## References


1 Cortes-Funes H, Martin C, Abratt R, Lund B. Safety profile of gemcitabine, a novel anticancer agent, in non-small cell lung cancer. *Anticancer Drugs* 1997; **8**:582–587.

2 Heinemann V, Wilke H, Mergenthaler HG, Clemens M, Konig H, Illiger HJ, *et al.* Gemcitabine and cisplatin in the treatment of advanced or metastatic pancreatic cancer. *Ann Oncol* 2000; **11**:1399–1403.

3 Tanaka T, Ikeda M, Okusaka T, Ueno H, Morizane C, Hagihara A, *et al.* Prognostic factors in Japanese patients with advanced pancreatic cancer treated with single-agent gemcitabine as first-line therapy. *Jpn J Clin Oncol* 2008; **38**:755–761.

4 Lee JO, Kim DY, Lim JH, Seo MD, Yi HG, Oh DY, *et al.* Palliative chemotherapy for patients with recurrent hepatocellular carcinoma after liver transplantation. *J Gastroenterol Hepatol* 2009; **24**:800–805.

5 Mackey JR, Yao SY, Smith KM, Karpinski E, Baldwin SA, Cass CE, *et al.* Gemcitabine transport in *xenopus* oocytes expressing recombinant plasma membrane mammalian nucleoside transporters. *J Natl Cancer Inst* 1999; **91**:1876–1881.

6 Ritzel MW, Ng AM, Yao SY, Graham K, Loewen SK, Smith KM, *et al.* Molecular identification and characterization of novel human and mouse concentrative $Na^+$-nucleoside cotransporter proteins (hCNT3 and mCNT3) broadly selective for purine and pyrimidine nucleosides (system cib). *J Biol Chem* 2001; **276**:2914–2927.

7 Mackey JR, Mani RS, Selner M, Mowles D, Young JD, Belt JA, *et al.* Functional nucleoside transporters are required for gemcitabine influx and manifestation of toxicity in cancer cell lines. *Cancer Res* 1998; **58**:4349–4357.

8 Van Rompay AR, Johansson M, Karlsson A. Substrate specificity and phosphorylation of antiviral and anticancer nucleoside analogues by human deoxyribonucleoside kinases and ribonucleoside kinases. *Pharmacol Ther* 2003; **100**:119–139.

9 Heinemann V, Xu YZ, Chubb S, Sen A, Hertel LW, Grindey GB, *et al.* Inhibition of ribonucleotide reduction in CCRF-CEM cells by 2′,2′-difluorodeoxycytidine. *Mol Pharmacol* 1990; **38**:567–572.

10 Plunkett W, Huang P, Gandhi V. Preclinical characteristics of gemcitabine. *Anticancer Drugs* 1995; **6 (Suppl 6)**:7–13.

11 Rha SY, Jeung HC, Choi YH, Yang WI, Yoo JH, Kim BS, *et al.* An association between *RRM1* haplotype and gemcitabine-induced neutropenia in breast cancer patients. *Oncologist* 2007; **12**:622–630.

12 Sugiyama E, Kaniwa N, Kim SR, Kikura-Hanajiri R, Hasegawa R, Maekawa K, *et al.* Pharmacokinetics of gemcitabine in Japanese cancer patients: the impact of a cytidine deaminase polymorphism. *J Clin Oncol* 2007; **25**:32–42.

13 Tanaka M, Javle M, Dong X, Eng C, Abbruzzese JL, Li D. Gemcitabine metabolic and transporter gene polymorphisms are associated with drug toxicity and efficacy in patients with locally advanced pancreatic cancer. *Cancer* 2010; **116**:5325–5335.

14 Chew HK, Doroshow JH, Frankel P, Margolin KA, Somlo G, Lenz HJ, *et al.* Phase II studies of gemcitabine and cisplatin in heavily and minimally pretreated metastatic breast cancer. *J Clin Oncol* 2009; **27**:2163–2169.

15 Nakamura Y. The BioBank Japan Project. *Clin Adv Hematol Oncol* 2007; **5**:696–697.

16 Ohnishi Y, Tanaka T, Ozaki K, Yamada R, Suzuki H, Nakamura Y. A high-throughput SNP typing system for genome-wide association studies. *J Hum Genet* 2001; **46**:471–477.

17 International HapMap Consortium. The International HapMap Project. *Nature* 2003; **426**:789–796.

18 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995; **57**: 289–300.

19 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**:559–575.

20 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**:263–265.

21 Levy-Strumpf N, Kimchi A. Death associated proteins (DAPs): from gene identification to the analysis of their apoptotic and tumor suppressive functions. *Oncogene* 1998; **17**:3331–3340.

22 Kimchi A. DAP genes: novel apoptotic genes isolated by a functional approach to gene cloning. *Biochim Biophys Acta* 1998; **1377**:F13–F33.

23 Larramendy ML, Niini T, Elonen E, Nagy B, Ollila J, Vihinen M, *et al.* Overexpression of translocation-associated fusion genes of *FGFRI*, *MYC*, *NPMI*, and *DEK*, but absence of the translocations in acute myeloid leukemia. A microarray analysis. *Haematologica* 2002; **87**:569–577.

24 Holleman A, den Boer ML, de Menezes RX, Cheok MH, Cheng C, Kazemier KM, *et al.* The expression of 70 apoptosis genes in relation to lineage, genetic subtype, cellular drug resistance, and outcome in childhood acute lymphoblastic leukemia. *Blood* 2006; **107**:769–776.

25 Zhang X, Yashiro M, Qiu H, Nishii T, Matsuzaki T, Hirakawa K. Establishment and characterization of multidrug-resistant gastric cancer cell lines. *Anticancer Res* 2010; **30**:915–921.

26 Muller T, Engels P, Fozard JR. Subtypes of the type 4 cAMP phosphodiesterases: structure, regulation and selective inhibition. *Trends Pharmacol Sci* 1996; **17**:294–298.

27 Torphy TJ. Phosphodiesterase isozymes: molecular targets for novel antiasthma agents. *Am J Respir Crit Care Med* 1998; **157**:351–370.

28 Tooker P, Yen WC, Ng SC, Negro-Vilar A, Hermann TW. Bexarotene (LGD1069, Targretin), a selective retinoid X receptor agonist, prevents and reverses gemcitabine resistance in NSCLC cells by modulating gene amplification. *Cancer Res* 2007; **67**:4425–4433.

29 Tibaldi C, Giovannetti E, Vasile E, Mey V, Laan AC, Nannizzi S, *et al.* Correlation of *CDA*, *ERCC1*, and *XPD* polymorphisms with response and survival in gemcitabine/cisplatin-treated advanced non-small cell lung cancer patients. *Clin Cancer Res* 2008; **14**:1797–1803.

30 Okazaki T, Javle M, Tanaka M, Abbruzzese JL, Li D. Single nucleotide polymorphisms of gemcitabine metabolic genes and pancreatic cancer survival and drug toxicity. *Clin Cancer Res* 2010; **16**:320–329.

31 Yonemori K, Ueno H, Okusaka T, Yamamoto N, Ikeda M, Saijo N, *et al.* Severe drug toxicity associated with a single-nucleotide polymorphism of the cytidine deaminase gene in a Japanese cancer patient treated with gemcitabine plus cisplatin. *Clin Cancer Res* 2005; **11**:2620–2624.

32 Soo RA, Wang LZ, Ng SS, Chong PY, Yong WP, Lee SC, *et al.* Distribution of gemcitabine pathway genotypes in ethnic Asians and their association with outcome in non-small cell lung cancer patients. *Lung Cancer* 2009; **63**:121–127.

33 Ueno H, Kaniwa N, Okusaka T, Ikeda M, Morizane C, Kondo S, *et al.* Homozygous *CDA*3* is a major cause of life-threatening toxicities in gemcitabine-treated Japanese cancer patients. *Br J Cancer* 2009; **100**:870–873.

*Human Molecular Genetics, 2012, Vol. 21, No. 9* **2102–2110**
*doi:10.1093/hmg/dds020*
*Advance Access published on January 27, 2012*

# Genome-wide association study for intracranial aneurysm in the Japanese population identifies three candidate susceptible loci and a functional genetic variant at *EDNRA*

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

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, [2]Laboratory for Statistical Analysis and [3]Laboratory for Genotyping Development, Center of Genomic Medicine, Institute of Physical and Chemical Research (CGM, RIKEN), Tokyo, Japan

Received October 21, 2011; Revised January 14, 2012; Accepted January 23, 2012

**Aneurysmal subarachnoid hemorrhage (aSAH) is the most serious subtype of stroke. Genetic factors have been known to play an important role in the development of intracranial aneurysm (IA), some of which further progress to subarachnoid hemorrhage (SAH). In this study, we conducted a genome-wide association study (GWAS) to identify common genetic variants that are associated with the risk of IA, using 1383 aSAH subjects and 5484 control individuals in the Japanese population. We selected 36 single-nucleotide polymorphisms (SNPs) that showed suggestive association ($P < 1 \times 10^{-4}$) in the GWAS as well as additional 7 SNPs that were previously reported to be associated with IA, and further genotyped an additional set of 1048 IA cases and 7212 controls. We identified an SNP, rs6842241, near *EDNRA* at chromosome 4q31.22 (combined *P*-value = $9.58 \times 10^{-9}$; odds ratio = 1.25), which was found to be significantly associated with IA. Additionally, we successfully replicated and validated rs10757272 on *CDKN2BAS* at chromosome 9p21.3 (combined *P*-value = $1.55 \times 10^{-7}$; odds ratio = 1.21) to be significantly associated with IA as previously reported. Furthermore, we performed functional analysis with the associated genetic variants on *EDNRA*, and identified two alleles of rs6841581 that have different binding affinities to a nuclear protein(s). The transcriptional activity of the susceptible allele of this variant was significantly lower than the other, suggesting that this functional variant might affect the expression of EDNRA and subsequently result in the IA susceptibility. Identification of genetic variants on *EDNRA* is of clinical significance probably due to its role in vessel hemodynamic stress. Our findings should contribute to a better understanding of physiopathology of IA.**

## INTRODUCTION

Intracranial aneurysms (IAs) are balloon-like dilations of the intracranial arterial wall in the brain. Rupture of IA causes subarachnoid hemorrhage (SAH), a serious subtype of stroke, which leads to fatality in ~50% of the cases and results in significant disability in 30% of the cases (1). The age- and sex-adjusted annual incidence and mortality rates of SAH were 23 and 9 per 100 000 for all ages, respectively, in Japan (1). The annual rupture risk of IA in Japan is relatively high at ~2.7% (2). The incidence of SAH is particularly higher in Finland and Japan than in the rest of the world (1,3–5).

Both environmental and genetic factors are known to be involved in the development of IA, and several studies have indicated that hypertension, hypercholesterolemia, cigarette smoking and female gender are risk factors for IA (6–10). Various Mendelian hereditary connective tissue disorders such as autosomal dominant polycystic kidney disease (11) and

*To whom correspondence should be addressed at: 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 354495372; Fax: +81 354495433; Email: yusuke@ims.u-tokyo.ac.jp

type-IV Ehlers–Danlos syndrome (12) are the inherited conditions that increase the risk of IA. Additionally, a positive family history for IA is considered to be an important risk factor, as the incidence of harboring IA for individuals who have at least two affected first-degree relatives was reported to range from 6 to 10% (6,13,14). With the high incidence of familial IA, several susceptible loci which include chromosomes 1p34.3–p36.13 (15), 7q11 (16), 19q13.3 (17) and Xp22 (18) were successfully identified to be linked to familial IA through the linkage analysis. All this evidence has strongly implied the role of genetic factors contributing to the physiopathology of IAs.

Since there is no clear symptom of IA, identification of risk factors, particularly genetic risk factors, that lead to an increase in the risk of the formation and rupture of IA, which result in aneurysmal subarachnoid hemorrhage (aSAH), is critically essential. With the current advances in biotechnology, it is now feasible to identify common genetic variants that are associated with polygenic diseases by means of genome-wide high-density SNP array. Two genome-wide association studies (GWAS) of multiple European populations have successfully identified common variants located on chromosomes 8q12.1, 9p21.3, 10q24.32, 13q13.1 and 18q11.2 that are associated with IA. Notably, these associated loci were successfully replicated in the Japanese population (19,20). Because of the complex linkage disequilibrium (LD) structures across different populations and potential interaction between genetic variants and environment factors, it is well known that the effect size of common genetic variants associated with the disease varies among different populations. Although a GWAS of IA in the Japanese population was reported previously, the association was not conclusive due to the lack of validation analysis as well as the insufficient statistical power of the study (21). Hence, we conducted an independent GWAS of IA with a larger samples size for the identification of genetic variants associated with IA in the Japanese population.

## RESULTS

To identify genetic variants associated with susceptibility to IA in the Japanese population, we performed a GWAS, using 1383 aSAH patients and 5484 control (Supplementary Material, Table S1) subjects, with Illumina OmniExpress BeadChip Kits that contained 733 202 SNPs. After quality check of the SNP genotyping data, a total of 565 149 autosomal SNPs were used for association analysis (Supplementary Material, Table S2).

Principal component analysis (PCA) revealed that all the subjects participating in this study were clustered in the Asian population (Supplementary Material, Fig. S1). The detailed PCA analysis on the basis of the genotype information from the cases and controls classified the sample populations mostly into two major clusters consisting of the Ryuukyu cluster (southern islands of Japan) and the Hondo cluster (mainland cluster) (Supplementary Material, Fig. S2). To avoid influences of population substructure in the sample populations as well as age and gender biases, the association study was performed by logistic regression analysis with associated eigenvectors, age and gender as covariates.

A quantile–quantile (Q–Q) plot for this GWAS based on 565 149 SNPs is shown in Supplementary Material, Figure S3. The genomic inflation factor ($\lambda_{GC}$) of the test statistic in this study was 1.055. Since it is known that the $\lambda_{GC}$ value increases with an increase of the sample size, we calculated the $\lambda_{GC}$ value adjusted to a sample size of 1000—which was 1.031, indicating a low possibility of false-positive association by population stratification. We subsequently applied the genomic control method to adjust the *P*-values and used the adjusted *P*-values ($P_{GC}$) for further analysis. The Manhattan plot shown in Figure 1 indicated no SNP to have achieved genome-wide significance level with the threshold at a $P_{GC}$ value of $<5 \times 10^{-8}$.

Six SNPs (rs10958409, rs9298506, rs1333040, rs11191514, rs1980781 and rs11661542) on chromosomes 8q11.23, 9p21.3, 10q24.32, 13q13.1 and 18q11.2, which were previously reported to be associated with IA in the European population, showed nominal association with *P*-values from $4.50 \times 10^{-2}$ to $9.52 \times 10^{-5}$ (Supplementary Material, Table S3). On chromosome 9p21.3, although the previously reported SNP, rs1333040, revealed a *P*-value of $3.79 \times 10^{-2}$, we identified another SNP, rs10757272, on this locus that showed stronger association with IA with a *P*-value of $7.75 \times 10^{-4}$. However, we observed no association with *P*-values of $>0.05$ for the five SNPs (rs7542311, rs358345, rs4628172, rs6461176, rs10217224) that were previously indicated in the GWAS in the Japanese population (Supplementary Material, Table S3).

To further validate a possible genetic variant(s) associated with IA, we selected a total of 64 SNPs showing suggestive association ($P_{GC} < 1 \times 10^{-4}$) with IA. After excluding SNPs that possess LD coefficient ($r^2$) of $>0.8$ within each LD block, we performed a replication study of 36 SNPs, using an independent set of samples consisting of 1048 IA patients and 7212 controls. In addition, we further analyzed seven previously reported SNPs, rs10958409, rs9298506, rs1333040, rs11191514, rs1980781 rs11661542 and rs10757272, that showed nominal association (*P*-value of $<0.05$) with IA in our first stage.

Among 43 SNPs in the replication study, two SNPs (rs6842241 on 4q31.22, rs10757272 on 9p21.3) were successfully replicated with Bonferroni-corrected *P*-value of $<1.16 \times 10^{-3}$ (0.05/43 independent tests) at the replication phase as shown in Table 1 and Supplementary Material, Table S4. The association of these two SNPs with IA was statistically significant, considering strict multiple testing with the Bonferroni correction. After evaluating the combined association of the discovery GWAS and replication stage using weighted inverse-variance meta-analysis, we identified an SNP, rs6842241, to have achieved the genome-wide significant level of association with IA in the Japanese population, yielding a combined *P*-value of $9.58 \times 10^{-9}$ (OR = 1.25; 95% CI = 1.16–1.34) without any significant heterogeneity (*P* for heterogeneity = 0.606 with $I^2 = 0.0\%$). This SNP is located within the regulatory region of the *EDNRA* gene on the chromosome locus 4q31.22. We also identified another *EDNRA* intronic SNP, rs17612742 ($r^2 = 0.99$ with rs6842241), from the GWAS to be significantly associated with IA. Imputation analysis of this locus identified two additional SNPs, rs6841581 and rs1878406, with an $r^2$-value of $>0.8$ showing similar levels of association with rs6842241 (Fig. 2A).

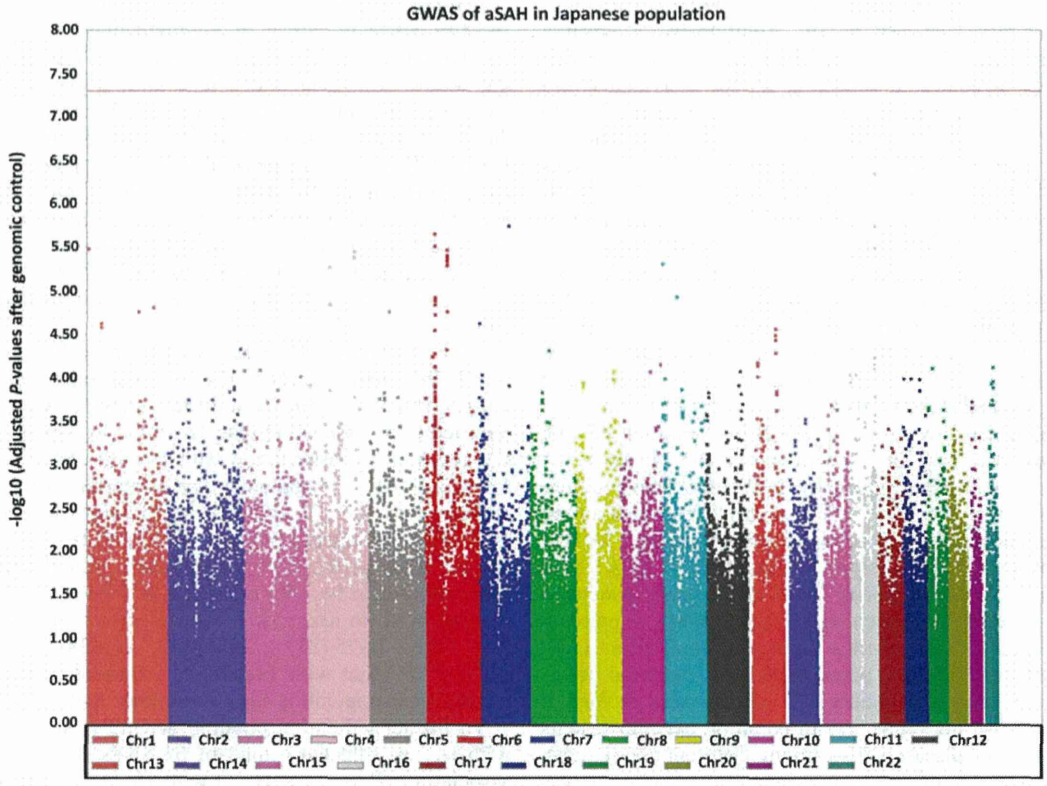


**Figure 1.** Manhattan plot of the GWAS of aSAH. A plot that utilized genomic-control adjustment *P*-values after eigenvectors, age and gender correction for 565 149 autosomal SNPs against their respective position on each chromosome.

Among the previously reported candidate loci (8q11.23, 9p21.3, 10q24.32, 13q13.1 and 18q11.2) associated with IA, the SNPs, rs1333040 and rs10757272 ($r^2 = 0.57$ with rs1333040), within the *CDKN2BAS* gene on a 9p21.3 region were replicated their significant association with IA in the Japanese population. In this study, rs10757272 revealed a combined *P*-value of $1.55 \times 10^{-7}$ (OR = 1.21; 95% CI = 1.13–1.30) (Table 1). The other SNP, rs1333040, on the same locus was also successfully replicated, but the effect of this variant was less significant with a *P*-value of $5.56 \times 10^{-5}$ (OR = 1.16, 95% CI = 1.09–1.25) than rs10757272 (Supplementary Material, Table S4). Imputation analysis on the 9p21.3 region indicated that most of the strongly associated SNPs were located at the 3′ end of *CDKN2BAS* (Fig. 2B).

After adjustment for known IA risk factors, which included smoking and hypertension, the association of rs6842241 and rs10757272 with IA remained strong ($2.40 \times 10^{-9}$, OR = 1.26, 95% CI = 1.17–1.36; $1.32 \times 10^{-8}$, OR = 1.23, 95% CI = 1.24–1.33) (Supplementary Material, Table S5).

Among the SNPs that were previously reported, rs11191514 on 10q24.32 and rs1980781 on 13q13.1 revealed less significant association with IA in the Japanese population; the combined analysis of the two stages showed a *P*-value of $9.68 \times 10^{-5}$ (OR = 1.17, 95% CI = 1.09–1.26) for rs11191514 on 10q24.32 and $9.20 \times 10^{-5}$ (OR = 1.17, 95% CI = 1.09–1.26) for rs1980781 on 13q13.1.

Although the SNP rs671 located in *ALDH2* on chromosome 12q24.12 was marginally replicated with a *P*-value of $5.56 \times 10^{-3}$, the meta-analysis combining the two stages revealed stronger association with a *P*-value of $2.63 \times 10^{-6}$ (OR = 1.24, 95% CI = 1.15–1.34). Owing to the functional relevance of this locus to IA, this SNP was considered to be significant in the susceptibility to IA in the Japanese population (Table 1).

Identification of novel genetic variants on the *EDNRA* gene revealed important insights into IA pathogenesis because of the biological function of this gene. Hence, we further investigated the role of genetic variants in the transcriptional level or the protein function. We identified four SNPs in an intron (rs17612742), or an upstream (rs6841581, 1878406 and rs6842241) of *EDNRA* to be significantly associated with IA through GWAS and imputation analysis. To examine the effect of these SNPs on the transcription, we performed electrophoretic mobility shift assays (EMSAs) and identified allelic differences in the binding affinity of a nuclear protein(s) from HEK293 cells, using the oligonucleotides corresponding to each allele of rs6841581 and rs1878406. The bands corresponding to the susceptible alleles G for rs6841581 and C for rs1878406 appeared to be strong, but those for protective alleles A and T appeared to be weak or undetectable (Fig. 3A). The different binding affinity between the two alleles for rs6841581 and those for rs1878406 were confirmed

**Table 1.** Association analysis of SNPs on chromosomes 4, 9 and 12 with IA

| Chr. | SNP | Chrloc. | Gene | Stage | Allele | | Case | | | Control | | | Risk allele frequency | | *P*-value[a] | OR | 95% confidence interval | | $P^{b}_{hetero}$ | $I^2$ (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 1 | 2 | 11 | 12 | 22 | 11 | 12 | 22 | Case | Control | | | L95 | U95 | | |
| 4 | rs6842241 | 148400819 | *EDNRA* | GWAS | A | C | 105 | 512 | 766 | 503 | 2356 | 2624 | 0.739 | 0.693 | 4.04E − 06 | 1.262 | 1.146 | 1.391 | | |
| | | | | Replication | A | C | 83 | 394 | 570 | 650 | 3110 | 3451 | 0.733 | 0.694 | 3.67E − 04 | 1.226 | 1.096 | 1.372 | | |
| | | | | Combined | A | C | 188 | 906 | 1336 | 1153 | 5466 | 6075 | 0.736 | 0.694 | 9.58E − 09 | 1.249 | 1.161 | 1.343 | 6.06E − 01 | 0.0 |
| 9 | rs10757272 | 22088260 | *CDKN2BAS* | GWAS | C | T | 137 | 571 | 674 | 663 | 2474 | 2346 | 0.694 | 0.654 | 7.75E − 04 | 1.175 | 1.072 | 1.289 | | |
| | | | | Replication | C | T | 97 | 463 | 481 | 931 | 3253 | 3024 | 0.684 | 0.645 | 5.45E − 05 | 1.245 | 1.119 | 1.384 | | |
| | | | | Combined | C | T | 234 | 1034 | 1155 | 1594 | 5727 | 5370 | 0.690 | 0.649 | 1.55E − 07 | 1.213 | 1.133 | 1.300 | 8.88E − 01 | 0.0 |
| Suggestive association | | | | | | | | | | | | | | | | | | | | |
| | | | | GWAS | T | C | 64 | 465 | 854 | 372 | 2083 | 3029 | 0.786 | 0.742 | 8.43E − 05 | 1.238 | 1.116 | 1.372 | | |
| 12 | rs671 | 112241766 | *ALDH2* | Replication | T | C | 67 | 343 | 638 | 491 | 2629 | 4091 | 0.772 | 0.750 | 5.56E − 03 | 1.180 | 1.050 | 1.326 | | |
| | | | | Combined | T | C | 131 | 808 | 1492 | 863 | 4712 | 7120 | 0.780 | 0.746 | 2.63E − 06 | 1.240 | 1.148 | 1.338 | 1.26E − 01 | 57.2 |

Chr., chromosome; Chrloc., chromosome location (bp); 11, homozygous genotype for allele 1; 12, heterozygous genotype; 22, homozygous genotype for allele 2; RAF, risk allele frequency; OR, odds ratio (calculated referred to risk allele).

[a]*P*-values of the GWAS were obtained from logistic regression analysis with eigenvectors, age and gender as covariates after genomic controls; *P*-values of the replication stage were obtained from logistic regression analysis with age and gender as covariates; *P*-values for combined were based on the weighted inverse-variance meta-analysis.

[b]$P_{hetero}$ values were obtained from heterogeneity analysis based on Cochran's $Q$ statistic and $I^2$ statistic (46).

by the competition assay in which non-labeled oligonucleotides were added at different concentrations; the increase of the non-labeled oligonucleotides inhibited the binding of the labeled oligonucleotides in a dose-dependent manner (Fig. 3B).

To identify whether these SNPs could affect the transcriptional activity of *EDNRA*, we performed reporter assays by inserting either of the oligonucleotides corresponding to the two alleles at the two SNP loci into luciferase-expressing vectors. Plasmids containing the susceptible allele G for the SNP rs6841581, which showed higher binding affinity to the nuclear protein(s), revealed significantly lower luciferase activity than the non-susceptible allele A (Fig. 3C), although no difference between the alleles was observed for the SNP rs1878406. It is notable that the non-susceptible allele A for rs6841581 revealed no enhancer activity in comparison with the mock (empty) vector, but plasmids containing the susceptible allele G revealed the suppressive effect on the activity (Fig. 3C). Taking together, our findings from the EMSA and reporter assays suggest that the 5′ flanking region including the SNP rs6841581 on *EDNRA* might function as a transcriptional repressor and that this SNP is likely to be a functional variant conferring IA susceptibility.

## DISCUSSION

To identify genetic variants associated with IA in the Japanese population, we performed a GWAS and a replication study with a total of 2431 aSAH/IA subjects and 12 696 control individuals. Among the identified SNPs, we did not observe significant differences in odds ratio (risk) between the aSAH patients in the discovery GWAS phase and the IA patients in the replication stage, indicating that the identified SNPs are likely to be associated with the risk of IA development.

The SNP rs6842241, which is located 1.25 kb upstream from the *EDNRA* gene encoding endothelin receptor A, revealed the most significant association with the combined *P*-value of $9.58 \times 10^{-9}$. Identification of this genetic variant in the regulatory region of *EDNRA* is clinically interesting since endothelin-1 (EDN-1) and its receptors, EDNRA and EDNRB, have been known to play a significant role in IA pathophysiology. EDN-1 is a potent vasoconstrictor produced by the endothelial cells in the vasculature system. The effect of EDN-1 is mediated by two major receptor subtypes, EDNRA and EDNRB, which activate a G-protein(s) and their second messenger system. EDNRA is located predominantly on vascular smooth muscle cells of the cerebrovascular system (22) and mediates vasoconstriction and proliferation (23). Accumulated evidence strongly implies that the EDN-1/EDNRA and EDNRB pathways are critically important to maintain the balance of vasoconstriction and vasodilatation in response to the hemodynamic stress. Previous reports suggested correlation of the elevated level of EDN-1 in the cerebrospinal fluid and plasma in patients with aSAH having persistent cerebral vasospasm (24,25). Interestingly, *EDNRA* variants were previously shown to be associated with a few cerebrovascular diseases including migraine (26), myocardiac infarction (27,28) and cystic fibrosis pulmonary disease (29). Furthermore, *EDNRA* variants were also associated with essential hypertension (30) and it is overexpressed in the arteries of

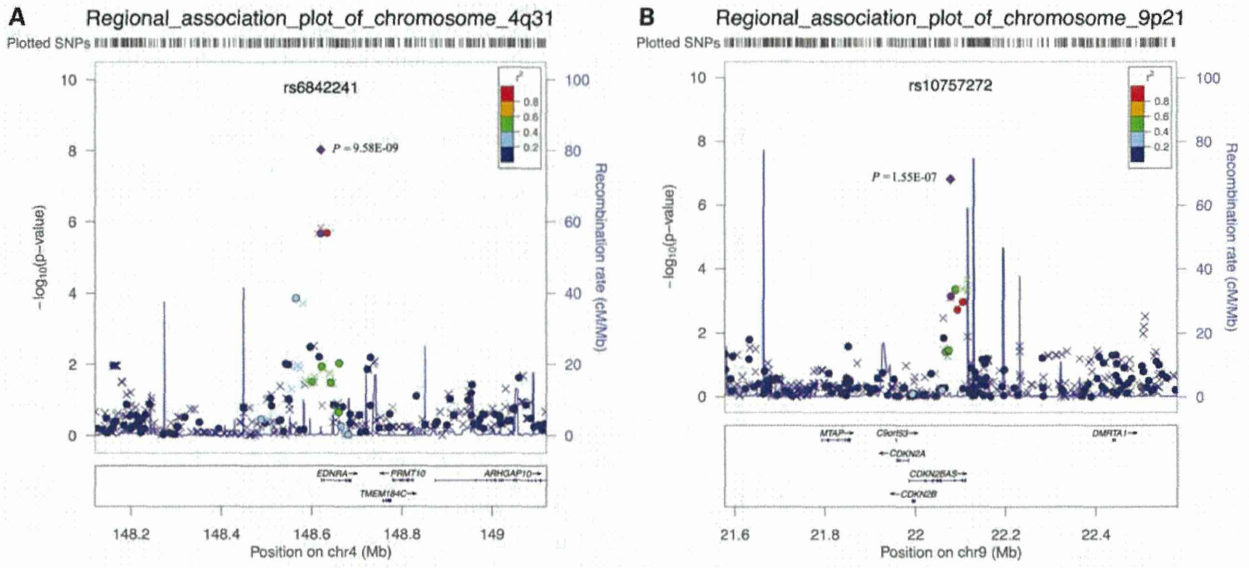


**Figure 2.** Regional association plots of two associated chromosome loci with IA after imputation analysis at 4q31 (**A**) and 9p21 (**B**) regions. The association of marker SNP is marked by a purple diamond before and after additional replication samples. SNPs from the GWAS are plotted as circles; imputed SNPs are plotted as crosses. The color intensity reflects the extent of LD with the marker SNP: red ($r^2 \geq 0.8$), orange ($0.6 \leq r^2 < 0.8$), green ($0.4 \leq r^2 < 0.6$), light blue ($0.2 \leq r^2 < 0.4$) and dark blue ($r^2 < 0.2$). Purplish blue lines represent local recombination rate. The SNP position is based on NCBI build 37.

hypertensive patients (31). The other three genes that reside at the same locus on chromosome 4q31.22 are *PRMT10* with an unknown function; *TMEM184C*, which is a possible tumor suppressor that may play some role in cell growth (32); and *ARHGAP10*, which is a Rho GTPase-activating protein 10 which stabilizes PAK and stimulates cell death (33).

With strong biological evidence for the role of *EDNRA* in IA pathogenesis, we carried out functional analysis of the identified SNPs from the GWAS and imputation analyses. It is well known that SNPs located in the transcriptional regulatory regions such as promoter and enhancer regions could affect the expression levels of the gene product through alteration of the binding affinity to a specific transcription factor(s) (34). The results of the EMSA and reporter assays indicated that the susceptible allele G of rs6841581 has higher affinity to the specific transcription factor(s) that might repress the transcriptional activity of *EDNRA*, compared with the non-susceptible allele A. We observed a different binding affinity to a nuclear protein(s) between the two alleles of rs1878406, and the reporter assay revealed no effect on the transcriptional activity between them. We suspect it might reflect that the 31 bp oligonucleotides may not be long enough to cover the enhancer- or repressor-binding regions. The other possibility is that a nuclear protein bound to this region has no stronger effect on the transcriptional regulation.

For SNPs on chromosome 9p21.3, we successfully validated the association of this locus with IA as reported previously. We identified a stronger association of an SNP, rs10757272 (combined $P = 1.55 \times 10^{-7}$), with IA than the SNP, rs1333040 (combined $P = 5.56 \times 10^{-5}$), that was previously identified in the study of the European population. Interestingly, the SNP rs10757272 was also shown to be associated with coronary artery disease (CAD) and platelet reactivity (a potential mechanism for increased vascular disease) in the European population (35,36). This finding suggests that rs10757272 or SNPs that are in high LD with it might be a common genetic risk factor(s) for multiple cardiovascular disorders. SNPs on a chromosome 9p21.3 region, which consist of $p15^{INK4b}$, $p16^{INK4a}$ and *CDKN2BAS*, have been indicated to be associated with several atherosclerotic vascular diseases such as CADs (37–39), stroke (40), myocardial infarction (38), abdominal aortic aneurysm (41) and IA (20). A recent study revealed that targeted deletion of the 9p21 CAD risk interval in a mouse model resulted in severely attenuated expression of two *CDKN2BAS*s' neighboring tumor-suppressor genes, $p15^{INK4b}$ and $p16^{INK4a}$, which subsequently affects the CAD progression by altering the dynamics of vascular cell proliferation (42). The findings of multiple vascular diseases associated with this locus have provided a new direction for the pathogenesis of these diseases.

Finally, although the association of rs671 (combined $P = 2.63 \times 10^{-6}$) on the *ALDH2* gene was marginally replicated in this study, the association of this locus remained to be of interest in the Japanese population. *ALDH2* belongs to the aldehyde dehydrogenase 2 family (mitochondrial), which is the second enzyme of the major oxidative pathway of alcohol metabolism. rs671 in this gene is the functional variant (Glu504Lys), and an A allele results in the inactivation of ALDH2, inducing 'alcohol flush'. The allelic frequency of this variant is uniquely high in the Asian population. In this study, the variant rs671A allele seemed to act as a protective allele, suggesting that individuals with the A allele might drink less amount of alcohol, which would result in the reduction of IA risk. Since alcohol drinking has been consistently indicated as one of the risk factors for IA, further validation of the association on this locus would be of medical importance.

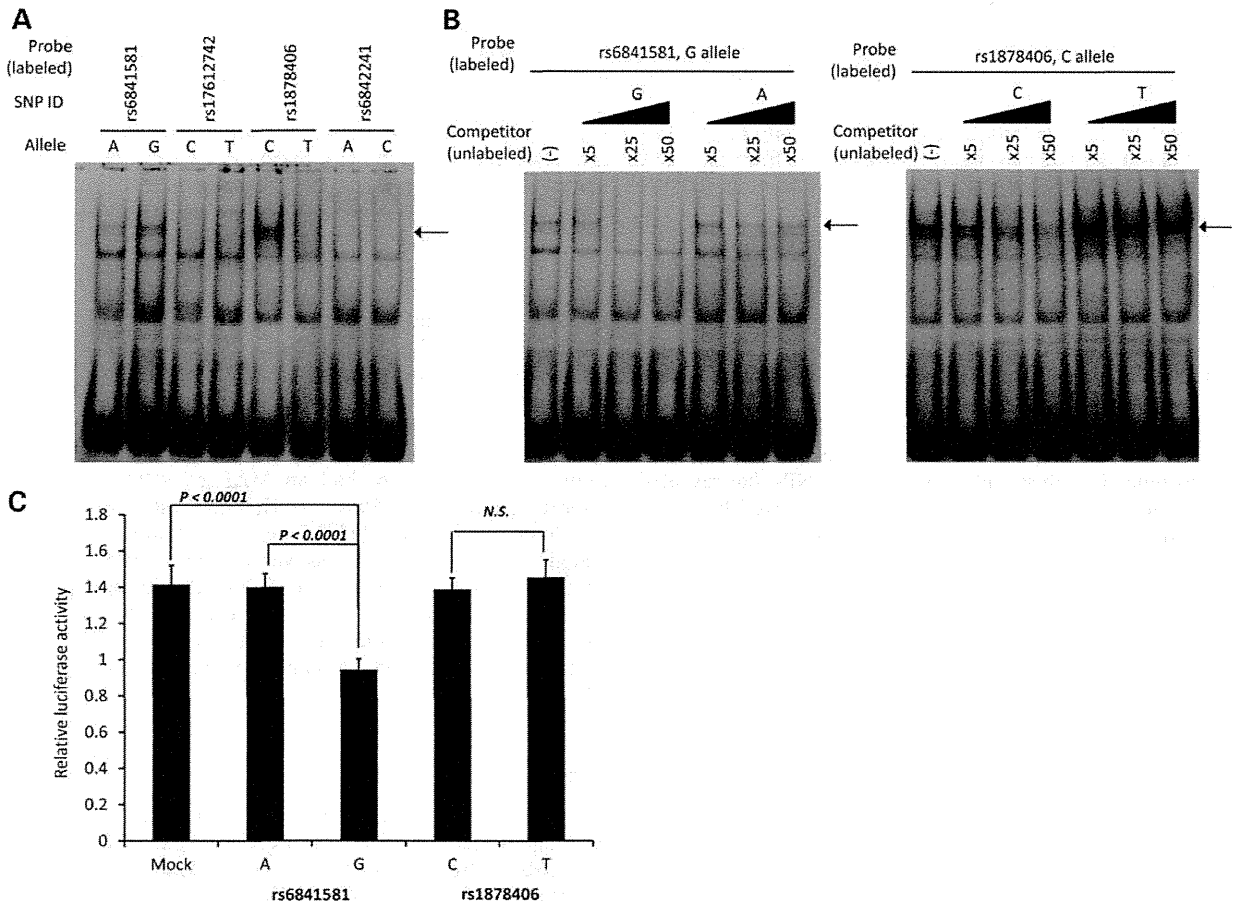


**Figure 3.** Functional analysis of SNPs at the *EDNRA* locus. (**A**) EMSA screening for variants associated with the binding of a nuclear protein(s) from nuclear extracts of HEK293 cells. The susceptible alleles G and C for rs6841581 and rs1878406, respectively, formed DNA–protein complexes, as pointed with arrows. (**B**) The specific interaction of labeled oligonucleotides corresponding to alleles G and C was completed in a dose-dependent manner using an unlabeled oligonucleotide with the G and C alleles, but not oligonucleotides corresponding to the A and T alleles of rs6841581 and rs1878406, respectively. (**C**) Differences in the transcriptional activity between the susceptible and non-susceptible alleles of rs6841581 and rs1878406 measured by dual-luciferase reporter assay. The values of the relative luciferase activity are shown with standard deviation after normalization with internal control *renilla* luciferase activity. The relative luciferase activity of the susceptible genotype G of rs6841581 was significantly lower than that of the non-susceptible allele A and mock (empty) vector ($P < 0.0001$, Student's *t*-test).

In conclusion, by a large-scale GWAS, we successfully identified two genomic loci, *EDNRA* (4q31.22) and *CDKN2BAS* (9p21.3), to be significantly associated with IA in the Japanese population. Owing to the complexity of detecting IA in the general population, the associations of genetic variants have not just served as a promising prediction tool to identify individuals who have a higher risk of IA, but have also provided a better understanding of the disease pathogenesis which subsequently leads to the development of clinical intervention for IA.

## MATERIALS AND METHODS

### Study population

All DNA samples for this study were recruited from the Biobank Japan Project that began in 2003 (http://biobankjp.org). The Biobank Japan Project has a collaborative network of 66 hospitals throughout Japan to collect DNAs and serum samples from nearly 300 000 cases with any of the 47 diseases we assigned. One of the major aims of this project was to identify common genetic variants that confer risk to common diseases, including metabolic diseases and cancers. As a discovery step of this study, we selected 1383 patients registered as aneurysmal SAH (aSAH) cases, which is caused by the rupture of IA. We selected aSAH patients for the discovery screening because aSAH is a severe form of IA in which the disease-associated variants are likely to be enriched. We verified the association results of the first set by utilizing an independent set of samples consisting of 1048 IA patients. Identification of IA in the case samples was done using computerized tomography angiogram, magnetic resonance angiogram or cerebral digital subtraction angiogram. We included 5484 and 7212 controls consisting of healthy volunteers from Midosuji Rotary Club, Osaka, Japan, Health Science Research Resource Bank and individuals in the Biobank

Japan who were registered not to have SAH/IA in the discovery and replication stages, respectively. The detailed demographic and clinical parameters of cases and controls are summarized in Supplementary Material, Table S1. Participants of this study provided written inform consent and this project was approved by the ethical committee from the Institute of Medical Sciences, the University of Tokyo and RIKEN Center of Genomic Medicine.

### Genotyping and quality control

For the GWAS discovery stage, we genotyped both case and control samples using Illumina OmniExpress BeadChip that contained a total of 733 202 SNPs. We performed standard SNP quality control by excluding SNPs with a call rate of $<0.99$, those deviated from the Hardy–Weinberg equilibrium ($P \leq 1.0 \times 10^{-6}$), non-polymorphic ones and those on the X chromosome. The cluster plots of top 100 SNPs that revealed the strongest association with aSAH were checked by visual observation to exclude SNPs with ambiguous patterns. The number of excluded SNPs in each quality control process is summarized in Supplementary Material, Table S2. We utilized the identity-by-state method to evaluate cryptic relatedness for each sample; samples that possess an average estimate value of $\geq 1.7$ were eliminated from subsequent analyses. Additionally, we examined population stratification by principal component analysis (PCA) using the EIGENSTRAT software v2.0 (http://genepath.med.harvard.edu/~reich/Software.htm). We first performed PCA, utilizing four populations in the HapMap database, which included Europeans (represented by Caucasian from UTAH, CEU), Africans (represented by Yoruba from Ibadan, YRI) and East Asians (represented by Japanese from Tokyo, JPT, and Han Chinese from Beijing, CHB) as reference populations for PCA. The top two principal components were utilized to produce a scatter plot for the identification of outliers who did not belong to the Asian cluster. To further investigate the population substructure in the sample population, we performed PCA using the genotype information of the case and control subjects in this study. The Q–Q plot that was generated between observed $P$-values against expected $P$-values and inflation factor ($\lambda$) values were used to evaluate the potential population substructure. Manhattan plot of the study was plotted using Haploview 4.1 (43).

For a replication study, a total of 36 SNPs that showed suggestive association with IA in the Japanese population ($P_{GC} < 1.0 \times 10^{-4}$) as well as 7 SNPs that were previously reported to be associated with IA (19,20) were selected for further evaluation with an independent set of 1048 IA cases and 7212 controls. We genotyped the cases with the multiplex-PCR Invader assay (44) and the control samples with Illumina OmniExpress BeadChip Kits. SNPs with a call rate of $<99\%$ and those that were deviated from the Hardy–Weinberg equilibrium ($P \leq 0.05$) were excluded for further analysis.

### Statistical analysis

The case–control association was evaluated using logistic regression analysis with associated eigenvectors, age and gender as covariates in the discovery (GWAS) and replication phases of this study. The $P$-values obtained from the discovery phase were subsequently corrected using the genomic control method (45), and the corrected $P$-values were used for further analysis. Meta-analysis for the combined analysis of discovery and replication phases was performed using the weighted inverse-variance method implemented in the METAL software (http://www.sph.umich.edu/csg/abecasis/Metal/index.html). $P$-values for the heterogeneity test are evaluated with Cochran's $Q$ statistic and $I^2$ statistic (46).

After the identification of candidate loci to be possibly associated with IA, imputation of the missing genotypes was performed with MACH 1.0 (http://www.sph.umich.edu/csg/abecasis/MACH/index.html). For imputation analysis, we included SNPs that were located $<500$ kb upstream or downstream of the marker SNP except SNPs that had a low genotyping rate ($<99\%$), showed deviations from Hardy–Weinberg equilibrium ($<1.0 \times 10^{-6}$) or had an MAF of $<0.01$. Genotype information from the Phase III HapMap database was used as reference. Using the MACH version 1.0 program, we estimated haplotypes, map crossover and error rates using 50 iterations of the Markov chain Monte Carlo algorithm. For imputation quality control, we excluded SNPs with $r^2$ values of $<0.3$. Regional association plots were generated using Locus Zoom (https://statgen.sph.umich.edu/locuszoom/genform.php?type=yourdata).

### Cell line

A human embryonic kidney cell line, HEK293, was purchased from the American Type Culture Collection (ATCC, Rockville, MD, USA). HEK293 cells were grown in Dulbecco's modified Eagle's medium (Invitrogen, Carlsbad, CA, USA; ATCC) supplemented with 10% bovine serum (GIBCO) and 1% antibiotic/antimycotic solution (Sigma-Aldrich, St Louis, MO, USA). The cells were maintained at 37°C in atmospheres of humidified air with 5% $CO_2$.

### Electrophoretic mobility shift assay

A nuclear fraction of HEK293 cells was extracted using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermo Scientific). Thirty-one base pairs of sense and antisense oligonucleotides corresponding to the genomic sequence which contained the SNPs of interest (rs6841581, rs17612742, rs1878406 and rs6842241) were synthesized (Supplementary Material, Table S6) and labeled with the DIG Gel Shift kit, second generation (Roche), according to the manufacturer's protocol. The reaction was started with pre-incubation of labeled oligonucleotide with poly[d(I-C)] and poly-L-lysine, and 10 μg of nuclear extract from HEK293, for 20 min at 25°C. For a competition assay, 5-fold, 25-fold or 50-fold excess of unlabeled oligonucleotide was added to nuclear extracts before adding the either of DIG-labeled probes. The protein–DNA complex was separated by electrophoresis on a 6% non-denaturing polyacrylamide gel with 0.5× Tris–borate EDTA buffer and transferred onto a nylon membrane. The protein complexes were visualized by autoradiography. All EMSAs were repeated twice to check for reproducibility.

### Dual-luciferase reporter assays

To construct luciferase reporter plasmids containing the SNPs of interest (rs6841581 and rs1878406), *Sma*I and *Bgl*II restriction enzyme sites were added to the 31 bp oligonucleotides that were used as a probe in the EMSA assay and either of the annealed double-stranded oligonucleotides was inserted into the upstream of the luciferase reporter gene in the pGL3 promoter (Promega). The sequences of the constructs were verified using the ABI3730 Genetic Analyzer (Applied Biosystems). After 24 h incubation of HEK293 cells ($2 \times 10^4$) on a 12-well plate, the cells were co-transfected with 400 ng of each reporter construct and 8 ng of the internal control pRL-TK (*renilla* luciferase), using the FuGene 6 transfection reagent (Roche). After 48 h incubation, the cells were lysed in passive lysis buffer and luciferase activities were measured using the Dual-Luciferase Reporter Assay System (Promega). The results were normalized by *renilla* luciferase activity.

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

## ACKNOWLEDGEMENTS

We express our heartfelt gratitude to all the patients who participated in this study. Our thankfulness also goes to the members of The Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for making this study possible. We extend our appreciation to Dr Yukinori Okada for his insightful comment on the statistical analysis and Dr Suyoun Chung for her advice in the functional analysis. We thank Miss Maiko Natsume for her helpful technical support. We would like to express our gratefulness to the staff of Biobank Japan for their outstanding assistance.

*Conflict of Interest statement*. The authors declare no competing financial interests.

## FUNDING

This work was conducted as part of the Biobank Japan Project that was supported by the Ministry of Education, Culture, Sports, Sciences and Technology from the Japanese Government. In addition, this work is performed in collaboration with the Center of Genomic Medicine, Institute of Physical and Chemical Research (CGM, RIKEN).

## REFERENCES

1. Inagawa, T. (2005) Trends in surgical and management outcomes in patients with aneurysmal subarachnoid hemorrhage in Izumo city, Japan, between 1980–1989 and 1990–1998. *Cerebrovasc. Dis.*, **19**, 39–48.
2. Morita, A., Fujiwara, S., Hashi, K., Ohtsu, H. and Kirino, T. (2005) Risk of rupture associated with intact cerebral aneurysms in the Japanese population: a systematic review of the literature from Japan. *J. Neurosurg.*, **102**, 601–606.
3. Sarti, C., Tuomilehto, J., Salomaa, V., Sivenius, J., Kaarsalo, E., Narva, E.V., Salmi, K. and Torppa, J. (1991) Epidemiology of subarachnoid hemorrhage in Finland from 1983 to 1985. *Stroke*, **22**, 848–853.
4. Ingall, T., Asplund, K., Mähönen, M. and Bonita, R. (2000) A multinational comparison of subarachnoid hemorrhage epidemiology in the WHO MONICA stroke study. *Stroke*, **31**, 1054–1061.
5. Ohkuma, H., Fujita, S. and Suzuki, S. (2002) Incidence of aneurysmal subarachnoid hemorrhage in Shimokita, Japan, from 1989 to 1998. *Stroke*, **33**, 195–199.
6. Schievink, W.I., Schaid, D.J., Michels, V.V. and Piepgras, D.G. (1995) Familial aneurysmal subarachnoid hemorrhage: a community-based study. *J. Neurosurg.*, **83**, 426–429.
7. Teunissen, L.L., Rinkel, G.J., Algra, A. and van Gijn, J. (1996) Risk factors for subarachnoid hemorrhage: a systematic review. *Stroke*, **27**, 544–549.
8. Connolly, E.S. Jr., Choudhri, T.F., Mack, W.J., Mocco, J., Spinks, T.J., Slosberg, J., Lin, T., Huang, J. and Solomon, R.A. (2001) Influence of smoking, hypertension, and sex on the phenotypic expression of familial intracranial aneurysms in siblings. *Neurosurgery*, **48**, 64–68; Discussion, 68–69.
9. Ronkainen, A., Miettinen, H., Karkola, K., Papinaho, S., Vanninen, R., Puranen, M. and Hernesniemi, J. (1998) Risk of harboring an unruptured intracranial aneurysm. *Stroke*, **29**, 359–362.
10. Inagawa, T. (2010) Risk factors for the formation and rupture of intracranial saccular aneurysms in Shimane, Japan. *World Neurosurg.*, **73**, 155–164; Discussion, e123.
11. Chapman, A.B., Rubinstein, D., Hughes, R., Stears, J.C., Earnest, M.P., Johnson, A.M., Gabow, P.A. and Kaehny, W.D. (1992) Intracranial aneurysms in autosomal dominant polycystic kidney disease. *N. Engl. J. Med.*, **327**, 916–920.
12. Schievink, W.I. (1997) Genetics of intracranial aneurysms. *Neurosurgery*, **40**, 651–662; Discussion, 662–653.
13. Ronkainen, A., Hernesniemi, J. and Ryynänen, M. (1993) Familial subarachnoid hemorrhage in east Finland, 1977–1990. *Neurosurgery*, **33**, 787–796; Discussion, 796–797.
14. Wang, P.S., Longstreth, W.T. and Koepsell, T.D. (1995) Subarachnoid hemorrhage and family history. A population-based case–control study. *Arch. Neurol.*, **52**, 202–204.
15. Nahed, B.V., Seker, A., Guclu, B., Ozturk, A.K., Finberg, K., Hawkins, A.A., DiLuna, M.L., State, M., Lifton, R.P. and Gunel, M. (2005) Mapping a Mendelian form of intracranial aneurysm to 1p34.3-p36.13. *Am. J. Hum. Genet.*, **76**, 172–179.
16. Onda, H., Kasuya, H., Yoneyama, T., Takakura, K., Hori, T., Takeda, J., Nakajima, T. and Inoue, I. (2001) Genomewide-linkage and haplotype-association studies map intracranial aneurysm to chromosome 7q11. *Am. J. Hum. Genet.*, **69**, 804–819.
17. Olson, J.M., Vongpunsawad, S., Kuivaniemi, H., Ronkainen, A., Hernesniemi, J., Ryynänen, M., Kim, L.L. and Tromp, G. (2002) Search for intracranial aneurysm susceptibility gene(s) using Finnish families. *BMC Med. Genet.*, **3**, 7.
18. Yamada, S., Utsunomiya, M., Inoue, K., Nozaki, K., Inoue, S., Takenaka, K., Hashimoto, N. and Koizumi, A. (2004) Genome-wide scan for Japanese familial intracranial aneurysms: linkage to several chromosomal regions. *Circulation*, **110**, 3727–3733.
19. Bilguvar, K., Yasuno, K., Niemela, M., Ruigrok, Y.M., von Und Zu Fraunberg, M., van Duijn, C.M., van den Berg, L.H., Mane, S., Mason, C.E., Choi, M. *et al.* (2008) Susceptibility loci for intracranial aneurysm in European and Japanese populations. *Nat. Genet.*, **40**, 1472–1477.
20. Yasuno, K., Bilguvar, K., Bijlenga, P., Low, S.K., Krischek, B., Auburger, G., Simon, M., Krex, D., Arlier, Z., Nayak, N. *et al.* (2010) Genome-wide association study of intracranial aneurysm identifies three new risk loci. *Nat. Genet.*, **42**, 420–425.
21. Akiyama, K., Narita, A., Nakaoka, H., Cui, T., Takahashi, T., Yasuno, K., Tajima, A., Krischek, B., Yamamoto, K., Kasuya, H. *et al.* (2010) Genome-wide association study to identify genetic variants present in Japanese patients harboring intracranial aneurysms. *J. Hum. Genet.*, **55**, 656–661.
22. Yu, J.C., Pickard, J.D. and Davenport, A.P. (1995) Endothelin ETA receptor expression in human cerebrovascular smooth muscle cells. *Br. J. Pharmacol.*, **116**, 2441–2446.
23. Janakidevi, K., Fisher, M.A., Del Vecchio, P.J., Tiruppathi, C., Figge, J. and Malik, A.B. (1992) Endothelin-1 stimulates DNA synthesis and proliferation of pulmonary artery smooth muscle cells. *Am. J. Physiol.*, **263**, C1295–C1301.
24. Fassbender, K., Hodapp, B., Rossol, S., Bertsch, T., Schmeck, J., Schütt, S., Fritzinger, M., Horn, P., Vajkoczy, P., Wendel-Wellner, M. *et al.*

(2000) Endothelin-1 in subarachnoid hemorrhage: an acute-phase reactant produced by cerebrospinal fluid leukocytes. *Stroke*, **31**, 2971–2975.
25. Juvela, S. (2000) Plasma endothelin concentrations after aneurysmal subarachnoid hemorrhage. *J. Neurosurg.*, **92**, 390–400.
26. Tzourio, C., El Amrani, M., Poirier, O., Nicaud, V., Bousser, M.G. and Alpérovitch, A. (2001) Association between migraine and endothelin type A receptor (ETA -231 A/G) gene polymorphism. *Neurology*, **56**, 1273–1277.
27. Nicaud, V., Poirier, O., Behague, I., Herrmann, S.M., Mallet, C., Troesch, A., Bouyer, J., Evans, A., Luc, G., Ruidavets, J.B. *et al.* (1999) Polymorphisms of the endothelin-A and -B receptor genes in relation to blood pressure and myocardial infarction: the Etude Cas-Témoins sur l'Infarctus du Myocarde (ECTIM) Study. *Am. J. Hypertens.*, **12**, 304–310.
28. Oguri, M., Kato, K., Yokoi, K., Itoh, T., Yoshida, T., Watanabe, S., Metoki, N., Yoshida, H., Satoh, K., Aoyagi, Y. *et al.* (2009) Association of genetic variants with myocardial infarction in Japanese individuals with metabolic syndrome. *Atherosclerosis*, **206**, 486–493.
29. Darrah, R., McKone, E., O'Connor, C., Rodgers, C., Genatossio, A., McNamara, S., Gibson, R., Stuart Elborn, J., Ennis, M., Gallagher, C.G. *et al.* (2010) EDNRA variants associate with smooth muscle mRNA levels, cell proliferation rates, and cystic fibrosis pulmonary disease severity. *Physiol. Genomics*, **41**, 71–77.
30. Benjafield, A.V., Katyk, K. and Morris, B.J. (2003) Association of EDNRA, but not WNK4 or FKBP1B, polymorphisms with essential hypertension. *Clin. Genet.*, **64**, 433–438.
31. Hasegawa, K., Fujiwara, H., Doyama, K., Inada, T., Ohtani, S., Fujiwara, T., Hosoda, K., Nakao, K. and Sasayama, S. (1994) Endothelin-1-selective receptor in the arterial intima of patients with hypertension. *Hypertension*, **23**, 288–293.
32. Akaishi, J., Onda, M., Okamoto, J., Miyamoto, S., Nagahama, M., Ito, K., Yoshida, A. and Shimizu, K. (2007) Down-regulation of an inhibitor of cell growth, transmembrane protein 34 (TMEM34), in anaplastic thyroid cancer. *J. Cancer Res. Clin. Oncol.*, **133**, 213–218.
33. Koeppel, M.A., McCarthy, C.C., Moertl, E. and Jakobi, R. (2004) Identification and characterization of PS-GAP as a novel regulator of caspase-activated PAK-2. *J. Biol. Chem.*, **279**, 53653–53664.
34. Chorley, B.N., Wang, X., Campbell, M.R., Pittman, G.S., Noureddine, M.A. and Bell, D.A. (2008) Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutat. Res.*, **659**, 147–157.
35. Erdmann, J., Grosshennig, A., Braund, P.S., König, I.R., Hengstenberg, C., Hall, A.S., Linsel-Nitschke, P., Kathiresan, S., Wright, B., Trégouët, D.A. *et al.* (2009) New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.*, **41**, 280–282.
36. Musunuru, K., Post, W.S., Herzog, W., Shen, H., O'Connell, J.R., McArdle, P.F., Ryan, K.A., Gibson, Q., Cheng, Y.C., Clearfield, E. *et al.* (2010) Association of single nucleotide polymorphisms on chromosome 9p21.3 with platelet reactivity: a potential mechanism for increased vascular disease. *Circ. Cardiovasc. Genet.*, **3**, 445–453.
37. W.T.C.C. Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.
38. Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A., Masson, G. *et al.* (2007) A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*, **316**, 1491–1493.
39. McPherson, R. (2010) Chromosome 9p21 and coronary artery disease. *N. Engl. J. Med.*, **362**, 1736–1737.
40. Anderson, C.D., Biffi, A., Rost, N.S., Cortellini, L., Furie, K.L. and Rosand, J. (2010) Chromosome 9p21 in ischemic stroke: population structure and meta-analysis. *Stroke*, **41**, 1123–1131.
41. Helgadottir, A., Thorleifsson, G., Magnusson, K.P., Gretarsdottir, S., Steinthorsdottir, V., Manolescu, A., Jones, G.T., Rinkel, G.J., Blankensteijn, J.D., Ronkainen, A. *et al.* (2008) The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat. Genet.*, **40**, 217–224.
42. Visel, A., Zhu, Y., May, D., Afzal, V., Gong, E., Attanasio, C., Blow, M.J., Cohen, J.C., Rubin, E.M. and Pennacchio, L.A. (2010) Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature*, **464**, 409–412.
43. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
44. Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. and Nakamura, Y. (2001) A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.*, **46**, 471–477.
45. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.
46. Higgins, J.P., Thompson, S.G., Deeks, J.J. and Altman, D.G. (2003) Measuring inconsistency in meta-analyses. *BMJ*, **327**, 557–560.

*Human Molecular Genetics, 2012, Vol. 21, No. 7* **1665–1672**
*doi:10.1093/hmg/ddr597*
*Advance Access published on December 16, 2011*

# A genome-wide association study identifies locus at 10q22 associated with clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients in Japanese

**Kazuma Kiyotani[1], Taisei Mushiroda[1], Tatsuhiko Tsunoda[2], Takashi Morizono[2], Naoya Hosono[3], Michiaki Kubo[3], Yusuke Tanigawara[4], Chiyo K. Imamura[4], David A. Flockhart[5], Fuminori Aki[6], Koichi Hirata[7], Yuichi Takatsuka[8], Minoru Okazaki[9], Shozo Ohsumi[10], Takashi Yamakawa[11], Mitsunori Sasa[12], Yusuke Nakamura[1,13,*] and Hitoshi Zembutsu[13]**

[1]Laboratory for Pharmacogenetics, [2]Laboratory for Medical Informatics and [3]Laboratory for Genotyping Development, RIKEN Center for Genomic Medicine, Yokohama 230-0045, Japan, [4]Department of Clinical Pharmacokinetics and Pharmacodynamics, School of Medicine, Keio University, Tokyo 160-8582, Japan, [5]Division of Clinical Pharmacology, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA, [6]Department of Surgery, Itoh Surgery and Breast Clinic, Kochi 780-0085, Japan, [7]1st Department of Surgery, Sapporo Medical University, Sapporo 060-8543, Japan, [8]Department of Breast Surgery, Kansai Rosai Hospital, Hyogo 060-0006, Japan, [9]Department of Surgery, Sapporo Breast Surgical Clinic, Sapporo 060-0006, Japan, [10]Department of Breast Oncology, Shikoku Cancer Center, Ehime 791-0280, Japan, [11]Department of Surgery, Yamakawa Breast Clinic, Kochi 780-8081, Japan, [12]Department of Surgery, Tokushima Breast Care Clinic, Tokushima 770-0052, Japan and [13]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan

Received September 24, 2011; Revised November 26, 2011; Accepted December 12, 2011

**Although many association studies of polymorphisms in candidate genes with the clinical outcomes of breast cancer patients receiving adjuvant tamoxifen therapy have been reported, genetic factors determining individual response to tamoxifen are not fully understood. To identify genetic polymorphisms associated with clinical outcomes of patients with tamoxifen treatment, we conducted a genome-wide association study (GWAS). We studied 462 Japanese patients with hormone receptor-positive, invasive breast cancer receiving adjuvant tamoxifen therapy. Of them, 240 patients were analyzed by genome-wide genotyping using the Illumina Human610-Quad BeadChips, and two independent sets of 105 and 117 cases were used for replication studies. In the GWAS, we detected significant associations with recurrence-free survival at 15 single-nucleotide polymorphisms (SNPs) on nine chromosomal loci (1p31, 1q41, 5q33, 7p11, 10q22, 12q13, 13q22, 18q12 and 19p13) that satisfied a genome-wide significant threshold (log-rank $P = 2.87 \times 10^{-9}$–$9.41 \times 10^{-8}$). Among them, rs10509373 in *C10orf11* gene on 10q22 was significantly associated with recurrence-free survival in the replication study (log-rank $P = 2.02 \times 10^{-4}$) and a combined analysis indicated a strong association of this SNP with recurrence-free survival in breast cancer patients treated with tamoxifen (log-rank $P = 1.26 \times 10^{-10}$). Hazard ratio per C allele of rs10509373 was 4.51 [95% confidence interval (CI), 2.72–7.51; $P = 6.29 \times 10^{-9}$]. In a combined analysis of rs10509373 genotype with previously identified genetic makers, *CYP2D6* and *ABCC2*, the number of risk alleles of these three genes had cumulative effects on recurrence-free survival among 345 patients receiving tamoxifen monotherapy (log-rank $P = 2.28 \times 10^{-12}$).**

*To whom correspondence should be addressed at: 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 354495372; Fax: +81 354495433; Email: yusuke@ims.u-tokyo.ac.jp

**In conclusion, we identified a novel locus associated with recurrence-free survival in Japanese breast cancer patients receiving adjuvant tamoxifen therapy.**

## INTRODUCTION

Tamoxifen has been the gold standard for endocrine treatment of patients with estrogen receptor (ER)-positive breast cancers. However, 30–50% of patients with adjuvant tamoxifen therapy experience relapse and subsequently die of the disease (1,2), indicating individual differences in responsiveness to tamoxifen.

Tamoxifen is metabolized to the highly active metabolites, 4-hydroxytamoxifen and 4-hydroxy-*N*-desmethyltamoxifen (endoxifen). It is reported that these metabolites are the active therapeutic moieties, having 100-fold greater affinity to ER and 30–100-fold greater potency in suppressing estrogen-dependent cell proliferation than those of tamoxifen (3–5). Inter-individual differences in the formation and elimination of these active metabolites could be one of the important factors affecting variability in the response to tamoxifen. Most previous reports focused on the genes involved in the pharmacokinetics of tamoxifen and its metabolites seek genetic variations which determine the individual response to tamoxifen. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6), which is the key enzyme responsible for the generation of endoxifen, is thought to be the most promising predictor of plasma concentration of endoxifen and clinical efficacy of tamoxifen in breast cancer patients (6–14). Schroth *et al.* (15) recently reported outstanding results in 1325 breast cancer patients, providing sufficiently powered evidence for an association between *CYP2D6* genotype and clinical outcomes in patients treated with tamoxifen in the adjuvant setting. Besides *CYP2D6*, several genes, such as *CYP2C19*, *CYP3A5*, sulfotransferase 1A1 (*SULT1A1*), UDP-glucuronosyltransferase 2B15 (*UGT2B15*) and ATP-binding cassette sub-family C member 2 (*ABCC2*), were reported as possible candidates associated with the clinical outcomes of tamoxifen therapy (7,10,14,16); however, associations of these candidate genes have not yet been sufficiently validated. Therefore, individual differences in responsiveness to tamoxifen still remain, even if the effects of genetic polymorphisms of *CYP2D6* were considered, suggesting the existence of other genetic factor(s).

In this study, to identify responsible loci for the clinical outcomes of tamoxifen therapy, we performed a genome-wide association study (GWAS) by genotyping over 610 000 single-nucleotide polymorphisms (SNPs) and identified the novel locus containing chromosome 10 open-reading frame 11 (*C10orf11*) gene associated with recurrence-free survival in the breast cancer patients treated with tamoxifen.

## RESULTS

### Patient characteristics

We recruited 462 Japanese patients with breast cancer receiving adjuvant tamoxifen therapy. Table 1 summarizes the characteristics of all of these patients who were pathologically diagnosed to have a hormone receptor-positive, invasive breast cancer. Their median age at the time of surgery was 51 years old (range, 27–84 years), the median follow-up period was 6.8 years (range, 0.6–23.5 years) and the median tamoxifen treatment period was 4.8 years (range, 0.6–6.3 years). Among the characteristics listed in Table 1, tumor size and nodal status showed significant associations with the recurrence-free survival [$P = 0.000215$; hazard ratio (HR), 1.71; 95% confidence interval (CI), 1.29–2.27; and $P = 0.0138$; HR, 1.83; 95% CI, 1.14–3.09, respectively] in the Cox proportional hazards analysis, whereas the other factors were not associated with recurrence-free survival (Supplementary Material, Table S1).

### Genome-wide association and replication studies

We conducted a GWAS of recurrence-free survival of 240 Japanese patients with breast cancer who received tamoxifen monotherapy using Illumina Human610-Quad BeadChips. After the standard quality control, association analysis was performed for 470 796 SNPs by the trend log-rank test. We generated a quantile–quantile plot (Supplementary Material, Fig. S1) and obtained the genomic control inflation factor ($\lambda_{GC}$) of 1.023, indicating a low possibility of false-positive associations resulting from population stratification. We detected significant associations with recurrence-free survival at 15 SNPs in nine genetic regions (1p31, 1q41, 5q33, 7p11, 10q22, 12q13, 13q22, 18q12 and 19p13) that satisfied a genome-wide significant threshold of $P < 1.06 \times 10^{-7}$ (Fig. 1). To further validate the results of GWAS, we carried out a replication study using an independent 105 breast cancer patients. We genotyped 9 of the 15 SNPs because 6 of them were highly linked to another SNP ($r^2 > 0.80$; Supplementary Material, Table S2). We found that rs10509373 in *C10orf11* gene on 10q22 was significantly associated with recurrence-free survival in the replication stage (log-rank $P = 4.18 \times 10^{-4}$; Table 2 and Fig. 2). The associations of the other SNPs were not replicated (Supplementary Material, Table S2). Furthermore, a combined result of the GWAS and first replication study strongly suggested an association of this locus with recurrence-free survival in breast cancer patients treated with tamoxifen (log-rank $P = 2.19 \times 10^{-10}$). We also genotyped rs10509373 using additional 117 samples, which include the patients receiving tamoxifen after chemotherapy and observed a significant association (log-rank $P = 1.86 \times 10^{-2}$). A combined *P*-value of all samples was $2.19 \times 10^{-10}$, suggesting the significant association with recurrence-free survival in breast cancer patients treated with tamoxifen (Fig. 2 and Supplementary Material, Table S3). In Cox proportional hazards analysis, *C10orf11* genotype (rs10509373) was an independent indicator of the recurrence-free survival after adjustment for tumor size and nodal status ($P = 6.28 \times 10^{-8}$; Table 2). The adjusted HRs of rs10509373

**Table 1.** Characteristics of patients

| Characteristic | No. of patients (%) GWAS | First replication | Second replication | Total |
|---|---|---|---|---|
| No. | 240 | 105 | 117 | 462 |
| Age at surgery (years) | | | | |
| Median | 51 | 50 | 48 | 51 |
| Range | 31–83 | 35–84 | 27–71 | 27–84 |
| Follow-up (years) | | | | |
| Median | 7.2 | 5.2 | 6.2 | 6.8 |
| Range | 1.1–23.5 | 0.6–19.3 | 1.0–15.5 | 0.6–23.5 |
| Tamoxifen treatment (years) | | | | |
| Median | 4.9 | 4.0 | 4.7 | 4.8 |
| Range | 1.0–6.1 | 0.6–6.0 | 0.7–6.3 | 0.6–6.3 |
| Menopausal status | | | | |
| Pre-menopause | 101 (42.1) | 40 (38.1) | 75 (64.1) | 216 (46.8) |
| Post-menopause | 131 (54.6) | 40 (38.1) | 35 (29.9) | 206 (44.6) |
| Unknown | 8 (3.3) | 25 (23.8) | 7 (6.0) | 40 (8.7) |
| Tumor size (cm) | | | | |
| ≤2 | 138 (57.5) | 57 (54.3) | 48 (41.0) | 243 (52.6) |
| 2.1–5 | 92 (38.3) | 34 (32.4) | 56 (47.9) | 182 (39.4) |
| >5 | 1 (0.4) | 2 (1.9) | 12 (10.3) | 15 (3.2) |
| Unknown | 9 (3.8) | 12 (11.4) | 1 (0.9) | 22 (4.8) |
| Nodal status | | | | |
| Negative | 193 (80.4) | 88 (83.8) | 74 (63.2) | 355 (76.8) |
| Positive | 44 (18.3) | 13 (12.4) | 41 (35.0) | 98 (21.2) |
| Unknown | 3 (1.3) | 4 (3.8) | 2 (1.7) | 9 (1.9) |
| ER status | | | | |
| Positive | 173 (72.1) | 87 (82.9) | 98 (83.8) | 358 (77.5) |
| Negative | 24 (10.0) | 2 (1.9) | 12 (10.3) | 38 (8.2) |
| Unknown | 43 (17.9) | 16 (15.2) | 7 (6.0) | 66 (14.3) |
| PR status | | | | |
| Positive | 167 (69.6) | 77 (73.3) | 87 (74.4) | 331 (71.6) |
| Negative | 28 (11.7) | 11 (10.5) | 22 (18.8) | 61 (13.2) |
| Unknown | 45 (18.8) | 17 (16.2) | 8 (6.8) | 70 (15.2) |
| Her-2 | | | | |
| Positive[a] | 3 (1.3) | 5 (4.8) | 6 (5.1) | 14 (3.0) |
| Negative | 82 (34.2) | 28 (26.7) | 60 (51.3) | 170 (36.8) |
| Unknown | 155 (64.6) | 72 (68.6) | 51 (43.6) | 278 (60.2) |
| Treatment | | | | |
| Tamoxifen alone | 240 (100.0) | 105 (100.0) | 0 (0.0) | 345 (76.7) |
| Tamoxifen + AC or EC | 0 (0.0) | 0 (0.0) | 41 (35.0) | 41 (8.9) |
| Tamoxifen + CMF | 0 (0.0) | 0 (0.0) | 32 (27.4) | 32 (6.9) |
| Tamoxifen + other chemotherapies | 0 (0.0) | 0 (0.0) | 44 (37.6) | 44 (9.5) |
| Events | | | | |
| No event | 210 (87.5) | 89 (84.8) | 98 (55.4) | 397 (85.9) |
| Locoregional events | 9 (3.8) | 0 (0.0) | 0 (0.0) | 9 (1.9) |
| Distant metastasis events | 12 (5.0) | 15 (14.3) | 17 (9.6) | 44 (9.5) |
| Contralateral breast events | 9 (3.8) | 1 (1.0) | 2 (1.1) | 12 (2.6) |

AC, adriamycin + cyclophosphamide; EC, epirubicin + cyclophosphamide; CMF, cyclophosphamide + methotrexate + 5-fluorouracil.
[a]Score of 3+ in immunohistochemistry.

C allele was 4.51 (95% CI, 2.72–7.51), suggesting that C allele was a risk allele for breast cancer recurrence.

To further identify SNPs associated with recurrence-free survival in patients receiving tamoxifen therapy, we genotyped 130 tag SNPs for fine mapping on chromosome 10q22 (Chr. 10: 77.35–78.70 Mb) where the most significant association with the recurrence-free survival was observed (Fig. 3 and Supplementary Material, Table S4). Although no SNPs showed a stronger association than the landmark SNP, rs10509373, fine mapping of this region indicated that a 172-kb linkage disequilibrium (LD) block (77.67–77.84 Mb) including *C10orf11* was likely to contain the genetic variant(s) associated with recurrence-free survival in patients receiving tamoxifen therapy.

### Combination analysis with previously identified gene loci

As we previously identified significant associations of *CYP2D6* and *ABCC2* rs3740065 genotypes with recurrence-free survival in patients treated with tamoxifen among the patients receiving tamoxifen monotherapy (Supplementary Material, Table S5) (7), we investigated a combined effect of *C10orf11* genotype in addition to *CYP2D6* and *ABCC2* genotypes on the recurrence-free survival by classifying the 345

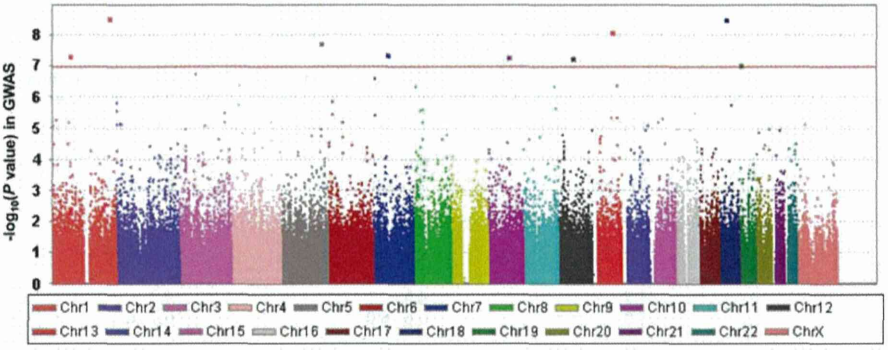


**Figure 1.** Results of the GWAS. Manhattan plot showing the $-\log_{10}$-transformed *P*-value of SNPs in the GWAS for 240 Japanese patients with breast cancer receiving tamoxifen monotherapy. The red line indicates the genome-wide significance level ($P = 1.06 \times 10^{-7}$).

**Table 2.** Association analysis of rs10509373 in C10orf11 with recurrence-free survival in breast cancer patients receiving tamoxifen therapy

| SNP | Chr | Chr location[a] | Allele (risk) | Study set | Risk allele frequency | | Log-rank *P* | Univariate | | Multivariate[b] | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Event | No event | | HR (95% CI)[c] | *P*-value | HR (95% CI)[c] | *P* value |
| rs10509373 | 10 | 77827578 | T/C (C) | GWAS | 0.117 | 0.021 | $5.19 \times 10^{-8}$ | 7.70 (3.25–18.22) | $3.41 \times 10^{-6}$ | 9.64 (3.85–24.12) | $1.29 \times 10^{-6}$ |
| | | | | First replication | 0.094 | 0.017 | $4.18 \times 10^{-4}$ | 7.93 (2.06–30.58) | $2.65 \times 10^{-3}$ | 5.96 (1.49–23.86) | $1.17 \times 10^{-2}$ |
| | | | | GWAS + first replication | 0.109 | 0.020 | $2.19 \times 10^{-10}$ | 7.34 (3.58–14.98) | $4.90 \times 10^{-8}$ | 9.18 (1.39–22.62) | $6.07 \times 10^{-9}$ |
| | | | | Second replication | 0.132 | 0.036 | $1.86 \times 10^{-2}$ | 2.72 (1.13–6.53) | $2.53 \times 10^{-2}$ | 2.92 (1.14–7.49) | $2.55 \times 10^{-2}$ |
| | | | | Combined replications | 0.114 | 0.027 | $2.02 \times 10^{-4}$ | 3.21 (1.65–6.22) | $5.67 \times 10^{-4}$ | 3.20 (1.53–6.69) | $1.97 \times 10^{-3}$ |
| | | | | Combined all | 0.115 | 0.024 | $1.26 \times 10^{-10}$ | 4.51 (2.72–7.51) | $6.29 \times 10^{-9}$ | 4.53 (2.62–7.83) | $6.28 \times 10^{-8}$ |

Chr, chromosome; CI, confidence interval; GWAS, genome-wide association study.
[a]Based on NCBI 36 genome assembly.
[b]Adjusted for tumor size and nodal status.
[c]HR per one allele.

patients into 6 groups (0, 1, 2, 3, 4 and 5 risk allele groups) according to the number of risk alleles of the three genes. Kaplan–Meier analysis revealed the number of risk alleles of these three genes to have cumulative effects on recurrence-free survival (log-rank $P = 2.24 \times 10^{-12}$; Fig. 4). In the Cox proportional hazards analysis of 345 patients, the *CYP2D6* and *ABCC2* genotypes showed similar associations with recurrence-free survival to those in previous analysis of 282 patients ($P = 1.99 \times 10^{-4}$ and $8.51 \times 10^{-4}$, respectively; Supplementary Material, Table S6) (7). In the multivariate analysis, rs10509373 in *C10orf11* still showed a significant association even after adjustment of *CYP2D6* and *ABCC2* genotypes in addition to tumor size and nodal status ($P = 4.74 \times 10^{-7}$; Supplementary Material, Table S6), indicating that rs10509373 is an independent risk factor of breast cancer recurrence. Furthermore, combined analysis of *C10orf11*, *CYP2D6* and *ABCC2* revealed that genotypes of the three genes have cumulative effects on recurrence-free survival ($P = 2.28 \times 10^{-12}$), and adjusted HR for risk of recurrence computed for patients carrying three or more risk alleles increased from 6.51-fold (three risk alleles) to 119.51-fold (five risk alleles) compared with those carrying one risk allele (Supplementary Material, Table S6). In the subgroup analysis of menopausal status, we identified the significant associations in both subgroups of pre- and postmenopausal patients, although the stronger association was observed in postmenopausal group than in the premenopausal patients (Supplementary Material, Table S7).

## DISCUSSION

This study represents the first GWAS which attempts to identify genetic variants associated with clinical outcomes of tamoxifen therapy and successfully revealed that a marker SNP, rs10509373, on chromosome 10q22 was significantly associated with recurrence-free survival in 462 Japanese patients with breast cancer receiving tamoxifen monotherapy. Furthermore, combined analysis of this SNP with previously identified predictors, *CYP2D6* and *ABCC2*, revealed that the number of risk alleles of the three genes have cumulative effects on recurrence-free survival in tamoxifen-treated breast cancer patients.

The most significantly associated SNP in this study, rs10509373 (combined $P = 1.26 \times 10^{-10}$), is located in a 172-kb LD block which contains the 3′ region of *C10orf11*

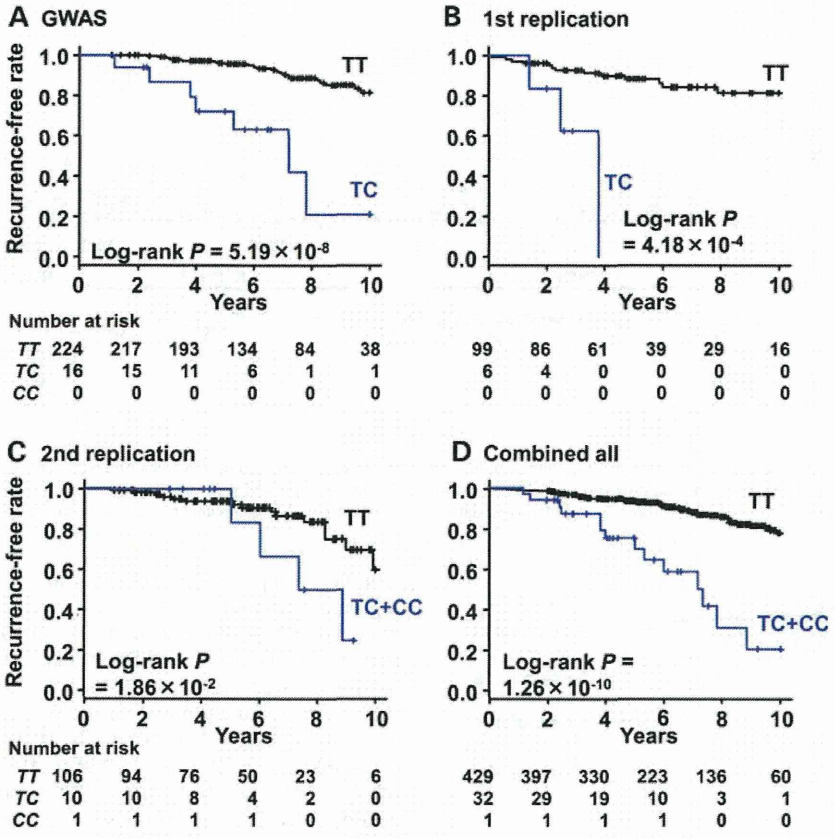


**Figure 2.** Kaplan–Meier estimates of recurrence-free survival for *C10orf11* rs10509373 genotype in 240 patients genotyped in the genome-wide association study (**A**), in 105 patients genotyped in the first replication study (**B**), in 117 patients genotyped in the second replication study (**C**) and in 462 patients in the combination analysis (**D**).

gene. The fine mapping of this region indicated that a peak association was located in intron 5 of *C10orf11* gene, suggesting that *C10orf11* is likely to be a causative gene to determine the clinical outcomes of breast cancer patients treated with tamoxifen. Because no associated SNPs were found in exon region by re-sequencing of *C10orf11* (Supplementary Material, Table S8), a genetic variant(s) within this LD block might alter C10orf11 transcriptional activity. C10orf11 protein, comprising 198 amino acids, is predicted to contain leucine-rich repeat domain and to have the capacity of protein binding in the SMART database (http://smart.embl-heidelberg.de/), although no report has clarified its function. It is reported that *C10orf11* region overlaps with ultraconserved elements (UCEs), perfectly constrained elements between the human, mouse and rat genomes (17–19). Their functional roles have not been completely elucidated yet; however, UCEs are thought to possess some essential functional properties. It is reported that paired box 2, encoded by *PAX2* gene on 10q24, which regulates *ERBB2* transcription and is involved in acquiring tamoxifen resistance (20), and special AT-rich sequence-binding protein-1 encoded by *SATB1* gene on 3p23, which delineates epigenetic modification and is associated with breast tumor growth and metastasis (21), are located in the UCE-rich regions (17). We further examined the association of pharmacokinetic data with *C10orf11* genotype; however, no significant difference was observed between *C10orf11* genotype and plasma levels of endoxifen and 4-hydroxytamoxifen in 98 breast cancer patients taking 20 mg/day tamoxifen (Supplementary Material, Fig. S2). According to our gene expression database (in-house), the C10orf11 transcript is expressed in breast cancer cells in clinical tissues. We examined the effects of rs10509373 on the expression levels of C10orf11 in peripheral blood mononuclear cells and brain using a public database SNPExpress (22). However, no significant associations were observed ($P = 0.63$ and 0.93, respectively) possibly because of the quite low expression of C10orf11 in these tissues. The association of *C10orf11* genotype was significant in the second replication samples, which include the patients receiving tamoxifen alone after chemotherapy (Table 2); however, neither *CYP2D6* nor *ABCC2* genotypes were significantly associated with clinical outcomes in the second replication samples as shown in our previous study (Supplementary Material, Table S5) (23). These lines of evidence might suggest that C10orf11 is involved in acquiring tamoxifen resistance or determining the characteristics of breast cancer, although further functional analysis will be needed to clarify the biological mechanisms which could have effects on the

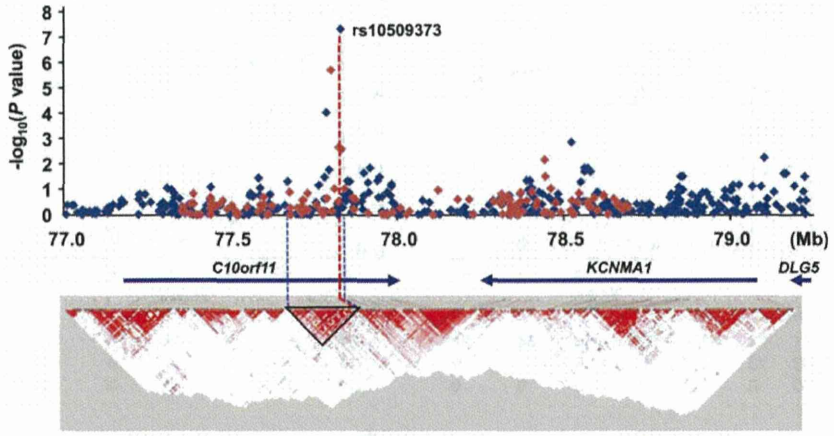


**Figure 3.** Association mapping and LD map of 10q22. Blue diamond dots represent $-\log_{10}$-transformed *P*-values of SNPs genotyped by Illumina Human610-Quad BeadChips in the GWAS, and red diamond dots show $-\log_{10}$-transformed *P*-values of the SNPs of fine mapping. Arrows indicate the position of known genes. The *D′*-based LD map (MAF $\geq$ 0.10) is drawn using genotype data of 240 patients with breast cancer enrolled in the GWAS.

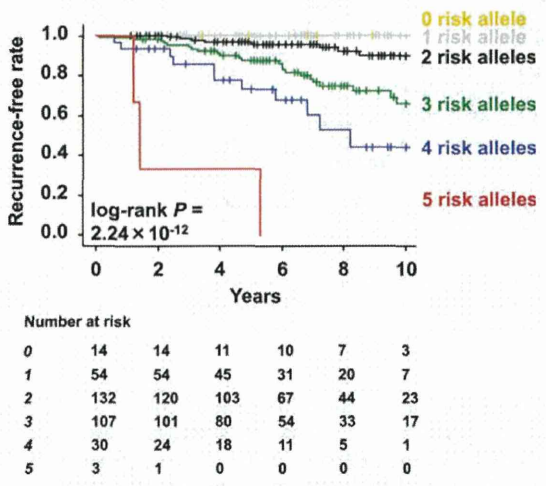


**Figure 4.** Combined effects of *C10orf11*, *CYP2D6* and *ABCC2* genotypes on clinical outcomes of tamoxifen monotherapy. Kaplan–Meier estimates of recurrence-free survival rate for combined effects of *C10orf11* rs10509373, *CYP2D6* and *ABCC2* rs3740065 genotypes. The 345 patients receiving tamoxifen monotherapy were classified into six groups (0, 1, 2, 3, 4 and 5 risk allele groups) based on the number of risk alleles of these three genes.

clinical outcomes of breast cancer patients receiving tamoxifen therapy.

Several research groups focused on the genes involved in the pharmacokinetics of tamoxifen or its metabolites have investigated genetic variants of *CYP2C19*, *CYP3A5*, *UGT2B15* and *SULT1A1* as candidate genes associated with clinical outcomes of tamoxifen therapy (7,10,14,16). In our GWAS, no SNP in these candidate genes showed significant association with recurrence-free survival in patients treated with tamoxifen (log-rank $P = 3.14 \times 10^{-2}$–$9.90 \times 10^{-1}$). According to the previous reports, the effect sizes of the above candidate genes were not so large, indicating that the sample size used in our study might not have enough power to detect associations of the SNPs with the tamoxifen efficacy.

Another group hypothesized that non-genomic steroid signaling and cross-talk with growth factor signaling pathways may contribute to the clinical outcomes of the patients treated with tamoxifen and reported that *TC21* promoter polymorphism was significantly associated with an unfavorable tamoxifen treatment outcome; however, no significant association was observed at SNPs in *TC21* gene in our GWAS (log-rank $P = 1.19 \times 10^{-1}$–$9.98 \times 10^{-1}$) (24). The *P*-values of the SNPs in the *ESR1*, *ESR2* and *PGR* genes, which encode ERα, ERβ and progesterone receptor (PR), respectively, ranged from $1.33 \times 10^{-2}$ to $9.88 \times 10^{-1}$, indicating no significant association.

In conclusion, our GWAS using 462 Japanese patients with breast cancer identified a new locus, containing the *C10orf11* gene, associated with the clinical outcomes of breast cancer patients treated with tamoxifen. These findings provide new insights into personalized selection of hormonal therapy for the patients with breast cancer. However, large-scale replication study and further functional analysis are required to verify our results and to clarify their biological mechanisms which have effects on the clinical outcomes of patients receiving tamoxifen therapy.

## MATERIALS AND METHODS

### Patients

A total of 462 patients with primary breast cancer (including the 282 patients reported previously (6,7)) were recruited at Shikoku-*10 collaborative group (Tokushima Breast Care Clinic, Yamakawa Breast Clinic, Shikoku Cancer Center, Kochi University Hospital, and Itoh Surgery and Breast Clinic), Kansai Rosai Hospital, Sapporo Breast Surgical Clinic and Sapporo Medical University Hospital. Of them, 240 patients who were recruited from September 2007 to September 2008 were used for a GWAS analysis, and the remaining 105 patients who were recruited from October 2008 to January 2010 were analyzed in a first replication study. All patients were Japanese women pathologically diagnosed with

ER-positive and/or PR-positive, invasive breast cancer who received adjuvant tamoxifen monotherapy without any other treatments after surgical treatment between 1986 and 2008. In addition, we analyzed 117 patients who had been treated with tamoxifen monotherapy after receiving chemotherapy as the second replication set (23). Data on primary breast cancer diagnosis or recurrence were confirmed from patients' medical record. Patients without recurrence were censored at the date of the last consultation. Recurrence-free survival time was defined as the time from surgical treatment to diagnosis of the recurrence of a breast cancer (locoregional, distant metastasis and contralateral breast events) or death. Patients received tamoxifen 20 mg/day for 5 years; tamoxifen was stopped at the time a recurrence was identified. ER and PR status were evaluated by enzyme immunoassay or immunohistochemistry. The cut-off for human epidermal growth factor receptor 2 overexpression was defined as 3+ immunohistochemical staining (25). Nodal status was determined according to the International Union against Cancer tumor-node-metastasis classification. This study was approved by the Institutional Review Board in the Institute of Medical Science, The University of Tokyo (Tokyo, Japan), and written informed consent was obtained from all patients.

### Genotyping and quality control

Genomic DNA was extracted from peripheral blood ($n = 424$) or frozen breast tissue ($n = 38$) using Qiagen DNA Extraction Kit (Qiagen, Valencia, CA, USA). In the GWAS, 240 patients were genotyped using the Illumina Human610-Quad BeadChip (Illumina, San Diego, CA, USA). Quality control of SNPs was achieved by excluding SNPs with low call rate (<99%) and SNPs with Hardy–Weinberg equilibrium $P$-value $<1.0 \times 10^{-6}$. SNPs with a minor allele frequency (MAF) <0.01 were also excluded from further analysis. A total of 470 796 SNPs passed the filters and were further analyzed. We used multiplex polymerase chain reaction-based Invader Assay (Third Wave Technologies, Madison, WI, USA) on ABI PRISM 7900HT (Applied Biosystems, Foster City, CA, USA) for the replication study and fine mapping (26). For fine mapping, tag SNPs were selected from the HapMap phase II JPT data (http://www.hapmap.org/) (27) with the following criteria: a pairwise $r^2 > 0.80$ and an MAF $\geq 0.01$ using Haploview software (28).

Genotyping of *CYP2D6* and *ABCC2* rs3740065 was performed using real-time Invader (Third Wave Technologies) and TaqMan assays (Applied Biosystems) as described previously (7,29,30). To evaluate the effects of *CYP2D6* alleles, we defined all of the decreased and null alleles (including **4*, **5*, **10*, **14*, **21* and **41*, and gene-duplication alleles, **10–*10* and **36–*36*) as allele '*V*', and alleles of **1* and duplicated **1–*1* as allele '*wt*' as described previously (7).

### Statistical analysis

Recurrence-free survival curves were estimated using the Kaplan–Meier method. Statistical significance of a relationship between clinical outcomes and genetic polymorphism was assessed by the trend log-rank test. The value of $\lambda_{GC}$ was calculated from the median of the trend log-rank test statistics (31). Cox proportional hazards analysis was used to identify significant prognostic clinical factors and to test for an independent contribution of genetic factors to recurrence-free survival. To examine potential confounding, age was treated as a continuous variable, tumor size was treated as an ordinal variable, and the other covariates were treated as categorical variables. Genotypes were analyzed by assigning an ordinal score to each genotype (0 for homozygous non-risk alleles, 1 for heterozygous risk alleles and 2 for homozygous risk alleles). Combination effects were investigated by adding up the number of risk alleles of *CYP2D6*, *ABCC2* and *C10orf11* genes. All polymorphisms evaluated in this study were tested for deviation from Hardy–Weinberg equilibrium with the use of a $\chi^2$-test. Statistical tests provided two-sided $P$-values, and a significance level of $P < 0.05$ was used. We used a significance level of $P < 1.06 \times 10^{-7}$ (0.05 of 470 796) in the GWAS and $5.56 \times 10^{-3}$ (0.05 of 9) in the replication study to adjust multiple testing by the strict Bonferroni correction. Statistical analyses were carried out using SPSS (version 17.0, SPSS, Chicago, IL, USA) and the R statistical environment version 2.9.2 (http://www.r-project.org/).

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

## ACKNOWLEDGEMENTS

We express our gratitude to all the study participants. We thank Yuka Kikuchi, Aiko Ohno, Shoko Higuchi and Kumi Matsuda for technical assistance. We thank all other members and staff for their contribution to the sample collection and the completion of our study.

*Conflict of Interest statement*. None declared.

## FUNDING

This work was supported mainly by a Grant-in-Aid for Leading Project of Ministry of Education, Culture, Sports, Science and Technology of Japan (to Y.N.). This work was also supported in part by Grant-in-Aids for Young Scientists (B) (22790179) of Ministry of Education, Culture, Sports, Science and Technology of Japan (to K.K.), the Kobayashi Institute for Innovative Cancer Chemotherapy (to K.K.), the Japan Research Foundation for Clinical Pharmacology (to K.K.), and the Takeda Science Foundation (to K.K.).

## REFERENCES

1. Early Breast Cancer Trialists' Collaborative Group. (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*, **365**, 1687–1717.
2. Early Breast Cancer Trialists' Collaborative Group. (1998) Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. *Lancet*, **351**, 1451–1467.
3. Borgna, J.L. and Rochefort, H. (1981) Hydroxylated metabolites of tamoxifen are formed *in vivo* and bound to estrogen receptor in target tissues. *J. Biol. Chem.*, **256**, 859–868.

4. Lien, E.A., Solheim, E., Lea, O.A., Lundgren, S., Kvinnsland, S. and Ueland, P.M. (1989) Distribution of 4-hydroxy-*N*-desmethyltamoxifen and other tamoxifen metabolites in human biological fluids during tamoxifen treatment. *Cancer Res.*, **49**, 2175–2183.
5. Johnson, M.D., Zuo, H., Lee, K.H., Trebley, J.P., Rae, J.M., Weatherman, R.V., Desta, Z., Flockhart, D.A. and Skaar, T.C. (2004) Pharmacological characterization of 4-hydroxy-*N*-desmethyl tamoxifen, a novel active metabolite of tamoxifen. *Breast Cancer Res. Treat.*, **85**, 151–159.
6. Kiyotani, K., Mushiroda, T., Sasa, M., Bando, Y., Sumitomo, I., Hosono, N., Kubo, M., Nakamura, Y. and Zembutsu, H. (2008) Impact of *CYP2D6*10* on recurrence-free survival in breast cancer patients receiving adjuvant tamoxifen therapy. *Cancer Sci.*, **99**, 995–999.
7. Kiyotani, K., Mushiroda, T., Imamura, C.K., Hosono, N., Tsunoda, T., Kubo, M., Tanigawara, Y., Flockhart, D.A., Desta, Z., Skaar, T.C. *et al.* (2010) Significant effect of polymorphisms in *CYP2D6* and *ABCC2* on clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. *J. Clin. Oncol.*, **28**, 1287–1293.
8. Goetz, M.P., Rae, J.M., Suman, V.J., Safgren, S.L., Ames, M.M., Visscher, D.W., Reynolds, C., Couch, F.J., Lingle, W.L., Flockhart, D.A. *et al.* (2005) Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J. Clin. Oncol.*, **23**, 9312–9318.
9. Goetz, M.P., Knox, S.K., Suman, V.J., Rae, J.M., Safgren, S.L., Ames, M.M., Visscher, D.W., Reynolds, C., Couch, F.J., Lingle, W.L. *et al.* (2007) The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res. Treat.*, **101**, 113–121.
10. Schroth, W., Antoniadou, L., Fritz, P., Schwab, M., Muerdter, T., Zanger, U.M., Simon, W., Eichelbaum, M. and Brauch, H. (2007) Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. *J. Clin. Oncol.*, **25**, 5187–5193.
11. Lim, H.S., Lee, J.H., Lee, S.K., Lee, S.E., Jang, I.J. and Ro, J. (2007) Clinical implications of CYP2D6 genotypes predictive of tamoxifen pharmacokinetics in metastatic breast cancer. *J. Clin. Oncol.*, **25**, 3837–3845.
12. Xu, Y., Sun, Y., Yao, L., Shi, L., Wu, Y., Ouyang, T., Li, J., Wang, T., Fan, Z., Fan, T. *et al.* (2008) Association between CYP2D6*10 genotype and survival of breast cancer patients receiving tamoxifen treatment. *Ann. Oncol.*, **19**, 1423–1429.
13. Jin, Y., Desta, Z., Stearns, V., Ward, B., Ho, H., Lee, K.H., Skaar, T., Storniolo, A.M., Li, L., Araba, A. *et al.* (2005) CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J. Natl Cancer Inst.*, **97**, 30–39.
14. Gjerde, J., Geisler, J., Lundgren, S., Ekse, D., Varhaug, J.E., Mellgren, G., Steen, V.M. and Lien, E.A. (2010) Associations between tamoxifen, estrogens, and FSH serum levels during steady state tamoxifen treatment of postmenopausal women with breast cancer. *BMC Cancer*, **10**, 313.
15. Schroth, W., Goetz, M.P., Hamann, U., Fasching, P.A., Schmidt, M., Winter, S., Fritz, P., Simon, W., Suman, V.J., Ames, M.M. *et al.* (2009) Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *J. Am. Med. Assoc.*, **302**, 1429–1436.
16. Wegman, P., Elingarami, S., Carstensen, J., Stal, O., Nordenskjold, B. and Wingren, S. (2007) Genetic variants of *CYP3A5*, *CYP2D6*, *SULT1A1*, *UGT2B15* and tamoxifen response in postmenopausal patients with breast cancer. *Breast Cancer Res.*, **9**, R7.
17. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. and Haussler, D. (2004) Ultraconserved elements in the human genome. *Science*, **304**, 1321–1325.
18. Deloukas, P., Earthrowl, M.E., Grafham, D.V., Rubenfield, M., French, L., Steward, C.A., Sims, S.K., Jones, M.C., Searle, S., Scott, C. *et al.* (2004) The DNA sequence and comparative analysis of human chromosome 10. *Nature*, **429**, 375–381.
19. Venkatesh, B., Kirkness, E.F., Loh, Y.H., Halpern, A.L., Lee, A.P., Johnson, J., Dandona, N., Viswanathan, L.D., Tay, A., Venter, J.C. *et al.* (2006) Ancient noncoding elements conserved in the human genome. *Science*, **314**, 1892.
20. Hurtado, A., Holmes, K.A., Geistlinger, T.R., Hutcheson, I.R., Nicholson, R.I., Brown, M., Jiang, J., Howat, W.J., Ali, S. and Carroll, J.S. (2008) Regulation of *ERBB2* by oestrogen receptor-PAX2 determines response to tamoxifen. *Nature*, **456**, 663–666.
21. Han, H.J., Russo, J., Kohwi, Y. and Kohwi-Shigematsu, T. (2008) SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature*, **452**, 187–193.
22. Heinzen, E.L., Ge, D., Cronin, K.D., Maia, J.M., Shianna, K.V., Gabriel, W.N., Welsh-Bohmer, K.A., Hulette, C.M., Denny, T.N. and Goldstein, D.B. (2008) Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol*, **6**, e1.
23. Kiyotani, K., Mushiroda, T., Hosono, N., Tsunoda, T., Kubo, M., Aki, F., Okazaki, Y., Hirata, K., Takatsuka, Y., Okazaki, M. *et al.* (2010) Lessons for pharmacogenomics studies: association study between *CYP2D6* genotype and tamoxifen response. *Pharmacogenet. Genomics*, **20**, 565–568.
24. Rokavec, M., Schroth, W., Amaral, S.M., Fritz, P., Antoniadou, L., Glavac, D., Simon, W., Schwab, M., Eichelbaum, M. and Brauch, H. (2008) A polymorphism in the *TC21* promoter associates with an unfavorable tamoxifen treatment outcome in breast cancer. *Cancer Res.*, **68**, 9799–9808.
25. Moeder, C.B., Giltnane, J.M., Harigopal, M., Molinaro, A., Robinson, A., Gelmon, K., Huntsman, D., Camp, R.L. and Rimm, D.L. (2007) Quantitative justification of the change from 10% to 30% for human epidermal growth factor receptor 2 scoring in the American Society of Clinical Oncology/College of American Pathologists Guidelines: tumor heterogeneity in breast cancer and its implications for tissue microarray based assessment of outcome. *J. Clin. Oncol.*, **25**, 5418–5425.
26. Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. and Nakamura, Y. (2001) A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.*, **46**, 471–477.
27. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
28. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
29. Hosono, N., Kato, M., Kiyotani, K., Mushiroda, T., Takata, S., Sato, H., Amitani, H., Tsuchiya, Y., Yamazaki, K., Tsunoda, T. *et al.* (2009) *CYP2D6* genotyping for functional-gene dosage analysis by allele copy number detection. *Clin. Chem.*, **55**, 1546–1554.
30. Kiyotani, K., Mushiroda, T., Kubo, M., Zembutsu, H., Sugiyama, Y. and Nakamura, Y. (2008) Association of genetic polymorphisms in *SLCO1B3* and *ABCC2* with docetaxel-induced leukopenia. *Cancer Sci.*, **99**, 967–972.
31. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.

特集
胆と膵 Vol. 34 (2) p. 143～148, 2013
胆道癌，膵癌に対する個別化治療の新展開

# ゲノムワイド関連解析による ジェムシタビン副作用関連遺伝子の同定

前佛　均[1]・清谷　一馬[2]・宇野　智子[3]・木村　康利[3]
薙田　泰誠[2]・光畑　直喜[4]・伊奈志乃美[5]・鬼原　史[3]
山上　裕機[5]・平田　公一[3]・中村　祐輔[1]

**要約：抗癌剤による副作用発現の有無は多くの要因が関係して規定されているものと考えられるが，遺伝的要因もその重要な因子の一つと考えられている。われわれはジェムシタビンにより引き起こされる重篤な骨髄抑制と関連する遺伝子多型（一塩基多型：SNP）を同定するため，164例のジェムザール単剤治療症例を用いてゲノムワイド関連解析および再現性確認のためのreplication studyを行った。その結果ジェムシタビンによる副作用と強い関連をもつ可能性の高い四つのSNPを含む遺伝領域を同定した（*DAPK1* 上のrs11141915：$P=1.27\times10^{-6}$，2q12に存在するrs1901440：$P=3.11\times10^{-6}$，*PDE4B* 上のrs12046844：$P=4.56\times10^{-5}$，3q29に存在するrs11719165：$P=5.98\times10^{-5}$）。同定された四つのSNPを用いて副作用リスクに働くと考えられるgenotypeの合計数に応じて各症例を点数化したところ，点数の高い症例では低い症例に比べて有意に副作用の発現率が高くなることが示された。今回同定された四つの遺伝子多型を用いたスコアリングシステムはジェムザールによる副作用の投与前診断に有用となる可能性が示された。**

Key words：ジェムシタビン，骨髄抑制，ゲノムワイド関連解析，遺伝子多型

## はじめに

現在胆膵悪性疾患をはじめ多くの悪性腫瘍に対する治療薬として適応となっているジェムシタビン（ジェムザール®）は骨髄抑制をはじめ，有害事象の発生頻度が決して少なくない薬剤であるが，その副作用の発現を規定する遺伝的要因についてはいまだ十分に解明されていないのが現状である。生命の設計図とも言われる人の遺伝情報（ゲノム配列）は同じ人間といえども個人間でわずかな違いが存在することが知られており，遺伝子多型（一塩基多型）と呼ばれる塩基配列の個人差を比較することで副作用の発現と関係する遺伝子を同定しようとする解析が進んできており，一部は日常臨床に応用されている。近年，ゲノム全体にわたり一塩基多型をgenotypingする技術が進歩し，ゲノムワイド関連解析（genome-wide association study，GWAS：「ジーワス」と呼ばれることが多い）という方法によりこれまで副作用との関連が全く知られていなかった新たな副作用関連遺伝子を発見する試みがなされるようになってきた。ジェムシタビンはSLC28A1，SLC28A3，SLC 29A1などの薬剤輸送タンパクを介して血中から細胞内に入り[1～3]，deoxycyti-

A Genome-wide Association Study Identifies Four Genetic Markers for Hematological Toxicities in Cancer Patients Receiving Gemcitabine Therapy
*Hitoshi Zembutsu et al*

1）東京大学医科学研究所ヒトゲノム解析センター（〒113-0033 文京区本郷7-3-1）
2）理化学研究所ゲノム医科学研究センター
3）札幌医科大学医学部外科学第一講座
4）呉共済病院泌尿器科
5）和歌山県立医科大学外科学第二講座

dine kinase (dCK), cytidine deaminase (CDA) などの酵素により代謝を受けることが知られていることから[4]，これらの既知遺伝子上の多型と副作用との関係を調べた報告はいくつか存在するが，現在のところ副作用と強い関連を示す遺伝子多型は同定されていない。本研究はゲノムワイド関連解析を通じジェムザールによる副作用と強い関係を有する遺伝子多型を同定することで，副作用予測診断へ応用することを目的として行われた[5]。

## Ⅰ．ジェムシタビンによる有害事象

ジェムシタビン単剤による抗腫瘍治療を受けた174症例を対象に解析を行った（表1）。174例中grade 3以上の白血球/好中球減少症をきたした54例をcase，副作用を示さなかった120例をcontrolとしcase-control studyを行った。解析は21例のcaseおよび58例のcontrolをゲノムワイド関連解析（GWAS）に用い，33例のcaseおよび62例のcontrolをGWAS結果の再現性確認のためのreplication studyに用いた。case-control間で有意な性差を認めず（$P>0.64$），年齢分布にも有意差を認めなかった（$P>0.53$）。疾患別では，半分以上の症例が膵癌(56.9%)でその他肺癌(20.1%)，胆管癌（18.4%）などであった。GWASで用いた症例と，再現性確認のためのreplication studyで用いたサンプル間で疾患分布に有意な差を認めなかった。

## Ⅱ．ゲノムワイド関連解析によるジェムシタビン副作用関連候補遺伝子の同定

ジェムシタビン投与により骨髄抑制（>grade 3）が認められた21例と，投与により有害事象を認めなかった58例を用いて，ゲノム全体にわたり（約610,000 SNP）遺伝子多型をスクリーニングした。得られた各症例の610,000 SNPのgenotype情報を用いてcase-control関連解析（Fisherの正確検定）を行った。その結果，もっとも副作用と強い関連を示した遺伝子多型（SNP）は$P=0.000006690$を示した。図1にゲノム全体にわたるマーカーSNPとジェムシタビン副作用との関連の強さをグラフで表したもの（マンハッタンプロット）を示すが，ジェムシタビンの副作用と関係するSNPはゲノム全体にわたり散在している可能性を示している。

## Ⅲ．ジェムシタビン副作用関連候補遺伝子のreplication study

ゲノムワイド関連解析の結果の再現性を確認するために，有意差上位100 SNPについて33例のcaseおよび62例のcontrolを用いて関連解析を行った。100 SNPに対するreplication studyの結果$P<0.05$を示す4 SNPが同定された（表2）。4 SNPとジェムシタビンによる骨髄抑制との関連はそれぞれ9番染色体上のrs11141915が$P=2.77\times10^{-3}$，2番染色体上のrs1901440は$P=1.82\times10^{-2}$，1番染色体上のrs12046844は$P=3.09\times10^{-2}$，3番染色体上のrs11719165は$P=4.61\times10^{-2}$を示した。さらにこの4 SNPについてGWASで用いたcaseおよびcontrol症例をそれぞれ加えて解析した結果，いずれもゲノムワイド有意水準である$1.07\times10^{-7}$に達するSNPは存在しなかったものの，9番染色体上のrs11141915は$P=1.27\times10^{-6}$，オッズ比4.10（95% CI：2.21-7.62），2番染色体上のrs1901440は$P=3.11\times10^{-6}$，オッズ比34.00（95% CI：4.29-269.48），1番染色体上のrs12046844は$P=4.56\times10^{-5}$，オッズ比4.13(95% CI：2.10-8.14)，3番染色体上のrs11719165は$P=5.98\times10^{-5}$，オッズ比2.60（95% CI：1.63-4.14）を示し，この4 SNPを含む遺伝的領域はジェムシタビンによる骨髄抑制と何らかの関連を示す結果となった。また，4遺伝領域の中で9番染色体上の領域については*DAPK1*，1番染色体上の領域については*PDE4B*という既知の遺伝子を含んでいた。

## Ⅳ．遺伝子多型情報を用いたジェムシタビンによる骨髄抑制予測診断モデル

ジェムシタビンによる骨髄抑制と関連が示唆された4SNPはmultiple logistic regression解析の結果それぞれ独立した副作用予測因子であったため，この4SNPを用いた骨髄抑制予測診断システムについて検討を行った。四つのSNPについて骨髄抑制リスクに働くと考えられるgenotypeを持っている場合，それぞれのSNPについて1点を与え，もっていない場合には0点として各症例合計点数別に骨髄抑制発現群（case）と副作用を認めなかった群（control）で分布を調べた結果が表3および図2である。スコア0または1を示した113例中骨髄抑制群は11.5%，スコア2については60.9%，スコア3については86.7%が骨髄抑制発現群が占めており，コントロール群に比べ有意に高いスコ