imputation

Imputation was performed by referring to the genotype data of Japanese (JPT) individuals as deposited in the Phase II HapMap database using MACH v1.0 (see URLs). Genotypes of SNPs that are located in the genomic region within 500 kb upstream or downstream of the marker SNP (the SNP that showed the strongest association with HRP breast cancer) were imputed. In the process of imputation, 50 Markov chain iterations were implemented. Imputed SNPs with an imputation quality score of $r^2 < 0.3$ were excluded from the subsequent analysis.

Re-sequencing analysis

Initially, we carried out SNP discovery by using DNA samples of 96 cases with HRP breast cancer. We designed 98 sets of primers (Supplementary Table 1) using the genomic sequence information from UCSC Genome Bioinformatics data base (NM_005067) to amplify the 22 353 bps (two exons, one intron, 5′-UTR and 3′-UTR) of the genomic region corresponding to the *SIAH2* (*intron of seven in absentia homolog 2*) gene. For each of the 96 DNA samples, PCRs were performed by using GeneAmp PCR system 9700 (Applied Biosystems, Foster City, CA, USA). We performed direct sequencing of the PCR products with the 96-capillary 3730 × l DNA Analyzer (Applied Biosystems) with Big Dye Terminators (Applied Biosystems) according to standard protocols. All amplified fragments were sequenced by two pairs of sequencing primers. Then SNPs were detected by Sequecher software v4.8 (Gene Codes, Ann Arbor, MI, USA).

RESULTS

To identify genetic variants susceptible to HRP breast cancer in the Japanese population, we performed a GWAS using 1086 female patients and 1816 female controls with Illumina HumanHap 610k BeadChip (Table 1). After the quality check of SNP genotyping data, a total of 453 627 SNPs were selected for further analysis. Principal component analysis revealed that all the subjects participating in this study were clustered in the Hapmap Asian population (Supplementary Figure 1S). A quantile–quantile plot for this GWAS is shown in Supplementary Figure 2S. The genomic inflation factor (λGC) of the test statistic in this study was 1.053 indicating a very low possibility of false-positive associations resulted from the population stratification. Although no SNP achieved genome-wide significance level, 46 SNPs in various chromosomes showed suggestive association (P-values $< 1 \times 10^{-4}$) as illustrated in Figure 1.

Among these 46 SNPs, we excluded SNPs possessing strong LD ($r^2 > 0.8$) and selected 33 SNPs for replication analysis as well as 4 additional SNPs that were previously reported their association with breast cancer and showed P-value of $< 1.0 \times 10^{-2}$ in GWAS analysis, using an independent set of 1653 female patients and 2797 female controls. Among 37 SNPs analyzed in the replication study, an SNP rs6788895 was successfully replicated with the P-value of $< 1.35 \times 10^{-3}$ even after the Bonferroni correction (0.05/37) as shown in Table 2 and Supplementary Table 2S. Combined analysis of the results of the GWAS and the replication study suggested strong association of the locus of the *SIAH2* gene on chromosome 3q25.1 (rs6788895, P_{combined} of 9.43×10^{-8} with OR of 1.22, 95% confidence interval 1.13–1.31) without any significant heterogeneity between the two studies ($P_{\text{heterogeneity}} = 2.33 \times 10^{-01}$).

The SNP rs6788895 was further examined its association with the subgroups of breast cancer, an invasive papilloductal breast cancer group and a HER2-negative breast cancer group, and found significant associations with them ($P_{\text{combined}} = 3.61 \times 10^{-07}$, 6.78×10^{-06}, OR = 1.23, 1.21, respectively) although they did not reach to the genome-wide significant level (Supplementary Table 3S). Imputation analysis of this locus identified nine additional SNPs in strong LD (r^2 of > 0.8) that showed similar levels of association with rs6788895 (Figure 2a). The subsequent logistic regression analysis revealed no significant association of these nine SNPs when we accounted the effect of SNP rs6788895. The haplotype analysis found no haplotype revealing stronger association than the single SNP (Supplementary Table 4S). Although *in silico* prediction of the functional effect of rs6788895 identified no possible biological effect, one SNP rs2018246 showing strong LD with rs6788895 ($r^2 = 0.94$), which was located about 0.7 kb upstream from the transcription initiation site of *SIAH2*, was indicated to be present within the binding site of multiple transcription factors such as STAT1, LEF1, PAX2, which were reported to have some implication to breast cancer.[13–16] The re-sequencing of 22 353 bps corresponding to the *SIAH2* gene identified 10 novel genetic variations in addition to 37 genetic variations reported previously. We further genotyped 22 of the 47 variations after the exclusion of SNPs showing strong LD with the marker SNP (r^2 of > 0.8). As a result, we identified no genetic variant showing significant association in HRP breast cancer (Supplementary Table 5S and Supplementary Table 6S)

Furthermore, we examined the association of 37 previously reported SNPs with the HRP breast cancer[17–26] using our sample sets (Supplementary Table 7S) and found very moderate association of four genetic variants, rs1292011, rs3803662, rs2981579 and rs3750817, with HRP breast cancer in the GWAS phase ($P_{GWAS} = 5.89 \times 10^{-02}$, 6.95×10^{-03}, 8.68×10^{-04} and 5.03×10^{-04}, respectively). Further analysis of these four SNPs identified significant

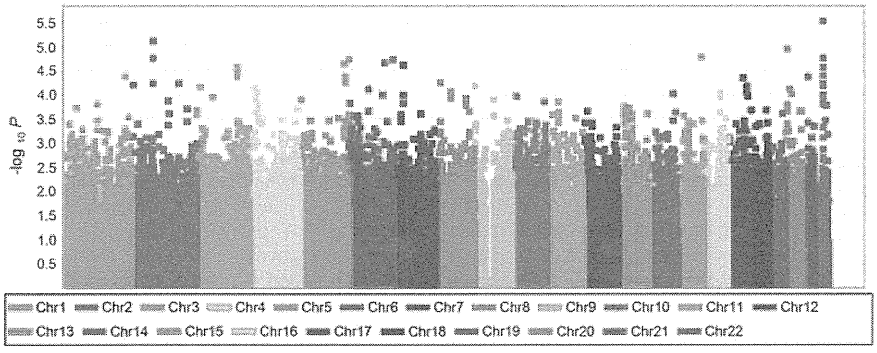

Figure 1 Manhattan plot for the genome-wide association study (GWAS) of hormonal receptor-positive breast cancer indicating −log10*P* of the Cochran–Armitage trend test for 453 627 single-nucleotide polymorphisms (SNPs) plotted against their respective positions on each chromosome.

Table 2 Association of SNP rs6788895 on chromosome 3q25.1 with hormonal receptor-positive breast cancer

					Case				*Control*							
Chr.	*Chrloc.*	*SNP*	*RA*	*Stage*	*11*	*12*	*22*	*RAF*	*11*	*12*	*22*	*RAF*	P_{assoc}[a]	*OR*	*(95% CI)*	*Phet*[b]
3	151950498	rs6788895	G	GWAS	106	456	524	0.69	242	832	742	0.64	2.34E–05	1.28	(1.14–1.43)	
				Rep	164	694	786	0.69	337	1265	1195	0.65	5.77E–04	1.18	(1.07–1.29)	2.33E–01
				Combined	270	1150	1310	0.69	579	2097	1937	0.65	9.43E–08	1.22	(1.13–1.31)	

Abbreviations: Chr., chromosome; chrloc., chromosomal location (bp); CI, confidence interval; GWAS, genome-wide association study; OR, odds ratio (calculated based on the risk allele); RA, risk allele; RAF, risk allele frequency; Rep, replication; SNP, single-nucleotide polymorphism; 11, homozygous non-risk genotype; 12, heterozygous genotype; 22, homozygous risk genotype.
[a]Passoc, *P*-value for the GWAS and replication study obtained from the Cochran–Armitage trend test and *P*-value for the combined study obtained from the simple combined test.
[b]Phet, *P*-value for heterogeneity test obtained from the Breslow–Day test.

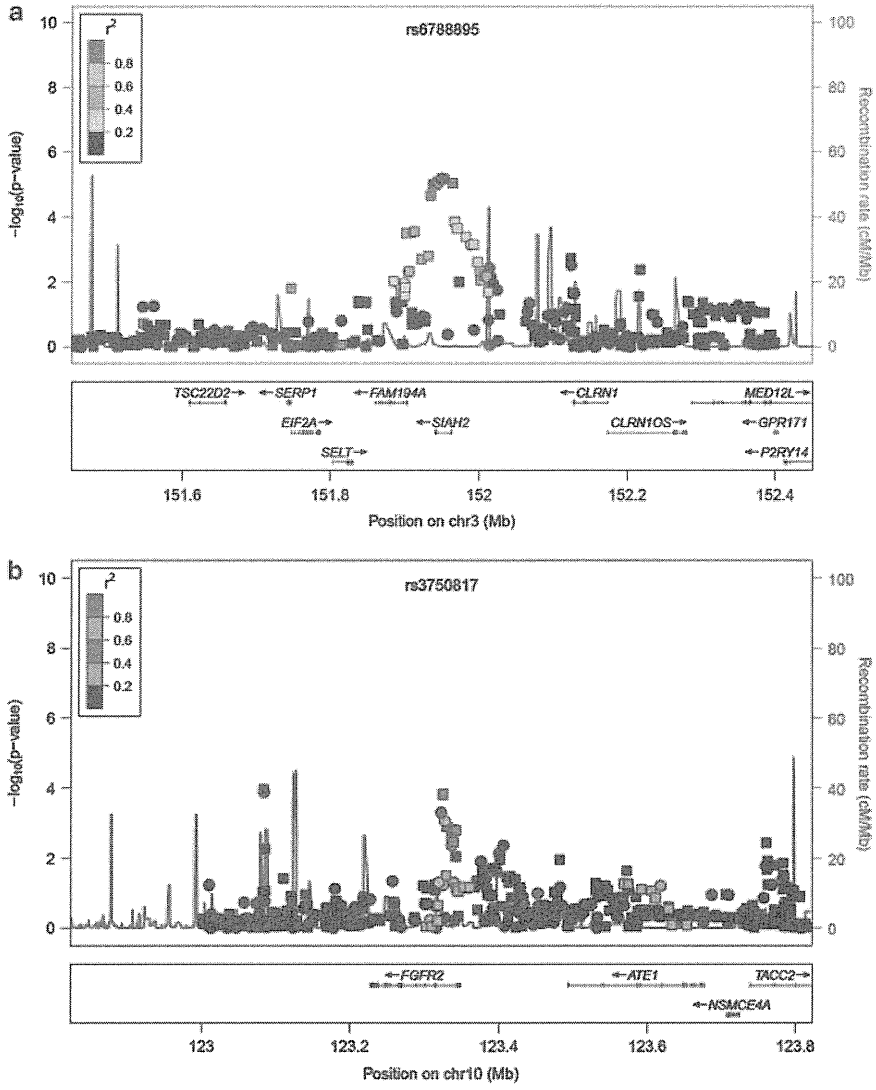

Figure 2 (**a**) Regional association plots of the locus associated with hormonal receptor-positive breast cancer on chromosomes 3q25.1 (*intron of seven in absentia homolog 2* (*SIAH2*)). (**b**) Regional association plots of the locus associated with hormonal receptor-positive breast cancer on chromosomes 10q26 (*fibroblast growth factor receptor 2* (*FGFR2*)). For each plot, $-\log_{10}P$ of the Cochran–Armitage trend test of single-nucleotide polymorphisms (SNPs) in the genome-wide association study (GWAS) was plotted against relative chromosomal locations. The square and rounded signs represent imputed and genotyped SNPs, respectively. All SNPs are color coded as red ($r^2 = 0.8$–1.0), orange ($r^2 = 0.6$–0.8), green ($r^2 = 0.4$–0.6), light blue ($r^2 = 0.4$–0.6), and dark blue ($r^2 < 0.2$) according to their pair wise r^2 to the marker SNP. The marker SNP is represented in purple color. SNP positions followed NCBI build 36 coordinates. Estimated recombination rates (cM/Mb) are plotted as a blue line.

replication of two SNPs, rs3750817 ($P_{\text{replication}} = 5.39 \times 10^{-5}$, OR $= 1.22$) and rs2981579 ($P_{\text{replication}} = 1.21 \times 10^{-3}$, OR $= 1.20$). Both SNPs are located within intron 2 of the fibroblast growth factor receptor 2 *(FGFR2)* genes. The combined analysis of the GWAS and replication phases of rs3750817 revealed strong association with $P_{\text{combined}} = 8.47 \times 10^{-08}$ (OR $= 1.22$) and that of rs2981579 was 1.77×10^{-06} (OR $= 1.20$) (Table 3). Imputation analysis of this locus identified three additional SNPs, rs9420318, rs11199914 and rs10736303 that showed similar levels of association with rs3750817 (Figure 2b).

DISCUSSION

We reported here GWA and replication studies using a total of 2730 female breast cancer cases and 4613 female controls in the Japanese population to identify common genetic variants susceptible to the HRP breast cancer. The SNP rs6788895 located in the intronic region of the *SIAH2* gene on chromosome 3q25.1 revealed a significant association with the HRP breast cancer (P_{combined} of 9.43×10^{-08} with OR of 1.22, 95% confidence interval of 1.13–1.31). We further examined the association of rs6788895 with the subgroups of breast cancer. The analysis of two histological subgroups, an invasive papilloductal breast cancer group and a HER2-negative breast cancer group, indicated suggestive associations with P_{combined} of 3.61×10^{-07} (OR $= 1.24$) and with P_{combined} of 6.78×10^{-06} (OR $= 1.21$), respectively (Supplementary Table 3S). However, rs6788895 showed no association in the GWAS with the hormonal receptor-negative group (P_{trend} of 1.03×10^{-01}) or with the HER2-positive breast cancer group (P_{trend} of 1.15×10^{-01}).

For further characterization of the chromosome 3q25.1 locus, we imputed genotypes of SNPs that were not genotyped in the GWAS

Table 3 rs2981579 and rs3750817 in different population

SNPs	Minor/major allele	MAF	OR	P-trend	Population
rs2981579 (FGFR2)	A/G	0.42	1.43	3.60×10^{-31}	UK[20]
rs2981579	A/G	0.44	1.31	2.60×10^{-09}	American[25]
rs2981579	A/G	0.47	1.20	1.77×0^{-06}	Japanese
rs3750817 (FGFR2)	T/C	0.49	1.22	8.47×10^{-08}	Japanese
rs3750817	T/C	0.37	0.78	8.20×10^{-08}	American[25]

Abbreviations: FGFR2, fibroblast growth factor receptor 2; MAF, minor allele frequency; OR, odds ratio (calculated based on the non susceptible allele) except rs3750817 in American population OR, calculated based on the susceptible allele); SNP, single-nucleotide polymorphism.

and then examined their associations with HRP breast cancer, but found no SNP showing stronger association than the marker SNP rs6788895 although several SNPs having strong LD with rs6788895 ($r^2 > 0.8$) showed similar levels of associations (Figure 2a). Previous reports implicated possible roles of *SIAH2* in breast carcinogenesis and described that SIAH2 expression was highly associated with estrogen receptor levels.[9,27–29] In addition, SIAH2 protein was indicated to have an essential role in the hypoxic response by regulating the hypoxia-inducible factor-α.[30]

Moreover, *SIAH2* was known to induce ubiquitin-mediated degradation of many substrates, including proteins involved in transcriptional regulation (POU2AF1, PML and NCOR1), a cell surface receptor (DCC) and an anti-apoptotic protein (BAG1). These proteins were reported to have some relations to breast cancer by different mechanisms.[31–35] Recent genetic studies showed that the chromosome 3q25.1 region might have a critical role in some estrogen-dependent diseases such as development of peritoneal leioyomatosis.[36,37]

We also examined the association of previously reported loci with the breast cancer[17–26] using our sample sets and found very moderate association of four genetic variants in our GWAS. Further analysis of these four SNPs identified significant replication of two SNPs, rs3750817 and rs2981579 ($P_{combined} = 8.47 \times 10^{-8}$ and 1.77×10^{-06} with OR = 1.22 and OR = 1.20, respectively). A T allele for rs3750817 is a protective allele for both Japanese and American populations with comparable ORs (Table 3).

For characterization of the chromosome 10q26 locus, we imputed genotypes of SNPs that were not genotyped in the GWAS, and examined the associations of these SNPs with HRP breast cancer. As a result, three additional SNPs, rs9420318, rs11199914 and rs10736303 were found to have similar levels of association with rs3750817 (Figure 2b). The most strongly associated SNPs are located in intron 2 of the *FGFR* gene. The intron 2 region contains a highly conserved region and possess the transcription factor binding sites possibly related to the estrogen receptor signaling pathway.[38] FGFR2 encodes a receptor tyrosine kinase and has an important role in human mammary epithelial-cell transformation,[39,40] suggesting that FGFR2 is a good candidate for breast cancer susceptibility. Subsequent functional analyses are thus essential to pinpoint the causal variants and genes associated with HRP breast cancer. In addition, because breast cancer is multi factorial disease, we could not exclude the possibility that some subjects with undiagnosed early stage of cancers or undiagnosed hormonal-dependent diseases or subject have diseases related to breast cancer might have been included as controls. Hence, this study might not have sufficient power to detect SNPs having very modest effects on susceptibility to HRP breast cancer. In conclusion, our findings, the verification of the association of the FGFR2 to the risk of breast cancer in the Japanese population and the novel identification of significant association of genetic variations in the *SIAH2* gene, should contribute to the better understanding of the susceptibility to HRP breast cancer.

URLS

The Leading Project for Personalized Medicine, http://biobankjp.org/;
EIGENSTRATsoftwarev2.0, http://genepath.med.harvard.edu/~reich/Software.htm;
R project v2.10.0, http://www.r-project.org/;
Haploview v4.1, http://www.broadinstitute.org/haploview/haploview;
MACH v1.0, http://www.sph.umich.edu/csg/yli/mach/index.html;
PLINK statistical software v1.06, http://pngu.mgh.harvard.edu/~purcell/plink/;
SNP info web server, http://manticore.niehs.nih.gov/index.html.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We extend our heartfelt gratitude to all individuals who kindly participated in the current study. We also thank the BioBank Japan for providing DNA samples and clinical information and the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for supporting this study. We extend our appreciation to Dr Siew-Kee Low for her useful discussion in the statistical analysis and Dr Osman Wael Mohammed for his insightful advices. Finally, we would like to express our gratefulness to technical staff of the Laboratory for Genotyping Development in RIKEN, Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, the University of Tokyo and technical staff of collaborative hospitals for their assistances.

1 Massarweh, S. & Schiff, R. Resistance to endocrine therapy in breast cancer: exploiting estrogen receptor/growth factor signaling crosstalk. *Endocr. Relat. Cancer* **13,** S15–S24 (2006).
2 Xu, W. -H., Shu, X. O., Long, J., Lu, W., Cai, Q., Zheng, Y. *et al.* Relation of FGFR2 genetic polymorphisms to the association between oral contraceptive use and the risk of breast cancer in Chinese women. *Am. J. Epidemiol.* **173,** 923–931 (2011).
3 Hu, Y.-H., Nagata, C., Shimizu, H., Kaneda, N. & Kashiki, Y. Association of body mass index, physical activity, and reproductive histories with breast cancer: a case-control study in Gifu, Japan. *Breast Cancer Res. Treat.* **43,** 65–72 (1997).
4 Tung, H. T., Tsukuma, H., Tanaka, H., Kinoshita, N., Koyama, Y., Ajiki, W. *et al.* Risk factors for breast cancer in Japan, with special attention to anthropometric measurements and reproductive history. *Jpn. J. Clin. Oncol.* **29,** 137–146 (1999).
5 Miyoshi, Y., Tanji, Y., Taguchi, T., Tamaki, Y. & Noguchi, S. Association of serum estrone levels with estrogen receptor-positive breast cancer risk in postmenopausal Japanese women. *Clin. Cancer Res.* **9,** 2229–2233 (2003).
6 Colditz, G. A. & Rosner, B. Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *Am. J. Epidemiol.* **152,** 950–964 (2000).
7 Yamashita, H., Iwase, H., Toyama, T., Takahashi, S., Sugiura, H., Yoshimoto, N. *et al.* Estrogen receptor-positive breast cancer in Japanese women: trends in incidence, characteristics, and prognosis. *Ann. Oncol.* **22,** 1318–1325 (2011).
8 Ao, A., Morrison, B. J., Wang, H., López, J. A., Reynolds, B. A. & Lu, J. Response of estrogen receptor-positive breast cancer tumorspheres to antiestrogen treatments. *PLoS ONE* **6,** e18810 (2011).
9 Jansen, M., Ruigrok-Ritstier, K., Dorssers, L. C., van Staveren, I. L., Look, M. P., Meijer-van Gelder, M. E. *et al.* Downregulation of SIAH2, an ubiquitin E3 ligase, is associated with resistance to endocrine therapy in breast cancer. *Breast Cancer Res. Treat.* **116,** 263–271 (2009).
10 Kiyotani, K., Mushiroda, T., Tsunoda, T., Morizono, T., Hosono, N., Kubo, M. *et al.* A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese. *Hum. Mol. Genet.* **21,** 1665–1672 (2012).
11 Prat, A. & Baselga, J. The role of hormonal therapy in the management of hormonal-receptor-positive breast cancer with co-expression of HER2. *Nat. Clin. Prac. Oncol.* **5,** 531–542 (2008).
12 Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. & Nakamura, Y. A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* **46,** 471–477 (2001).

13 Chan, S., Vermi, W., Luo, J., Lucini, L., Rickert, C., Fowler, A. *et al.* STAT1-deficient mice spontaneously develop estrogen receptor alpha-positive luminal mammary carcinomas. *Breast Cancer. Res.* **14,** R16 (2012).
14 Holmes, K. A., Song, J. S., Liu, X. S., Brown, M. & Carroll, J. S. Nkx3-1 and LEF-1 function as transcriptional inhibitors of estrogen receptor activity. *Cancer Res.* **68,** 7380–7385 (2008).
15 Huang, F. -I., Chen, Y. -L., Chang, C. -L., Yuan, R. -H. & Jeng, Y. -M. Hepatocyte growth factor activates Wnt pathway by transcriptional activation of LEF1 to facilitate tumor invasion. *Carcinogenesis* **33,** 1142–1148 (2012).
16 Beauchemin, D., Lacombe, C. & Van Themsche, C. PAX2 is activated by estradiol in breast cancer cells of the luminal subgroup selectively, to confer a low invasive phenotype. *Mol. Cancer* **10,** 148 (2011).
17 Walsh, T., Casadei, S., Coats, K. H., Swisher, E., Stray, S. M., Higgins, J. *et al.* Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA* 295, 1379–1388, 2006).
18 Sehrawat, B., Sridharan, M., Ghosh, S., Robson, P., Cass, C. E., Mackey, J. R. *et al.* Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. *Hum. Genet.* **130,** 529–537 (2011).
19 Higginbotham, K. S. P., Breyer, J. P., Bradley, K. M., Schuyler, P. A., Plummer, Jr W. D., Freudenthal, M. E. *et al.* A multistage association study identifies a breast cancer genetic locus at NCOA7. *Cancer Res.* **71,** 3881–3888 (2011).
20 Prentice, R. L., Huang, Y., Hinds, D. A., Peters, U., Pettinger, M., Cox, D. R. *et al.* Variation in the FGFR2 gene and the effects of postmenopausal hormone therapy on invasive breast cancer. *Cancer Epidem. Biomarker. Prev.* **18,** 3079–3085 (2009).
21 Ghoussaini, M., Fletcher, O., Michailidou, K., Turnbull, C., Schmidt, M. K., Dicks, E. *et al.* Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat. Genet.* **44,** 312–318 (2012).
22 Rebbeck, T. R., DeMichele, A., Tran, T. V., Panossian, S., Bunin, G. R., Troxel, A. B. *et al.* Hormone-dependent effects of FGFR2 and MAP3K1 in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. *Carcinogenesis* **30,** 269–274 (2009).
23 Yoshimoto, N., Nishiyama, T., Toyama, T., Takahashi, S., Shiraki, N., Sugiura, H. *et al.* Genetic and environmental predictors, endogenous hormones and growth factors, and risk of estrogen receptor-positive breast cancer in Japanese women. *Cancer Sci.* **102,** 2065–2072 (2011).
24 Fanale, D., Amodeo, V., Corsini, L. R., Rizzo, S., Bazan, V. & Russo, A. Breast cancer genome-wide association studies: there is strength in numbers. *Oncogene* **31,** 2121–2128 (2012).
25 Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M. *et al.* Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat. Genet.* **42,** 504–507 (2010).
26 Hamaguchi, M., Nishio, M., Toyama, T., Sugiura, H., Kondo, N., Fujii, Y. *et al.* Possible difference in frequencies of genetic polymorphisms of estrogen receptor α, estrogen metabolism and P53 genes between estrogen receptor-positive and -negative breast cancers. *Jpn. J. Clin. Oncol.* **38,** 734–742 (2008).
27 Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A. *et al.* Molecular portraits of human breast tumours. *Nature* **406,** 747–752 (2000).
28 van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536 (2002).
29 Roh, M. S., Hong, S. H., Jeong, J. S., Kwon, H. C., Kim, M. C., Cho, S. H. *et al.* Gene expression profiling of breast cancers with emphasis of β-catenin regulation. *J. Korean Med. Sci.* **19,** 275–282 (2004).
30 Chan, P., Möller, A., Liu, M. C., Sceneay, J. E., Wong, C. S., Waddell, N. *et al.* The expression of the ubiquitin ligase SIAH2 (seven in absentia homolog 2) is mediated through gene copy number in breast cancer and is associated with a basal-like phenotype and p53 expression. *Breast Cancer Res.* **13,** R19 (2011).
31 Koren, R., Dekel, Y., Sherman, E., Weissman, Y., Dreznik, Z., Klein, B. *et al.* The expression of DCC protein in female breast cancer. *Breast Cancer Res. Treat.* **80,** 215–220 (2003).
32 Zhen-huan, Z., Yamashita, H., Toyama, T., Yamamoto, Y., Kawasoe, T., Ibusuki, M. *et al.* Nuclear corepressor 1 expression predicts response to first-line endocrine therapy for breast cancer patients on relapse. *Chin. Med. J.* **122,** 1764–1768 (2009).
33 Plevova, P., Bouchal, J., Fiurásková, M., Foretová, L., Navrátilová, M., Zapletalová, J. *et al.* PML protein expression in hereditary and sporadic breast cancer. *Neoplasma* **54,** 263–268 (2007).
34 Yang, X., Hao, Y., Ding, Z., Pater, A. & Tang, S.-C. Differential expression of antiapoptotic gene BAG-1 in human breast normal and cancer cell lines and tissues. *Clin. Cancer Res.* **5,** 1816–1822 (1999).
35 Baysal, B. E. A high-resolution integrated map spanning the SDHD gene at 11q23: a 1.1-Mb BAC contig, a partial transcript map and 15 new repeat polymorphisms in a tumour-suppressor region. *Eur. J. Human Genet.* **9,** 121–129 (2001).
36 Ordulu, Z., Dal C., P., Chong, W. W., Choy, K. W., Lee, C., Muto, M. G. *et al.* Disseminated peritoneal leiomyomatosis after laparoscopic supracervical hysterectomy with characteristic molecular cytogenetic findings of uterine leiomyoma. *Genes Chromosom. Cancer* **49,** 1152–1160 (2010).
37 Cha, P. -C., Takahashi, A., Hosono, N., Low, S. K., Kamatani, N., Kubo, M. *et al.* A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids. *Nat. Genet.* **43,** 447–450 (2011).
38 Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoute, J. *et al.* Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38,** 1289–1297 (2006).
39 Moffa, A. B. & Ethier, S. P. Differential signal transduction of alternatively spliced FGFR2 variants expressed in human mammary epithelial cells. *J. Cell. Physiol.* **210,** 720–731 (2007).
40 Koziczak, M., Holbro, T. & Hynes, N. E. Blocking of FGFR signaling inhibits breast cancer cell proliferation through downregulation of D-type cyclins. *Oncogene* **23,** 3501–3508 (2004).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

Clinical Cancer Research

A Genome-Wide Association Study Identifies Novel Loci for Paclitaxel-Induced Sensory Peripheral Neuropathy in CALGB 40101

R. Michael Baldwin[1], Kouros Owzar[5,6], Hitoshi Zembutsu[11], Aparna Chhibber[1], Michiaki Kubo[11], Chen Jiang[6], Dorothy Watson[6], Rachel J. Eclov[1], Joel Mefford[2], Howard L. McLeod[7], Paula N. Friedman[8], Clifford A. Hudis[9], Eric P. Winer[10], Eric M. Jorgenson[3,4], John S. Witte[2,4], Lawrence N. Shulman[10], Yusuke Nakamura[11], Mark J. Ratain[8], and Deanna L. Kroetz[1,4]

Abstract

Purpose: Sensory peripheral neuropathy is a common and sometimes debilitating toxicity associated with paclitaxel therapy. This study aims to identify genetic risk factors for the development of this toxicity.

Experimental Design: A prospective pharmacogenetic analysis of patients with primary breast cancer, randomized to the paclitaxel arm of CALGB 40101, was used to identify genetic predictors of the onset and severity of sensory peripheral neuropathy. A genome-wide association study in 855 subjects of European ancestry was conducted and findings were replicated in additional European ($n = 154$) and African American ($n = 117$) subjects.

Results: A single nucleotide polymorphism in *FGD4* was associated with the onset of sensory peripheral neuropathy in the discovery cohort [rs10771973; HR, 1.57; 95% confidence interval (CI), 1.30–1.91; $P = 2.6 \times 10^{-6}$] and in a European (HR, 1.72; 95% CI, 1.06–2.80; $P = 0.013$) and African American (HR, 1.93; 95% CI, 1.13–3.28; $P = 6.7 \times 10^{-3}$) replication cohort. There is also evidence that markers in additional genes, including *EPHA5* (rs7349683) and *FZD3* (rs10771973), were associated with the onset or severity of paclitaxel-induced sensory peripheral neuropathy.

Conclusions: A genome-wide association study has identified novel genetic markers of paclitaxel-induced sensory peripheral neuropathy, including a common polymorphism in *FGD4*, a congenital peripheral neuropathy gene. These findings suggest that genetic variation may contribute to variation in development of this toxicity. Validation of these findings may allow for the identification of patients at increased risk of peripheral neuropathy and inform the use of an alternative to paclitaxel and/or the clinical management of this toxicity. *Clin Cancer Res; 18(18); 5099–109. ©2012 AACR.*

Authors' Affiliations: Departments of [1]Bioengineering and Therapeutic Sciences,[2]Epidemiology and Biostatistics, [3]Neurology, and [4]Institute for Human Genetics, University of California San Francisco, San Francisco, California; [5]Department of Biostatistics and Bioinformatics and [6]CALGB Statistical Center, Duke University, Durham; [7]Department of Pharmacotherapy and Experimental Therapeutics, University of North Carolina, Chapel Hill, North Carolina; [8]Department of Medicine, University of Chicago, Chicago, Illinois; [9]Memorial Sloan-Kettering Cancer Center, New York, New York; [10]Dana-Farber Cancer Institute, Boston, Massachusetts; and [11]University of Tokyo and Riken Center for Genomic Medicine, Yokohama, Japan

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

R.M. Baldwin and K. Owzar contributed equally to this work.

Corresponding Author: Deanna L. Kroetz, Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, 1550 4th Street RH584E, San Francisco, CA 94158-2911. Phone: 415-476-1159; Fax: 415-514-4361; E-mail: deanna.kroetz@ucsf.edu

doi: 10.1158/1078-0432.CCR-12-1590

Introduction

Paclitaxel is a useful microtubule-stabilizing agent with efficacy in the treatment of many cancers. It is effective for the treatment of breast cancer in the metastatic, adjuvant, and neoadjuvant settings (1, 2). Sensory peripheral neuropathy remains a significant issue in the clinical use of this agent. More than 50% of patients experience some degree of sensory peripheral neuropathy during their course of paclitaxel treatment, with 5% to 30% experiencing grade 3 or 4 toxicity (3, 4). Paclitaxel-induced sensory peripheral neuropathy is dose-, treatment schedule-, and infusion time–dependent (3). Cumulative dose is a significant predictor of sensory peripheral neuropathy, as is underlying diabetes and concurrent or previous administration of other drugs associated with this toxicity. A recent study suggests that mild to moderate symptoms of sensory peripheral neuropathy can persist for up to 2 years following completion of paclitaxel treatment (5). Long-term neuropathy is

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

Translational Relevance

Paclitaxel is widely used in the treatment of many cancers, including breast cancer. Treatment with paclitaxel is often limited by the development of peripheral neuropathies that can significantly impact a patient's quality of life. Biomarkers for the prediction of paclitaxel-induced peripheral neuropathy could be used to optimize the use of paclitaxel. A genome-wide genotyping approach in women receiving single-agent paclitaxel as adjuvant therapy for breast cancer identified several novel genetic loci implicated in paclitaxel-induced sensory peripheral neuropathy. In particular, a common genetic variant in *FGD4*, a causal gene for the congenital peripheral neuropathy Charcot–Marie–Tooth disease, was associated with increased onset of neuropathy in both Europeans and African Americans. This variant and others identified in these studies could be validated as genetic predictors of paclitaxel-induced sensory peripheral neuropathy. The genetic variants identified in these studies will also lead to investigations into novel pathways for this common chemotherapy-induced toxicity.

particularly concerning for patients with primary breast cancer, such as those evaluated in the current study, as more than 80% will be long-term survivors whose quality of life will be compromised. Significant sensory peripheral neuropathy during paclitaxel treatment can lead to dose reductions and treatment suspension, possibly resulting in suboptimal disease treatment and the potential for an increased likelihood of relapse. A predictive marker for this dose-limiting toxicity would enable studies to identify whether an individualized assessment of adverse event risk could be useful in the clinical decision making process. It could also provide a possible target for therapeutic interventions.

Substantial interindividual differences in the prevalence, reported and objective severity, and onset of peripheral neuropathy is consistent with an underlying genetic susceptibility to this toxicity. CALGB 40101 is a phase III randomized study comparing cyclophosphamide and doxorubicin versus single-agent paclitaxel as adjuvant therapy for patients with breast cancer at relatively low risk for relapse. In addition, the study compared short versus longer therapy of each regimen as a 2 × 2 factorial design. A pharmacogenetic companion study (CALGB 60202) was included in this trial to prospectively evaluate germline determinants of interindividual differences in response and toxicity. An initial analysis of treatment outcome in CALGB 40101 has shown no difference in response between the 4 and 6 cycle treatment arms (6); additional analyses of response await complete follow-up data. The goal of this present study was to identify genetic markers predictive of sensory peripheral neuropathies in the paclitaxel treatment arm of CALGB 40101 and to further our understanding of the underlying mechanism of injury and repair. Herein, we report the results of a genome-wide association study (GWAS) of 1,040 paclitaxel-treated women to identify novel germline susceptibility loci associated with the development of sensory peripheral neuropathies. This represents the largest prospective breast cancer pharmacogenetic study of paclitaxel treatment toxicities to date and provides a paradigm for the identification of genetic markers with potential clinical application in personalized medicine.

Materials and Methods

Participants

All study participants were enrolled in CALGB 40101 and gave their additional consent to participate in the pharmacogenetic companion study (CALGB 60202). CALGB 40101 was open from May 15, 2002 until July 30, 2010. The final total accrual was 3,873 patients. Patients eligible for the treatment protocol were females with histologically confirmed invasive carcinoma of the breast and 0 to 3 axillary nodes positive for cancer. Eastern Cooperative Oncology Group (ECOG) performance status of 0–1, adequate organ function, and absence of congestive heart failure or myocardial infarction in the previous 6 months were required. Enrollment was required within 84 days of breast surgery (either modified radical mastectomy or lumpectomy) and the treatment began within 7 days of registration. Patients with locally advanced, inflammatory, or metastatic breast cancer or involvement of dermal lymphatics were ineligible. Patients were disease-free from any prior malignancies for at least 5 years. Previous trastuzumab, chemotherapy, or hormonal therapy, with the exception of tamoxifen, for the current malignancy was not permitted nor was anthracycline treatment for any previous disease. Patients who received tamoxifen or any other selective estrogen receptor modulators (SERM) for prevention or other indications (e.g., osteoporosis) were eligible. Treatment with tamoxifen, other SERMs, or exogenous hormones (e.g., hormone replacement therapy, oral contraceptives, raloxifene) was discontinued before enrollment. Trastuzumab was recommended for patients with HER2-positive disease. Patients could also enroll in adjuvant studies of bisphosphonates or hormonal therapies (e.g., ovarian suppression concurrent with chemotherapy). All patients provided written informed consent for both the treatment and companion protocols that met state, federal, and institutional guidelines.

Treatment

Patients were randomly assigned with equal probability to 4 or 6 cycles of cyclophosphamide/doxorubicin (AC) or paclitaxel. A full description of the study design is included in a recent publication describing the initial analysis of treatment response (6). The first 570 patients were treated with AC every 3 weeks, or paclitaxel weekly for 12 or 18 weeks. Thereafter, both regimens were administered every 2 weeks for 4 or 6 cycles. Pharmacogenetic samples were collected only from patients enrolled on every 2-week regimens, who received dose-dense paclitaxel for 4 or 6 cycles. Paclitaxel was given for more than 3 hours at 175 mg/m^2 when given every 2 weeks. The 6 cycle treatment arms for

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

Table 1. Patient demographics

		Randomized[a]	Post Quality Control[b]	Discovery[c]	European Replication	African American Replication
Sample size		1,940	1,023	855	154	117
Age	Mean (SD)	53.4 (9.6)	53.4 (9.6)	53.7 (9.6)	55.2 (9.4)	54.1 (9.2)
Self-reported race and ethnicity	White (Non-Hispanic/Non-Latino)	1,434	788	772	143	1[d]
	White (Hispanic or Latino)	63	15	1	—	—
	White (Unknown)	115	77	72	11	—
	Black or African American (Non-Hispanic/Non-Latino)	204	85	—	—	101
	Black or African American (Hispanic or Latino)	10	2	—	—	1[d]
	Black or African American (Unknown)	17	13	1	—	13
	Asian	29	10	—	—	—
	Native Hawaiian or Pacific Islander	2	2	—	—	—
	American Indian or Alaska Native	18	7	2	—	—
	Multiple	2	—	—	—	—
	Unknown (Non-Hispanic/Non-Latino)	9	4	4	—	—
	Unknown (Hispanic or Latino)	24	13	—	—	—
	Unknown (Unknown)	13	7	3	—	1[d]
Menopausal status	Post	1,176 (61%)	609 (60%)	513 (60%)	99 (64%)	81 (69%)
	Pre	764 (39%)	414 (40%)	342 (40%)	55 (36%)	36 (31%)
ER/PR Status	ER+/PR+	1,059 (55%)	546 (53%)	475 (56%)	108 (70%)	35 (30%)
	ER+/PR−	223 (11%)	119 (12%)	98 (11%)	15 (10%)	18 (15%)
	ER+/PR unknown	3 (<1%)	2 (<1%)	2 (<1%)	1 (1%)	0 (<1%)
	ER-/PR+	24 (1%)	15 (1%)	13 (2%)	1 (1%)	2 (2%)
	ER-/PR−	629 (32%)	341 (33%)	267 (31%)	29 (19%)	62 (53%)
	ER unknown/PR unknown	2 (<1%)	—	—	—	—
HER2 status	Positive	1,505 (78%)	790 (77%)	660 (77%)	120 (78%)	88 (75%)
	Negative	361 (19%)	195 (19%)	163 (19%)	27 (18%)	26 (22%)
	Unknown	74 (4%)	38 (4%)	32 (4%)	7 (5%)	3 (3%)
Assigned number of cycles	4	1,151 (59%)	572 (56%)	471 (55%)	139 (90%)[e]	75 (64%)
	6	789 (41%)	451 (44%)	384 (45%)	15 (10%)	42 (36%)

[a] Randomized refers to all patients enrolled in CALGB 40101 and assigned to the paclitaxel treatment arm.

[b] Post quality control refers to patients with whole genome data passing quality control (n = 1,029) and excluding patients without evaluable phenotype data (n = 6).

[c] Discovery cohort is all patients with Northwestern European ancestry and evaluable phenotype data.

[d] Identified using principal components analysis of whole genome data.

[e] This reflects the early closure of the 6 cycle arm of the study.

both drugs were closed after enrolling 3,172 patients. Arms were stratified by menopausal, estrogen receptor (ER), progesterone receptor (PR), and HER2 status. Patient demographics are shown in Table 1. Premedication recommendations for the initial dose were 12.5 to 50 mg diphenhydramine and 50 mg ranitidine, 300 mg cimetidine, or 20 mg famotidine administered i.v. 30 to 60 minutes before paclitaxel. Dexamethasone was given as a 10 mg i.v. dose within 60 minutes of paclitaxel or alternatively, as a 10 mg or 20 mg oral dose more than one hour before paclitaxel. To facilitate the 14-day dosing schedule, filgrastim was recommended on days 3 to 10 of each cycle (5 μg/kg rounded to either 300 or 480 μg). Sargramostim (250–500 μg/m^2, days 3–10) or pegfilgrastim (6 mg s.c., 24–36 hours after paclitaxel) could be used in place of filgrastim. The treating physician could omit granulocyte colony-stimulating factor (G-CSF) treatment when confident neutrophils would recover within 14 days; however, if treatment could not be delivered on schedule, then a G-CSF was required in subsequent cycles. Erythropoetin was permitted at the discretion of the treating physician. Patients positive for HER2 by either

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

immunohistochemical 3^+ staining or gene amplification by FISH could initiate adjuvant trastuzumab concurrent with paclitaxel (weekly administration) or at the completion of paclitaxel (weekly or every 3 weeks). Weekly trastuzumab consisted of a 4 mg/kg i.v. loading dose followed by weekly doses of 2 mg/kg and the 3-week schedule of a loading dose of 8 mg/kg and 6 mg/kg every 3 weeks for a total duration of one year.

Genotyping and quality control

A summary of the steps included in sample and single nucleotide polymorphism (SNP) quality control and in principal components analysis (PCA) is illustrated in Supplementary Fig. S1. A total of 1,040 paclitaxel-treated patients with informed consent and a DNA sample (obtained from peripheral blood) available as of July 1, 2009 were included in the primary study. Genomic DNA was genotyped using the HumanHap610-Quad Genotyping BeadChip (Illumina), which interrogated 592,532 SNPs. Subjects with call rates less than 0.98 ($n = 5$) or with suboptimal genotype clustering performance ($n = 1$) were excluded followed by reassessment of genotypes within the remaining subjects. SNPs with call rates less than 0.95, poor genotype clustering performance, more than 1 replicate or Mendelian discordance, relative minor allele frequency (MAF) less than 0.005, nondiploid (e.g., Y or mitochondrial chromosomes), or deemed unreliable by Illumina ($n = 4{,}106$; Tech Note: Infinium Genotyping Data Analysis, 2007) were excluded, leaving 572,745 SNPs. Identity-by-descent (IBD) analysis verified the absence of closely related individuals (proportion IBD > 0.15) and identified one unintended duplicate pair, which was removed and later confirmed to be due to a DNA plating error (PLINK version 1.07; ref. 7). Evaluation of X-chromosome heterozygosity identified 3 genetic males that were also removed and similarly confirmed to be due to a DNA plating error (8). PCA, as implemented by EIGENSOFT version 3.0, was used to visualize the genetic ancestry of the 1,029 individuals passing quality control (9). PCA was conducted using genotypes from study subjects combined with genotypes of unrelated individuals from the HapMap Project representing Northwest European (CEU, $n = 73$), African (YRI, $n = 77$), and Chinese (CHB, $n = 75$) ancestries and genotyped using the same platform by Illumina (Supplementary Fig. S2; ref. 10). To address the potential bias arising from population stratification, we chose to focus our primary analysis on individuals of Northern European descent. A second PCA was conducted using only 1,029 study subjects. Mean values for the first 3 eigenvectors within all patients self-declaring "White" race and "Non-Hispanic" ethnicity were determined. "Genetic Northwest Europeans" (herein called Europeans) were defined as individuals with each of their first 3 eigenvectors within 2 SDs of each mean value irrespective of self-declared race and ethnicity. A total of 859 individuals were identified and identical results were obtained when repeated with the inclusion of HapMap individuals (data not shown). These 859 individuals were the focus of the primary analysis (Supplementary Fig. S2).

Imputation of genotypes was conducted within the 859 Europeans using MACH 1.0 (11) and reference haplotypes from unrelated CEU individuals from either HapMap (r22) or the 1000 Genomes Project (June 2010 release). Before imputation, study genotypes were more stringently filtered and limited to autosomal SNPs with MAF 0.01 or more and exact Hardy–Weinberg P values ≥ 0.001 in control subjects. To address any potential stranding inconsistencies between study genotypes and the reference haplotypes, all symmetric SNPs (A/T or C/G) with MAF more than 0.40, and therefore difficult to resolve, were removed leaving 548,596 and 547,465 SNPs for imputation using the HapMap and 1000 Genomes reference haplotypes, respectively. Imputed SNPs with MAF less than 0.01 or R^2 less than 0.5 were excluded. Genotyping within the replication cohorts (described below) was conducted using TaqMan Allelic Discrimination assays (Applied Biosystems), and individual assays are shown in Supplementary Tables S1 and S2.

One hundred fifty nine self-declared "White" individuals with either "Non-Hispanic" or "Unknown" ethnicity, who enrolled in the CALGB 40101 pharmacogenetic companion study subsequent to the genotyping of the original 1,040 subjects, were used as a replication cohort. Within the discovery set, these criteria accurately identified 98.7% of the 859 Europeans with a false-positive rate of 2.4%. An additional 100 individuals of African ancestry were also identified from within the group of 1,029 individuals passing sample quality control. African ancestry was defined using individuals who self-declared "Black/African American" race with either "Non-Hispanic" or "Unknown" ethnicity. Any individual with their first 3 eigenvectors within 3 SDs of each eigenvector mean value were considered to be of African descent. This self-declared race/ethnicity criteria identified 94.2% of individuals with African ancestry and incorrectly identified 2.0%. The final African American replication cohort consisted of the 100 patients of African descent with genome-wide data and an additional 20 self-declared "Black/African American" individuals with either "Non-Hispanic" or "Unknown" ethnicity, who enrolled after the original genotyping.

Statistical analysis

The primary objective was the identification of SNPs associated with the occurrence of sensory peripheral neuropathy. The analyses were carried out using 2 complementary endpoints: (i) the cumulative dose level triggering the first grade II or higher treatment related sensory peripheral neuropathy episode and (ii) the maximum observed treatment-related sensory peripheral neuropathy grade. The adverse events were graded according to the National Cancer Institute Common Toxicity Criteria for Adverse Events (NCI-CTCAE) version 2.0. The timing of sensory peripheral neuropathy was assessed with a time-to-event approach in which an event was defined as the first incidence of a grade II or higher neuropathy and the time as cumulative paclitaxel exposure (mg/m^2). For patients not experiencing any event, the total study paclitaxel drug exposure was used. These patients are effectively right-censored at the cumulative dose

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

level. The marginal associations were tested using the Cox score test (12). The severity of sensory peripheral neuropathy, defined as the maximum grade neuropathy observed during paclitaxel treatment or within 30 days following the last dose, was evaluated using ordinal logistic regression. Cumulative dose (mg/m^2) was log-transformed and incorporated into the ordinal regression model. For both cases, the marginal null sampling distribution was approximated using asymptotics. These analyses were powered for an additive genetic model. To minimize type I error due to sparseness, SNPs within the European discovery set were constrained to relative MAF of 1% or more and the observation of a minimum of 2 minor allele homozygous genotypes leaving 521,600 evaluable SNPs. Imputed genotypes were represented as allele dosages bound between 0.0 and 2.0. All analyses were conducted using the *R* statistical environment version 2.12 with the cumulative dose-to-event and ordinal analyses implemented using functions from the survival and MASS extension packages (13–16). Quantile-quantile plots of the marginal asymptotic *P* values were evaluated for potential remaining population stratification or inflation of significance levels. Each SNP with a marginal *P* value $\leq 10^{-5}$ was evaluated further for potential errors by checking its MAF (vs. HapMap), Hardy–Weinberg Equilibrium within unaffected subjects, and potentially informative missing rates; they were also visually inspected for genotype clustering performance.

On the basis of the combined results of the time-to-event and ordinal regression analyses of the 859 European patients, a replication plan delineating SNPs, regression model, genetic model (the most plausible model suggested from Kaplan–Meier estimates), and effect direction for one-sided testing was drafted *a priori* to any data collection within the replication cohorts. Three SNPs from the genes *FZD3*, *EPHA5*, and *FGD4* (rs7001034, rs7349683, and rs10771973) were selected for replication based on marginal significance levels, biologic plausibility, and estimated effect size (as detailed in the Results). An additional 10 SNPs with *P* values $< 10^{-5}$ and/or previously implicated in congenital sensory peripheral neuropathies (*NDRG1*) were also evaluated with the specified limitation of being constrained to exploratory analyses. Genotypes for the *FZD3* SNP rs7001034 were captured indirectly using a proxy SNP (rs7833751; $R^2 = 1.0$ CEU HapMap r27) due to the absence of acceptable TaqMan assays to evaluate the locus directly. Because of the impracticality of capturing the *FZD3* linkage disequilibrium (LD) block to the same extent as the European group, this locus was not evaluated in the African ancestry replication group. Direct sequencing was used to capture the *FGD4* rs10771973 genotypes within the replication cohorts. To limit the overall type I error rate for the validation study at the one-sided 0.05 level, we tested each of the 3 SNPs at the marginal 0.01 level. Because the *FGD4* locus replicated in both populations and there are significant differences in LD structure between the European and African American populations, an additional 4 coding region SNPs were chosen from the approximately 30 kb LD block containing rs10771973 to further extend this finding. In addition, to evaluate the independence of the identified association in rs10771973, the time-to-event analysis was repeated with rs10771973 as a covariate. This analysis was conducted using the *R* extension package GenABEL (13). A haplotype-based association test was also conducted for the 3 genes containing the top hits (*EPHA5*, *FZD3*, and *FGD4*), using all genotyped SNPs within 100 kb of the transcription start and stop sites for each gene. Phase for each SNP set was estimated using fastPHASE v 1.1 in all samples combined (17). Haplotype block boundaries using the method of Gabriel and colleagues were generated in Haploview v4.2 using HapMap v3 r2 CEU samples (18). For each haplotype block that included an allele with a per SNP association signal of less than 10^{-3}, individual haplotypes were extracted from fastPHASE output, and haplotypes with frequency less than 5% were combined. Association with outcome was analyzed on a per haplotype basis using time-to-event or maximum grade as described above.

Results

Of the 859 individuals with European ancestry randomly assigned to paclitaxel treatment, 4 withdrew before study treatment and were therefore excluded (Supplementary Fig. S1). Patient characteristics of the CALGB 40101 paclitaxel treatment arm, the genotyped samples, and the discovery and replication cohorts are listed in Table 1. The menopausal, ER, PR and HER2 status, and the assigned number of cycles were not different between the genotyped paclitaxel cohort and the European discovery cohort. The genotyped sample was also representative of all patients randomized to paclitaxel treatment in CALGB 40101. One exception is a fewer number of samples from the 6 cycle paclitaxel arm in the European replication cohort, which reflects the early closure of the 6 cycle arm and the later study enrollment of this group of patients. Peripheral sensory neuropathy was the major dose limiting toxicity in the paclitaxel arm, and the distribution of toxicity grades within the 855 patients in the primary analysis, stratified for number of treatment cycles assigned, is shown in Table 2. Sensory peripheral neuropathy was dose dependent with 17% of the patients randomized to 4 cycles of paclitaxel experiencing a grade II or greater event as compared with 33% of those randomized to 6 cycles of treatment. The cumulative incidence of sensory peripheral neuropathy was similar between the entire cohort randomized to paclitaxel treatment and the discovery set (Supplementary Fig. S3), and between the discovery set and both replication groups (data not shown). There was no effect of age on cumulative dose triggering a grade II or greater peripheral neuropathy event (data not shown).

Among the SNPs analyzed in the GWAS for association with the initial onset of sensory peripheral neuropathy, none reached genome-wide significance although 7 had a marginal significance level of $P < 10^{-5}$ (Table 3 and Supplementary Fig. S4). Inspection of the quantile-quantile plot of the marginal *P* values (Supplementary Fig. S5A) indicates the absence of any remaining population substructure ($\lambda = 1.01$). Of these top SNPs, biologic relevance

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

Table 2. Incidence of sensory peripheral neuropathies in study groups

	Sensory peripheral neuropathy grade					
	0	1	2	3	4	Event rate[b]
Discovery set						
4 Cycles	181 (38%)[a]	209 (44%)	66 (14%)	15 (3%)	0 (0%)	17%
6 Cycles	99 (26%)	160 (42%)	81 (21%)	44 (11%)	0 (0%)	33%
European replication						
4 Cycles	44 (32%)	67 (48%)	24 (17%)	4 (3%)	0 (0%)	20%
6 Cycles	6 (40%)	4 (27%)	3 (20%)	2 (13%)	0 (0%)	33%
African American Replication						
4 Cycles	23 (31%)	33 (44%)	12 (16%)	7 (9%)	0 (0%)	25%
6 Cycles	9 (21%)	19 (45%)	7 (17%)	6 (14%)	1 (2%)	33%

[a] Number of patients and percentage (in parentheses) of all patients in the discovery or replication cohort assigned to 4 or 6 cycles of dose dense paclitaxel.

[b] Event rate is the incidence of a grade 2 or greater sensory peripheral neuropathy.

was apparent for polymorphisms in *EPHA5* (rs7349683; per allele HR, 1.63; 95% CI, 1.34–1.98; $P = 9.6 \times 10^{-7}$; Fig. 1A) and *FGD4* (rs10771973; per allele HR, 1.57; 95% CI, 1.30–1.91; $P = 2.6 \times 10^{-6}$; Fig. 1B). *EPHA5* encodes an ephrin receptor gene implicated in the process of neuronal regeneration following nerve injury and *FGD4* encodes a Rho-GTPase guanine nucleotide exchange factor previously implicated in congenital peripheral neuropathies (19–22). The *FGD4* (Table 3; Supplementary Fig. S6A and S6B) and *EPHA5* (Table 3; Supplementary Fig. S7A and S7B) SNPs were tested in replication cohorts, and association for the former was confirmed in both the European and African American samples (Europeans: rs10771973; per allele HR, 1.72; 95% CI, 1.06–2.80; $P = 0.013$; African Americans: rs10771973; per allele HR, 1.93; 95% CI, 1.13–3.28; $P = 6.7 \times 10^{-3}$). Considering the high minor allele frequency of this risk allele in Europeans, 42% of patients are expected to have a 1.6-fold increased risk and 9% a 2.6-fold increased risk of peripheral neuropathy; in African Americans (MAF 17%), the increased risk is 1.9- and 3.7-fold, respectively. Inspection of the Kaplan–Meier genotype stratified time to neuropathy distributions suggests that an allele dose–effect assumption for *FGD4* rs10771973 is appropriate (Fig. 1B).

No haplotypes in *FGD4* or *EPHA5* showed stronger association with time to sensory peripheral neuropathy than the single SNP analyses in these regions (data not shown). After conditioning the time-to-event analysis on rs10771973, no other genotyped markers at the *FGD4* locus showed association with time to peripheral neuropathy (data not shown). Using imputation to infer additional untyped markers and visualizing the LD structure within the HapMap CEU population revealed an approximately 30 kb region of high LD within the *FGD4* locus showing a strong and reproducible association with the onset of sensory peripheral neuropathy (Supplementary Fig. S8). Approximately, 16 SNPs are strongly linked ($R^2 \geq 0.80$) with rs10771973, 5 of which are synonymous variants within the coding region.

Ordinal logistic regression analyses were used to identify SNPs associated with the severity of sensory peripheral neuropathy. Four SNPs were associated with toxicity grade with a significance level of $P < 1 \times 10^{-5}$ (Table 4 and Supplementary Fig. S4). As with the Cox analysis, a quantile-quantile plot of the normalized marginal P values (Supplementary Fig. S5B) suggests the absence of any remaining population substructure ($\lambda = 0.986$). A SNP within the Frizzled 3 homolog *WNT* signaling receptor gene (*FZD3*) met the threshold of genome-wide significance (rs7001034; $P = 3.1 \times 10^{-9}$; OR, 0.57; 95% CI, 0.48–0.69) and showed a clear relationship between allele dosage and sensory peripheral neuropathy grade (Fig. 2). However, none of these top SNPs from the ordinal regression analysis replicated in either the European or African American populations (Table 4).

Discussion

A small subset of patients exposed to paclitaxel have significant and occasionally protracted neuropathy that has a major impact on quality of life. If we could prospectively identify these patients before administration of paclitaxel, they might be otherwise equally well served with alternative nonpaclitaxel containing regimens. Using a genome-wide association study of CALGB 40101, we have identified several genetic loci associated with the onset or severity of paclitaxel-induced sensory peripheral neuropathy. One of these novel markers associated with early-onset, paclitaxel-induced sensory peripheral neuropathy (*FGD4*, rs10771973) was replicated in both Europeans and African Americans and resides within a gene with a clearly established role in the hereditary peripheral neuropathy Charcot–Marie–Tooth disease (CMT). These findings will inform studies to test the application of genetic markers for optimization of paclitaxel selection, dosing, and adverse event management. Several features of the study design and analysis support the robustness of our findings, including

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

Table 3. Top SNPs from cumulative dose to event analysis

				Discovery ($n = 855$)			European replication ($n = 154$)[d]			African American replication ($n = 117$)[d]		
SNP	Chr	Gene[a]	Alleles	MAF[b]	HR (95% CI)	P[c]	MAF	HR (95% CI)	P	MAF	HR (95% CI)	P
rs7349683	4	*EPHA5*	C/T	0.36	1.63 (1.34–1.98)	9.6×10^{-7}	0.32	0.96 (0.57–1.60)	0.43	0.13	1.16 (0.55–2.42)	0.35
rs4737264	8	*XKR4*	A/C	0.22	1.68 (1.36–2.09)	1.9×10^{-6}	0.24	1.84 (1.02–3.33)	0.021	0.18	1.23 (0.69–2.21)	0.24
rs10771973	12	*FGD4*	G/A	0.31	1.57 (1.30–1.91)	2.6×10^{-6}	0.33	1.72 (1.06–2.80)	0.013	0.17	1.93 (1.13–3.28)	6.7×10^{-3}
rs16948748	17	*PITPNA*	T/G	0.04	2.37 (1.63–3.44)	2.7×10^{-6}	0.02	2.65[e] (0.63–11.1)	0.083	0.07	1.07[e] (0.41–2.77)	0.45
rs16916932	10	*CACNB2*	C/T	0.06	2.08 (1.51–2.87)	4.3×10^{-6}	0.06	0.38[e] (0.09–1.58)	0.082	0.08	1.13[e] (0.47–2.74)	0.39
rs17781082	12	*GRIP1/CAND1*	C/T	0.42	1.60 (1.31–1.96)	4.3×10^{-6}	0.43	1.22 (0.74–1.99)	0.22	0.21	1.32 (0.76–2.30)	0.16
rs1903216	3	*BCL6/*	G/A	0.48	1.59 (1.30–1.95)	5.6×10^{-6}	0.41	2.08[f] (0.99–4.37)	0.024	0.03	3.02[f] (1.04–8.73)	0.016
rs2233335	8	*NDRG1*	T/G	0.38	0.65 (0.52–0.80)	5.2×10^{-5}	0.39	0.94 (0.58–1.53)	0.41	0.23	1.40 (075–2.60)	0.14

[a] Intergenic SNPs are denoted by the closest flanking annotated gene(s).
[b] MAF was calculated within the indicated cohort.
[c] *P* values are 2-sided for discovery analysis and one-sided for replication.
[d] As stated in the replication plan, analyses were exploratory for all except the *EPHA5* and *FGD4* SNPs.
[e] Analysis assumed a dominant model.
[f] Analysis assumed a recessive model.

the prospective design, a large cohort of patients with primary breast cancer who are chemotherapy naive and treated with single-agent paclitaxel, careful collection of sensory peripheral neuropathy and covariate data, strict censoring for dose and cycle reductions for other adverse reactions and preexisting neuropathy, and the use of cumulative dose to the initial incidence of grade 2 toxicity to account for the established effect of total drug exposure on sensory peripheral neuropathy.

The current finding that *FGD4* plays a role in the development of paclitaxel-induced sensory peripheral neuropathy and/or the repair response of peripheral nerves following paclitaxel injury is consistent with the known functions of the gene. *FGD4* encodes for the protein FGD1-related F-actin binding protein (Frabin), and previous studies have shown specific point mutations in *FGD4* can cause the congenital peripheral neuropathy CMT (CMT4H; refs. 21–24). The disease is characterized by a slow progressive demyelination of peripheral sensory and motor neurons accompanied by distal muscle weakness and atrophy, sensory loss, hyporeflexia, and skeletal deformity (25). Paclitaxel-induced peripheral neuropathy shares some of these characteristics, including sensory loss and secondary demyelination (26–28). Frabin is a guanine nucleotide exchange factor for cdc42, a Rho-GTPase that regulates cellular morphogenesis, including myelination. Several hypotheses have been proposed to explain how mutations in *FGD4* might lead to demyelinating CMT4H disease, including disruption of the actin/microtubule cytoskeleton, loss of c-*Jun*-NH-terminal kinase (JNK) activation signals, and disruption of phosphoinositide signaling pathways, all of which could affect Schwann cell myelination and/or the bidirectional communication between Schwann cells and axons (21).

The observed association between the *FGD4* SNP rs10771973 and paclitaxel-induced sensory peripheral neuropathy is consistent with the hypothesis that common *FGD4* polymorphisms subtly affect the development and/or maintenance of Schwann cell function. In this case, carriers of common *FGD4* polymorphisms would have preexisting subclinical abnormalities and a predisposition for toxicity. This is supported by increased risk for paclitaxel-induced sensory peripheral neuropathy in asymptomatic patients with diabetes, previous platinum drug exposures and alcohol use (3), and early Schwann cell activation in response to paclitaxel administration (29). Alternatively, *FGD4* polymorphisms could lead to impaired repair processes such as Schwann cell remyelination and/or axonal regeneration after paclitaxel exposure. Genetic variation in *FGD4* could also directly affect the response of Schwann cells to axonal injury via its ability to activate JNK (30). A neuronal protective role for activated JNK in cultured dorsal root ganglion cells exposed to oxaliplatin has been reported (31). Whether changes in frabin activity or expression lead to a decreased neuronal regenerative capacity and/or an increased sensitivity to paclitaxel-induced sensory peripheral neuropathy requires further study. Interestingly, *FGD4* was identified through a genome-wide siRNA screen in lung

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

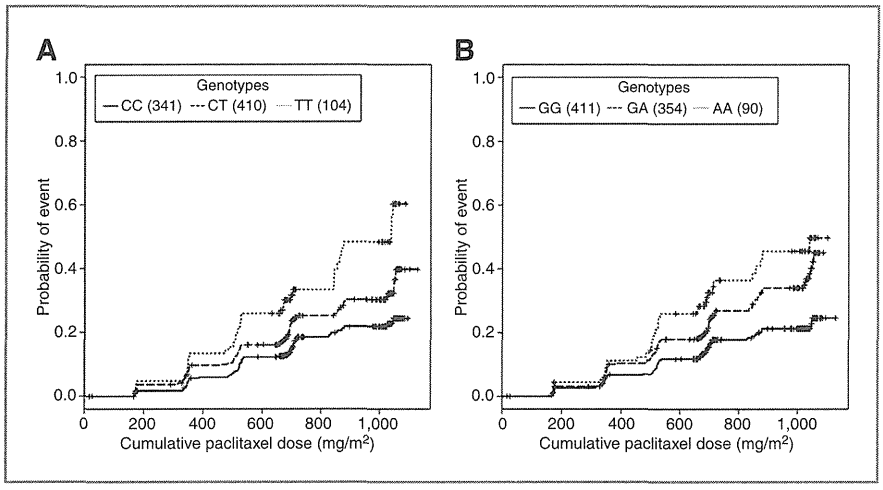

Figure 1. The *EPHA5* rs7349683 C>T and *FGD4* rs10771973 G>A polymorphisms are associated with an increased probability of developing paclitaxel-induced grade 2 or greater sensory peripheral neuropathy. The probability of the first instance of grade 2 or greater neuropathy as a function of cumulative paclitaxel dose (corrected for body surface area) is shown for each genotype. Results are shown for (A) rs7239683 (per allele HR = 1.63; 95% CI, 1.34–1.98; $P = 9.6 \times 10^{-7}$) and (B) rs10771973 (per allele HR, 1.57; 95% CI, 1.30–1.91; $P = 2.6 \times 10^{-6}$) in the discovery set. The number of individuals with each genotype is noted in parentheses.

cancer cell lines as a paclitaxel chemosensitizer. The chemosensitizing properties of *FGD4* are related, at least in part, to its ability to prevent mitotic progression (32). Whether a similar mechanism is involved in the repair response to paclitaxel-induced peripheral neuropathy is unknown.

The *FGD4* rs10771973 SNP is located in the intronic region and is in tight LD with a number of other SNPs. Computational analysis of the genomic region surrounding this SNP found that rs10771972, another intronic SNP in high LD with rs10771973 in both the European and African populations, is predicted to alter conserved transcription factor–binding sites for Myc-Max and USF (data not shown). One could speculate that disruption of either one or both of these transcription factor–binding sites in patients carrying the rs10771973 SNP could lead to altered expression and therefore function of FGD4/Frabin.

The other 2 top hits from the genome-wide analysis are also of potential interest for the paclitaxel-induced sensory peripheral neuropathy phenotype. In the time-to-toxicity analysis, the most significant SNP was in *EPHA5*, which encodes for an ephrin receptor involved in axonal guidance and regeneration following injury. Recent studies have shown that in mice, EphA5 mRNA is rapidly upregulated in response to a sciatic nerve lesion (20) and that EphA5 signaling during synaptogenesis is transduced via cdc42 (19), the Rho-GTPase involved in Frabin signaling. A common SNP in *FZD3* reached genome-wide significance in the ordinal analysis. *FZD3* encodes a Wnt receptor with reported roles in neurite outgrowth (33). In light of the biologic relevance of *EPHA5* and *FZD3* and the limited size of the replication cohorts available for these studies, it will be necessary to further explore the role of these 2 genes in larger populations of paclitaxel-treated patients. Additional studies are also warranted for other top hits, including rs2233335 in the N-myc downstream-regulated gene 1 (*NDRG1*; Supplementary Table S1). Rare mutations in *NDRG1* are also associated with a different subtype of CMT (CMT4D; ref. 34).

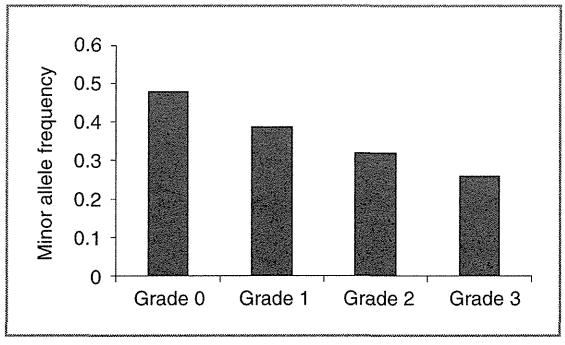

Figure 2. Association of FZD3 SNP rs7001034 with sensory peripheral neuropathy. The minor allele frequency of rs7001034 in the European discovery cohort is expressed as a function of maximal grade of sensory peripheral neuropathy in 855 individuals.

Until the availability of genome-wide approaches for identifying genetic predictors of paclitaxel-induced peripheral neuropathy, candidate gene approaches focused mostly on drug metabolizing enzymes and transporters implicated in paclitaxel exposure. These candidate gene studies yielded no replicated associations of SNPs with paclitaxel-induced sensory peripheral neuropathy, and most were complicated by a very small number of subjects, a retrospective analysis of toxicity, and chemotherapy with multiple agents (35–38). In the current analysis, no significant associations were observed for any SNPs residing in the candidate genes known to influence paclitaxel exposure (Supplementary Table S3), providing further evidence that factors contributing to the function and repair of peripheral nerves are more important than alterations in paclitaxel pharmacokinetics for determining genetic susceptibility to this toxicity. Interestingly, recent analyses of peripheral neuropathy induced by treatment with bortezomib, thalidomide, and vincristine have provided evidence that genes involved in repair mechanisms, inflammation, peripheral nervous system development, and

Published OnlineFirst July 27, 2012; DOI:10.1158/1078-0432.CCR-12-1590

Table 4. Top SNPs from ordinal analysis

				Discovery ($n = 855$)			European Replication ($n = 154$)[d]			African American Replication ($n = 117$)[d]		
SNP	Chr	Gene[a]	Alleles	MAF[b]	OR (95% CI)	P[c]	MAF	OR (95% CI)	P	MAF	OR (95% CI)	P
rs7001034[e]	8	*FZD3*	G/T	0.40	0.57 (0.48–0.69)	3.1×10^{-9}						
rs7833751[f]	8	*FZD3*	G/T	0.40	0.58 (0.49–0.70)	7.5×10^{-9}	0.41	1.67[f] (0.88–3.19)	0.058			
rs5934683	X	*/SHROOM2*	C/T	0.36	1.61 (1.33–1.93)	6.0×10^{-7}	0.33	0.99[g] (0.54–1.82)	0.49	0.30	1.96[g] (0.49–7.89)	0.17
rs2941627	8	*/ZFPM2*	A/G	0.13	1.91 (1.40 -2.51)	3.5×10^{-6}	0.15	1.37[g] (0.68–2.73)	0.19	0.25	0.49[g] (0.24–0.97)	0.021
rs7973533	12	*/BCAT1*	T/G	0.46	0.66 (0.55–0.79)	8.4×10^{-6}	0.44	0.75 (0.47–1.19)	0.11	0.27	1.13 (0.65–1.94)	0.33

[a] Intergenic SNPs are denoted by the closest flanking annotated gene.
[b] MAF was calculated within the indicated cohort.
[c] *P* values are 2-sided for discovery analysis and one-sided for replication.
[d] As stated in the replication plan, all analyses were exploratory except for the *FZD3* SNP.
[e] There was no available TaqMan assay for rs7001034 to use in replication studies.
[f] This SNP tags rs7001034 in the European population but not in the African American population.
[g] Analysis assumed a dominant model.

mitochondrial dysfunction could influence an individual patient's risk of developing toxicity (39–42). However, there was no overlap of implicated genes with the current study (Supplementary Table S3), suggesting that the mechanisms underlying this common toxicity might be drug specific.

To assess the potential translational implications of this finding to clinical practice, we estimated the cumulative dose level triggering an event for each *FGD4* rs10771973 genotype. Considering the data in Fig. 1B, to control the probability of experiencing a neuropathy event at a critical threshold of 33%, the tolerated cumulative dose level for patients with 2 copies of the risk allele is 710 mg/m^2. The corresponding expected critical dose level for patients with one copy of the risk allele is increased to 877 mg/m^2. Patients with no copies of the risk allele are expected to tolerate more than 1047 mg/m^2, corresponding to the full dose of paclitaxel for 6 cycles. If these thresholds are prospectively validated and further refined in follow-up studies, they may be used to estimate tolerable dose levels based on genotype and to tailor the treatment regimen.

While this pharmacogenetic study has several advantages over previous studies on paclitaxel pharmacogenetics, including a large cohort of treatment-naive patients receiving single-agent paclitaxel and a genome-wide approach to discovery, several limitations also exist. The most significant limitation is the sole use of the NCI-CTC for assessment of sensory peripheral neuropathy. It is widely recognized that detailed patient-reported symptom data and a quality of life assessment more accurately describes this phenotype and that physician-reported NCI-CTC grading underreports peripheral neuropathy (43–45). However, it remains difficult to apply these techniques across the multiple sites and large sample sizes required for the sufficient power for pharmacogenetic analyses. In a recent phase III study of 1,060 women treated with taxanes, the Patient Neurotoxicity Questionnaire and the Functional Assessment of Cancer Therapy-General were administered to only the first 300 patients in the study (46). The only use of patient-reported toxicity data and symptom measurements for pharmacogenetic analysis of taxane peripheral neuropathy is limited by the very small sample size of the study (38). While it will be important in follow-up studies to validate these findings using additional instruments, it should be noted that despite its limitations, the NCI-CTC scores are widely accepted for primary evaluation of treatment toxicity in large phase III studies such as CALGB 40101. A second limitation of the current study is the small sample size of the replication cohorts, a common issue confronting almost all pharmacogenetic studies (47).

In summary, our findings support the use of prospective pharmacogenetic analyses of well-phenotyped data sets collected under controlled clinical trial settings and unbiased genome-wide genetic approaches for the identification of novel genes involved in drug efficacy and toxicity. Using a prospective design for validation and replication and a well-controlled single-agent clinical study, we have identified an SNP in *FGD4* associated with increased risk of developing paclitaxel-induced sensory peripheral neuropathy. The