Predictive Biomarkers and Personalized Medicine

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 2.6×10^{-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 6.7×10^{-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

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

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. Premedica tion 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 pegfilgras tim (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

immunohistochemical 3^+ staining or gene amplification by FISH could initiate adjuvant trastuzumab concurrent with paclitaxel (weekly administration) or at the com pletion 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 Sup plementary 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 Genotyp ing 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 repre senting Northwest European (CEU, n 73), African (YRI, n 77), and Chinese (CHB, n 75) ancestries and geno typed 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 individ ual 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 Ameri can" race with either "Non Hispanic" or "Unknown" eth nicity. 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 neu ropathy. The analyses were carried out using 2 complemen tary endpoints: (i) the cumulative dose level triggering the first grade II or higher treatment related sensory peripheral neuropathy episode and (ii) the maximum observed treat ment related sensory peripheral neuropathy grade. The adverse events were graded according to the National Can cer 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

level. The marginal associations were tested using the Cox score test (12). The severity of sensory peripheral neurop athy, 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 incor porated 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 obser vation of a minimum of 2 minor allele homozygous geno types 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 stratifi cation 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 mar ginal 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 con strained 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 Euro pean group, this locus was not evaluated in the African ancestry replication group. Direct sequencing was used to capture the *FGD4* rs10771973 genotypes within the repli cation 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 signif icant 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 meno pausal, 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 enroll ment of this group of patients. Peripheral sensory neurop athy 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 experienc ing a grade II or greater event as compared with 33% of those randomized to 6 cycles of treatment. The cumulative inci dence of sensory peripheral neuropathy was similar between the entire cohort randomized to paclitaxel treat ment 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 Sup plementary Fig. S4). Inspection of the quantile quantile plot of the marginal *P* values (Supplementary Fig. S5A) indicates the absence of any remaining population sub structure (λ 1.01). Of these top SNPs, biologic relevance

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 9.6×10^{-7}; Fig. 1A) and *FGD4* (rs10771973; per allele HR, 1.57; 95% CI, 1.30 1.91; P 2.6×10^{-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 6.7×10^{-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 asso ciation 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 asso ciation 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 periph eral 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 quan tile quantile plot of the normalized marginal P values (Supplementary Fig. S5B) suggests the absence of any remaining population substructure (λ 0.986). A SNP within the Frizzled 3 homolog *WNT* signaling receptor gene (*FZD3*) met the threshold of genome wide significance (rs7001034; P 3.1×10^{-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 prospec tively 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 neuropa thy. 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 estab lished role in the hereditary peripheral neuropathy Char cot 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

Table 3. Top SNPs from cumu at ve dose to event ana ys s

				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)	$9\,6 \times 10^{-7}$	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)	$1\,9 \times 10^{-6}$	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)	$2\,6 \times 10^{-6}$	0 33	1 72 (1 06–2 80)	0 013	0 17	1 93 (1 13–3 28)	$6\,7 \times 10^{-3}$
rs16948748	17	*PITPNA*	T/G	0 04	2 37 (1 63–3 44)	$2\,7 \times 10^{-6}$	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)	$4\,3 \times 10^{-6}$	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)	$4\,3 \times 10^{-6}$	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)	$5\,6 \times 10^{-6}$	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)	$5\,2 \times 10^{-5}$	0 39	0 94 (0 58–1 53)	0 41	0 23	1 40 (075–2 60)	0 14

[a] ntergen c SNPs are denoted by the c osest flank ng annotated gene(s)
[b] MAF was ca cu ated w th n the nd cated cohort
[c] P va ues are 2-s ded for d scovery ana ys s and one-s ded for rep cat on
[d] As stated n the rep cat on p an ana yses were exp oratory for a except the *EPHA5* and *FGD4* SNPs
[e] Ana ys s assumed a dom nant mode
[f] Ana ys s assumed a recess ve mode

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 cumu lative 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 devel opment of paclitaxel induced sensory peripheral neuropa thy and/or the repair response of peripheral nerves follow ing 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, sen sory loss, hyporeflexia, and skeletal deformity (25). Pacli taxel induced peripheral neuropathy shares some of these characteristics, including sensory loss and secondary demy elination (26 28). Frabin is a guanine nucleotide exchange factor for cdc42, a Rho GTPase that regulates cellular mor phogenesis, 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 neu ropathy 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 pro cesses 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 periph eral neuropathy requires further study. Interestingly, *FGD4* was identified through a genome wide siRNA screen in lung

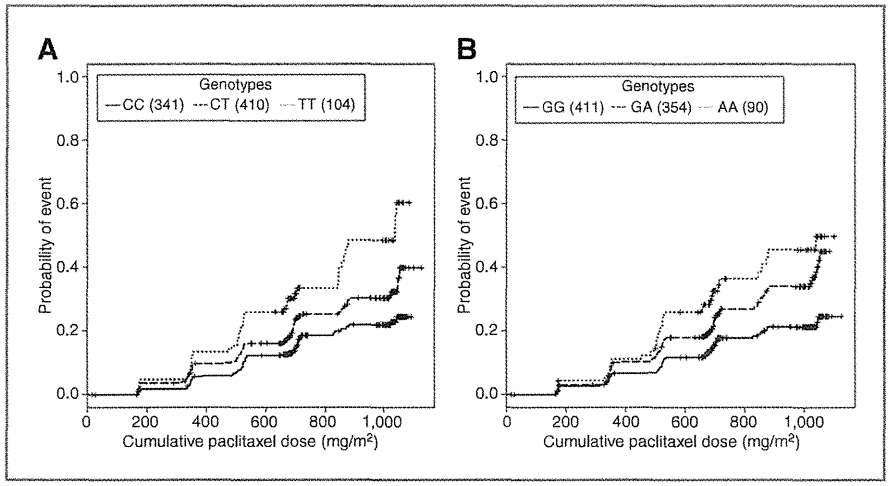


Figure 1. The *EPHA5* rs7349683 C>T and *FGD4* rs10771973 G>A polymorphisms are associated with an increased probability of developing paclitaxel induced grade 2 or greater sensory peripheral neuropathy. The probability of the first instance of grade 2 or greater neuropathy as a function of cumulative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

cancer cell lines as a paclitaxel chemosensitizer. The che mosensitizing 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

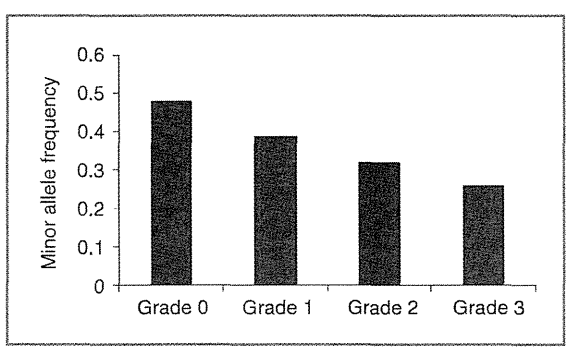


Figure 2. Association of FZD3 SNP rs7001034 with sensory peripheral neuropathy. The minor allele frequency of rs7001034 in the European discovery cohort is expressed as a function of maximal grade of sensory peripheral neuropathy in 855 individuals.

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 com mon 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).

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 signif icant associations were observed for any SNPs residing in the candidate genes known to influence paclitaxel expo sure (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 sus ceptibility to this toxicity. Interestingly, recent analyses of peripheral neuropathy induced by treatment with borte zomib, thalidomide, and vincristine have provided evi dence that genes involved in repair mechanisms, inflam mation, peripheral nervous system development, and

Table 4. Top SNPs from ord na ana ys s

				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)	$3 1 \times 10^{-9}$						
rs7833751[f]	8	*FZD3*	G/T	0 40	0 58 (0 49–0 70)	$7 5 \times 10^{-9}$	0 41	1 67[f] (0 88–3 19)	0 058			
rs5934683	X	*/SHROOM2*	C/T	0 36	1 61 (1 33–1 93)	$6 0 \times 10^{-7}$	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)	$3 5 \times 10^{-6}$	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)	$8 4 \times 10^{-6}$	0 44	0 75 (0 47–1 19)	0 11	0 27	1 13 (0 65–1 94)	0 33

[a] ntergen c SNPs are denoted by the c osest flank ng annotated gene
[b] MAF was ca cu ated w th n the nd cated cohort
[c] *P* va ues are 2-s ded for d scovery ana ys s and one-s ded for rep cat on
[d] As stated n the rep cat on p an a ana yses were exp oratory except for the *FZD3* SNP
[e] There was no ava ab e TaqMan assay for rs7001034 to use n rep cat on stud es
[f] Th s SNP tags rs7001034 n the European popu at on but not n the Afr can Amer can popu at on
[g] Ana ys s assumed a dom nant mode

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 receiv ing 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 diffi cult 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 Neurotox icity Questionnaire and the Functional Assessment of Can cer 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 pharmaco genetic 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 unbi ased 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

involvement of *FGD4* in CMT disease, a congenital peripheral neuropathy, provides strong evidence for the biologic significance of this finding. The fact that a common *FGD4* SNP is associated with an increased risk of paclitaxel induced sensory peripheral neuropathy in patients with both European and African ancestry makes it of potentially broad clinical significance. Additional SNPs in *EPHA5* and *FZD3* were also identified as potential risk factors for the onset and severity of sensory peripheral neuropathy. Additional samples for extension and validation of these findings are currently being collected in ongoing CALGB clinical trials of paclitaxel in the setting of metastatic breast cancer.

Disclosure of Potential Conflicts of Interest

The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute. No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: R.M. Baldwin, K. Owzar, H.L. McLeod, P.N. Friedman, C.A. Hudis, E.P. Winer, L. Shulman, M.J. Ratain, D. Kroetz
Development of methodology: R.M. Baldwin, K. Owzar, H. Zembutsu, M. Kubo, E. Jorgenson, J.S. Witte, D. Kroetz
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R.M. Baldwin, H. Zembutsu, R.J. Eclov, C.A. Hudis, E.P. Winer, L. Shulman, Y. Nakamura, D. Kroetz
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R.M. Baldwin, K. Owzar, H. Zembutsu, A. Chhibber, M. Kubo, C. Jiang, R.J. Eclov, J. Mefford, H.L. McLeod, C.A. Hudis, E. Jorgenson, J.S. Witte, L. Shulman, D. Kroetz
Writing, review, and/or revision of the manuscript: R.M. Baldwin, K. Owzar, A. Chhibber, C. Jiang, H.L. McLeod, P.N. Friedman, C.A. Hudis, E.P. Winer, E. Jorgenson, J.S. Witte, L. Shulman, M.J. Ratain
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): R.M. Baldwin, K. Owzar, M. Kubo, D. Watson, H.L. McLeod, P.N. Friedman, C.A. Hudis
Study supervision: C.A. Hudis, L. Shulman, Y. Nakamura, M.J. Ratain, D. Kroetz

Grant Support

The research for CALGB 60202 and 40101 was supported, in part, by grants from the National Cancer Institute (CA31946) to the Cancer and Leukemia Group B (to M.M. Bertagnolli) and to the CALGB Statistical Center (to D.J. Sargent, CA33601). This work was also supported in part by NIH grants GM61390 and GM61393, and the Biobank Japan Project funded by the Japanese Ministry of Education, Culture, Sports, and Science and Technology. This work is part of the NIH Pharmacogenomics Research Network-RIKEN Center for Genomic Medicine Global Alliance. A. Chhibber and R.J. Eclov were supported in part by NIH Training Grant T32 GM007175.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 15, 2012; revised June 27, 2012; accepted July 17, 2012; published OnlineFirst July 27, 2012.

References

1. Bishop JF, Dewar J, Toner GC, Smith J, Tattersall MH, Olver IN, et al. Initial paclitaxel improves outcome compared with cmfp combination chemotherapy as front line therapy in untreated metastatic breast cancer. J Clin Oncol 1999;17:2355 64.
2. Buzdar AU, Singletary SE, Theriault RL, Booser DJ, Valero V, Ibrahim N, et al. Prospective evaluation of paclitaxel versus combination chemotherapy with fluorouracil, doxorubicin, and cyclophosphamide as neoadjuvant therapy in patients with operable breast cancer. J Clin Oncol 1999;17:3412 7.
3. Lee JJ, Swain SM. Peripheral neuropathy induced by microtubule stabilizing agents. J Clin Oncol 2006;24:1633 42.
4. Rowinsky EK, Donehower RC. Paclitaxel (taxol). N Engl J Med 1995; 332:1004 14.
5. Hershman DL, Weimer LH, Wang A, Kranwinkel G, Brafman L, Fuentes D, et al. Association between patient reported outcomes and quantitative sensory tests for measuring long term neurotoxicity in breast cancer survivors treated with adjuvant paclitaxel chemotherapy. Breast Cancer Res Treat 2011;125:767 74.
6. Shulman LN, Cirrincione CT, Berry DA, Becker HP, Perez EA, O'Regan RO, et al. Six cycles of doxorubicin and cyclophosphamide or paclitaxel are not superior to 4 cycles as adjuvant chemotherapy for breast cancer in women with 0 3 positive axillary nodes: Calgb 40101. J Clin Oncol. 2012 Jul 23. [Epub ahead of print].
7. Purcell S, Neale B, Todd Brown K, Thomas L, Ferreira MA, Bender D, et al. Plink: A tool set for whole genome association and population based linkage analyses. Am J Hum Genet 2007;81: 559 75.
8. Weale ME. Quality control for genome wide association studies. Methods Mol Biol 2010;628:341 72.
9. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome wide association studies. Nat Genet 2006;38:904 9.
10. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million snps. Nature 2007;449:851 61.
11. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. Mach: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010;34:816 34.
12. Cox DR. Regression models and life tables. J Royal Stat Soc Series B 1972;34:187 220.
13. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. Genabel: An r library for genome wide association analysis. Bioinformatics 2007;23:1294 6.
14. Team RDC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2010.
15. Therneau TM, Grambsch PM. Modeling survival data: Extending the cox model. New York: Springer Verlag; 2000.
16. Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New York: Springer; 2002.
17. Scheet P, Stephens M. A fast and flexible statistical model for large scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 2006;78:629 44.
18. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225 9.
19. Akaneya Y, Sohya K, Kitamura A, Kimura F, Washburn C, Zhou R, et al. Ephrin a5 and epha5 interaction induces synaptogenesis during early hippocampal development. PLoS One 2010;5:e12486.
20. Barrette B, Calvo E, Vallieres N, Lacroix S. Transcriptional profiling of the injured sciatic nerve of mice carrying the wld(s) mutant gene: Identification of genes involved in neuroprotection, neuroinflammation, and nerve regeneration. Brain Behav Immun 2010;24:1254 67.
21. Delague V, Jacquier A, Hamadouche T, Poitelon Y, Baudot C, Boccaccio I, et al. Mutations in fgd4 encoding the rho gdp/gtp exchange factor frabin cause autosomal recessive charcot marie tooth type 4h. Am J Hum Genet 2007;81:1 16.
22. Stendel C, Roos A, Deconinck T, Pereira J, Castagner F, Niemann A, et al. Peripheral nerve demyelination caused by a mutant rho gtpase guanine nucleotide exchange factor, frabin/fgd4. Am J Hum Genet 2007;81:158 64.
23. Fabrizi GM, Taioli F, Cavallaro T, Ferrari S, Bertolasi L, Casarotto M, et al. Further evidence that mutations in fgd4/frabin cause charcot marie tooth disease type 4h. Neurology 2009;72:1160 4.
24. Houlden H, Hammans S, Katifi H, Reilly MM. A novel frabin (fgd4) nonsense mutation p.R275x associated with phenotypic variability in cmt4h. Neurology 2009;72:617 20.

25. Patzko A, Shy ME. Update on charcot marie tooth disease. Curr Neurol Neurosci Rep 2011;11:78 88.
26. Lipton RB, Apfel SC, Dutcher JP, Rosenberg R, Kaplan J, Berger A, et al. Taxol produces a predominantly sensory neuropathy. Neurology 1989;39:368 73.
27. Mielke S, Sparreboom A, Mross K. Peripheral neuropathy: A persisting challenge in paclitaxel based regimes. Eur J Cancer 2006;42:24 30.
28. Sahenk Z, Barohn R, New P, Mendell JR. Taxol neuropathy. Electro diagnostic and sural nerve biopsy findings. Arch Neurol 1994;51: 726 9.
29. Cavaletti G, Tredici G, Braga M, Tazzari S. Experimental peripheral neuropathy induced in adult rats by repeated intraperitoneal admin istration of taxol. Exp Neurol 1995;133:64 72.
30. Parkinson DB, Bhaskaran A, Arthur Farraj P, Noon LA, Woodhoo A, Lloyd AC, et al. C jun is a negative regulator of myelination. J Cell Biol 2008;181:625 37.
31. Scuteri A, Galimberti A, Maggioni D, Ravasi M, Pasini S, Nicolini G, et al. Role of mapks in platinum induced neuronal apoptosis. Neurotoxicol ogy 2009;30:312 9.
32. Whitehurst AW, Bodemann BO, Cardenas J, Ferguson D, Girard L, Peyton M, et al. Synthetic lethal screen identification of chemosensi tizer loci in cancer cells. Nature 2007;446:815 9.
33. Endo Y, Beauchamp E, Woods D, Taylor WG, Toretsky JA, Uren A, et al. Wnt 3a and dickkopf 1 stimulate neurite outgrowth in ewing tumor cells via a frizzled3 and c jun n terminal kinase dependent mecha nism. Mol Cell Biol 2008;28:2368 79.
34. Kalaydjieva L, Gresham D, Gooding R, Heather L, Baas F, de Jonge R, et al. N myc downstream regulated gene 1 is mutated in hered *itary motor and sensory neuropathy lom. Am J Hum Genet 2000;67:* 47 58.
35. Bergmann TK, Green H, Brasch Andersen C, Mirza MR, Herrstedt J, Holund B, et al. Retrospective study of the impact of pharmacogenetic variants on paclitaxel toxicity and survival in patients with ovarian cancer. Eur J Clin Pharmacol 2011;67:693 700.
36. Leskela S, Jara C, Leandro Garcia LJ, Martinez A, Garcia Donas J, Hernando S, et al. Polymorphisms in cytochromes p450 2c8 and 3a5 are associated with paclitaxel neurotoxicity. Pharmacogenomics J 2011;11:121 9.
37. Marsh S, Paul J, King CR, Gifford G, McLeod HL, Brown R. Pharma cogenetic assessment of toxicity and outcome after platinum plus taxane chemotherapy in ovarian cancer: The scottish randomised trial in ovarian cancer. J Clin Oncol 2007;25:4528 35.
38. Sissung TM, Mross K, Steinberg SM, Behringer D, Figg WD, Sparre boom A, et al. Association of *abcb1* genotypes with paclitaxel medi ated peripheral neuropathy and neutropenia. Eur J Cancer 2006;42: 2893 6.
39. Broyl A, Corthals SL, Jongen JL, van der Holt B, Kuiper R, de Knegt Y, et al. Mechanisms of peripheral neuropathy associated with bortezo mib and vincristine in patients with newly diagnosed multiple myeloma: a prospective analysis of data from the hovon 65/gmmg hd4 trial. Lancet Oncol 2010;11:1057 65.
40. Cavaletti G, Alberti P, Marmiroli P. Chemotherapy induced peripheral neurotoxicity in the era of pharmacogenomics. Lancet Oncol 2011;12:1151 61.
41. Favis R, Sun Y, van de Velde H, Broderick E, Levey L, Meyers M, et al. Genetic variation associated with bortezomib induced peripheral neu ropathy. Pharmacogenet Genom 2011;21:121 9.
42. Johnson DC, Corthals SL, Walker BA, Ross FM, Gregory WM, Dickens NJ, et al. Genetic factors underlying the risk of thalidomide related neuropathy in patients with multiple myeloma. J Clin Oncol 2011;29: 797 804.
43. Cleeland CS, Farrar JT, Hausheer FH. Assessment of cancer related neuropathy and neuropathic pain. Oncologist 2010;15 Suppl 2:13 8.
44. Kuroi K, Shimozuma K. Neurotoxicity of taxanes: Symptoms and quality of life assessment. Breast Cancer 2004;11:92 9.
45. Postma TJ, Heimans JJ. Grading of chemotherapy induced peripheral *neuropathy. Ann Oncol 2000;11:509 13.*
46. Shimozuma K, Ohashi Y, Takeuchi A, Aranishi T, Morita S, Kuroi K, et al. Taxane induced peripheral neuropathy and health related quality of life in postoperative breast cancer patients undergoing adjuvant chemotherapy: N sas bc 02, a randomized clinical trial. Support Care Cancer. 2012 May 15. [Epub ahead of print].
47. Motsinger Reif AA, Jorgenson E, Relling MV, Kroetz DL, Weinshilboum R, Cox NJ, et al. Genome wide association studies in pharmacoge nomics: successes and lessons. Pharmacogenet Genom 2010. 2010 Jul 15. [Epub ahead of print].

Journal of Human Genetics (2012) 57, 235 237

www.nature.com/jhg

ORIGINAL ARTICLE

A genome-wide association study identifies SNP in *DCC* is associated with gallbladder cancer in the Japanese population

Pei-Chieng Cha[1], Hitoshi Zembutsu[1], Atsushi Takahashi[2], Michiaki Kubo[3], Naoyuki Kamatani[4] and Yusuke Nakamura[1]

Gallbladder cancer (GC) is a relatively uncommon cancer with higher incidence in certain areas including Japan. Because of the difficulty in diagnosis, prognosis of GC is very poor. To identify genetic determinants of GC, we conducted a genome-wide association study (GWAS) in 41 GC patients and 866 controls. Association between each single-nucleotide polymorphism (SNP) with GC susceptibility was evaluated by multivariate logistic regression analysis conditioned on age and gender of subjects. SNPs that showed suggestive association ($P<1\times10^{-4}$) with GC were further examined in 30 cases and 898 controls. SNP rs7504990 in the *DCC* (*deleted in colorectal cancer*, 18q21.3) that encodes a netrin 1 receptor achieved a combined *P*-value of 7.46×10^{-8} (OR=6.95; 95% CI=3.43–14.08). Subsequent imputation analysis identified multiple SNPs with similarly strong associations in an adjacent genomic region, where loss of heterozygosity was reported in GC and other cancers. Reduced expression of *DCC* was indicated to be associated with the poorly differentiated histological type, increased proliferation and metastasis through loss of adhesiveness. However, due to the limited sample size investigated here, further replication study and functional analysis would be necessary to further confirm the result of the association.
Journal of Human Genetics (2012) **57**, 235–237; doi:10.1038/jhg.2012.9; published online 9 February 2012

Keywords: *DCC*; GWAS; gallbladder cancer; SNP

INTRODUCTION

Gallbladder cancer (GC) is the most common malignancy of the biliary tract and the fifth most common cancer of the gastrointest inal tract.[1] With an overall 5 year survival rate of <5%, GC is a highly lethal malignancy with very poor prognosis.[2] Most of GCs were diagnosed at a very late stage because of the lack of symptoms and non specific symptoms of early stage tumors.[1,3] Although the incidence of GC is relatively rare compared with other cancers, higher incidences of GC have been reported in certain geographical regions including India, Pakistan, Ecuador, Korea and Japan.[4] In addition, the prevalence of GC is known to be three times higher in women than in men.[5] Although several clinical risk factors of GC such as gallstones, cholecystitis, porcelain gallbladder, gallbladder polyps, anomalous panreatobiliary duct junction and obesity have been indicated,[1,6] etiology of GC is largely unknown. Only few somatic genetic changes including mutations in K ras, TP53 and *p16^{Ink4}/CDKN2* as well as loss of heterozygosity at several chromo somal regions harboring known or putative tumor suppressor genes have been reported in GCs.[5,7] Hence, identification of novel genetic factors associated with susceptibility to GC should provide new insights into pathogenesis and novel therapeutic interventions of GC. Here, we report a genome wide association study (GWAS) that aims to identify genetic factors associated with GC susceptibility.

MATERIALS AND METHODS

Subjects

In the GWAS, 41 patients who were diagnosed to have GC and 866 control subjects were examined. All cases were registered into Biobank Japan supported by the Ministry of Education, Culture, Sports, Science and Technology, Japan (http://www.biobankjp.org/). Controls consisted of healthy volunteers from Osaka Midosuji Rotary Club. In the replication study, 30 GC patients and 898 controls were investigated. All cases and controls in the replication study were obtained from Biobank Japan. All control subjects do not have medical history of GC or gallstones, cholecystitis and other known confounding diseases for GC. Among the 71 cases, 45 were adenocarcimas and one was a squamous cell carcinoma, and the histological information for the remaining ones was not obtained. All subjects had given written informed consent to participate in the study in accordance with the process approved by Ethical Committee at the Institute of Medical Science of the University of Tokyo and Center for Genomic Medicine of RIKEN. Demographical information of subjects was summarized in Supplementary Table 1.

[1]Institute of Medical Science, Laboratory of Molecular Medicine, Human Genome Center, 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 and [4]Research Groups for Pharmacogenomics, RIKEN Center for Genomic Medicine, Yokohama, Japan
Correspondence: Professor Y Nakamura, Laboratory of Molecular Medicine, Human Genome Center, The University of Tokyo, 4 6 1 Shirokanedai, Minato ku, Tokyo 108 8639, Japan.
E mail: yusuke@ims.u tokyo.ac.jp
Received 28 October 2011; revised 4 January 2012; accepted 13 January 2012; published online 9 February 2012

Genotyping and quality control

All cases and controls in the GWAS were genotyped by using the Illumina HumanHap550 Genotyping BeadChip (San Diego, CA, USA). QC of genotyping data was performed whereby subjects with a call rate of <98%, single nucleotide polymorphisms (SNPs) with a call rate of <99% or minor allele frequency of <0.01, as well as SNPs with a Hardy Weinberg equilibrium test's *P* value of $<1\times10^{-6}$ were excluded from the subsequent statistical analysis. In the replication study, cases and controls were genotyped by using the Illumina HumanHap610 Genotyping BeadChip (San Diego).

Statistical analysis

Association of each SNP with susceptibility to GC was evaluated by logistic regression analysis conditioned on age and gender of subjects. The significance of association was evaluated based on the minimum *P* value among the additive, dominant and recessive model of inheritance. SNPs that showed a minimum *P* value of $<1\times10^{-4}$ in the GWAS were considered as showing suggestive association with GC and were examined in additional subjects. QC and statistical analysis were performed by using the PLINK statistical software (v1.06) (http://pngu.mgh.harvard.edu/~purcell/plink/).[8]

Imputation

Imputation analysis was performed based on genotype of Japanese (JPT) individuals in the Phase II HapMap database (release 24) by using software MACH v1.0 (http://www.sph.umich.edu/csg/yli/mach/index.html).[9] SNPs located in the genomic region within 1500 kb upstream and downstream of the marker SNP, which showed the strongest association with GC, was imputed by implementing 50 Markov Chain iterations. As a QC measure, imputed SNPs with an imputation quality score of <0.3 were excluded from the subsequent association analysis. Pair wise linkage disequilibrium (r^2) between each SNP with the marker SNP was estimated by using the PLINK statistical software (v1.06); whereas regional association plot was generated by the R program v2.10.0 (http://www.r project.org/). Possible functional consequences of SNPs were predicted *in silico* by using the SNPinfo web server (http://manticore.niehs.nih.gov/cgi bin/snpinfo/snpfunc.cgi). eQTL analysis was performed based on data available from the Sanger Institute GENEVAR project[10] for lymphoblastoid cell lines from the four HapMap populations.

RESULTS

In this GWAS, 425 706 SNPs with a total genotyping rate of >99% in 41 GC patients and 866 controls were analyzed after QC of the genotyping data. A genomic inflation factor (λ) in the quantile quantile plot (Supplementary Figure 1a) was calculated to be 0.9903, implying low possibility of false positive associations resulting from the population stratification or cryptic relatedness. The Manhattan plot (Supplementary Figure 1b) indicated that none of the genotyped SNPs achieved genome wide significant association ($P<1\times10^{-7}$) with GC. However, 130 SNPs showed suggestive association with a minimum *P* value of $<1\times10^{-4}$. These SNPs were examined in the additional 30 GC patients and 898 controls, and the genotyping results are shown in Supplementary Table 2. Among these SNPs, SNP rs7504990 that is located in the *DCC* (*deleted in colorectal cancer*) achieved a minimum *P* value of 9.67×10^{-5} under the recessive model of inheritance in the replication study. As illustrated in Table 1, this SNP achieved a *P* value of 7.46×10^{-8} (OR=6.95; 95% CI=3.43 14.08) in the combined study. Genotype AA the risk genotype, was found to be enriched in cases than in the controls, with a genotype frequency of 0.19 versus 0.04. Interestingly, the SNP that showed the second lowest *P* value, SNP rs4078288, is also located in the *DCC* (Table 1). To further characterize the association of the *DCC* region with the GC risk, genotypes of SNPs located in the genomic region within 1500 kb upstream and downstream of the marker SNP rs7504990 were imputed. Association analysis for the imputed and genotyped SNPs with GC susceptibility was then conducted. Figure 1 illustrates the regional plot for association results of the investigated

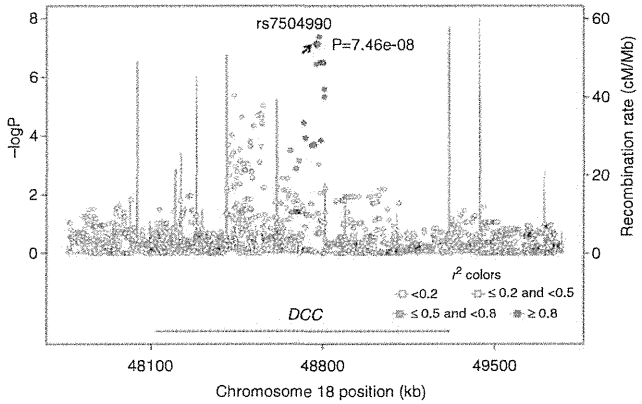


Figure 1 Regional plot for associations of SNPs located within 1500 kb upstream and downstream of the marker SNP, rs7504990, with gallbladder cancer (GC). The $\log_{10}(P)$ values of SNPs were plotted against relative chromosomal locations. Diamond and circle signs represent genotyped and imputed SNPs, respectively. All SNPs are color coded as red (0.8 1.0), orange (0.5 0.8), yellow (0.2 0.5) and white (<0.2), according to their pairwise r^2 to the marker SNP. The marker SNP is indicated by an arrow and the combined *P* value of the marker SNP is represented by a blue diamond sign. SNP positions followed NCBI build 36 coordinates. Estimated recombination rates (cM/Mb) were plotted in dark blue line.

SNPs. We observed that multiple SNPs in the genomic region adjacent to the marker SNP also showed associations, as strong as or stronger than the marker SNP, with GC susceptibility. SNPs that showed *P* values of $<1\times10^{-6}$ with GC susceptibility were listed in Supplementary Table 3. All these SNPs are located in intronic regions of the *DCC*. Online prediction by using the SNPinfo web server (http://manticore.niehs.nih.gov/cgi bin/snpinfo/snpfunc.cgi) revealed that none of these SNPs was likely to alter transcription factor binding site. In addition, none of them was linked to exonic SNPs of the *DCC* with an r^2 value of 0.8 or over. We examined eQTL data for all SNPs in Supplementary Table 3 that showed strong association with GC ($P<1\times10^{-6}$) and found that none of them potentially altered expression of *DCC* in the four investigated HapMap populations.

DISCUSSION

Through the current GWAS and the subsequent replication study, we report here that SNP rs7504990 in the *DCC* showed genome wide significant association with GC susceptibility ($P_{Combined}=7.46\times10^{-8}$; OR=6.95; 95% CI=3.43 14.08) in the Japanese population. Subsequent imputation and association analysis revealed additional SNPs located adjacent to the marker SNP rs7504990 in the *DCC* to be associated with GC susceptibility. Regional association plot indicated that SNPs showing strong association with GC are located in one linkage disequilibrium block that was not disrupted by recombination event. Results from subsequent *in silico* analysis and eQTL analysis are concordant, whereby none of these SNPs potentially alter transcription factor binding site nor possibly alter the expression of *DCC*.

DCC encodes a netrin 1 receptor, which is a transmembrane protein that is a member of the immunoglobulin superfamily of cell adhesion molecules.[11] Loss of heterozygosity and microsatellite instability in the chromosome 18q21 region that contains the *DCC* have been observed in multiple cancer types, particularly cancers in the digestive organ including colon,[12] stomach,[13] esophagus,[14] pancreas,[15] and gallbladder.[16] In addition, reduced or loss of the *DCC* expression has been associated with the poorly differentiated histological type, increased proliferation and metastasis through the loss of adhesiveness.[17]

Table 1 SNPs in *DCC* associated with GC

Chromosome no.	Chromosome location	SNP	A1	Stages	Cases p(11)	Cases p(12)	Cases p(22)	Cases MAF	Controls p(11)	Controls p(12)	Controls p(22)	Controls MAF	Minimum P	OR	L95	U95
18	48771774	rs7504990	A	GWAS	7	11	23	0.30	28	297	533	0.21	9.84E 05	6.99	2.63	18.58
				Replication	6	12	8	0.46	38	272	541	0.20	9.67E 05	8.13	2.83	23.30
				Combined	13	23	31	0.37	66	569	1074	0.21	7.46E 08	6.95	3.43	14.08
18	48801249	rs4078288	G	GWAS	7	11	23	0.30	27	291	539	0.20	5.00E 05	7.71	2.87	20.67
				Replication	5	13	8	0.44	38	270	543	0.20	8.70E 04	6.63	2.18	20.17
				Combined	12	24	31	0.36	65	561	1082	0.20	3.19E 07	6.66	3.22	13.77

Abbreviations: A1, minor allele; CHR: chromosome; GC, gallbladder cancer; L95, lower boundary of 95% confidence interval; MAF, minor allele frequency; OR, odds ratio; p(11), number of subjects with homozygous genotypes for minor allele; p(12), number of subjects with heterozygous genotypes; p(22), number of subjects with homozygous genotypes for major allele; SNP, single-nucleotide polymorphism; U95, upper boundary of 95% confidence interval.

Although the pathophysiological role of *DCC* in gallbladder carcino genesis has not been clarified, *DCC* was demonstrated to induce apoptosis in the absence of its ligand netrin 1.[18] Furthermore, enforced expression of netrin 1 in mouse gastrointestinal tract has been found to induce spontaneous formation of hyperplastic and neoplastic lesions, highlighting the potential role of *DCC* as a tumor suppressor gene.[19]

Our finding that SNPs in *DCC* were associated with GC suscept ibility supports findings of several recent studies, which also pointed out a tumor suppressing role of the gene in gallbladder carcinogenesis. For instances, previous reports have revealed the incidences of loss of heterozygosity at *DCC* in GC to be as high as 30 45%.[16,19,20] Chromosomal loss in human genome often implies that the affected region may harbor a tumor suppressor gene, where the loss of which could lead to carcinogenesis. Nevertheless, both *in silico* and eQTL analysis indicated that SNPs in *DCC* showed no association with expression of the gene, implying other unknown mechanisms might be in action. Subsequent fine mapping and resequencing for this region and functional analysis would be necessary to clarify associa tion between *DCC* and GC.

In the current study, because of the limited number of subjects examined, we might not have enough statistical power to detect other genetic variants with modest or weak effects on susceptibility to GC. Hence, some SNPs of clinical importance might be missed in this study. Further replication study involving a larger number of samples and functional analysis of the *DCC* is urgently needed for validation of the association of *DCC* with the risk of GC.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We express our heartfelt gratitude to all the subjects who participated in this study, and the members of the Rotary Club of Osaka Midosuji District 2660 Rotary International in Japan for supporting our study. We would also like to acknowledge members of the Laboratory for Genotyping Development, RIKEN Center for Genomic Medicine for their excellent technical assistance, and the BioBank Japan for providing DNA samples and clinical information for this study. We also thank Drs Jin Young Jang, Yoon Sup Song, Jae Hyun Park and Siew Kee Low for their kind support and helpful discussion. This work was supported by Leading Project for Personalized Medicine in Ministry of Education, Culture, Sports, Science and Technology, Japan.

1 Misra, S., Chaturvedi, A., Misra, N. C. & Sharma, I. D. Carcinoma of the gallbladder. *Lancet. Oncol.* **4,** 167 176 (2003).
2 Piehler, J. M. & Crichlow, R. W. Primary carcinoma of the gallbladder. *Surg. Gynecol. Obstet.* **147,** 929 942 (1978).
3 Reid, K. M., Ramos-De la Medina, A. & Donohue, J. H. Diagnosis and surgical management of gallbladder cancer: a review. *J. Gastrointest. Surg.* **11,** 671 681 (2007).
4 Randi, G., Franceschi, S. & La Vecchia, C. Gallbladder cancer worldwide: geographical distribution and risk factors. *Int. J. Cancer.* **118,** 1591 1602 (2006).
5 Lazcano-Ponce, E., Miquel, J., Muñoz, N., Herrero, R., Ferrecio, C., Wistuba, I. *et al.* Epidemiology and molecular pathology of gallbladder cancer. *CA Cancer J. Clin.* **51,** 349 364 (2001).
6 Gourgiotis, S., Kocher, H. M., Solaini, L., Yarollahi, A., Tsiambas, E. & Salemis, N. S. Gallbladder cancer. *Am. J. Surg.* **196,** 252 264 (2008).
7 Goldin, R. & Roa, J. Gallbladder cancer: a morphological and molecular update. *Histopathology* **55,** 218 229 (2009).
8 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559 575 (2007).
9 International HapMap Consortium. The International HapMap Project. *Nature* **426,** 789 796 (2003).
10 Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315,** 848 853 (2007).
11 Fearon, E. R., Cho, K. R., Nigro, J. M., Kern, S. E., Simons, J. W., Ruppert, J. M. *et al.* Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* **247,** 49 56 (1990).
12 Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M. *et al.* Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319,** 525 532 (1988).
13 Uchino, S., Tsuda, H., Noguchi, M., Yokota, J., Terada, M., Saito, T. *et al.* Frequent loss of heterozygosity at the DCC locus in gastric cancer. *Cancer Res.* **52,** 3099 3102 (1992).
14 Miyake, S., Nagai, K., Yoshino, K., Oto, M., Endo, M. & Yuasa, Y. Point mutations and allelic deletion of tumor suppressor gene DCC in human esophageal squamous cell carcinomas and their relation to metastasis. *Cancer Res.* **54,** 3007 3010 (1994).
15 Hohne, M. W., Halatsch, M. E., Kahl, G. F. & Weinel, R. J. Frequent loss of expression of the potential tumor suppressor gene DCC in ductal pancreatic adenocarcinoma. *Cancer Res.* **52,** 2616 2619 (1992).
16 Wistuba, I. I., Sugio, K., Hung, J., Kishimoto, Y., Virmani, A. K., Roa, I. *et al.* Allele-specific mutations involved in the pathogenesis of endemic gallbladder carcinoma in Chile. *Cancer Res.* **55,** 2511 2515 (1995).
17 Yoshiya, G., Takahata, T., Hanada, N., Suzuki, K., Ishiguro, A., Saito, M. *et al.* Influence of cancer-related gene polymorphisms on clinicopathological features in colorectal cancer. *J. Gastroenterol. Hepatol.* **23,** 948 953 (2008).
18 Mehlen, P., Rabizadeh, S., Snipas, S. J., Assa-Munt, N., Salvesen, G. S. & Bredesen, D. E. The DCC gene product induces apoptosis by a mechanism requiring receptor proteolysis. *Nature* **395,** 801 804 (1998).
19 Mazelin, L., Bernet, A., Bonod-Bidaud, C., Pays, L., Arnaud, S., Gespach, C. *et al.* Netrin-1 controls colorectal tumorigenesis by regulating apoptosis. *Nature* **431,** 80 84 (2004).
20 Hidaka, E., Yanagisawa, A., Sakai, Y., Seki, M., Kitagawa, T., Setoguchi, T. *et al.* Losses of heterozygosity on chromosomes 17p and 9p/18q may play important roles in early and advanced phases of gallbladder carcinogenesis. *J. Cancer Res. Clin. Oncol.* **125,** 439 443 (1999).

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

A genome-wide association study identifies four genetic markers for hematological toxicities in cancer patients receiving gemcitabine therapy

Kazuma Kiyotani[a], Satoko Uno[c], Taisei Mushiroda[a], Atsushi Takahashi[b], Michiaki Kubo[d], Naoki Mitsuhata[e], Shinomi Ina[f], Chikashi Kihara[g], Yasutoshi Kimura[g], Hiroki Yamaue[f], Koichi Hirata[g], Yusuke Nakamura[a,c] and Hitoshi Zembutsu[c]

Objective **Genetic factors are thought to be one of the causes of individual variability in the adverse reactions observed in cancer patients who received gemcitabine therapy. However, genetic factors determining the risk of adverse reactions of gemcitabine are not fully understood.**

Patients and methods **To identify a genetic factor(s) determining the risk of gemcitabine-induced leukopenia/neutropenia, we conducted a genome-wide association study, by genotyping over 610 000 single nucleotide polymorphisms (SNPs), and a replication study in a total of 174 patients, including 54 patients with at least grade 3 leukopenia/neutropenia and 120 patients without any toxicities.**

Results **We identified four loci possibly associated with gemcitabine-induced leukopenia/neutropenia [rs11141915 in *DAPK1* on chromosome 9q21, combined $P=1.27\times10^{-6}$, odds ratio (OR)=4.10; rs1901440 on chromosome 2q12, combined $P=3.11\times10^{-6}$, OR=34.00; rs12046844 in *PDE4B* on chromosome 1p31, combined $P=4.56\times10^{-5}$, OR=4.13; rs11719165 on chromosome 3q29, combined $P=5.98\times10^{-5}$, OR=2.60]. When we examined the combined effects of these four SNPs, by classifying patients into four groups on the basis of the total number of risk genotypes of these four SNPs, significantly higher risks of gemcitabine-induced leukopenia/neutropenia were observed in the patients having two and three risk genotypes ($P=6.25\times10^{-10}$, OR=11.97 and $P=4.13\times10^{-9}$, OR=50.00, respectively) relative to patients with zero or one risk genotype.**

Conclusion **We identified four novel SNPs associated with gemcitabine-induced severe leukopenia/neutropenia. These SNPs might be applicable in predicting the risk of hematological toxicity in patients receiving gemcitabine therapy. *Pharmacogenetics and Genomics* 00:000–000 © 2012 Wolters Kluwer Health | Lippincott Williams & Wilkins.**

Pharmacogenetics and Genomics 2012, 00:000 000

Keywords: adverse drug reaction, myelosuppression, pancreatic cancer, single nucleotide polymorphism

[a]Laboratory for Pharmacogenetics, [b]Laboratory for Statistical Analysis, [c]Laboratory for Genotyping Development, RIKEN Center for Genomic Medicine, Yokohama, [d]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, [e]Kure Kyosai Hospital, Hiroshima, [f]Second Department of Surgery, Wakayama Medical University, Wakayama and [g]First Department of Surgery, Sapporo Medical University, Sapporo, Japan

Correspondence to Yusuke Nakamura, MD, PhD, 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
Tel: +81 3 5449 5372; fax: +81 3 5449 5433;
e mail: yusuke@ims.u tokyo.ac.jp

Kazuma Kiyotani and Satoko Uno contributed equally to this study

Received 1 July 2011 Accepted 19 October 2011

Introduction

Gemcitabine (2′,2′-difluorodeoxycytidine) is a deoxycytidine analogue that is used for the treatment of patients with various solid tumors, including pancreatic and non-small-cell lung cancers, as a single agent or in combination with platinum agents [1,2]. Dose-limiting adverse drug reactions (ADRs) of gemcitabine are known to be emesis and hematological toxicities, including neutropenia, leukopenia, anemia, and thrombocytopenia [1]. The use of gemcitabine is often limited by these unpredictable dose-limiting toxicities. A large interindividual variation has been noted in the toxicities of gemcitabine, and the frequency of severe leukopenia/neutropenia was reported to be 13–35% [3,4].

Gemcitabine is transported into cells by centrative nucleoside transporters (CNT1 and CNT3; also known as solute carrier (SLC) 28A1 and SLC28A3, respectively) and an equilibrative nucleoside transporter (ENT1; SLC29A1) [5–7], activated by intracellular phosphorylation by deoxycytidine kinase (dCK) to form gemcitabine monophosphate [8], and incorporated into DNA as its triphosphate to inhibit DNA synthesis. It has also been

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website (*www.pharmacogeneticsandgenomics.com*).

1744 6872 © 2012 Wolters Kluwer Health | Lippincott Williams & Wilkins

DOI: 10.1097/FPC.0b013e32834e9eba

suggested that the diphosphate and triphosphate of gemcitabine inhibit ribonucleotide reductase [9]. Gemcitabine is rapidly metabolized to the inactive metabolite 2′,2′-difluorodeoxyuridine by cytidine deaminase (CDA), and excreted into the urine [10]. A number of genetic polymorphisms have been reported in these enzymes and transporters; therefore, genetic variations in these genes may influence the pharmacokinetics and pharmacodynamics of gemcitabine, resulting in differences in the toxicity and efficacy of gemcitabine among individuals. Several case–control association studies focusing on gemcitabine-metabolizing enzymes and transporters as candidates have been carried out [11–14]; however, useful genetic markers to predict toxicities of gemcitabine have not been identified yet. In this study, to identify genetic loci related to the gemcitabine-induced hematological toxicities, we carried out a genome-wide association study (GWAS) by genotyping over 610 000 single nucleotide polymorphisms (SNPs), and identified the loci that are likely to be associated with risk of severe leukopenia/neutropenia in the patients treated with gemcitabine monotherapy.

Materials and methods

Patients

Most of the patients were registered in the BioBank Japan Project (*http://www.biobankjp.org/*) [15], which was started in 2003 by a collaborative network of 66 hospitals across Japan, with the goal of collecting genomic DNA, serum, and clinical information from 300 000 patients who have at least one of the 47 diseases. Among the patients registered in BioBank from June 2003 to December 2008, 461 patients received gemcitabine treatment. Of them, 52 patients (11.3%) experienced grade 3 or 4 leukopenia/neutropenia (ADR group), whereas 86 patients revealed no adverse events (non-ADR group), and the remaining patients had grade 1 or 2 leukopenia/neutropenia and/or other ADRs. We selected 79 patients treated with gemcitabine monotherapy, consisting of 21 ADR patients and 58 non-ADR patients, for the GWAS (Table 1). In a replication study, 33 ADR and 62 non-ADR patients were collected from Sapporo Medical University Hospital, Wakayama Medical University Hospital, Kure Kyosai Hospital, and Hakodate Kyoukai Hospital, as well as BioBank Japan Project from January 2009 to December 2010. Clinical information including drug use and ADRs was collected from medical records. The grades of toxicities were given according to the National Cancer Institute – Common Toxicity Criteria, version 2. As a general control population, we used healthy volunteers from the Midousuji Rotary Club, Osaka, Japan. All participants were of Japanese origin and provided written informed consent. This project was approved by the Institutional Review Board in the Institute of Medical Science, The University of Tokyo, Wakayama Medical University, Kure Kyosai Hospital, and Sapporo Medical University.

Genotyping and quality control

In the GWAS, 79 patients were genotyped using the Illumina Human610-Quad BeadChip (Illumina, San Diego, California, USA). We applied SNP quality control (call rate of ≥ 0.99 in both ADR and non-ADR groups, and a Hardy–Weinberg equilibrium P-value of $> 1.0 \times 10^{-6}$ in the non-ADR group); 470 064 SNPs in autosomal chromosomes passed the quality control filters. SNPs with a minor allele frequency of less than 0.01 were also excluded from further analysis. Of the SNPs analyzed in the GWAS, we selected 70 SNPs of the top 100 SNPs, after considering linkage disequilibrium (LD; $r^2 < 0.8$), for a replication study. In the replication study, we carried out a multiplex polymerase chain reaction-based Invader assay (Third Wave Technologies, Madison, Wisconsin, USA) on ABI PRISM 7900HT (Applied Biosystems, Foster City, California, USA) [16].

Imputation

Imputation was performed by referring to the genotype data of Japanese (JPT) individuals, as deposited in the Phase II HapMap database [17] using MACH v1.0 (*http://www.sph.umich.edu/csg/yli/mach/index.html*). Genotypes

Table 1 Patients' characteristics

Stage	Platform	Source	No. of samples	Female (%)	Age (mean ± SD)	Cancer types, *N*			
						Pancreatic	Lung	Bile duct	Others
GWAS									
ADR	Illumina HumanHap610 Quad	BioBank Japan	21	45.0	64.8 ± 10.9	12	6	1	2
Non ADR	Illumina HumanHap610 Quad	BioBank Japan	58	41.8	64.0 ± 8.7	23	19	10	1
Replication study									
ADR	Invader assay	BioBank Japan, Sapporo Medical University, Wakayama Medical University, Kure Kyosai Hospital	33	35.5	64.2 ± 9.9	28	3	4	3
Non ADR	Invader assay	BioBank Japan, Sapporo Medical University, Wakayama Medical University, Kure Kyosai Hospital	62	30.2	64.9 ± 9.0	36	7	17	2

ADR, adverse drug reaction; GWAS, genome wide association study.

of SNPs that are located in the genomic region within 500 kb upstream or downstream of the marker SNP, which showed the strongest association, at each locus were imputed using genotype data of the GWAS. In the process of imputation, 50 Markov chain iterations were implemented. Imputed SNPs with an imputation quality score of r^2 less than 0.3 were excluded from the subsequent analysis.

Statistical analysis

In the GWAS and the replication study, Fisher's exact test was applied to three genetic models: an allele frequency model, a dominant-inheritance model, and a recessive-inheritance model. Odds ratios (ORs) and confidence intervals (CIs) were calculated for the genetic model with the lowest *P*-value, using a nonrisk allele or a nonrisk genotype as a reference. We used a significance level of 1.07×10^{-7} (0.05/470 064) in the GWAS and 7.14×10^{-4} (0.05/70) in the replication study for adjustment of multiple testing by the strict Bonferroni correction. For combination analysis, the genotype count of the replication study was added to that of the GWAS. The difference in the distribution of age was assessed by the Mann–Whitney *U*-test, and the differences in the sex and cancer types were evaluated by Fisher's exact test. For the prediction scoring system of severe leukopenia/neutropenia induced by gemcitabine, we assigned a score of 1 to individuals homozygous for the risk allele and 0 to individuals with the other genotypes (homozygous and heterozygous for the nonrisk allele), and summed up the scores for each gene to obtain individuals' scores. On the basis of this system, each patient was classified into any of the five groups (group 0, 1, 2, 3, or 4). False discovery rate values were calculated using the Benjamini–Hochberg method to evaluate the probability of false-positive associations [18]. Population stratification for the GWAS data was examined by principal component analysis (PCA) using EIGENSTRAT software v2.0. The four HapMap populations, namely Europeans (CEU), Africans (YRI), and East-Asians (Japanese and Han Chinese, denoted JPT + CHB), were used as reference groups in the PCA. All the statistical analyses were carried out using R statistical environment version 2.12.1 (*http://www.r project.org/*) or PLINK version 1.06 [19]. Haploview software was used to analyze LD values and to draw the LD map [20].

Results

Toxicity of gemcitabine

A total of 174 patients, including 54 ADR patients and 120 non-ADR patients who were treated with gemcitabine monotherapy, were analyzed in this study (Table 1). The distributions of sex (percentage of female patients) were 45.0 and 41.8% in the ADR and non-ADR groups, respectively, in the GWAS (*P* 1.00), and 35.5 and 30.2%, respectively, in the replication samples (*P* 0.636). There was no significant difference in the age distributions (mean ± SD) between ADRs and non-ADRs (64.8 ± 10.9 vs. 64.0 ± 8.7, *P* 0.988 in the GWAS; 64.2 ± 9.9 vs. 64.9 ± 9.0, *P* 0.527 in the replication study). More than half of the patients had pancreatic cancer (*N* 99, 56.9%). The remaining subjects were patients with lung cancer (*N* 35, 20.1%), bile duct cancer (*N* 32, 18.4%), and other cancers (*N* 8, 4.6%). No significant difference in these cancer types was observed between ADR and non-ADR patients in both the GWAS and the replication study (*P* 0.159 and 0.125, respectively).

Genome-wide association and replication studies

We conducted a GWAS of 21 ADRs and 58 non-ADRs in the Japanese patients who received gemcitabine monotherapy using Illumina Human610-Quad BeadChip (Illumina, San Diego, California, USA). After the standard quality control, association analysis was carried out for 470 064 SNPs by Fisher's exact test on the basis of three genetic models: allelic, dominant, and recessive. PCA in the GWAS and HapMap samples showed no evidence of population stratification between the ADR and non-ADR groups (Supplementary Fig. 1, SDC-1, *http://links.lww.com/FPC/A357*). The top 100 SNPs which revealed the smallest *P*-values showed possible associations (2.12×10^{-4} to 6.69×10^{-6}; Supplementary Table 1, SDC-2, *http://links.lww.com/FPC/A358* and Supplementary Fig. 2, SDC-3, *http://links.lww.com/FPC/A359*). To validate the results of the GWAS analysis, we carried out a replication study using 95 independent patients, including 33 ADR and 62 non-ADR patients. Of the top 100 SNPs, we selected and genotyped 70 SNPs since 30 SNPs were highly linked ($r^2 > 0.8$) to another SNP. In the replication study, we identified four SNPs with associations of *P* less than 0.05 (rs11141915 on chromosome 9q21, *P* 2.77×10^{-3}; rs1901440 on chromosome 2q21, *P* 1.82×10^{-2}; rs12046844 on chromosome 1p31, *P* 3.09×10^{-2}; rs11719165 on chromosome 3q29, *P* 4.61×10^{-2}; Table 2). A combined result of the two studies suggested possible associations with loci rs11141915 (combined *P* 1.27×10^{-6}, OR 4.10, 95% CI: 2.21–7.62), rs1901440 (combined *P* 3.11×10^{-6}, OR 34.00, 95% CI: 4.29–269.48), rs12046844 (combined *P* 4.56×10^{-5}, OR 4.13, 95% CI: 2.10–8.14), and rs11719165 (combined *P* 5.98×10^{-5}, OR 2.60, 95% CI: 1.63–4.14; Table 2), although none of them reached the genome-wide significance (*P* 1.07×10^{-7}) when we considered the number of SNPs analyzed in the GWAS. To further characterize the four loci, which were associated with gemcitabine-induced severe myelosuppression, we imputed genotypes of SNPs that were not genotyped in the GWAS but are located within 500 kb upstream or downstream of the four marker SNPs, and examined the associations of these SNPs. Although several SNPs were indicated to be possibly associated with the gemcitabine-induced myelosuppression (Fig. 1), no SNP showed a lower *P* value than the marker SNP. Imputation analysis revealed that rs11141915 represented an associated region spanning 87 kb (chr.9: 89.39–89.48 Mb), which is located in the *DAPK1* gene,

Table 2 Summary of assoc at on resu ts of the genome-w de assoc at on study and rep cat on study

SNP	Chromosome	Chromosome location[a]	Gene	Allele 1/2 (risk)	Stage	ADR				Non-ADR				*P*-value			False discovery rate	Odds ratio (95% CI)[b]
						11	12	22	RAF	11	12	22	RAF	Allelic	Dominant	Recessive		
rs11141915	9	89425614	*DAPK1*	G/T (T)	GWAS	0	3	18	0.07	7	30	21	0.38	1.27×10^{-4}	1.80×10^{-1}	1.04×10^{-4}		7.94 (2.32–27.25)
					Follow-up	0	11	22	0.17	8	31	23	0.38	2.77×10^{-3}	4.73×10^{-2}	9.23×10^{-3}	0.342	3.05 (1.45–6.41)
					Combined	0	14	40	0.13	15	61	44	0.38	1.27×10^{-6}	6.11×10^{-3}	6.91×10^{-6}		4.10 (2.21–7.62)
rs1901440	2	134154429	No gene	A/C (C)	GWAS	11	3	7	0.40	31	27	0	0.23	4.42×10^{-2}	1.00×10^{-0}	4.01×10^{-5}		60.52 (5.45–632.87)
					Follow-up	20	8	5	0.27	42	19	1	0.17	1.30×10^{-1}	5.05×10^{-1}	1.82×10^{-2}	0.488	10.89 (1.22–97.64)
					Combined	31	11	12	0.32	73	46	1	0.20	1.44×10^{-2}	7.39×10^{-1}	3.11×10^{-6}		34.00 (4.29–269.48)
rs12046844	1	66010967	*PDE4B*	T/C (C)	GWAS	1	5	15	0.83	12	32	14	0.52	3.93×10^{-4}	1.67×10^{-1}	1.95×10^{-4}		7.86 (2.56–24.12)
					Follow-up	4	10	19	0.73	7	34	21	0.61	1.50×10^{-1}	1.00×10^{-0}	3.09×10^{-2}	0.542	2.65 (1.11–6.31)
					Combined	5	15	34	0.77	19	66	35	0.57	3.05×10^{-4}	3.43×10^{-1}	4.56×10^{-5}		4.13 (2.10–8.14)
rs11719165	3	196067377	No gene	T/C (C)	GWAS	2	10	9	0.33	26	27	5	0.68	1.15×10^{-4}	3.49×10^{-3}	1.21×10^{-3}		4.27 (2.01–9.05)
					Follow-up	8	16	9	0.48	24	31	7	0.64	4.61×10^{-2}	1.78×10^{-1}	8.12×10^{-2}	0.600	1.87 (1.02–3.42)
					Combined	10	26	18	0.43	50	58	12	0.66	5.98×10^{-5}	3.26×10^{-3}	3.66×10^{-4}		2.60 (1.63–4.14)

ADR, adverse drug reaction CI, confidence interval GWAS, genome-wide association study RAF, risk allele frequency SNP, single nucleotide polymorphism.
[a]On the basis of NCBI 36 genome assembly.
[b]Odds ratios were shown for the model with minimum *P*-values.

encoding death-associated protein kinase 1 (Fig. 1a). rs1901440 on chromosome 2q12 was located in an associated region consisting of a 141-kb LD block (spanning from 134.14 to 134.28 Mb), which contains no known genes (Fig. 1b). Imputation analysis of chromosome 1p31 identified a 63-kb associated region (chr.1: 66.05–66.11 Mb) represented by rs12046844. This region contains the 5′ region of the *PDE4B* gene, encoding phosphodiesterase 4B (Fig. 1c). The rs11719165 locus on chromosome 3q29 was in an associated region spanning 104 kb (chr.3: 196.04–196.14 Mb), which is an intergenic region (Fig. 1d).

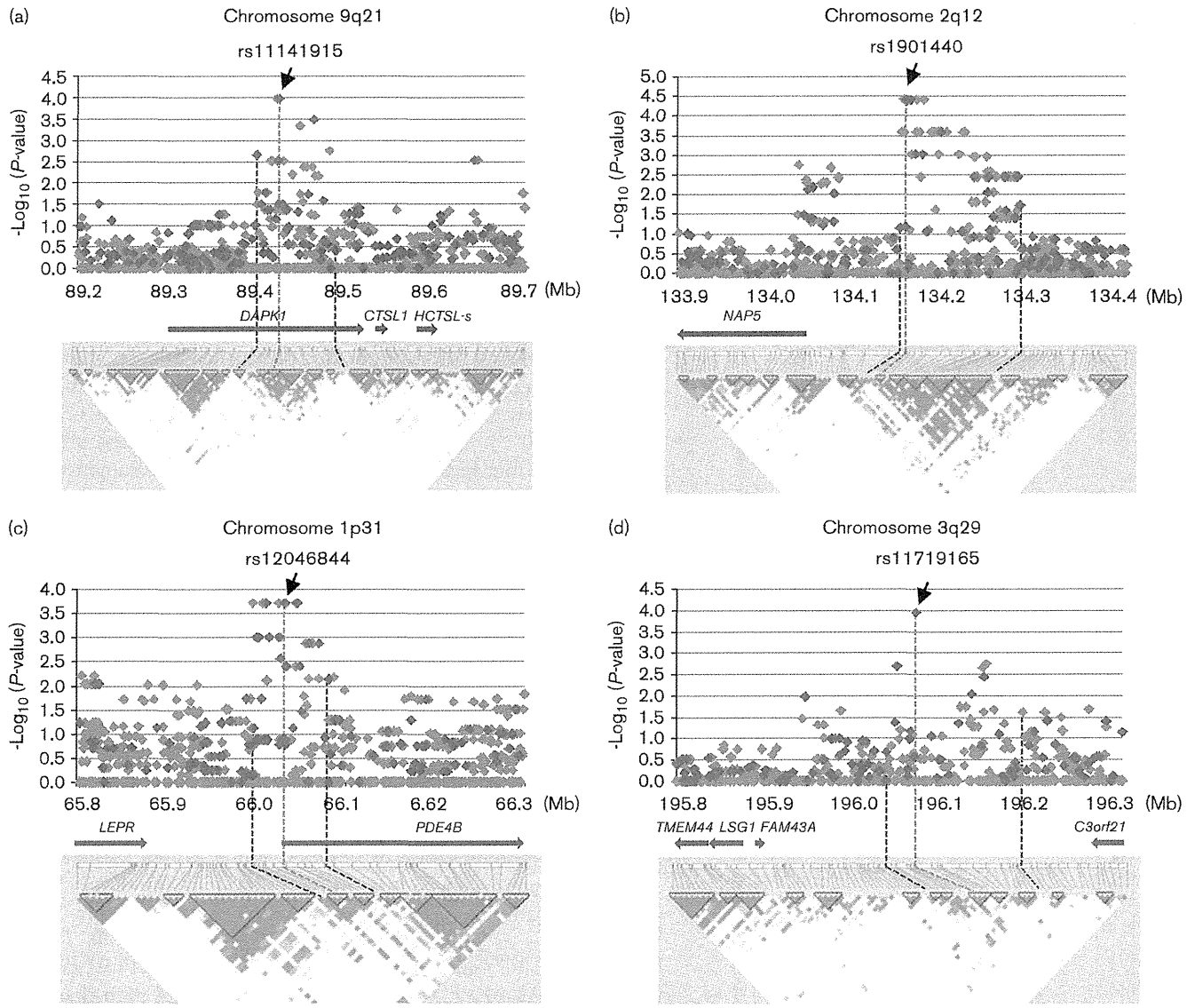
Predictive scoring system for gemcitabine-induced leukopenia/neutropenia

The four SNPs identified by the combined study were independent predictors of gemcitabine-induced toxicities when analyzed by multiple logistic regression ($P < 3.11 \times 10^{-3}$). Therefore, we investigated combined effects of the four loci on the risk of severe leukopenia/neutropenia using a scoring system. For the prediction scoring system, each patient was scored according to the total number of risk genotypes of the four SNP loci; a score of 1 was provided to individuals homozygous for the risk allele and 0 to those with the other genotypes, because the recessive-inheritance model revealed a lower *P*-value than the dominant-inheritance model in each of the four SNP loci. In this population, no patient had the four risk genotypes (score 4). The proportion of patients with gemcitabine-induced leukopenia/neutropenia was significantly increased in groups with higher prediction scores (trend test P 1.31×10^{-14}); the incidences of grade 3/4 leukopenia/neutropenia were 11.5% (13/113) in the combined group of scores 0 and 1, 60.9% (28/46) in the score 2 group, and 86.7% (13/15) in the score 3 group (Table 3). Correspondingly, the OR in the score 3 group was as high as 50.00 (95% CI: 10.13–246.90, P 4.13×10^{-9}) and that of the score 2 group was 11.97 (95% CI: 5.23–27.37, P 6.25×10^{-10}), compared with that in the group of scores 0 and 1. We also examined the distribution of prediction scores in the general control using the four SNPs, which could predict the risk of adverse events in patients treated with gemcitabine. In the general Japanese population, the frequencies of individuals with the scores 0, 1, 2, and 3 were 29.0, 45.3, 20.8, and 4.9%, respectively. This information would be useful to predict the number of patients classified into a group of high, moderate, or low risk before initiation of chemotherapy in the clinical setting.

Discussion

This study is the first GWAS that attempted to identify genetic variants associated with severe hematological toxicity induced by gemcitabine monotherapy, and identified four possible markers, rs11141915, rs1901440, rs12046844, and rs11719165, on chromosomal regions of 9q12, 2q12, 1p31, and 3q29, respectively. Furthermore, the combined analysis of the four SNP loci revealed that

Fig. 1



Regional association plots and linkage disequilibrium (LD) maps of the four loci associated with gemcitabine-induced severe hematological toxicity. *P*-value plots, genomic structures, and LD maps of chromosome regions 9q12 (a), 2q12 (b), 1p31 (c), and 3q29 (d). Blue, diamond-shaped dots represent $-\log_{10}$-transformed minimum *P*-values of single nucleotide polymorphisms genotyped using an Illumina Human610-Quad BeadChip in the genome-wide association study, and red, diamond-shaped dots show $-\log_{10}$-transformed minimum *P*-values of the imputed single nucleotide polymorphisms. Blue arrows indicate the position of known genes. The *D'*-based LD map (minor allele frequency ≥ 0.15) is drawn using genotype data of 79 patients enrolled in the genome-wide association study.

Table 3 Prediction scores of gemcitabine-induced sever leukopenia/neutropenia using rs11141915, rs1901440, rs12046844, and rs11719165

Score	ADR, *N* (%) (*N*=54)	Non ADR, *N* (%) (*N*=120)		Odds ratio (95% CI) *P* value	General control, *N* (%) (*N*=934)
0	4 (7.4%)	50 (41.7%)	]	1.00 (reference)	271 (29.0%)
1	9 (16.7%)	50 (41.7%)	]		423 (45.3%)
2	28 (51.9%)	18 (15.0%)		11.97 (5.23 27.37) 6.25×10^{-10}	194 (20.8%)
3	13 (24.1%)	2 (1.7%)		50.00 (10.13 246.90) 4.13×10^{-9}	46 (4.9%)
			(trend test)	9.91 (5.56 17.67) 1.31×10^{-14}	

ADR, adverse drug reaction; CI, confidence interval.

the number of risk genotypes of the SNPs has cumulative effects on the risk of gemcitabine-induced severe hematological toxicity.

Since this study was conducted with a small number of samples, the statistical power was calculated to be 12 and 43%, respectively, to detect an effect with an OR of 5.0 or greater for an allele with 30% frequency at the genome-wide significance level (α 1.0×10^{-7}) and the significance level of possible association (α 1.0×10^{-5}), suggesting that an SNP with a relatively small OR could easily be missed in our GWAS screening. In addition, the associations of SNPs detected in this study were marginal, which did not reach the genome-wide significance level; therefore, further replication studies are required.

The SNP showing the lowest *P* value in this study, rs11141915 (P 1.27×10^{-6}, OR 4.10), is located in intron 3 of the *DAPK1* gene. DAPK1 is a member of a serine/threonine kinase family that mediates the γ-interferon-induced cell death and also mediates apoptosis induced by tumor necrosis factor-α [21,22]. DAPK1 is reported to be expressed in bone marrow and peripheral blood [23,24]. It has also been reported that the expression of DAPK1 is associated with the resistance to an anticancer drug, irinotecan, in gastric cancer cell lines [25]. Although further analysis is required to clarify the functional importance of DAPK1 in gemcitabine-induced leukopenia/neutropenia, the difference in expression of this protein in the bone marrow might cause interindividual differences in toxicities induced by gemcitabine.

rs12046844 (P 4.56×10^{-5}, OR 4.13; the third locus) was located in a 63-kb region containing the 5′ region of the *PDE4B* gene (Fig. 1c). PDE4B is a phosphodiesterase isozyme in various leukocytes, including neutrophils and monocytes, and plays a key role in the regulation of inflammatory cell activation [26,27]. The expression of PDE4B was increased in non-small-cell lung cancer cells that acquired gemcitabine resistance, and was restored by the treatment with bexarotene, a selective retinoid X receptor agonist, which has the potential of resensitizing gemcitabine-resistant tumor cells [28]. These lines of evidence suggest that PDE4B may regulate the sensitivity of cells to gemcitabine. The second locus (rs1901440 on chromosome 2q12) and the fourth locus (rs11719165 on chromosome 3q29) were in regions containing no reported genes (Fig. 1b and d). According to the UCSC database (*http://genome.ucsc.edu/*), several expressed sequence tags were mapped in these regions, but none of them has open reading frames. Hence, further studies will be required to clarify their functional associations with myelosuppression in patients treated with gemcitabine.

The variants of candidate genes, which are involved in the metabolism and transport of gemcitabine, such as *CDA*, *dCK*, *SLC28A1*, *SLC28A3*, and *SLC29A1*, as well as those that are target molecules of gemcitabine, including ribonucleotide reductase M1 subunit (*RRM1*), *RRM2*, and *RRM2B*, have been suggested to be associated with clinical outcomes and adverse events in gemcitabine therapy [11–14,29–32]. In our GWAS, no SNP in the pharmacokinetics-related candidate genes showed a significant association with the risk of severe leukopenia/neutropenia in the patients receiving gemcitabine monotherapy ($P \geq 1.07 \times 10^{-2}$). We also investigated the association between *CDA*3* (208 G > A; Ala70Thr), which was reported to be associated with higher gemcitabine concentration in plasma, and higher risks of gemcitabine-induced toxicity [12,13,33]. However, no significant association was observed between *CDA*3* and the risk of gemcitabine-induced leukopenia/neutropenia in this study (combined P 5.71×10^{-1}; Supplementary Table 2). For the pharmacodynamics-related candidate genes, we found possible associations between SNPs in *RRM1* (P 1.03×10^{-3}) and *RRM2B* (P 4.78×10^{-4}) in the GWAS samples. However, in the subsequent analysis of the replication sample set, we obtained no supportive results (Supplementary Table 2, SDC-4, *http://links.lww.com/FPC/A360*). Further validation studies using a large number of patients will be required to clarify the effects of these candidate SNPs on the risk of gemcitabine-induced severe leukopenia/neutropenia.

In conclusion, our GWAS using a total of 174 Japanese patients receiving gemcitabine monotherapy has identified four novel candidate loci, 1p31, 2q12, 3q29, and 9q12, which are associated with the risk of gemcitabine-induced grade 3 or 4 leukopenia/neutropenia. Furthermore, the combined analysis of the four SNP loci revealed that the number of risk genotypes of the SNPs was significantly associated with increasing risk of gemcitabine-induced severe hematological toxicity, although replication and validation studies are required. Detailed information about regimen and dosage of gemcitabine is not available in this study; however, further analysis considering this information could improve personalized selection of gemcitabine-based chemotherapy for patients with cancer.

Acknowledgements

The authors express their heartfelt gratitude to all the study participants. They thank Yuka Kikuchi, Aiko Ohno, and Kumi Matsuda for technical assistance. They also thank all other members and staff for their contribution to the sample collection and the completion of their study. This study was supported by a Grant-in-Aid for Leading Project of Ministry of Education, Culture, Sports, Science and Technology of Japan.

Conflicts of interest

There are no conflicts of interest.