[6] Grothey A. Oxaliplatin-safety profile: neurotoxicity. Semin Oncol 2003; 30:5–13.
[7] Grothey A. Clinical management of oxaliplatin-associated neurotoxicity. Clin Colorectal Cancer 2005;5(Suppl. 1):S38–46.
[8] Graham J, Mushin M, Kirkpatrick P. Oxaliplatin. Nat Rev Drug Discov 2004;3:11–2.
[9] Land SR, Kopec JA, Cecchini RS, Ganz PA, Wieand HS, Colangelo LH. Neurotoxicity from oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: NSABP C-07. J Clin Oncol 2007;25:2205–11.
[10] Ishimoto TM, Ali-Osman F. Allelic variants of the human glutathione S-transferase P1 gene confer differential cytoprotection against anticancer agents in *Escherichia coli*. Pharmacogenetics 2002;12:543–53.
[11] Zimniak P, Nanduri B, Pikula S, Bandorowicz-Pikula J, Singhal SS, Srivastava SK, et al. Naturally occurring human glutathione S-transferase GSTP1-1 isoforms with isoleucine and valine in position 104 differ in enzymic properties. Eur J Biochem 1994;224:893–9.
[12] Ali-Osman F, Akande O, Antoun G, Mao JX, Buolamwini J. Molecular cloning, characterization, and expression in *Escherichia coli* of full-length cDNAs of three human glutathione S-transferase Pi gene variants. Evidence for differential catalytic activity of the encoded proteins. J Biol Chem 1997;272: 10004–12.
[13] Watson MA, Stewart RK, Smith GB, Massey TE, Bell DA. Human glutathione S-transferase P1 polymorphisms: relationship to lung tissue enzyme activity and population frequency distribution. Carcinogenesis 1998;19:275–80.
[14] Stoehlmacher J, Park DJ, Zhang W, Groshen S, Tsao-Wei DD, Yu MC, et al. Association between glutathione S-transferase P1, T1, and M1 genetic polymorphism and survival of patients with metastatic colorectal cancer. J Natl Cancer Inst 2002;94:936–42.
[15] Stoehlmacher J, Park DJ, Zhang W, Yang D, Groshen S, Zahedy S. A multivariate analysis of genomic polymorphisms: prediction of clinical outcome to 5-FU/oxaliplatin combination chemotherapy in refractory colorectal cancer. Br J Cancer 2004;91:344–54.
[16] Booton R, Ward T, Heighway J, Ashcroft L, Morris J, Thatcher N. Glutathione-S-transferase P1 isoenzyme polymorphisms, platinum-based chemotherapy, and non-small cell lung cancer. J Thorac Oncol 2006;1:679–83.
[17] Lecomte T, Landi B, Beaune P, Laurent-Puig P, Loriot MA. Glutathione S-transferase P1 polymorphism (Ile$^{105}$Val) predicts cumulative neuropathy in patients receiving oxaliplatin-based chemotherapy. Clin Cancer Res 2006;12:3050–6.
[18] Ruzzo A, Graziano F, Kawakami K, Watanabe G, Santini D, Catalano V, et al. Pharmacogenetic profiling and clinical outcome of patients with advanced gastric cancer treated with palliative chemotherapy. J Clin Oncol 2006; 24:1883–91.
[19] Goekkurt E, Hoehn S, Wolschke C, Wittmer C, Stueber C, Hossfeld DK, et al. Polymorphisms of glutathione S-transferases (GST) and thymidylate synthase (TS)—novel predictors for response and survival in gastric cancer patients. Br J Cancer 2006;94:281–6.
[20] Oldenburg J, Kraggerud SM, Cvancarova M, Lothe RA, Fossa SD. Cisplatin-induced long-term hearing impairment is associated with specific glutathione S-transferase genotypes in testicular cancer survivors. J Clin Oncol 2007;25: 708–14.
[21] Ruzzo A, Graziano F, Kawakami K, Watanabe G, Santini D, Catalano V, et al. Pharmacogenetic profiling in patients with advanced colorectal cancer treated with first-line FOLFOX-4 chemotherapy. J Clin Oncol 2007;25:1247–54.
[22] Marsh S, Paul J, King CR, Gifford G, McLeod HL, Brown R. Pharmacogenetic assessment of toxicity and outcome after platinum plus taxane chemotherapy in ovarian cancer: the Scottish Randomised Trial in Ovarian Cancer. J Clin Oncol 2007;25:4528–35.
[23] Le Morvan V, Smith D, Laurand A, Brouste V, Bellott R, Soubeyran I, et al. Determination of ERCC2 Lys751Gln and GSTP1 Ile$^{105}$Val gene polymorphisms in colorectal cancer patients: relationships with treatment outcome. Pharmacogenomics 2007;8:1693–703.
[24] Gamelin L, Capitain O, Morel A, Dumont A, Traore S, Anne le B, et al. Predictive factors of oxaliplatin neurotoxicity: the involvement of the oxalate outcome pathway. Clin Cancer Res 2007;13:6359–68.
[25] Pare L, Marcuello E, Altes A, del Rio E, Sedano L, Salazar J, et al. Pharmacogenetic prediction of clinical outcome in advanced colorectal cancer patients receiving oxaliplatin/5-fluorouracil as first-line chemotherapy. Br J Cancer 2008;99:1050–5.
[26] Kweekel DM, Gelderblom H, Antonini NF, Van der Straaten T, Nortier JW, Punt CJ, et al. Glutathione-S-transferase pi (GSTP1) codon 105 polymorphism is not associated with oxaliplatin efficacy or toxicity in advanced colorectal cancer patients. Eur J Cancer 2009;45:572–8.
[27] Goekkurt E, Al-Batran SE, Hartmann JT, Mogck U, Schuch G, Kramer M, et al. Pharmacogenetic analyses of a phase III trial in metastatic gastroesophageal adenocarcinoma with fluorouracil and leucovorin plus either oxaliplatin or cisplatin: a study of the Arbeitsgemeinschaft Internistische Onkologie. J Clin Oncol 2009;27:2863–73.
[28] Grothey A, McLeod HL, Green EM, Sargent DJ, Fuchs C, Ramanathan RK, et al. Glutathione S-transferase P1 I105V (GSTP1 I105V) polymorphism is associated with early onset of oxaliplatin-induced neurotoxicity. J Clin Oncol 2005;23:3509.
[29] Adelsberger H, Quasthoff S, Grosskreutz J, Lepier A, Eckel F, Lersch C. The chemotherapeutic oxaliplatin alters voltage-gated Na(+) channel kinetics on rat sensory neurons. Eur J Pharmacol 2000;406:25–32.
[30] Grolleau F, Gamelin L, Boisdron-Celle M, Lapied B, Pelhate M, Gamelin E. A possible explanation for a neurotoxic effect of the anticancer agent oxaliplatin on neuronal voltage-gated sodium channels. J Neurophysiol 2001;85:2293–7.
[31] Holmes RP, Assimos DG. Glyoxylate synthesis, and its modulation and influence on oxalate synthesis. J Urol 1998;160:1617–24.
[32] Innocenti F, Undevia SD, Iyer L, Chen PX, Das S, Kocherginsky M, et al. Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. J Clin Oncol 2004;22:1382–8.
[33] Han JY, Lim HS, Shin ES, Yoo YK, Park YH, Lee JE, et al. Comprehensive analysis of UGT1A polymorphisms predictive for pharmacokinetics and treatment outcome in patients with non-small-cell lung cancer treated with irinotecan and cisplatin. J Clin Oncol 2006;24:2237–44.
[34] Minami H, Sai K, Saeki M, Saito Y, Ozawa S, Suzuki K, et al. Irinotecan pharmacokinetics/pharmacodynamics and UGT1A genetic polymorphisms in Japanese: roles of UGT1A1*6 and *28. Pharmacogenet Genom 2007; 17:497–504.
[35] Cheeseman SL, Joel SP, Chester JD, Wilson G, Dent JT, Richards FJ. A 'modified de Gramont' regimen of fluorouracil, alone and with oxaliplatin, for advanced colorectal cancer. Br J Cancer 2002;87:393–9.
[36] Levi F, Perpoint B, Garufi C, Focan C, Chollet P, Depres-Brummer P, et al. Oxaliplatin activity against metastatic colorectal cancer. A phase II study of 5-day continuous venous infusion at circadian rhythm modulated rate. Eur J Cancer 1993;29A:1280–4.
[37] Kurata S, Kanagawa T, Yamada K, Torimura M, Yokomaku T, Kamagata Y, et al. Fluorescent quenching-based quantitative detection of specific DNA/RNA using a BODIPY((R)) FL-labeled probe or primer. Nucleic Acids Res 2001;29:E34.
[38] Matsumoto N, Kakihara F, Kimura S, Kurebayashi Y, Hirai M, Yohda M, et al. Single nucleotide polymorphism genotyping of CYP2C19 using a new automated system. Anal Biochem 2007;370:121–3.
[39] Kamatani N, Sekine A, Kitamoto T, Iida A, Saito S, Kogame A, et al. Large-scale single-nucleotide polymorphism (SNP) and haplotype analyses, using dense SNP maps, of 199 drug-related genes in 752 subjects: the analysis of the association between uncommon SNPs within haplotype blocks and the haplotypes constructed with haplotype-tagging SNPs. Am J Hum Genet 2004;75:190–203.
[40] Levi F, Zidani R, Misset J. Randomised multicentre trial of chronotherapy with oxaliplatin, fluorouracil, and folinic acid in metastatic colorectal cancer. International Organization for Cancer Chronotherapy. Lancet 1997;350: 681–6.

*Human Molecular Genetics, 2010, Vol. 19, No. 12* **2516–2523**
*doi:10.1093/hmg/ddq123*
*Advance Access published on March 29, 2010*

# The *FOXE1* locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl

**Meiko Takahashi[1,2,†], Vladimir A. Saenko[3,†], Tatiana I. Rogounovitch[4], Takahisa Kawaguchi[1,2], Valentina M. Drozd[5], Hisako Takigawa-Imamura[1], Natallia M. Akulevich[4], Chanavee Ratanajaraya[1], Norisato Mitsutake[4], Noboru Takamura[4], Larisa I. Danilova[6], Maxim L. Lushchik[5], Yuri E. Demidchik[7], Simon Heath[8], Ryo Yamada[1], Mark Lathrop[8,9], Fumihiko Matsuda[1,2,*] and Shunichi Yamashita[3,4]**

[1]Center for Genomic Medicine and [2]Institut National de la Santé et de la Recherche Médicale (INSERM) Unit U852, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan, [3]Department of International Health and Radiation Research and [4]Department of Molecular Medicine, Atomic Bomb Disease Institute, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki 852-8523, Japan, [5]Department of Thyroid Disease Research, [6]Department of Endocrinology and [7]Belarusian Medical Academy for Postgraduate Education, Minsk 220013, Republic of Belarus, [8]Centre National de Génotypage, Institut Génomique, Commissariat à l'Énergie Atomique, Evry 91000, France and [9]Fondation Jean Dausset-CEPH, Paris 75010, France

Received January 18, 2010; Revised and Accepted March 17, 2010

**Papillary thyroid cancer (PTC) among individuals exposed to radioactive iodine in their childhood or adolescence is a major internationally recognized health consequence of the Chernobyl accident. To identify genetic determinants affecting individual susceptibility to radiation-related PTC, we conducted a genome-wide association study employing Belarusian patients with PTC aged 0–18 years at the time of accident and age-matched Belarusian control subjects. Two series of genome scans were performed using independent sample sets, and association with radiation-related PTC was evaluated. Meta-analysis by the Mantel–Haenszel method combining the two studies identified four SNPs at chromosome 9q22.33 showing significant associations with the disease (Mantel–Haenszel *P*: mhp = $1.7 \times 10^{-9}$ to $4.9 \times 10^{-9}$). The association was further reinforced by a validation analysis using one of these SNP markers, rs965513, with a new set of samples (overall mhp = $4.8 \times 10^{-12}$, OR = 1.65, 95% CI: 1.43–1.91). Rs965513 is located 57-kb upstream to *FOXE1*, a thyroid-specific transcription factor with pivotal roles in thyroid morphogenesis and was recently reported as the strongest genetic risk marker of sporadic PTC in European populations. Of interest, no association was obtained between radiation-related PTC and rs944289 (mhp = 0.17) at 14p13.3 which showed the second strongest association with sporadic PTC in Europeans. These results show that the complex pathway underlying the pathogenesis may be partly shared by the two etiological forms of PTC, but their genetic components do not completely overlap each other, suggesting the presence of other unknown etiology-specific genetic determinants in radiation-related PTC.**

## INTRODUCTION

The Chernobyl accident in April 1986 led to radioactive contamination of vast territories in Belarus, Ukraine and Russia. Millions of residents were exposed to a wide spectrum of radionuclides of which $^{131}I$ was the major dose-forming isotope for the thyroid. A sharp increase in thyroid cancer incidence among those exposed in childhood or adolescence has

*To whom correspondence should be addressed at: Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Yoshida Konoe-cho, Sakyo, Kyoto 606-8501, Japan. Tel: +81 757539313; Fax: +81 757539314; Email: fumi@genome.med.kyoto-u.ac.jp
†These authors contributed equally to this work.

been reported since the early 1990s. Its specific temporal and geographic distribution was suggestive of a common causative event in the development of the malignancy (1,2), which was later proved to be internal exposure to $^{131}I$ through its incorporation into food chains of pastured cows and further consumption of fresh milk (3). In 2002, the number of diagnosed thyroid cancers in the three most affected countries approached 5000 of which an estimate of 75% could be attributed to Chernobyl radiation (2,4).

Among the variety of histological types of thyroid cancer, only papillary thyroid carcinoma (PTC) displays evident radiation dose–response and accounts for ~95% cases in the Chernobyl aftermath (3,5,6). Radiation is the only known environmental risk factor for PTC seen both after external exposure (7) and internal irradiation (5). The risk for thyroid cancer in the individuals exposed to radiation at young age remains elevated throughout their lifespan. Although a role of predisposing factors commonly associated with sporadic PTC to the female sex is less relevant in cases of radiation-related PTC, a female to male ratio of 1.6 to 1 has been reported (8). Furthermore, radiation-related PTC is also variable in terms of the duration period of latency, the earliest of which is reported to be 4 years (1). It also remains unclear why, notwithstanding the appreciably comparable thyroid radiation doses in Chernobyl PTC patients and in healthy individuals of the same age and of the same settlements (3,9,10), thyroid malignancy develops only in a small fraction of those exposed. Thus, while radiation dose and young age at exposure are well-established risk factors for PTC, observations are suggestive of an existence of genetic factors and complex gene–environment interactions that may modulate individual radiation sensitivity and susceptibility to radiation-related PTC.

In order to identify genetic determinants that modify individual predisposition to radiation-related thyroid malignancy, we conducted a genome-wide association (GWA) study. Two series of genome scans were performed using two independent sample sets consisting of childhood PTC patients of Belarus and control subjects, followed by a validation study using a third set of case and control samples. A total of 667 patients diagnosed for PTC in 1989–2009, and 827 age-matched controls from the same regions were recruited, comprising the largest collection of patients analyzed to date. In addition, genome scan results of 448 Russian DNA samples were also included as general population controls.

## RESULTS

### GWA study

In the initial genome scan (termed as Study 1), a total of 532 024 autosomal SNP markers of 187 PTC patients and 172 controls were chosen for a case–control association study after quality control of the genotyping results (Table 1). The average call rates per SNP marker and per DNA sample were 0.999 and 0.999, respectively. No strong deviation of inflation factor was observed between the case and control groups (genomic control inflation factor $\lambda = 1.08$, Supplementary Material, Fig. S1a). A statistical analysis comparing genotype distributions did not find SNP markers that showed genome-wide significance. In the subsequent genome scan (termed as Study 2), 214 cases were examined in the association analysis after quality control, and genotype distributions of 509 610 SNP markers were compared with those of 448 Russian population controls. In Study 2, the average call rates per SNP marker and per DNA sample were 0.998 and 0.980, respectively. A slight inflation of genomic control λ-value was observed between the case and control groups (genomic control inflation factor $\lambda = 1.14$, Supplementary Material, Figs S1b and S2), which is most likely due to within-Russia substructures in the Russian population controls. Again, there were no SNP markers that showed genome-wide significance.

A meta-analysis was undertaken through integration of the genotypes obtained in Study 1 and Study 2. Association with radiation-related PTC was evaluated using the Mantel–Haenszel method for 506 840 SNP markers that passed quality control in both studies. The distribution of the mhp-values along the chromosomes is shown in Figure 1. A slight inflation of λ-value was observed between case and control ($\lambda = 1.11$, Supplementary Material, Fig. S1c). A cluster of four SNPs at chromosome 9q22.23 showed genome-wide significance ($P < 5.0 \times 10^{-8}$), namely, rs925489, rs7850258, rs965513 and rs10759944 with meta-analysis $P$-values of $1.7 \times 10^{-9}$, $1.7 \times 10^{-9}$, $4.9 \times 10^{-9}$ and $3.5 \times 10^{-9}$, respectively (Fig. 2 and Table 2). These markers are in strong linkage disequilibrium (LD) to each other (pairwise $D' > 0.999$, $r^2 > 0.999$). Although there were no neighboring SNPs showing stronger signals (Supplementary Material, Table S1), nine other markers at the same chromosomal locus showed suggestive association signals ($mhp = 5.2 \times 10^{-4}$ to $1.4 \times 10^{-6}$) (Table 2). In addition, we examined the association by pooling genotypes obtained in Studies 1 and 2. After correction for population stratification using Eigenstrat as well as for residual inflation by the genomic control method, all four markers that showed genome-wide significance in the meta-analysis were slightly below the level of genome-wide significance (rs7850258: $P = 1.5 \times 10^{-7}$, rs925489: $P = 1.5 \times 10^{-7}$, rs10759944: $P = 2.4 \times 10^{-7}$, rs965513: $P = 3.2 \times 10^{-7}$).

SNP markers located on the X chromosome were tested for association in a separate analysis. Cases and controls were sub-grouped into males and females and association analysis was carried out. As a result, none of the markers showed genome-wide significance ($mhp > 3.6 \times 10^{-5}$ for males, $mhp > 1.6 \times 10^{-5}$ for females).

### Validation study

The 425-kb region between rs4742698 and rs4618817 encompassing these markers was evaluated for LD structure with the genotyping results of Study 1 and Study 2. Three LD blocks were identified: block A between rs4742698 and rs16924042, block B between rs1512261 and rs10818094 and block C between rs7871887 and rs4618817. All of the four most significant markers are in block B (Fig. 2). There are eight genes that have been localized in the vicinity of these SNPs: *TMOD1* (Entrez Gene ID: 7111), *C9orf97* (ID: 158427), *NCBP1* (ID: 4686), *XPA* (ID: 7507), *KRT18P13* (ID: 392371), *FOXE1* (ID: 2304), *C9orf156* (ID: 51531) and

**Table 1.** Specification of the DNA samples used for the study

| Study | Sample set | Classification | Number | Age at exposure | | Age at diagnosis | |
|---|---|---|---|---|---|---|---|
| | | | | Range | Mean ± SD | Range | Mean ± SD |
| Study 1 | PTC1 | Cases | 187 | 0 – 17 | 3.0 ± 3.8 | 3 – 20 | 10.0 ± 4.1 |
| | CTR1 | Controls | 172 | 0 – 17 | 1.5 ± 2.8 | — | — |
| Study 2 | PTC2 | Cases | 214 | 0 – 17 | 5.8 ± 5.2 | 2 – 22 | 13.9 ± 5.5 |
| | CTR2 | Controls[a] | 448 | — | — | — | — |
| Study 3 | PTC3 | Cases | 259 | 0 – 18 | 6.8 ± 5.5 | 3 – 22 | 16.5 ± 4.4 |
| | CTR3 | Controls | 648 | 0 – 26 | 6.2 ± 5.9 | — | — |

[a]Russian population controls from other genetic studies.

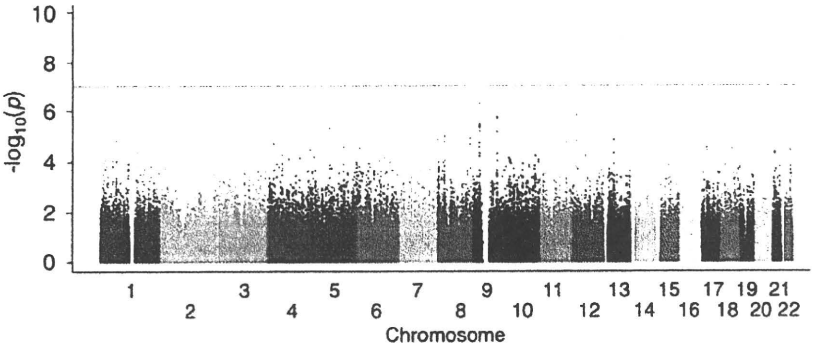


**Figure 1.** Manhattan plot of the combined GWAS results for Studies 1 and 2. *P*-values calculated by the Trend $\chi^2$ test for 506 840 autosomal SNPs are plotted in $-\log_{10}$ scale with respect to their chromosomal positions. The horizontal line indicates Bonferroni-adjusted $P = 9.6 \times 10^{-8}$.

*HEMGN* (ID: 55363), but none of these genes reside in block B. Seven out of the nine markers showing suggestive association signals are located at either 5′ or 3′ flanking region of the *FOXE1* gene in block C. In addition, an imputation analysis was performed for SNP markers in blocks A, B and C using genotypes of International HapMap Project as reference. We identified three additional SNPs in block B, namely rs7030280, rs10983700 and rs1588635, located approximately 9–11 kb centromeric to rs925489, showing similar levels of association (imputed $P = 2.8 \times 10^{-9}$ for rs7030280 and rs10983700, imputed $P = 3.7 \times 10^{-9}$ for rs1588635) (Supplementary Material, Table S2). No other SNP markers in block A or block C reached genome-wide significance.

This region at 9q22.23 containing the *FOXE1* (or *TTF2*) gene which encodes a thyroid-specific transcription factor was recently identified as a chromosomal locus strongly associated with predisposition to sporadic thyroid cancer in an Icelandic study (11). Among the seven SNPs showing significant associations ($P < 2.8 \times 10^{-9}$) in the Icelandic sporadic PTC patients, rs965513 was strongest ($P = 6.8 \times 10^{-20}$, OR = 1.77) (Table 2). We therefore selected rs965513 located ~57 kb upstream to *FOXE1* for further genotyping by Taqman using an independent sample set (termed as Study 3) of 259 cases and 648 controls (Table 1). The strong association ($P = 2.0 \times 10^{-4}$) was reproduced and was further reinforced when the genotypes of the three studies were combined for meta-analysis (mhp = $4.8 \times 10^{-12}$, OR = 1.65, 95% CI: 1.43–1.91).

Very recently, another genetic study focusing on 97 candidate genes mediating thyroid carcinogenesis identified rs1867277 in the 5′-UTR of *FOXE1* as a genetic determinant for sporadic PTC ($P = 5.9 \times 10^{-9}$, OR = 1.49, 95% CI: 1.30–1.70) (12). Since rs1867277 was not examined in our study, we designed a Taqman probe and genotyped 660 PTC cases (PTC1, PTC2 and PTC3) and 820 Belarusian controls (CTR1 and CTR3) (Table 1). A significant association was obtained with a *P*-value of $4.5 \times 10^{-7}$ and OR of 1.48 (95% CI: 1.27–1.71).

## Genotyping of rs944289 at chromosome 14q13.3

Rs944289 at chromosome 14q13.3 showed the second strongest association with sporadic PTC in the Icelandic population ($P = 2.5 \times 10^{-8}$, OR = 1.44, 95% CI: 1.26–1.63) (11). This SNP is located in a 249-kb LD region which does not contain any known genes, but it lies close to *TTF1* (ID: 7080), another thyroid-specific transcription factor gene. We investigated whether rs944289 showed significant association in our genome scan results. Of our interest, it failed to show any association with radiation-related PTC ($P = 0.23$ in Study 1, $P = 0.43$ in Study 2 and mhp = 0.17 by meta-analysis) (Table 2).

## Correlation between rs965513 genotypes and disease latency

It is considered that thyroid cancer requires an induction and latency period of at least 10 years after exposure to ionizing radiation (13). We divided the 660 case samples into two groups depending on the date of diagnosis being either within, or more than, 10 years since radiocontamination.

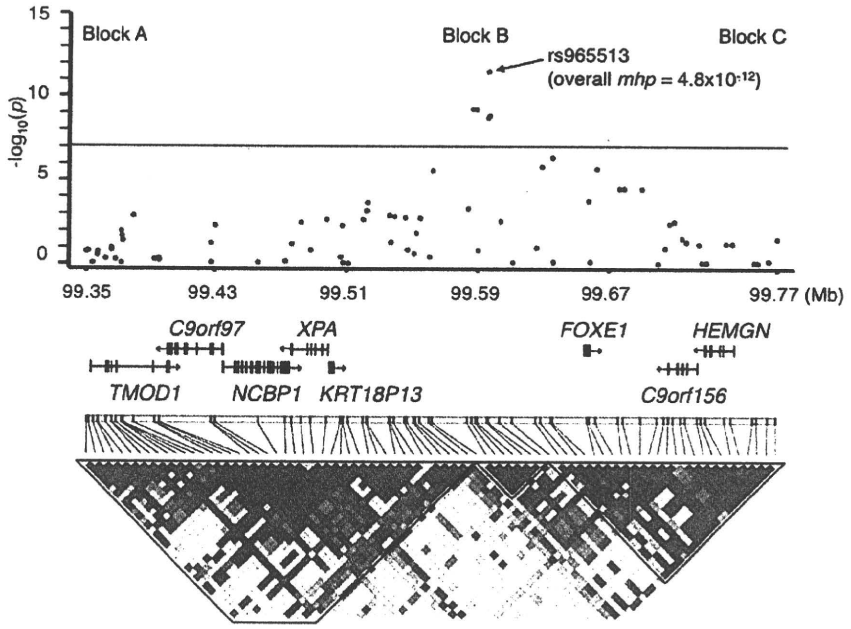


**Figure 2.** A schematic organization of the human *FOXE1* locus at 9q22.23 with the genome scanning results. Mhp-values calculated by the Trend $\chi^2$ test in $-\log_{10}$ scale were plotted in red circles for SNPs located in the 425-kb region between rs4742698 and rs4618817 at chromosome 9q22.23. The blue circle indicates mhp-value of rs965513 by meta-analysis using the combined results of Study 1 to Study 3. The structure and orientation of eight genes in the region were shown below the plots with their transcriptional orientations according to NCBI Reference Sequence Build 36.3. LD blocks were generated according to pairwise LD estimates of the SNPs located within the region using the genome scan results of Study 1 and Study 2.

For the early-onset group, there were 178 patients aged 3–25 years (mean age ± SD: 11.2 ± 4.3 years) who were diagnosed within the first 10 years (before 1997), with latency of 7.0 ± 1.9 years. For the late-onset group, there were 482 samples aged 10–39 years (22.9 ± 7.5 years) who were diagnosed after 1997, with latency of 16.4 ± 3.8 years. Looking at the results for rs965513, there was a much stronger association observed for the early-onset cases ($P = 2.0 \times 10^{-9}$, OR = 1.97, 95% CI: 1.58–2.47) than the late-onset cases ($P = 6.0 \times 10^{-8}$, OR = 1.52, 95% CI: 1.31–1.77) when compared with 1268 controls (CTR1 to CTR3). However, there was no statistical significance to prove the stronger impact of rs965513 on the early-onset of PTC ($p$-heterogeneity = 0.063).

## DISCUSSION

In this study, we have undertaken a GWA study of radiation-related PTC employing Belarusian patients and control subjects. We identified four markers in strong LD at chromosome 9q22.23 that were significantly associated with the disease. The strong association was further evident by selecting one of these markers, rs965513, with the genotyping of an independent set of samples by Taqman (overall mhp = $4.8 \times 10^{-12}$, OR = 1.65, 95% CI: 1.43–1.91). Rs965513 was recently identified as a genetic risk factor for sporadic PTC in individuals of European descent (11) and is located within an LD block which lies centromeric to *FOXE1*.

Another recent report showed a strong association of rs1867277 at the 5′-UTR of *FOXE1* with the risk of differentiated thyroid cancer, in particular with the classic variant of PTC. *FOXE1* is a thyroid-specific DNA binding protein recognizing binding sites on thyroglobulin and thyroperoxidase genes expressed in thyroid follicular cells (14,15). Although the precise role of *FOXE1* in PTC remains to be fully established, this study provides further evidence of *FOXE1* involvement in thyroid carcinogenesis. Rs1867277 is so far the only functional variant associated with sporadic PTC identified within the *FOXE1* gene, and the risk allele (A) augmented *FOXE1* transcription by creation of a binding site for USF1 and USF2 transcription factors. The fact that stronger association signals were observed for SNPs outside block C containing *FOXE1* in both the Icelandic and Belarusian studies may indicate the existence of DNA sequences in block B with unknown function acting cooperatively with rs1867277. Certainly, however, we cannot rule out the involvement of other genes in the region.

Although the association of rs965513 with PTC was stronger in the early-onset cases than in the late-onset cases, the difference was not statistically significant ($p$-heterogeneity = 0.063). Short latency was reported to be often associated with more aggressive tumors with prominent local invasion and distant metastases (16). However, it is difficult to directly associate our results to such morphological features since the environmental background of patients, including individual thyroid radiation dose and detailed clinical information are not available.

Individual susceptibility to thyroid cancer is considered to be complex involving the interaction of low-penetrance genes and the environment. Here we provide the first evidence

**Table 2.** Results of association analysis for SNP markers at 9q22.33 and 14q13.3 using the Chernobyl childhood thyroid cancer cohort

| Marker | Allele[a] | | Chr | Position | Statistics by study | | | | | Study 1 + 2 | | Study 1 + 2 + 3 | | Gudmundsson *et al.* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ref. | Var | | | Study | Freq var. Case | Control | Trend $P^b$ | OR (95% CI)[c] | Mhp[d] | OR (95% CI)[c] | Mhp[d] | OR (95% CI)[c] | Freq var. Case | Control | $P$-value[e] | OR (95% CI)[c] |
| rs1512261 | G* | T | 9 | 99562351 | 1 | 0.500 | 0.419 | 0.029 | 1.39 (1.03, 1.87) | $6.9 \times 10^{-6}$ | 1.53 (1.27, 1.84) | | | | | | |
| | | | | | 2 | 0.509 | 0.391 | $5.2 \times 10^{-5}$ | 1.62 (1.28, 2.04) | | | | | | | | |
| rs1877432 | G* | A | 9 | 99583701 | 1 | 0.706 | 0.610 | 0.010 | 1.53 (1.12, 2.09) | $5.2 \times 10^{-4}$ | 1.40 (1.16, 1.69) | | | | | | |
| | | | | | 2 | 0.697 | 0.628 | 0.016 | 1.36 (1.06, 1.74) | | | | | | | | |
| rs925489 | C* | T | 9 | 99586421 | 1 | 0.487 | 0.334 | $3.3 \times 10^{-5}$ | 1.89 (1.40, 2.55) | $1.7 \times 10^{-9}$ | 1.79 (1.48, 2.16) | | | | | | |
| | | | | | 2 | 0.481 | 0.349 | $6.0 \times 10^{-6}$ | 1.73 (1.37, 2.18) | | | | | | | | |
| rs7850258 | A* | G | 9 | 99588834 | 1 | 0.487 | 0.334 | $3.3 \times 10^{-5}$ | 1.89 (1.40, 2.55) | $1.7 \times 10^{-9}$ | 1.79 (1.48, 2.16) | | | | | | |
| | | | | | 2 | 0.481 | 0.349 | $6.0 \times 10^{-6}$ | 1.73 (1.37, 2.18) | | | | | | | | |
| rs965513 | A* | G | 9 | 99595930 | 1 | 0.487 | 0.334 | $3.3 \times 10^{-5}$ | 1.89 (1.40, 2.55) | $4.9 \times 10^{-9}$ | 1.76 (1.45, 2.12) | $4.8 \times 10^{-12}$ | 1.65 (1.43, 1.91) | 0.490 | 0.352 | $6.8 \times 10^{-20}$ | 1.77 (1.57, 2.00) |
| | | | | | 2 | 0.476 | 0.352 | $1.7 \times 10^{-5}$ | 1.68 (1.33, 2.12) | | | | | | | | |
| | | | | | 3 | 0.462 | 0.367 | $2.0 \times 10^{-4}$ | 1.48 (1.20,1.83) | | | | | | | | |
| rs10759944 | A* | G | 9 | 99596793 | 1 | 0.487 | 0.334 | $3.3 \times 10^{-5}$ | 1.89 (1.40, 2.55) | $3.5 \times 10^{-9}$ | 1.77 (1.46, 2.14) | | | 0.490 | 0.352 | $1.7 \times 10^{-19}$ | 1.77 (1.57, 2.01) |
| | | | | | 2 | 0.479 | 0.352 | $1.2 \times 10^{-5}$ | 1.69 (1.34, 2.14) | | | | | | | | |
| rs7848973 | A* | G | 9 | 99628660 | 1 | 0.503 | 0.392 | 0.0032 | 1.56 (1.16, 2.10) | $2.5 \times 10^{-6}$ | 1.56 (1.29, 1.87) | | | | | | |
| | | | | | 2 | 0.502 | 0.393 | $2.0 \times 10^{-4}$ | 1.56 (1.24, 1.97) | | | | | | | | |
| rs7024345 | A* | G | 9 | 99635059 | 1 | 0.380 | 0.305 | 0.038 | 1.39 (1.02, 1.90) | $1.4 \times 10^{-6}$ | 1.63 (1.33, 1.98) | | | 0.387 | 0.285 | $1.9 \times 10^{-12}$ | 1.58 (1.39, 1.80) |
| | | | | | 2 | 0.397 | 0.273 | $4.0 \times 10^{-6}$ | 1.75 (1.37, 2.23) | | | | | | | | |
| rs1443434 | G* | T | 9 | 99657300 | 1 | 0.487 | 0.392 | 0.012 | 1.47 (1.09, 1.97) | $2.6 \times 10^{-4}$ | 1.41 (1.17, 1.70) | | | 0.488 | 0.385 | $2.8 \times 10^{-9}$ | 1.52 (1.32, 1.74) |
| | | | | | 2 | 0.477 | 0.398 | 0.0070 | 1.37 (1.09, 1.73) | | | | | | | | |
| rs907580 | T* | C | 9 | 99662418 | 1 | 0.374 | 0.314 | 0.10 | 1.30 (0.96, 1.78) | $5.7 \times 10^{-6}$ | 1.58 (1.30, 1.92) | | | 0.395 | 0.281 | $1.1 \times 10^{-14}$ | 1.66 (1.46, 1.89) |
| | | | | | 2 | 0.396 | 0.273 | $4.2 \times 10^{-6}$ | 1.74 (1.37, 2.23) | | | | | | | | |
| rs925487 | C* | T | 9 | 99676219 | 1 | 0.463 | 0.372 | 0.015 | 1.45 (1.08, 1.96) | $4.5 \times 10^{-5}$ | 1.47 (1.22, 1.77) | | | 0.472 | 0.359 | $2.6 \times 10^{-13}$ | 1.60 (1.41, 1.81) |
| | | | | | 2 | 0.465 | 0.369 | $9.2 \times 10^{-4}$ | 1.48 (1.17, 1.88) | | | | | | | | |
| rs10984103 | A* | C | 9 | 99679096 | 1 | 0.463 | 0.372 | 0.015 | 1.45 (1.08, 1.96) | $4.6 \times 10^{-5}$ | 1.47 (1.22, 1.77) | | | 0.472 | 0.359 | $2.2 \times 10^{-13}$ | 1.59 (1.41, 1.81) |
| | | | | | 2 | 0.465 | 0.369 | $9.5 \times 10^{-4}$ | 1.48 (1.17, 1.87) | | | | | | | | |
| rs7866436 | G* | A | 9 | 99689917 | 1 | 0.465 | 0.369 | 0.010 | 1.49 (1.10, 2.00) | $5.2 \times 10^{-5}$ | 1.47 (1.22, 1.76) | | | | | | |
| | | | | | 2 | 0.463 | 0.372 | 0.0016 | 1.46 (1.15, 1.84) | | | | | | | | |
| rs944289 | C | T* | 14 | 35718997 | 1 | 0.626 | 0.580 | 0.23 | 1.21 (0.90, 1.63) | 0.17 | 1.13 (0.95, 1.36) | | | 0.644 | 0.558 | $2.5 \times 10^{-8}$ | 1.44 (1.26, 1.63) |
| | | | | | 2 | 0.607 | 0.584 | 0.43 | 1.10 (0.87, 1.40) | | | | | | | | |

SNP markers in blocks B and C with $P < 1 \times 10^{-3}$ are shown for 9q22.23. Rs944289 on 14q13.3 which also showed significant association in the Icelandic study was included. A complete list of the markers in the 425-kb region with statistical results is shown in Supplementary Material, Table S1.

[a]The reference (ref.) and variant (var.) alleles refer to NCBI Build 36.3 and the risk allele is indicated with an asterisk.

[b]The $P$-values using Trend $\chi^2$ test are shown.

[c]Odds ratio (OR) is calculated for the risk allele with a confidence interval (CI) of 95%.

[d]The Trend $\chi^2$ Mantel–Haenszel $P$-values are shown.

[e]The $P$-values using a standard likelihood ratio $\chi^2$ statistic are shown.

that the risk of developing PTC after internal radiation exposure is largely associated with the genetic determinant conferring risk for human thyroid malignancies in the general population. However, *FOXE1* is unlikely to be the only key player in radiation-related thyroid carcinogenesis and it remains to be established whether or not radiation-related PTC has other etiology-specific genetic components for inherited predisposition. Rs944289 at chromosome 14q13.3 strongly associated with sporadic PTC in the Icelandic population was not significant in our results. Moreover, in our GWA study, two additional SNPs with meta-analysis *P*-value being smaller than $1 \times 10^{-6}$ were identified, of which one was on chromosome 9p and the other on chromosome 12p. Since neither of these chromosomal loci have been identified as being associated with sporadic PTC, they may be potential candidates for susceptibility loci specific to radiation-related PTC. These observations clearly suggest that different genetic components are involved in carcinogenesis of sporadic and radiation-related PTC.

Only a few case–control studies to identify genetic risk factors of radiation-related thyroid cancer have been reported to date. Three studies included Chernobyl PTC (17–19) and thyroid cancers in an occupationally exposed cohort (20). A recent article examined the genetic determinants in the patients with radiation-related thyroid nodules (21). The possibilities of association between the risk for PTC after radiation exposure and *TP53* (ID: 7157) (17,18), *RET* (ID: 5979) (20) or *XRCC1* (ID: 7515) (20) were demonstrated. However, most of these studies had a limited sample size and insufficient gene coverage. Apart from the *TP53* Arg72Pro polymorphism (rs1042522) being associated with the risk of radiation-related PTC in adult patients (17,18), the findings were not replicated in independent sample sets. None of the SNP markers that were significant in the above studies were on the Illumina array. According to HAPMAP, rs25487 (*XRCC1*) and rs1800858 (*RET*) are in complete LD ($D' = 1$, $r^2 = 1$) with rs1799778 and rs2505535, respectively, which are both on the array. However, the associations were negative for both markers in our study ($P = 0.94$ for rs1799778 and $P = 0.03$ for rs2505535).

## MATERIALS AND METHODS

### Study populations

A total of 667 patients (174 males and 493 females, sex ratio 0.35) diagnosed for thyroid cancer in 1989–2009 were recruited. Inclusion criteria for cases were as follows: (i) age at the time of Chernobyl accident 0–18 years old, including those *in utero*, in April–June 1986, who were (ii) residing at the time in the radiocontaminated regions of Belarus and (iii) histologically verified diagnosis of PTC. Demographic and diagnostic information was retrieved from Thyroid Cancer Center (Minsk, Belarus). At the moment of exposure, 378 patients were residents of Gomel region of Belarus which is the most radiocontaminated area in the country, 195 patients were from Brest region, 10 from Mogilev region and 84 were from other radiocontaminated regions of the country.

As control subjects, a total number of 620 healthy individuals (165 males and 455 females, sex ratio 0.36) were recruited. Inclusion criteria for controls were: (i) age at the time of accident between 0 and 18 years old, including those *in utero*, in April–June 1986, who were (ii) residing at the time in the radiocontaminated regions of Belarus, (iii) euthyroid state and (iv) no thyroid cancer by the time of sampling (February 2006 to April 2009). At the time of possible radiation exposure, 574 healthy participants were residents of Brest region, 34 of Gomel region, 11 of Mogilev region and one individual from another region. According to the radioecological and radiation epidemiology studies, all cases and 620 controls are considered to have received thyroid doses ranging 21–1500 mGy (22,23). Additional DNA samples of 207 individuals who were: (i) born after 1987 (79 samples), (ii) older than 18 years of age at the time of accident (three samples) or (iii) considered to have been exposed to a negligible amount, if any, of radiation according to their residential information (125 samples), were also utilized for the studies as representative Belarusian population controls. Demographic and residential information was obtained by personal inquiry, and peripheral blood samples were collected in the contaminated regions during bi-annual thyroid screening programs (which also included neck ultrasound and consultation of endocrinologist) of Belarusian population. Euthyroid state was confirmed by laboratory tests being $1.64 \pm 1.57$ μU/ml for thyrotropin (normal range 0.5–5.0 μU/ml) and $1.17 \pm 0.28$ ng/dl for free thyroxin (normal range 0.7–1.55 ng/dl) in the whole control group. The absence of thyroid cancer was met by selecting only those individuals without detectable thyroid nodules on ultrasound. For Study 2, the genotypes of 448 Russian controls were used as population controls (24). The Institutional Review Board and the Ethics Committee of each institution approved the protocols used. All participants were fully informed of the purpose and procedures, and a written consent was obtained.

### DNA preparation

DNA was extracted from peripheral blood mononuclear cells using Puregene kit (Qiagen, Germantown, MD, USA) according to the manufacturer's protocol. DNA concentration and purity were measured with a Nanodrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). The samples were stored at −80°C until use.

### GWA study

Two series of genome scans were performed using two independent sample sets. 194 cases and 179 controls, and 214 cases and 448 Russian population controls, were used in the first and second genome scans (Study 1 and Study 2), respectively. Validation of genome scan results (Study 3) was performed by Taqman analysis using a third independent sample set consisting of 259 cases and 648 controls.

*Study 1: genome scan.* A total of 567 512 autosomal SNPs were genotyped in 194 thyroid cancer patients and 179 controls with Illumina Human610-Quad BeadChip on a BeadStation 500G Genotyping System, and genotype calls were generated and summary files were made using the Bead Studio version 3.1.3.0 software package (Illumina, Inc., San

Diego, CA, USA). Quality control procedures were systematically performed for the genome scan results. Initially, two control samples with call rates being smaller than 90% were removed from the analysis. Subsequently, degrees of kinship between individuals were examined by Pi-hat in PLINK, a multidimensional scaling method (25). For seven pairs of cases and five pairs of controls showing high degrees of kinship (PI-HAT > 0.3), the sample with the lower call rate was excluded. Principal component analysis by 'smartpca' in EIGENSOFT (26) including HAPMAP phase II samples confirmed no deviation in all DNA samples from Caucasian population. Following the quality control for SNP markers, a total of 35 488 markers were excluded due to low call rates (lower than 95%), a low minor allele frequency (smaller than 0.01) or significant distortion from Hardy–Weinberg equilibrium ($P$-value smaller than $10^{-7}$). After these steps, 532 024 SNP markers of 187 PTC patients (mean age ± SD: 3.0 ± 3.8 years) and 172 controls (1.5 ± 2.8 years) were used for statistical analyses.

Association of SNP markers on the X chromosome was examined in a separate analysis. The same criteria for QC were applied and 16 448 SNP markers were used to test disease association between 58 cases and 60 controls for males, and 128 cases and 111 controls for females.

*Study 2: genome scan.* In 214 thyroid cancer patients (mean age ± SD: 5.8 ± 5.2 years), 567 512 autosomal SNPs were genotyped using the same SNP arrays as those used in Study 1. Genotype calls of 448 Russian DNA samples were used as population-based controls. The same exclusion criteria as Study 1 were applied for the quality control, but no DNA samples were removed from the analysis. After removing 57 902 SNP markers that fit the exclusion criteria, a total of 509 610 SNP markers were used for statistical analyses. Analysis of the X chromosome was performed as described for Study 1, in 52 cases and 235 controls for males and in 161 cases and 213 controls for females.

*Study 3: validation analysis.* Validation of genome scan results was carried out in 259 cases (mean age ± SD: 6.8 ± 5.5 years) and 648 controls (mean age ± SD: 6.2 ± 5.9 years) using the Taqman SNP assays (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's guidelines. A predesigned and functionally tested probe was used for rs965513 (C_1593670_20, Applied Biosystems), and a custom designed probe by the same producer was used for rs1867277.

### Statistical analysis

A case–control association in each study was examined using trend $\chi^2$ test to compare genotypic distributions between cases and controls (27). Population stratification was assessed by the genomic control method (28). Meta-analysis of genome scan results was carried out with trend mode of the Mantel–Haenszel method (29), by combining the genotypes of Study 1 and Study 2 for 506 840 autosomal SNP markers that passed quality control in both studies. The genotypes for the autosomal SNPs obtained in Studies 1 and 2 were pooled, and population stratification was corrected by Eigenstrat (26) followed by the genomic control method. Meta-analysis of 16 448 SNP markers on the X chromosome was performed for Study 1 males and Study 2 males, as well as for Study 1 females and Study 2 females.

The overall significance level of rs965513 was calculated by meta-analysis using the Mantel–Haenszel method, combining the genotypes of Study 1 to Study 3. The LD structure was derived using the genotypes of Study 1 and Study 2 using the Haploview software (30) by calculating pairwise LD indices ($D'$ and $r^2$) between SNP markers in the region.

Imputation of missing genotypes was performed using MACH 1.0 (http://www.sph.umich.edu/csg/abecasis/MaCH/index.html). The genotype data of CEU (CEPH European) obtained from the Phase III HapMap database (draft2) were used as reference and the 425-kb region between rs4742698 and rs4618817 was examined for association. In the process of imputation, 50 Markov chain iterations were implemented.

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

## ACKNOWLEDGEMENTS

We would like to thank all the study participants who made this project possible. The Russian Control Data came from three studies: KMSU: A.V. Polonikov, V.P. Ivanov, M.A. Solodilova; TOMSK: M.B. Freidin, V.P. Puzyrev, L.M. Ogorodova; UFA: E.K. Khusnutdinova, A.S. Karunas, Y.Y. Fedorova. We also thank M. Kokubo, M. Palomares, M. Aksornworanart and M. Mizutani for technical assistance, and A. Yoshizumi, H. Uneme and K. Hirosawa for informatics management.

*Conflict of Interest statement.* None declared.

## FUNDING

This work was supported in part by Nagasaki University Global COE Program and by 'Grants-in-Aid for Young Scientists' from the Ministry of Education, Culture, Sports, Science and Technology (Japan). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

## REFERENCES

1. Kazakov, V.S., Demidchik, E.P. and Astakhova, L.N. (1992) Thyroid cancer after Chernobyl. *Nature*, **359**, 21.
2. Bennett, B., Repacholi, M. and Carr, Z. (2006) *Report of the UN Chernobyl Forum Expert Group 'Health'*, WHO Press, Geneva.
3. Cardis, E., Kesminiene, A., Ivanov, V., Malakhova, I., Shibata, Y., Khrouch, V., Drozdovitch, V., Maceika, E., Zvonova, I., Vlassov, O. *et al.* (2005) Risk of thyroid cancer after exposure to 131I in childhood. *J. Natl Cancer Inst.*, **97**, 724–732.
4. Tronko, M.D., Howe, G.R., Bogdanova, T.I., Bouville, A.C., Epstein, O.V., Brill, A.B., Likhtarev, I.A., Fink, D.J., Markov, V.V., Greenebaum, E. *et al.* (2006) A cohort study of thyroid cancer and other thyroid diseases after the chornobyl accident: thyroid cancer in Ukraine detected during first screening. *J. Natl Cancer Inst.*, **98**, 897–903.
5. Williams, D. (2002) Cancer after nuclear fallout: lessons from the Chernobyl accident. *Nat. Rev. Cancer*, **2**, 543–549.

6. Tronko, M.D., Brenner, A.V., Olijnyk, V.A., Robbins, J., Epstein, O.V., McConnell, R.J., Bogdanova, T.I., Fink, D.J., Likhtarev, I.A., Lubin, J.H. *et al.* (2006) Autoimmune thyroiditis and exposure to iodine 131 in the Ukrainian cohort study of thyroid cancer and other thyroid diseases after the Chornobyl accident: results from the first screening cycle (1998–2000). *J. Clin. Endocrinol. Metab.*, **91**, 4344–4351.
7. Ron, E., Lubin, J.H., Shore, R.E., Mabuchi, K., Modan, B., Pottern, L.M., Schneider, A.B., Tucker, M.A. and Boice, J.D. Jr (1995) Thyroid cancer after exposure to external radiation: a pooled analysis of seven studies. *Radiat. Res.*, **141**, 259–277.
8. Pacini, F., Vorontsova, T., Demidchik, E.P., Molinaro, E., Agate, L., Romei, C., Shavrova, E., Cherstvoy, E.D., Ivashkevitch, Y., Kuchinskaya, E. *et al.* (1997) Post-Chernobyl thyroid carcinoma in Belarus children and adolescents: comparison with naturally occurring thyroid carcinoma in Italy and France. *J. Clin. Endocrinol. Metab.*, **82**, 3563–3569.
9. Jacob, P., Bogdanova, T.I., Buglova, E., Chepurniy, M., Demidchik, Y., Gavrilin, Y., Kenigsberg, J., Kruk, J., Schotola, C., Shinkarev, S. *et al.* (2006) Thyroid cancer among Ukrainians and Belarusians who were children or adolescents at the time of the Chernobyl accident. *J. Radiol. Prot.*, **26**, 51–67.
10. Likhtarev, I., Bouville, A., Kovgan, L., Luckyanov, N., Voilleque, P. and Chepurny, M. (2006) Questionnaire- and measurement-based individual thyroid doses in Ukraine resulting from the Chornobyl nuclear reactor accident. *Radiat. Res.*, **166**, 271–286.
11. Gudmundsson, J., Sulem, P., Gudbjartsson, D.F., Jonasson, J.G., Sigurdsson, A., Bergthorsson, J.T., He, H., Blondal, T., Geller, F., Jakobsdottir, M. *et al.* (2009) Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. *Nat. Genet.*, **41**, 460–464.
12. Landa, I., Ruiz-Llorente, S., Montero-Conde, C., Inglada-Perez, L., Schiavi, F., Leskela, S., Pita, G., Milne, R., Maravall, J., Ramos, I. *et al.* (2009) The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. *PLoS Genet.*, **5**, e1000637.
13. UNSCEAR (1994) In. *United Nations Scientific Committee on the Effects of Atomic Radiation, 1994 Report to the General Assembly*, New York.
14. Zannini, M., Avantaggiato, V., Biffali, E., Arnone, M.I., Sato, K., Pischetola, M., Taylor, B.A., Phillips, S.J., Simeone, A. and Di Lauro, R. (1997) TTF-2, a new forkhead protein, shows a temporal expression in the developing thyroid which is consistent with a role in controlling the onset of differentiation. *EMBO J.*, **16**, 3185–3197.
15. Chadwick, B.P., Obermayr, F. and Frischauf, A.M. (1997) FKHL15, a new human member of the forkhead gene family located on chromosome 9q22. *Genomics*, **41**, 390–396.
16. Williams, E.D., Abrosimov, A., Bogdanova, T., Demidchik, E.P., Ito, M., LiVolsi, V., Lushnikov, E., Rosai, J., Sidorov, Y., Tronko, M.D. *et al.* (2004) Thyroid carcinoma after Chernobyl latent period, morphology and aggressiveness. *Br. J. Cancer*, **90**, 2219–2224.
17. Rogounovitch, T.I., Saenko, V.A., Ashizawa, K., Sedliarou, I.A., Namba, H., Abrosimov, A.Y., Lushnikov, E.F., Roumiantsev, P.O., Konova, M.V., Petoukhova, N.S. *et al.* (2006) TP53 codon 72 polymorphism in radiation-associated human papillary thyroid cancer. *Oncol. Rep.*, **15**, 949–956.
18. Akulevich, N.M., Saenko, V.A., Rogounovitch, T.I., Drozd, V.M., Lushnikov, E.F., Ivanov, V.K., Mitsutake, N., Kominami, R. and Yamashita, S. (2009) Polymorphisms of DNA damage response genes in radiation-related and sporadic papillary thyroid carcinoma. *Endocr. Relat. Cancer*, **16**, 491–503.
19. Stephens, L.A., Powell, N.G., Grubb, J., Jeremiah, S.J., Bethel, J.A., Demidchik, E.P., Bogdanova, T.I., Tronko, M.D. and Thomas, G.A. (2005) Investigation of loss of heterozygosity and SNP frequencies in the RET gene in papillary thyroid carcinoma. *Thyroid*, **15**, 100–104.
20. Lonn, S., Bhatti, P., Alexander, B.H., Pineda, M.A., Doody, M.M., Struewing, J.P. and Sigurdson, A.J. (2007) Papillary thyroid cancer and polymorphic variants in TSHR- and RET-related genes: a nested case–control study within a cohort of U.S. radiologic technologists. *Cancer Epidemiol. Biomarkers Prev.*, **16**, 174–177.
21. Sigurdson, A.J., Land, C.E., Bhatti, P., Pineda, M., Brenner, A., Carr, Z., Gusev, B.I., Zhumadilov, Z., Simon, S.L., Bouville, A. *et al.* (2009) Thyroid nodules, polymorphic variants in DNA repair and RET-related genes, and interaction with ionizing radiation exposure from nuclear tests in Kazakhstan. *Radiat. Res.*, **171**, 77–88.
22. Kruk, J.E., Prohl, G. and Kenigsberg, J.I. (2004) A radioecological model for thyroid dose reconstruction of the Belarus population following the Chernobyl accident. *Radiat. Environ. Biophys.*, **43**, 101–110.
23. Bouville, A., Likhtarev, I.A., Kovgan, L.N., Minenko, V.F., Shinkarev, S.M. and Drozdovitch, V.V. (2007) Radiation dosimetry for highly contaminated Belarusian, Russian and Ukrainian populations, and for less contaminated populations in Europe. *Health Phys.*, **93**, 487–501.
24. Heath, S.C., Gut, I.G., Brennan, P., McKay, J.D., Bencko, V., Fabianova, E., Foretova, L., Georges, M., Janout, V., Kabesch, M. *et al.* (2008) Investigation of the fine structure of European populations with applications to disease association studies. *Eur. J. Hum. Genet.*, **16**, 1413–1429.
25. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
26. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
27. Yamada, R. and Okada, Y. (2009) An optimal dose-effect mode trend test for SNP genotype tables. *Genet. Epidemiol.*, **33**, 114–127.
28. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.
29. Woolson, R.F. and Bean, J.A. (1982) Mantel–Haenszel statistics and direct standardization. *Stat. Med.*, **1**, 37–39.
30. 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.

Genetic Epidemiology 34:543–551 (2010)

# Estimation of *P*-value of MAX Test with Double Triangle Diagram for 2 × 3 SNP Case-Control Tables

**Katsura Hirosawa, Takahisa Kawaguchi, Fumihiko Matsuda, and Ryo Yamada***

*Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan*

Single nucleotide polymorphisms (SNPs) are the most popular markers in genetic epidemiology. Multiple tests have been applied to evaluate genetic effect of SNPs, such as Pearson's test with two degrees of freedom, three tests with one degree of freedom ($\chi^2$ tests for dominant and recessive modes and Cockran-Armitage trend test for additive mode) as well as MAX3 test and MAX test, which are combination of four tests mentioned earlier. Because MAX test is a combination of Pearson's test of two degrees of freedom and two tests of one degree of freedom, the probability density function (pdf) of MAX statistics does not match pdf of $\chi^2$ distribution of either one or two degrees of freedom. In order to calculate *P*-value of MAX test, we introduced a new diagram, Double Triangle Diagram, which was an extension of de Finetti diagram in population genetics which characterized all of the tests for 2 × 3 tables. In the diagram the contour lines of MAX statistics were consisted of elliptic curves and two tangent lines to the ellipses in the space. We normalized the ellipses into regular circles and expressed *P*-value of MAX test in an integral form. Although a part of the integral was not analytically solvable, it was calculable with arbitrary accuracy by dividing the area under pdf into finite rectangles. We confirmed that *P*-values from our method took uniform distribution from 0 to 1 in three example marginal count sets and concluded that our method was appropriate to give *P*-value of MAX test for 2 × 3 tables. *Genet. Epidemiol.* 34:543–551, 2010. 

**Key words: SNP; MAX test; association study; trend test**

*Correspondence to: Ryo Yamada, Yoshida-konoecho, Sakyo-ku, Kyoto 606-8501, Japan. E-mail: ryamada@genome.med.kyoto-u.ac.jp
Received 9 December 2009; Revised 16 March 2010; Accepted 29 March 2010
Published online 17 August 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/gepi.20510

## INTRODUCTION

Genetic epidemiology has been one of the most active research fields in genetics. Since the human genome project published the reference sequence of human beings, genome-wide case-control association studies (GWAS) have been carried out on a large scale with remarkable results. In GWAS, single nucleotide polymorphisms (SNP) have been used as principal genetic markers. In individuals, SNPs have two alleles, major (M) and minor (m), and three genotypes, MM, Mm and mm. Therefore, case-control studies in GWAS consist of 2 × 3 contingency tables for the two groups (case and control) and three genotypes. The technology of molecular genetics has been progressing very rapidly and SNPs are no longer the only genetic markers to be tested in GWAS studies [Balding, 2006]. However, the importance of 2 × 3 tables has not become obsolete, because any genetic factor in DNA can be evaluated with 2 × 3 tables in case-control studies.

For 2 × 3 tables, Pearson's test of two degrees of freedom can be applied. When three categories are in order, the Cockran-Armitage trend test (CAT) of one degree of freedom is the best choice. In many cases in genetics, it is reasonable to consider that the risk of the heterozygous type (Mm) is between the risks of two homozygous types (MM and mm). Therefore, CAT has been frequently used for analyzing the additive effect, which considers the middle category as the average of the other two categories. However, dominant and recessive effects are also well known in genetics and these effects are tested frequently with 2 × 2 tables in which the risk of the heterozygous type is considered the same as the risk of two homozygous types [Balding, 2006; Cochran, 1954]. Sometimes the MAX3 test is used, which consists of three tests (CAT and dominant and recessive tests) and adopts the maximum of the three as its statistical value. In fact, MAX3 test was used with successful identification of disease-associated markers in a genome-wide association study (GWAS) [Sladek et al., 2007]. Alternatively, the MAX test or the optimal mode trend test (OMTT test) can be used [Campbell, 2005]. The MAX method or the OMTT method tests all modes between dominant and recessive including the additive mode. The MAX and OMTT methods are conceptually the same and the OMTT offered exact calculation of *P*-value of the test. In fact, all the abovementioned tests are trend tests with different types of scores [Yamada and Okada, 2009; Zheng et al., 2009]. Because the MAX test best represents the genetic hypothesis in many situations and has the highest power among these tests under the hypothesis, it seems to be the best test for 2 × 3 table tests for SNP genetic studies. However, the *P*-value of the MAX test is not analytically calculable, which is a drawback of the test. Although we previously proposed a method to calculate the exact *P*-value [Yamada and Okada, 2009], it requires a high computational load.

A method to approximation of *P*-value of the MAX test was proposed by Li et al. with good performance [Li et al., 2008]. In this paper, we introduce a diagram to display a $2 \times 3$ table test in which the contour lines of the MAX test are drawn as a combination of an ellipse and its tangent lines, and we propose a method to estimate the *P*-value of the MAX test using the diagram.

# DOUBLE TRIANGLE DIAGRAM: A GEOMETRIC LAYOUT OF $2 \times 3$ TABLES IN TWO-DIMENSIONAL SPACE

## DOUBLE TRIANGLE DIAGRAM AS AN EXTENSION OF THE DE FINETTI DIAGRAM

A de Finetti diagram is a ternary plot to graph the genotype frequencies of populations, where there are two alleles and the population is diploid. The diagram locates the conditions of genotype frequencies in an equilateral triangle. It is based on Viviani's theorem that at any point within the triangle, given the three lines from that point that are perpendicular to the sides of the triangle, the sum of the lengths of the lines is a fixed value, regardless of the position of the point [Cannings and Edwards, 1968].

Because the marginal counts of a $2 \times 3$ table are given, the sum of three genotypes is fixed for both cases and controls. Therefore, three genotype counts of each group can be plotted as a point in an equilateral triangle. In our double triangle diagram, two triangles for cases and controls are drawn and $2 \times 3$ tables are plotted as described below.

Let $\tau = \{\{n_{11}, n_{12}, n_{13}\}, \{n_{21}, n_{22}, n_{23}\}\}$ denote the observed table and its cell counts and $m = \{n_{1\cdot}, n_{2\cdot}, n_{\cdot 1}, n_{\cdot 2}, n_{\cdot 3}, n_{\cdot\cdot}\}$ denote its marginal counts. Let $t_e = \{\{e_{11}, e_{12}, e_{13}\}, \{e_{21}, e_{22}, e_{23}\}\}$ denote the expected table and its counts under the null assumption and $\tau - t_e = \{\{d_{11}, d_{12}, d_{13}\}, \{d_{21}, d_{22}, d_{23}\}\}$ denote the difference between $\tau$ and $t_e$. Then, we have the following tables.

| | *AA* | *Aa* | *aa* | *Sum* |
|---|---|---|---|---|
| Case | $n_{11}$ | $n_{12}$ | $n_{13}$ | $n_{1\cdot}$ |
| Control | $n_{21}$ | $n_{22}$ | $n_{23}$ | $n_{2\cdot}$ |
| Total | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $n_{\cdot 3}$ | $n_{\cdot\cdot}$ |

$n_{ij} \geq 0$

The difference between $\tau$ and $t_e$ is as follows:

| | *AA* | *Aa* | *aa* | Sum |
|---|---|---|---|---|
| Case | $e_{11}$ | $e_{12}$ | $e_{13}$ | $n_{1\cdot}$ |
| Control | $e_{21}$ | $e_{22}$ | $e_{23}$ | $n_{2\cdot}$ |
| Total | $n_{\cdot 1}$ | $n_{\cdot 2}$ | $n_{\cdot 3}$ | $n_{\cdot\cdot}$ |

$e_{ij} = n_{i\cdot} n_{\cdot j} / n_{\cdot\cdot}$.

| | *AA* | *Aa* | *aa* |
|---|---|---|---|
| Case | $d_{11}$ | $d_{12}$ | $d_{13} = -(d_{11} + d_{12})$ |
| Control | $d_{21} = -d_{11}$ | $d_{22} = -d_{12}$ | $d_{23} = -d_{13} = d_{11} + d_{12}$ |

$d_{ij} = n_{ij} - e_{ij} \geq -e_{ij}$

$\sum_{i=1}^{2} d_{ij} = 0, \sum_{j=1}^{3} d_{ij} = 0.$

We introduce following coordinates:

$$(x, y) = \left( d_{12}, \frac{1}{\sqrt{3}}(d_{11} - d_{13}) \right). \quad (1)$$

These co-ordinates are based on the idea that geometric arrangement of multiple categories as below. Two categories are expressed as two vectors in opposite directions to each other. They are in one-dimensional space. Three categories are expressed as three vectors in two-dimensional space and their three tips are the vertices of a regular triangle. Four categories make a regular tetrahedron. In general, $k$ categories make $k-1$ simplex in $k-1$ dimensional space, which has $k$ vertices. Equation (1) is one way to give Cartesian coordinates to vertices of regular triangle (2 simplex for 3 categories). With these co-ordinates,

| | *AA* | *Aa* | *aa* |
|---|---|---|---|
| Case | $d_{11} = -\frac{1}{2}x + \frac{\sqrt{3}}{2}y$ | $d_{12} = x$ | $d_{13} = -\frac{1}{2}x - \frac{\sqrt{3}}{2}y$ |
| Control | $d_{21} = -d_{11} = \frac{1}{2}x - \frac{\sqrt{3}}{2}y$ | $d_{22} = -d_{12} = -x$ | $d_{23} = -d_{13} = \frac{1}{2}x + \frac{\sqrt{3}}{2}y$ |

Because $n_{ij} \geq 0$ and because $d_{1j} + d_{2j} = 0$ and $d_{1j} \geq -e_{1j}$ and $d_{2j} \geq -e_{2j}$, therefore $d_{1j} \geq -e_{1j}$ and $-d_{2j} = d_{1j} \leq -e_{2j}$,

$$-e_{1j} \leq d_{1j} \leq e_{2j},$$

which can be re-written as,

$$\frac{1}{\sqrt{3}}x - \frac{2}{\sqrt{3}}e_{11} \leq y \leq \frac{1}{\sqrt{3}}x + \frac{2}{\sqrt{3}}e_{21}$$

$$-e_{12} \leq x \leq e_{22}$$

$$-\frac{1}{\sqrt{3}}x - \frac{2}{\sqrt{3}}e_{23} \leq y \leq -\frac{1}{\sqrt{3}}x + \frac{2}{\sqrt{3}}e_{13}$$

These three equations demarcate the field with three sets of parallel lines which make two equilateral triangles.

Figure 1A is the diagram of the following table, in which the case-to-control ratio is 1.5 and the total allele frequency is 0.2 and the samples are in Hardy-Weinberg equilibrium. The size of triangles is proportional to the sample size of groups and the larger and the smaller triangles represent controls and cases, respectively.

| | *AA* | *Aa* | *aa* | *Sum* |
|---|---|---|---|---|
| Case | $n_{11} = 150$ | $n_{12} = 520$ | $n_{13} = 330$ | $n_{1\cdot} = 1{,}000$ |
| Control | $n_{21} = 250$ | $n_{22} = 680$ | $n_{23} = 570$ | $n_{2\cdot} = 1{,}500$ |
| Total | $n_{\cdot 1} = 400$ | $n_{\cdot 2} = 1{,}200$ | $n_{\cdot 3} = 900$ | $n_{\cdot\cdot} = 2{,}500$ |

# DISTRIBUTION OF TEST STATISTICS IN DOUBLE-TRIANGLE DIAGRAM

## CONTOUR LINES OF TEST STATISTICS

For $2 \times 3$ contingency tables in SNP case-control association studies, the above-mentioned multiple tests, Pearson's genotype test of two degrees of freedom, the three tests of one degree of freedom for additive, dominant and recessive mode, the MAX test [Zheng et al., 2009]

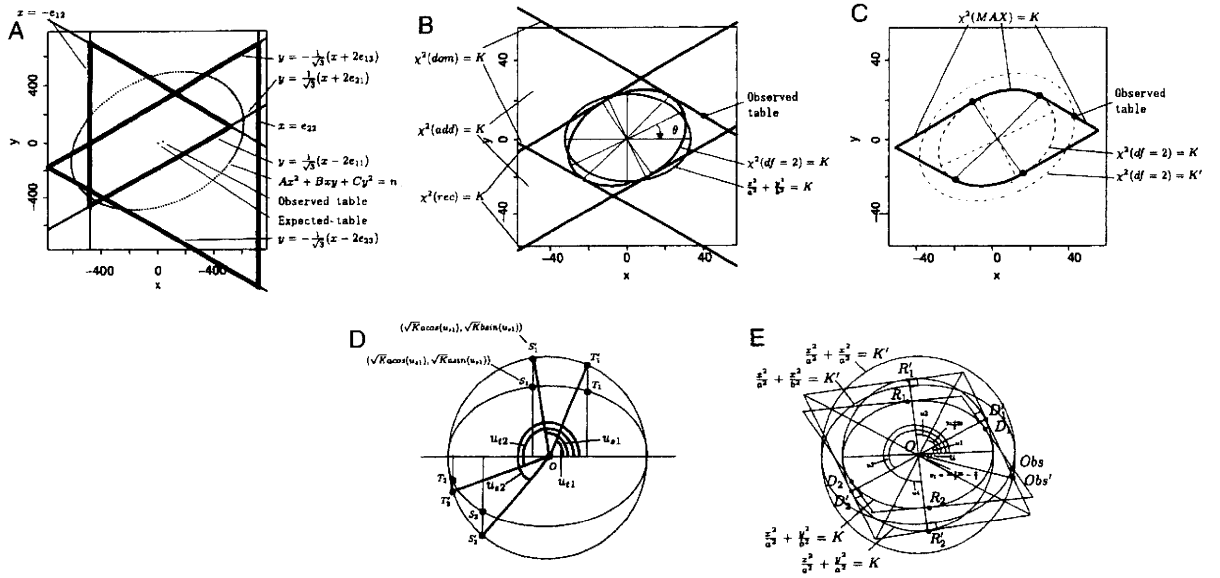


**Fig. 1. Double triangle diagrams and contour lines of statistical tests. (A) The space of the tables sharing the marginals. The origin (0,0) represents the expected table and the neighboring point is the table given in the text. Six lines demarcate the space where tables exist, which share the marginals with the observed table. The overlap of two triangles is a pentagon. The dashed ellipse is the contour curve of $\chi^2$(df2) = $n$. (B) The central area of (A) is enlarged. The solid ellipse of $\chi^2$(df2) = $K$ with its major and minor axis is shown. The solid ellipse is rotated by θ to the dashed ellipse, which is in the normalized form, $(x^2/a^2)+(y^2/b^2) = K$. Three lines, $\chi^2$(dom) = $K$, $\chi^2$(add) = $K$ and $\chi^2$(rec) = $K$ are tangent to $\chi^2$(df2) = $K$. The gradients of the lines are $-(1/\sqrt{3})$, 0 and $1/\sqrt{3}$, respectively. The observed table is indicated by a dot on $\chi^2$(dom) = $K$ and outside of the ellipse, $\chi^2$(df = 2) = $K$. (C) The contour line of $\chi^2$(MAX) = $K$ consists of $\chi^2$(df = $K$), $\chi^2$(dom) = $K$ and $\chi^2$(rec) = $K$. In the observed table, $\chi^2$(df2) = $K'$ is more than $K$. The contour line of the ellipse, $\chi^2$(df2) = $K'$, is indicated by a larger dashed ellipse. Because $\chi^2$(dom) and $\chi^2$(df2) of the observed table are $K$ and $K'$, respectively, the dot of the observed table is the intersection of $\chi^2$(dom) = $K$ and $\chi^2$(df2) = $K'$. Description on $K'$ appears in the section "Geometric evaluation of $\chi^2$(MAX). Therefore ignore $K'$ when this figure was indicated earlier in the main text. (D) The ellipse $(x^2/a^2)+(y^2/b^2) = K$ is enlarged by $\frac{a}{b}$ in the $y$-axis direction and it becomes a circle, $(x^2/a^2)+(y^2/b^2) = K$. The coordinates of $S_i$, $S_i'$, $T_i$ and $T_i'$ are parameterized with $a$, $b$, $K$, $K'$ and $u_i$. The ratio of the area of the sector in the ellipse and the corresponding sector in the circle is $b/a$. In the figure, $u_{s1} - u_{t1} = u_{s2} - u_{t2}$. Therefore, the areas of the two sectors, $S_1OT_1$ and $S_2T_2$, are the same. (E) The ellipses, $(x^2/a^2)+(y^2/b^2) = K$ and $(x^2/a^2)+(y^2/b^2) = K'$, and the corresponding circles, enlarged by $a/b$ in the $y$-axis direction, are drawn in the left and right sides, respectively.. Four tangent points on the ellipse $(x^2/a^2)+(y^2/b^2) = K$, $D_1$, $D_2$, $R_1$ and $R_2$ and their corresponding points on the circle, $D_1'$, $D_2'$, $R_1'$ and $R_2'$, are plotted. The $D$s and $R$s are on the lines of $\chi^2$(dom) = $K$ and $\chi^2$(rec) = $K$, respectively. Four tangent lines to the ellipse form a parallelogram. The enlargement in the y-axis moves the lines tangent to the ellipse to the lines tangent to the corresponding circle. The tangent lines to the circle form a rhombus and the radii to $D$'s and $R$'s, and the tangent lines are perpendicular. Two diagonals quadrisect the rhombus. Obs and Obs' are the point of the observed table and its corresponding point in the enlarged circle, respectively. Because the observed table's $\chi^2$(MAX) is $K$ and its $\chi^2$(df = 2) is $K'$, Obs is the intersection of the straight part of the contour line of $\chi^2$(MAX) = $K$ and the ellipse $(x^2/a^2)+(y^2/b^2) = K'$. Lines Obs'$D_1$ and OD$_1$ are perpendicular and angle Obs'$D_1 = u_1 - u$. The length is $D_1O = \sqrt{K}a$ and the distance is $Obs'O = \sqrt{K'}a$. Therefore, $\sqrt{K'}a/\sqrt{K}a = \cos(u_1 - u)$.**

[or the OMTT test; Yamada and Okada, 2009], can be applied. As reported, these tests are all trend tests with different types of scores, as expressed below.

$$\chi^2(\mathrm{df2}) = \max(Y^2(\{0, r, 1\}); -\infty \le r \le \infty)$$
$$\chi^2(\mathrm{dom}) = Y^2(\{0, 0, 1\})$$
$$\chi^2(\mathrm{rec}) = Y^2(\{0, 1, 1\})$$
$$\chi^2(\mathrm{add}) = Y^2(\{0, 0.5, 1\})$$
$$\chi^2(\mathrm{MAX}) = \max(Y^2(\{0, r, 1\}); 0 \le r \le 1)$$

where

$$Y^2(\{0, r, 1\}) = \frac{n_{..}^2}{n_{1.}n_{2.}} \frac{(d_{11}(-W)+d_{12}(r-W)+d_{13}(1-W))^2}{n_{.1}(-W)^2+(n_{.2}(r-W)^2+n_{.3}(1-W))^2}$$
$$W = \frac{n_{.2} \times r + n_{.3}}{n_{..}}$$

Excluding $\chi^2$(MAX), they are expressed with $x$ and $y$ as shown below.

$$\chi^2(\mathrm{df2}) = \frac{n_{..}^2}{n_{1.}n_{2.}}\left(\frac{1}{4}\left(\frac{1}{n_{.1}}+\frac{1}{n_{.3}}\right)+\frac{1}{n_{.2}}\right) x^2+\frac{\sqrt{3}}{2}\left(\frac{1}{n_{.3}}-\frac{1}{n_{.1}}\right)xy+\frac{3}{4}\left(\frac{1}{n_{.1}}+\frac{1}{n_{.3}}\right)y^2)$$
$$\chi^2(\mathrm{dom}) = \frac{n_{..}^2}{n_{1.}n_{2.}}\left(\frac{1}{n_{.1}+n_{.2}}+\frac{1}{n_{.3}}\right)\left(-\frac{1}{2}x-\frac{\sqrt{3}}{2}y\right)^2 \quad (2)$$
$$\chi^2(\mathrm{rec}) = \frac{n_{..}^2}{n_{1.}n_{2.}}\left(\frac{1}{n_{.1}}+\frac{1}{n_{.2}+n_{.3}}\right)\left(-\frac{1}{2}x+\frac{\sqrt{3}}{2}y\right)^2$$
$$\chi^2(\mathrm{add}) = \frac{n_{..}^2}{n_{1.}n_{2.}}\frac{n_{.1}n_{.2}n_{.3}}{\frac{1}{4}\left(\frac{1}{n_{.1}}+\frac{1}{n_{.3}}\right)+\frac{1}{n_{.2}}}\left(\frac{\sqrt{3}}{2}y\right)^2$$

Let $K$ denote $\chi^2$(MAX) of the observed table. Figure 1B indicates the contour lines of $\chi^2$(df2) = $K$, $\chi^2$(dom) = $K$, $\chi^2$(rec) = $K$ and $\chi^2$(add) = $K$. Figure 1C indicates the contour lines of $\chi^2$(MAX) = $K$. The contour line of $\chi^2$(df2) is an ellipse and the contour lines of $\chi^2$(dom), $\chi^2$(rec) and $\chi^2$(add) are pairs of parallel lines. The lines of $\chi^2$(dom) and $\chi^2$(rec) are parallel to the lines of the triangles. The lines of $\chi^2$(add) are horizontal. The contour lines of $\chi^2$(MAX) consist of the elliptic curve of $\chi^2$(df2) and the straight lines of $\chi^2$(dom) and $\chi^2$(rec).

The contour lines of $\chi^2$(dom) = $K$, $\chi^2$(rec) = $K$ and $\chi^2$(add) = $K$ are tangent to the ellipse, which can be shown by the simple transformation of equations (not shown).

## ELLIPSE NORMALIZATION

In general, the ellipse can be normalized by rotation.

The ellipse of $\chi^2(\mathrm{df2}) = K$ is normalized to $(x^2/a^2)+(y^2/b^2) = K$, $(a \geq b)$ by rotating $\theta$ in the clockwise direction as shown below (Fig. 1B).

$$\begin{aligned}
&\frac{x^2}{a^2}+\frac{y^2}{b^2} = K\\
&a = \sqrt{\frac{2}{A+C-\sqrt{B^2+(A-C)^2}}}\\
&b = \sqrt{\frac{2}{A+C+\sqrt{B^2+(A-C)^2}}}\\
&\theta = \frac{1}{2}\sin^{-1}\frac{B}{A-C}\\
&A = \frac{n^2}{n_{1\cdot}n_{2\cdot}}\left(\frac{1}{4}\left(\frac{1}{n_{\cdot1}}+\frac{1}{n_{\cdot3}}\right)+\frac{1}{n_{\cdot2}}\right)\\
&B = \frac{n^2}{n_{1\cdot}n_{2\cdot}}\frac{\sqrt{3}}{2}\left(\frac{1}{n_{\cdot3}}-\frac{1}{n_{\cdot1}}\right)\\
&C = \frac{n^2}{n_{1\cdot}n_{2\cdot}}\frac{3}{4}\left(\frac{1}{n_{\cdot1}}+\frac{1}{n_{\cdot3}}\right)
\end{aligned} \tag{3}$$

When the ellipse is normalized, the coordinates of the points on the ellipse are given as $(\sqrt{K}a \times \cos(u), \sqrt{K}b \times \sin(u))$.

We enlarge the figure by $a/b$ in the $y$-axis direction, and the ellipse becomes a regular circle. The coordinates of the points change from $(\sqrt{K}a \times \cos(u), \sqrt{K}b \times \sin(u))$ to $(\sqrt{K}a \times \cos(u), \sqrt{K}a \times \sin(u))$ (Fig. 1D).

The tangent lines to the ellipse change their gradients but remain tangent to the circle (Fig. 1E).

# GEOMETRIC EVALUATION OF $X^2$(MAX)

The contour line of $\chi^2$(MAX) = $K$ consists of the tangent lines of the dominant and recessive models and the elliptic curve (see Fig. 1E). The four tangent points, $R_1$, $R_2$, $D_1$ and $D_2$ are on the ellipse and their locations are given as $(\sqrt{K}a\cos(u_i), \sqrt{K}b\sin(u_i))$. The observed table is on the larger ellipse and its location is given as $(\sqrt{K'}a\cos(u), \sqrt{K'}b\sin(u))$. The observed table is indicated as Obs at the intersection of the ellipse and the tangent line from one of the four tangent points, $D_1$. After enlargement in the $y$-axis direction, Obs and $D_1$ are moved to Obs' and $D_1$, respectively. Because the radius and tangent line of a regular circle are perpendicular, the line Obs'$-D_1$ is perpendicular to the radius to $D_1$. Therefore,

$$\frac{\sqrt{K}a}{\sqrt{K'}a} = \cos(u_i - u),$$

and,

$$K'(u|u_i) = \frac{K}{(\cos(u_i-u))^2}.$$

We will use $K'(u)$ instead of $K'(u \mid u_i)$ for simplicity.

The $\chi^2$(df2) values of points on $\chi^2$(MAX) = $K$, $K'(u)$ are

$$K'(u) = \begin{cases} K & (u_1 \leq u \leq u_2, u_3 \leq u \leq u_4) \\ \min\left(\dfrac{K}{(\cos(u_i-u))^2}\right) & \text{otherwise} \end{cases} \tag{4}$$

# GEOMETRIC CALCULATION OF *P*-VALUE OF MAX TEST

The area of the ellipse, $(x^2/a^2)+(y^2/b^2) = K$, is $A = \pi Kab$ (see Fig. 1D).

Assume two sectors of the ellipse, $S_1OT_1$ and $S_2OT_2$ $(S_i = (\sqrt{K}a\cos(u_{si}), \sqrt{K}b\sin(u_{si}))$ and $T_i = (\sqrt{K}a\cos(u_{ti}), \sqrt{K}b\sin(u_{ti})))$. When $u_{s1} - u_{t1} = u_{s2} - u_{t2}$, the area of the sectors, $((u_{s1} - u_{t1})/2\pi)A$ and $((u_{s2} - u_{t2})/2)A$, are equal.

Let $\Pr(x,y)$ denote the probability density function (pdf) of $\chi^2(\mathrm{df} = 2)$ in two-dimensional space. The $P$-value of $\chi^2(\mathrm{df} = 2) = K$, $P_{\mathrm{df2}}(K) = e^{-\frac{K}{2}}$ is given as,

$$P_{\mathrm{df2}}(K) = \int_{\frac{x^2}{a^2}+\frac{y^2}{b^2} \geq K} \Pr(x,y)\,\mathrm{d}x\mathrm{d}y. \tag{5}$$

Because $\Pr(x_i,y_i) = \Pr(x_j,y_j)$, when

$$\frac{x_i^2}{a^2}+\frac{y_i^2}{b^2} = \frac{x_j^2}{a^2}+\frac{y_j^2}{b^2}$$

the integral of $\Pr(x,y)$ in the sector $S_1OT_1$ is,

$$\begin{aligned}
&\int_{\frac{x^2}{a^2}+\frac{y^2}{b^2} \geq K, u_{t1} \leq u \leq u_{s1}} \Pr(x,y)\frac{u_{s1}-u_{t1}}{2\pi}\,\mathrm{d}x\mathrm{d}y\\
&= P_{\mathrm{df2}}(K)\frac{u_{s1}-u_{t1}}{2\pi}.
\end{aligned} \tag{6}$$

This can be applied to $\mathrm{d}u$ as,

$$\int_{\frac{x^2}{a^2}+\frac{y^2}{b^2} \geq K(u_t), u_t \leq u \leq u_t+\mathrm{d}u} \Pr(x,y)\frac{\mathrm{d}u}{2\pi}\mathrm{d}x\mathrm{d}y = P_{\mathrm{df2}}(K(u_t))\frac{\mathrm{d}u}{2\pi}. \tag{7}$$

For the points on the contour line $\chi^2$(MAX) = $K$, $\chi^2(\mathrm{df} = 2)$ is not constant because it is a function of $u$, $K'(u)$ (Equation (4)). Figure 2A,B shows $K'(u)$ and $P_{\mathrm{df2}}(K(u)) = e^{-\frac{K(u)}{2}}$.

The $P$-value of the MAX test when $\chi^2$(MAX) = $K$, $P_{\mathrm{MAX}}(K)$ is geometrically defined as,

$$P_{\mathrm{MAX}}(K) = \int_0^{2\pi} P(K'(u))\frac{\mathrm{d}u}{2\pi}. \tag{8}$$

Because the four tangent lines form a rhombus in the enlarged coordinates where the ellipse is the regular circle,

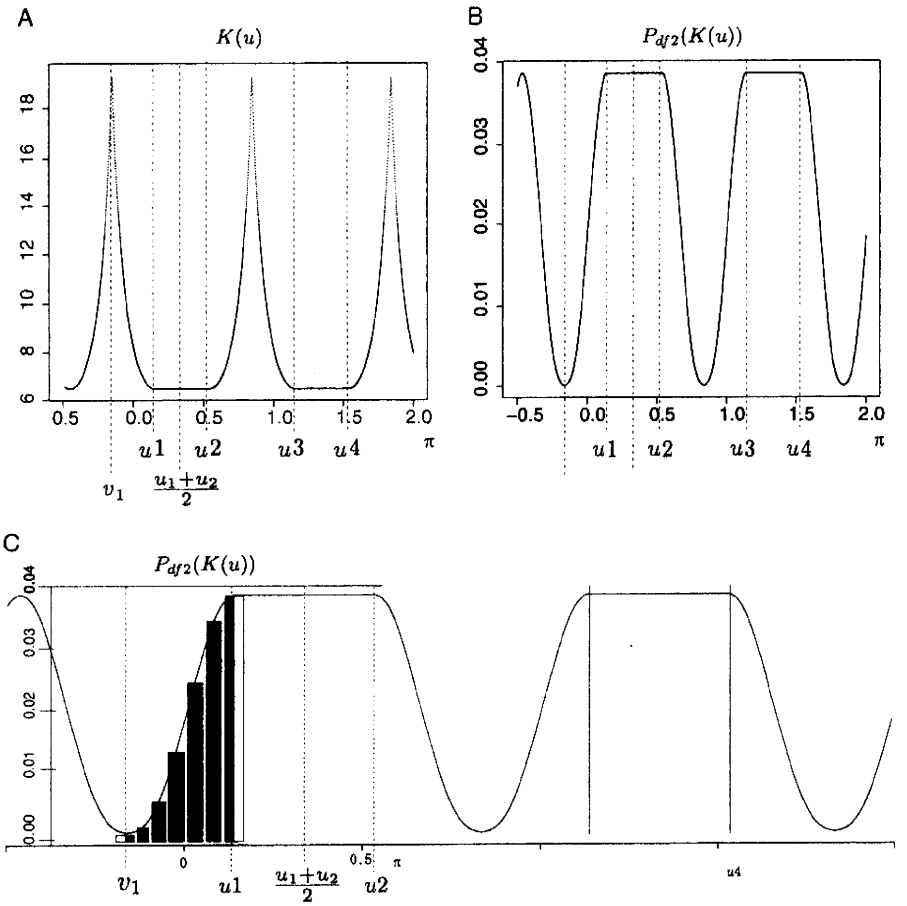


**Fig. 2. Plots of $\chi^2$(df = 2) and $P_{df2}(K)$ on the contour line of $\chi^2$(MAX) = $K$ to the circular angles. (A) $K'(u)$, $\chi^2$(df2) values of points on the contour line $\chi^2$(MAX) = $K$ = 6.51 are plotted along the axis of $\frac{u}{\pi}$. The line is cyclic. Between $u_1$ and $u_2$ and between $u_3$ and $u_4$, the line is flat at 6.51; otherwise, $K'(u)$ is a curve larger than the value of the flat segments. (B) $P_{df2}(K(u))$ were plotted on the same horizontal axis. The flat segments correspond to the flat segments in (A). Otherwise, the line is a curve smaller than the value of the flat segments. (C) A quarter from $v_1 = \frac{u_1+u_2}{2} - \frac{\pi}{2}$ to $\frac{u_1+u_2}{2}$ of (B) is enlarged. The black bars represent the area of equation (9) for estimating $P_{MAX}(K)$. The height of the bars is $P_{MAX}(K(u)$ of the midpoint. The width of the right- and left-most black bars are half of the others.**

and because a rhombus consists of four congruent regular triangles, a quarter of the rhombus can be considered. The first quarter of the rhombus of $\chi^2(\mathrm{MAX}) = K$ is parameterized with $u$ from $v_1 = ((u_1+u_2)/2) - \pi/2$ to $v_1+(\pi/2) = (u_1+u_2)/2$ (Fig. 1E). Therefore,

$$P_{\mathrm{MAX}}(K) = 4\int_{v_1}^{v_1+\frac{\pi}{2}} P_{\mathrm{df2}}(K'(u))\frac{\mathrm{d}u}{2\pi}$$
$$= 4\int_{v_1}^{u_1} P_{\mathrm{df2}}\left(\frac{K}{(\cos(u_1-u))^2}\right)\frac{\mathrm{d}u}{2\pi} + \int_{u_1}^{v_1+4\frac{\pi}{2}} P_{\mathrm{df2}}(K)\frac{\mathrm{d}u}{2\pi}.$$

The first term of the right side is

$$4\int_{v_1}^{u_1} P_{\mathrm{df2}}\left(\frac{K}{(\cos(u_1-u))^2}\right)\frac{\mathrm{d}u}{2\pi} = \frac{2}{\pi}\int_{v_1}^{u_1} \mathrm{e}^{-\frac{K}{2(\cos(u_1-u))^2}}\mathrm{d}u.$$

This integral cannot be analytically solved. The second term of the right side is

$$4\int_{u_1}^{v_1+\frac{\pi}{2}} P_{\mathrm{df2}}(K)\frac{\mathrm{d}u}{2\pi} = \frac{2}{\pi}P(K)\left(v_1+\frac{\pi}{2}-u_1\right) = \frac{1}{\pi}P(K)(u_2-u_1).$$

Therefore,

$$P(\chi^2(\mathrm{MAX}) = K) = \frac{1}{\pi}\left(2\int_{v_1}^{u_1} \mathrm{e}^{-\frac{K}{2(\cos(u_1-u))^2}}\frac{\mathrm{d}u}{2\pi} + \mathrm{e}^{-\frac{K}{2}}(u_2-u_1)\right).$$

Although the first term cannot be analytically solved, it can be calculated by summing the thin rectangles, as expressed below (Fig. 2C):

$$\frac{u_1-t_1}{2\pi N}\left(P_{\mathrm{df2}}(K'(u_1-t_1)) + P_{\mathrm{df2}}(K'(u_1)) \times \frac{1}{2} + \sum_{i=1}^{N-1} P_{\mathrm{df2}}\left(K'\left(u_1-\left(1-\frac{i}{N}\right)t\right)\right)\right). \quad (9)$$

The precision can be adjusted by changing $N$. In the following calculation, the first $N$ started with 2 and it was repeatedly doubled until the difference of the estimated $P_{MAX}(K)$ by each update of $N$ was less than $P_{df2}(K) \times 10^{-3}$. For example, when $P_{df2}(K) \times 10^{-4}$, calculation is continued until the difference between iterations $\chi$becomes less than $10^{-4-3} = 10^{-7}$.

## VALIDATION OF THE METHOD TO CALCULATE $P(\chi^2(MAX) = K)$

Previous reports have confirmed that $\chi^2(MAX)$ ranks the observed tables appropriately under the condition that $r$ is in the hypothesized range [Yamada and Okada, 2009; Zheng et al., 2009]. Because $P$-values need to be observed in uniform distribution from 0 to 1 when tests are repeated under the null hypothesis, the appropriateness of the method to estimate $P_{MAX}(K)$ proposed above is confirmed by observing that the distribution of $P(\chi^2(MAX))$ is uniform for multiple tables sampled under the null hypothesis. As Equation (4) indicates, $\chi^2(df2)$ is a function of $u_i$, which determines the location of the tangent points. Therefore, the eccentricity of the ellipse, $\sqrt{1-(b/a)^2}$, and the angle $\theta$ affect the estimation of $P(\chi^2(MAX)) = K$. Both $b/a$ and $\theta$ are parameterized only by $n_j$ based on Equation (3) (details are not shown). So, we selected three marginal count sets as examples $(n_{\cdot 1}, n_{\cdot 2}, n_{\cdot 3}) = (3333, 3333, 3334), (100, 9000, 900)$ and $(9000, 100, 900)$, with $(n_{1\cdot}, n_{2\cdot}) = (5000, 5000)$, for evaluation of the method to estimate $P_{MAX}(K)$. Figure 3A shows the double triangle diagrams of the three examples. First, (3333,3333,3334) has similar values for the three categories and its diagram gives an ellipse that is almost a regular circle and the fraction where $\chi^2(MAX) = \chi(df = 2)$ is approximately one-third. In the second example, (100,9000,900) has a very large value of $n_{\cdot 2}$ and its diagram gives an ellipse that is long in the vertical axis and the fraction where $\chi^2(MAX) = \chi(df = 2)$ is very small. In the third example, (9000,100,900) has a very large value of $n_{\cdot 1}$ instead and its diagram gives an ellipse that is long in the horizontal axis and the fraction where $\chi^2(MAX) = \chi(df = 2)$ is very large. A total of 1,000 tables that had the marginal counts were randomly sampled for each example and $P$-$P$ plots of $P_{MAX}(K)$ and $P_{df2}(K)$ were drawn on a linear scale and logarithmic scale (Fig. 3A). The $P$-$P$ plots for $P_{MAX}$ and $P_{df2}$ indicate that they are uniformly distributed. There is no difference among the examples, which indicates that the difference of eccentricity or rotation of the ellipses does not affect the distributions. Figure 3B displays the relationship among $P_{df2}$, $P_{MAX}$, $P_{dom}$, $P_{rec}$ and $P_{add}$ for the 1,000 table samples of (3333,3333,3334). The correlation between $P_{df2}$ and $P_{MAX}$ is the strongest. Although three tests of one degree of freedom show considerable correlation with $P_{MAX}$, a fraction of the tables show a substantial difference.

## COMPARISON OF THE POWER OF NEW METHOD WITH PEARSON'S TEST

We compared power of our method with Pearson's test and asymptotic estimation of $P$-value of MAX by Zang

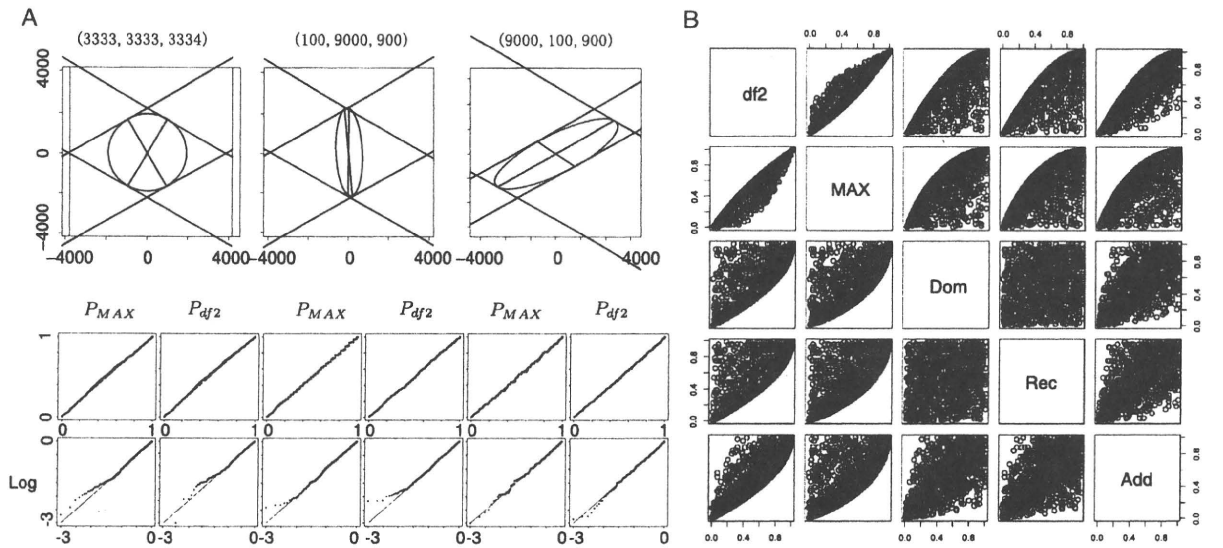


**Fig. 3. Distribution of $P_{MAX}$ for three sets of marginal counts. (A) A double triangle diagram and four $P$-$P$ plots below it are drawn for three marginal counts, (3333,3333,3334), (100,9000,900) and (9000,100,900), as indicated. The two darker lines in each double triangle diagram indicate the lines connecting tangent points. They separate the sections where the contour line of $\chi^2(MAX) = K$ corresponds to the ellipse, the elliptic sections (the upper and the lower sections) and the sections where the contour line of $\chi^2(MAX) = K$ corresponds to the tangent lines (the right and left sections). The angles where the contour line of $\chi^2(MAX) = K$ is a part of ellipse is middle, narrow and wide for three examples, respectively. For each marginal count set, 1000 tables were randomly simulated under the null hypothesis. (B)Four $P$-$P$ plots of $P_{MAX}$ and $P_{df2}$ were drawn below each diagram. The left-side plots are $P_{MAX}$ and the right-side plots are $P_{df2}$. The upper plots are on a linear scale and the lower plots are on a logarithmic scale.**

et al. in "Rassoc" package in CRAN (http://cran.r-project.org/), for six genetic models [Zang et al., 2010]. For each model, Hardy-Weinberg equilibrium was assumed in a population and allele frequency of risk allele in the population and prevalence of phenotype were set at 0.3 and 0.01, respectively. Genotypic relative risks for each model were given as (1.5,1.5,1), (1.5,1.25,1), (1.5,1,1) for dominant, additive and recessive models, respectively. A model between dominant and additive (half dominant model) and a model between additive and recessive (half recessive model) were also defined with (1.5,1.375,1) and (1.5,1.125,1). Relative risk of the sixth model (heterozygote-specific model) was given as (1,1.5,1). Five hundreds of cases and five hundreds of controls were randomly sampled from the population and 1,000 $2 \times 3$ tables were created for each model and tested with three tests. The result was shown in Table I. Powers of proposed method and MAX3 with asymptotic $P$-value were almost identical for all models. Power of Pearson's test was less powerful than the other two for all models.

## APPLICATION TO REAL GENOTYPE DATA

In response to Decision Letter, we applied our proposing method to two types of real GWAS study data. The first data were 17 SNPs that were reported with statistical significance in three papers and that were used by Li et al. to evaluate their method to approximate $P$-value of MAX test [Li et al., 2008]. The second data were 10,000 SNPs among WTCCC study for rheumatoid arthritis [The Wellcome Trust Case Control Consortium, 2007]. The ten thousands SNPs were selected from the top of the list of markers in the order of chromosomal location. For the first 17 SNPs, we applied our new method ($P_{MAX}$) and asymptotic $P$-value estimation of MAX3 test [($P_{MAX3asy}$ Zang et al., 2010 ("Rassoc" package in CRAN (http://cran.r-project.org/)], and the exact $P$ of the MAX or the OMTT [Yamada and Okada, 2009] ($P_{OMITex}$) and they were shown in Table II with $P$-value based on Rhombus formula $P_{rhombus}$. In the report by Li et al., they compared $P_{rhombus}$ with empirical $P$-values of bootstraps and permutations. In this report, we adopted $P_{OMITex}$, instead. As shown in Table II, all methods gave similar values for all SNPs.

For 10,000 SNPs in WTCCC study, we compared $P_{MAX}$, $P_{MAX3asy}$ and $P_{OMITex}$. Figure 4A, B shows their co-plots in regular and logarithmic scales, respectively. $P_{MAX}$ showed stronger correlation with $P_{OMITex}$ than $P_{MAX3asy}$. Figure 4C–F shows relation of difference between $P_{MAX}$ and $P_{OMITex}$ or $P_{MAX3asy}$ and $P_{OMITex}$ with allele frequency or the minimum value of $2 \times 3$ table cells. The differences between $P_{MAX}$ and $P_{OMITex}$ or $P_{MAX3asy}$ and $P_{OMITex}$ were both larger when allele frequency of minor allele was smaller and the minimum value of table was smaller. When the minimum value of table was very small, any asymptotic method deviates from the exact method and this was why some of $P_{MAX}$ were relatively more deviated from $P_{OMITex}$. However, excluding these exceptions, $P_{MAX}$ tended to give closer value to $P_{OMITex}$ than $P_{MAX3asy}$ than $P_{MAXasy}$ regardless of the minimum of table cells. It

**TABLE I. Power comparison of proposed method and Pearson's test and MAX3**

| | Recessive | | | Half recessive | | | Additive | | | Half dominant | | | Dominat | | | Heterozygout-specric | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cut off | Proposed | Pearson's | MAX3 | Proposed | Pearson's | MAX3 | Proposed | Pearson's | MAX3 | Proposed | Pearson's | MAX3 | Proposed | Pearson's | MAX3 | Proposed | Pearson's | MAX3 |
| 0.01 | 198 | 157 | 204 | 224 | 186 | 217 | 332 | 288 | 330 | 460 | 406 | 468 | 651 | 599 | 660 | 445 | 392 | 461 |
| 0.001 | 54 | 37 | 54 | 63 | 48 | 55 | 130 | 102 | 125 | 200 | 170 | 203 | 342 | 322 | 378 | 196 | 166 | 211 |
| 0.0001 | 14 | 11 | 15 | 14 | 12 | 14 | 41 | 31 | 45 | 72 | 60 | 74 | 189 | 164 | 191 | 65 | 55 | 78 |
| 0.00001 | 5 | 4 | 4 | 4 | 1 | 4 | 11 | 9 | 11 | 26 | 21 | 30 | 82 | 60 | 93 | 21 | 17 | 22 |
| 0.000001 | 1 | 0 | 1 | 0 | 0 | 0 | 3 | 3 | 3 | 8 | 6 | 9 | 22 | 18 | 25 | 8 | 6 | 8 |
| 0.0000001 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 2 | 9 | 8 | 11 | 2 | 2 | 2 |

Number of tables with $P$ less than cut-off values in 1,000 simulations.

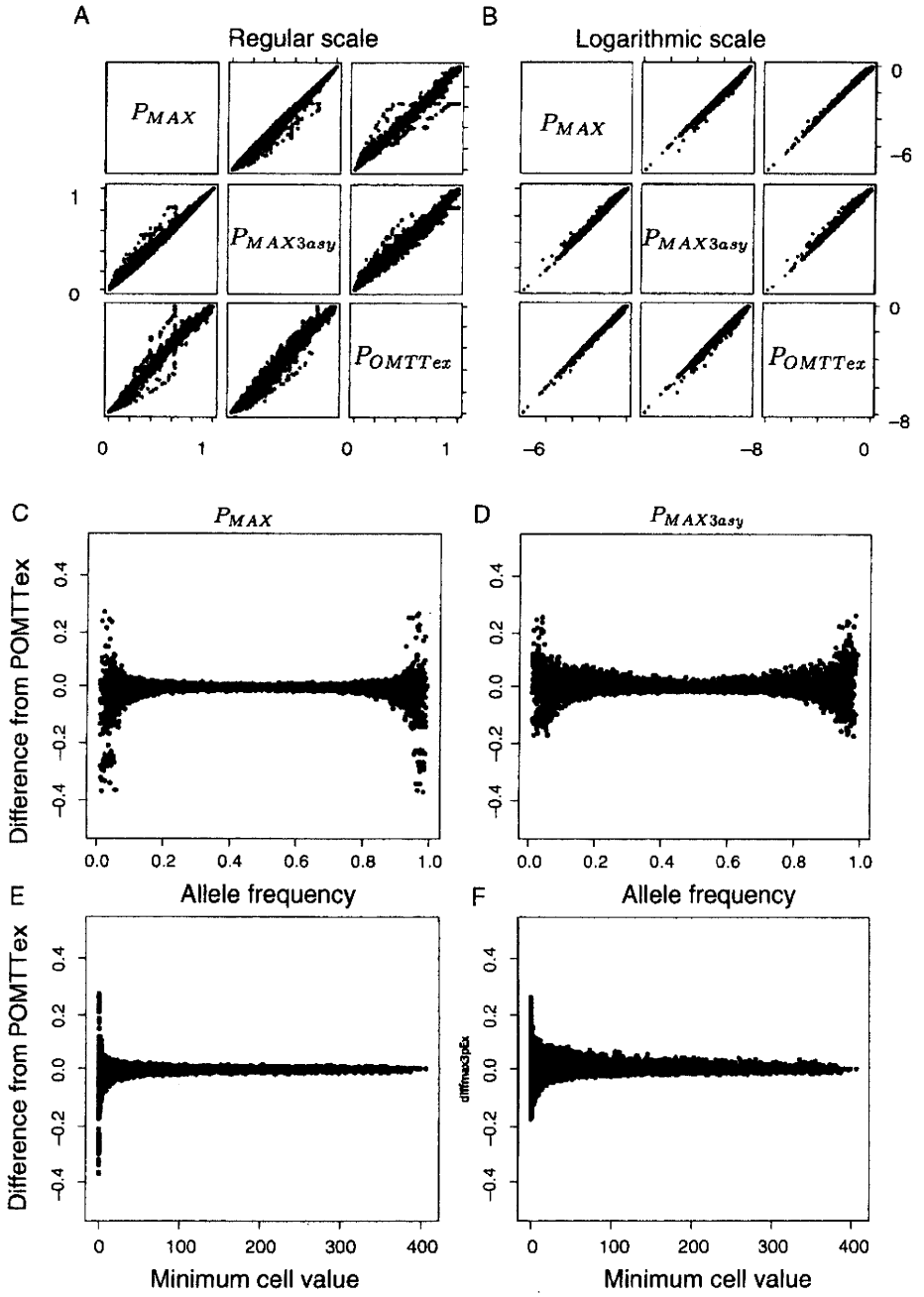


**Fig. 4.** $P_{MAX}$, $P_{MAX3asy}$ and $P_{OMITex}$ of 10,000 SNP tables in a GWAS were compared. (A) and (B) are co-plots among $P_{MAX}$, $P_{MAX3asy}$ and $P_{OMITex}$ in regular and logarithmic scale. All three were well correlated and $P_{MAX}$ and $P_{OMITex}$ were better. (C) and (D) Differences of $P_{MAX}$ and $P_{MAX3asy}$ from $P_{OMITex}$ were plotted along allele frequency. (E) and (F) Differences of $P_{MAX}$ and $P_{MAX3asy}$ from $P_{OMITex}$ were plotted along the minimum value of $2 \times 3$ table cells.

seemed reasonable that $P_{MAX}$ performed relatively better with tables of low minor allele frequency or of small minimal value of cells, because the ellipses of those tables tend to have high eccentricity, to which our method was designed to correct.

## DISCUSSION

In this paper, the de Finetti diagram, a diagram for genotype frequency of diallelic markers in population genetics, was applied to $2 \times 3$ contingency tables for case-control

**TABLE II. *P*-values of identified SNPs in GWASs of diabetes, breast, and prostate cancers**

| SNPID | A11 | A12 | A22 | B11 | B12 | B22 | PMAX | PMAXasy | PRhombus | POMTTex |
|---|---|---|---|---|---|---|---|---|---|---|
| 8 confirmed SNPs associated with Type 2 diabetes | | | | | | | | | | |
| rs7903146 | 197 | 348 | 149 | 335 | 254 | 65 | 6.26E-19 | 0.00E+00 | 1.58E-18 | 3.73E-19 |
| rs13266634 | 54 | 229 | 411 | 53 | 293 | 307 | 2.40E-05 | 1.89E-05 | 1.84E-05 | 2.27E-05 |
| rs1111875 | 77 | 302 | 315 | 119 | 308 | 227 | 8.44E-06 | 7.04E-06 | 6.78E-06 | 7.74E-06 |
| rs7923837 | 66 | 300 | 328 | 116 | 296 | 242 | 1.52E-06 | 2.36E-06 | 2.28E-06 | 1.43E-06 |
| rs7480010 | 301 | 327 | 66 | 363 | 246 | 45 | 2.56E-05 | 2.24E-05 | 2.18E-05 | 2.46E-05 |
| rs3740878 | 25 | 273 | 386 | 65 | 249 | 353 | 2.43E-05 | 1.89E-05 | 1.84E-05 | 1.94E-05 |
| rs11037909 | 25 | 274 | 387 | 65 | 251 | 353 | 2.44E-05 | 1.89E-05 | 1.85E-05 | 2.00E-05 |
| rs1113132 | 25 | 271 | 390 | 63 | 251 | 355 | 5.30E-05 | 4.22E-05 | 4.12E-05 | 4.11E-05 |
| 6 reported SNPs associated with breast cancer | | | | | | | | | | |
| rs10510126 | 955 | 180 | 10 | 854 | 272 | 14 | 2.21E-06 | 1.41E-06 | 1.42E-06 | 1.63E-06 |
| rs12505080 | 608 | 477 | 50 | 628 | 408 | 99 | 1.06E-04 | 8.46E-05 | 8.27E-05 | 1.02E-04 |
| rs17157903 | 777 | 316 | 18 | 862 | 220 | 26 | 8.41E-05 | 6.17E-05 | 6.20E-05 | 7.56E-05 |
| rs1219648 | 352 | 543 | 250 | 433 | 538 | 170 | 4.08E-06 | 4.99E-06 | 4.80E-06 | 3.80E-06 |
| rs7696175 | 353 | 605 | 187 | 396 | 496 | 249 | 2.30E-03 | 2.07E-03 | 1.98E-03 | 2.37E-03 |
| rs2420946 | 357 | 546 | 242 | 440 | 537 | 165 | 4.83E-06 | 5.34E-06 | 5.14E-06 | 4.73E-06 |
| 3 reported SNPs associated with prostate cancer | | | | | | | | | | |
| rs1447295 | 25 | 283 | 864 | 10 | 218 | 929 | 1.17E-04 | 1.09E-04 | 1.10E-04 | 1.00E-04 |
| rs6983267 | 351 | 598 | 223 | 277 | 579 | 301 | 1.95E-05 | 2.16E-05 | 2.06E-05 | 1.93E-05 |
| rs7837688 | 861 | 283 | 27 | 939 | 206 | 11 | 9.10E-06 | 6.66E-06 | 6.67E-06 | 9.06E-06 |

association tests with SNPs, and a novel diagram, called the double triangle diagram, was proposed. Using the new diagram, test statistics of $2 \times 3$ tables were geometrically described. Given that the tests for $2 \times 3$ tables, Pearson's test, dominant test, recessive test, additive test and MAX test are all trend tests with different scores, the results were interpreted within this context. The occurrence probability distribution in the diagram space was elliptic, and the eccentricity and rotation were functions of the marginal counts. Once the distribution of the probability density was expressed in algebraic geometry, it was easy to transform the ellipse into a regular circle. In the normalized figure, integration of the probability was easy even when the integral could not be solved analytically. Subsequently, the integration of probability of the MAX test was implemented and validated by a simulation in which the geometrically estimated *P*-values of the MAX test were in a uniform distribution. Although our method performed well in terms of type I error and power and gave close value to the exact MAX test, there were limitations. In recent GWASs, tests of SNP data have to consider covariates in many cases and actually the rhombus method proposed by Li et al. [2008] was designed to be able to adjust for the covariates. However, our approach was not applicable to such conditions. Besides the handling of covariates, we evaluated our geometric approach only for $2 \times 3$ tables and no further extensions to other statistical tests of higher dimensions and more complexed data structure were investigated.

The web tool to estimate the *P*-values of the MAX test is available from the author's web site along with the R source code (http://www.genome.med.kyoto-u.ac.jp/wiki_tokyo/index.php/EllipseMAXP).

## REFERENCES


Balding DJ. 2006. A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781–791.

Campbell M. 2005. $\chi 2$ test for linear trend—what's that? Midwifery 32:127–130.

Cannings C, Edwards AWF. 1968. Natural selection and the de Finetti diagram. Ann Hum Gen 31:421–428.

Cochran WG. 1954. Some methods for strengthening the common $\chi 2$ tests. Biometrics 10:417–451.

Li Q, Zheng G, Li Z, Yu K. 2008. Efficient approximation of P-value of the maximum of correlated tests, with applications to genome-wide association studies. Ann Hum Genet 72:397–406.

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445:881–885.

The Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678.

Yamada R, Okada Y. 2009. An optimal dose-effect mode trend test for SNP genotype tables. Genet Epidemiol 33:114–127.

Zang Y, Wing F K, Zheng G. 2010. Simple algorithms to calculate asymptotic null distributions of robust tests in case-control genetic association studies in R. J Stat 33:1–24.

Zheng G, Joo J, Yand Y. 2009. Pearson's test, trend test, and MAX are all trend tests with different types of scores. Ann Hum Genet 73:133–140.

Cell Metabolism
Article

# A Liver-Derived Secretory Protein, Selenoprotein P, Causes Insulin Resistance

Hirofumi Misu,[1,10] Toshinari Takamura,[1,10,*] Hiroaki Takayama,[1] Hiroto Hayashi,[1] Naoto Matsuzawa-Nagata,[1] Seiichiro Kurita,[1] Kazuhide Ishikura,[1] Hitoshi Ando,[1] Yumie Takeshita,[1] Tsuguhito Ota,[1] Masaru Sakurai,[1] Tatsuya Yamashita,[1] Eishiro Mizukoshi,[1] Taro Yamashita,[1] Masao Honda,[1] Ken-ichi Miyamoto,[2,3] Tetsuya Kubota,[4] Naoto Kubota,[4] Takashi Kadowaki,[4] Han-Jong Kim,[5] In-kyu Lee,[5] Yasuhiko Minokoshi,[6] Yoshiro Saito,[7] Kazuhiko Takahashi,[8] Yoshihiro Yamada,[9] Nobuyuki Takakura,[9] and Shuichi Kaneko[1]

[1]Department of Disease Control and Homeostasis
[2]Department of Hospital Pharmacy
[3]Department of Medicinal Informatics
Kanazawa University Graduate School of Medical Science, Kanazawa, Ishikawa 920-8641, Japan
[4]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, University of Tokyo, Tokyo 113-8655, Japan
[5]Section of Endocrinology, Department of Internal Medicine, Kyungpook National University Hospital, School of Medicine, Kyungpook National University, Jungu, Daegu 700-412, Korea
[6]Division of Endocrinology and Metabolism, Department of Developmental Physiology, National Institute for Physiological Sciences, Okazaki, Aichi 444-8585, Japan
[7]Department of Medical Life Systems, Faculty of Medical and Life Sciences, Doshisha University, Kyotanabe, Kyoto 610-0394, Japan
[8]Department of Nutritional Biochemistry, Hokkaido Pharmaceutical University, Otaru, Hokkaido 047-0264, Japan
[9]Department of Signal Transduction, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan
[10]These authors contributed equally to this work
*Correspondence: ttakamura@m-kanazawa.jp
DOI 10.1016/j.cmet.2010.09.015

## SUMMARY

**The liver may regulate glucose homeostasis by modulating the sensitivity/resistance of peripheral tissues to insulin, by way of the production of secretory proteins, termed hepatokines. Here, we demonstrate that selenoprotein P (SeP), a liver-derived secretory protein, causes insulin resistance. Using serial analysis of gene expression (SAGE) and DNA chip methods, we found that hepatic SeP mRNA levels correlated with insulin resistance in humans. Administration of purified SeP impaired insulin signaling and dysregulated glucose metabolism in both hepatocytes and myocytes. Conversely, both genetic deletion and RNA interference-mediated knockdown of SeP improved systemic insulin sensitivity and glucose tolerance in mice. The metabolic actions of SeP were mediated, at least partly, by inactivation of adenosine monophosphate-activated protein kinase (AMPK). In summary, these results demonstrate a role of SeP in the regulation of glucose metabolism and insulin sensitivity and suggest that SeP may be a therapeutic target for type 2 diabetes.**

## INTRODUCTION

Insulin resistance is an underlying feature of people with type 2 diabetes and metabolic syndrome (Saltiel and Kahn, 2001), but is also associated with risk for cardiovascular diseases (Després et al., 1996) and contributes to the clinical manifestations of nonalcoholic steatohepatitis (Ota et al., 2007). In an insulin-resistant state, impaired insulin action promotes hepatic glucose production and reduces glucose uptake by peripheral tissues, resulting in hyperglycemia. The molecular mechanisms underlying insulin resistance are not fully understood, but are now known to be influenced by the secretion of tissue-derived factors, traditionally considered separate from the endocrine system. Recent work in obesity research, for example, has demonstrated that adipose tissues secrete a variety of proteins, known as adipocytokines (Friedman and Halaas, 1998; Maeda et al., 1996; Scherer et al., 1995; Steppan et al., 2001; Yang et al., 2005), which can either enhance or impair insulin sensitivity, thereby contributing to the development of insulin resistance.

SeP (in humans encoded by the *SEPP1* gene) is a secretory protein primarily produced by the liver (Burk and Hill, 2005; Carlson et al., 2004). It contains ten selenocysteine residues and functions as a selenium supply protein (Saito and Takahashi, 2002). However, the role of SeP in the regulation of glucose metabolism and insulin sensitivity has not yet been established. Furthermore, the clinical significance of SeP in human diseases has not been well defined, although studies of SeP knockout mice showed SeP deficiency to be associated with neurological injury and low fertility (Hill et al., 2003; Schomburg et al., 2003).

The liver plays a central role in glucose homeostasis and is also the site for the production of various secretory proteins. For example, recent work in our laboratory has revealed that genes encoding secretory proteins are abundantly expressed in the livers of people with type 2 diabetes (Misu et al., 2007). Moreover, genes encoding angiogenic factors, fibrogenic factors, and redox-associated factors were differentially expressed in the livers of people with type 2 diabetes (Takamura et al., 2004; Takeshita et al., 2006), possibly contributing to the pathophysiology of

type 2 diabetes and its clinical manifestations. On the basis of these findings, we hypothesize that, analogous to adipose tissues, the liver may also contribute to the development of type 2 diabetes and insulin resistance, through the production of secretory proteins, termed hepatokines.

## RESULTS

### Identification of a Hepatic Secretory Protein Involved in Insulin Resistance

To identify hepatic secretory proteins involved in insulin resistance, we performed liver biopsies in humans and conducted a comprehensive analysis of gene expression profiles, using two distinct methods. First, we obtained human liver samples from five patients with type 2 diabetes and five nondiabetic subjects who underwent surgical procedures for malignant tumors, and we subjected them to serial analysis of gene expression (SAGE) (Velculescu et al., 1995). Consequently, we identified 117 genes encoding putative secretory proteins with expression levels in people with type 2 diabetes, 1.5-fold or greater higher than those in normal subjects. Next, we obtained ultrasonography-guided percutaneous needle liver biopsies from ten people with type 2 diabetes and seven normal subjects (Table S1 available online), and we subjected them to DNA chip analysis to identify genes whose hepatic expression was significantly correlated with insulin resistance (Table S2). We performed glucose clamp experiments on these human subjects and measured the metabolic clearance rate (MCR) of glucose (glucose infusion rate divided by the steady-state plasma glucose concentration) as a measure of systemic insulin sensitivity. As a result, we found that *SEPP1* expression levels were upregulated 8-fold in people with type 2 diabetes compared with normal subjects, as determined by SAGE (Table S2). Additionally, there was a negative correlation between hepatic *SEPP1* messenger RNA (mRNA) levels and the MCR of glucose, indicating that elevated hepatic *SEPP1* mRNA levels were associated with insulin resistance (Figure 1A). As a corollary, we found a positive correlation between the levels of hepatic *SEPP1* mRNA and postloaded or fasting plasma glucose (Figures 1B and 1C).

### Elevation of SeP in Type 2 Diabetes

To characterize the role of SeP in the development of insulin resistance, we measured serum SeP levels in human samples (Table S3), using enzyme-linked immunosorbent assays (ELISA), as described previously (Saito et al., 2001). Consistent with elevated hepatic *SEPP1* mRNA levels, we found a significant positive correlation between serum SeP levels and both fasting plasma glucose and hemoglobin $A_{1c}$ ($HbA_{1c}$) levels (Figures 1D and 1E). $HbA_{1c}$ is a clinical marker of protein glycation due to hyperglycemia, and elevated $HbA_{1c}$ levels generally reflect poor glucose control over a 2–3 month period. Additionally, serum levels of SeP were significantly elevated in people with type 2 diabetes compared with normal subjects (Figure 1F and Table S4). Similar to data derived from clinical specimens, in rodent models of type 2 diabetes, including OLETF rats and KKAy mice, hepatic *Sepp1* mRNA and serum SeP levels were elevated (Figures 1G–1J and Table S5).

### SeP Expression in Hepatocytes Is Regulated by Glucose, Palmitate, and Insulin

To clarify the pathophysiology contributing to the hepatic expression of SeP in type 2 diabetes, we investigated the effects of nutrient supply on *Sepp1* mRNA expression in cultured hepatocytes. We found that the addition of glucose or palmitate upregulated *Sepp1* expression, whereas insulin downregulated it in a dose- and time-dependent manner (Figures 2A, 2C, 2E, and 2F). Similar effects on SeP protein levels were observed in primary mouse hepatocytes (Figures 2B, 2D, and 2G). Consistent with the negative regulation of *Sepp1* by insulin in hepatocytes, *Sepp1* mRNA levels were elevated in the livers of fasting C57BL6J mice, compared with those that had been fed (Figure 2H). Thus, multiple lines of evidence suggest that elevated SeP is associated with the development of insulin resistance.

### SeP Impairs Insulin Signaling and Dysregulates Glucose Metabolism In Vitro

Because there is no existing cell culture or animal model in which SeP is overexpressed, we purified SeP from human plasma using chromatographic methods (Saito et al., 1999; Saito and Takahashi, 2002) to examine the effects of SeP on insulin-mediated signal transduction. Treatment of primary hepatocytes with purified SeP induced a reduction in insulin-stimulated phosphorylation of insulin receptor (IR), and Akt (Figures 3A and 3B). SeP exerts its actions through an increase in cellular glutathione peroxidase (Saito and Takahashi, 2002). Coadministration of BSO, a glutathione synthesis inhibitor, rescued cells from the inhibitory effects of SeP (Figure 3C). Moreover, SeP increased phosphorylation of IRS1 at Ser307, the downregulator of tyrosine phosphorylation of IRS (Figure S1A). Similar effects of SeP were also observed in C2C12 myocytes (Figure S1B). Next, we assessed whether SeP dysregulated cellular glucose metabolism. In H4IIHEC hepatocytes, treatment with SeP upregulated mRNA expression of *Pck1* and *G6pc*, key gluconeogenic enzymes, resulting in a 30% increase in glucose release in the presence of insulin (Figures 3D–3F). Treatment with SeP alone had no effects on the levels of mRNAs encoding gluconeogenic enzymes or on glucose production in the absence of insulin, suggesting that SeP modulates insulin signaling. Additionally, treatment with SeP induced a reduction in insulin-stimulated glucose uptake in C2C12 myocytes (Figure 3G). These in vitro experiments indicate that, at physiological concentrations, SeP impairs insulin signal transduction and dysregulated cellular glucose metabolism.

### SeP Impairs Insulin Signaling and Disrupts Glucose Homeostasis In Vivo

To examine the physiological effects of SeP in vivo, we treated female C57BL/6J mice with two intraperitoneal injections of purified human SeP (1 mg/kg body weight), 12 and 2 hr before the experiments. Injection of purified human SeP protein resulted in serum levels of 0.5–1.5 μg/mL (data not shown). These levels correspond to the incremental change of SeP serum levels in people with normal glucose tolerance to those with type 2 diabetes (Saito et al., 2001). Glucose and insulin tolerance tests revealed that treatment of mice with purified SeP induced glucose intolerance and insulin resistance (Figures 3H and 3I). Blood insulin levels were significantly elevated in