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]*, Shunichi Yamashita[3,4].

1. Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan.

2. Institut National de la Sante et de la Recherche Medicale (INSERM) Unit U852, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan.

3. Department of International Health and Radiation Research, Atomic Bomb Disease Institute, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki 852-8523, Japan.

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, Belarusian Medical Academy for Postgraduate Education, Minsk 220013, Republic of Belarus.

6. Department of Endocrinology, Belarusian Medical Academy for Postgraduate Education, Minsk 220013, Republic of Belarus.

7. Belarusian Medical Academy for Postgraduate Education, Minsk 220013, Republic of Belarus.

8. Centre National de Genotypage, Institut Genomique, Commissariat a l'Energie Atomique, Evry 91000, France.

9. Fondation Jean Dausset-CEPH, Paris 75010 France.

[+] These authors contributed equally to this work.

*Correspondence should be addressed to:

Fumihiko Matsuda,

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

Yoshida Konoe-cho, Sakyo, Kyoto, 606-8501, Japan.

Telephone: +81-75-753-9313

Fax: +81-75-753-9314

e-mail: fumi@genome.med.kyoto-u.ac.jp

Abstract

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 x 10^{-9} ~ 4.9 x 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 x 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

3

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 been reported since the early 1990's. 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 approximately 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

Genome-wide association study

In the initial genome scan (termed as Study1), 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 λ=1.08, Supplementary Figure 1a). A statistical analysis comparing genotype distributions did not find SNP markers that showed genome-wide significance. In the subsequent genome scan (termed as Study2), 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 Study2, 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 λ=1.14, Supplementary Figure 1b, Supplementary Figure 2), 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 Study1 and Study2. 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 (λ=1.11, Supplementary Figure 1c). A cluster of four SNPs at chromosome 9q22.23 showed genome-wide significance ($p<5.0$ x 10^{-8}), namely, rs925489, rs7850258, rs965513 and rs10759944 with meta-analysis *p*-values of 1.7 x 10^{-9}, 1.7 x 10^{-9}, 4.9 x 10^{-9} and 3.5 x 10^{-9}, respectively (Figure 2, 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 Table 1), nine other markers at the same chromosomal locus showed suggestive association signals (*mhp*=5.2 x 10^{-4} ~ 1.4 x 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 x 10^{-7}, rs925489: *p*=1.5 x 10^{-7}, rs10759944: *p*=2.4 x 10^{-7}, rs965513: *p*=3.2 x 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>3.6x10^{-5} for males, mhp>1.6x10^{-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 Study1 and Study2. Three LD blocks were identified: blockA between rs4742698 and rs16924042, blockB between rs1512261 and rs10818094, and blockC between rs7871887 and rs4618817. All of the four most significant markers are in blockB (Figure 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 *HEMGN* (ID: 55363), but none of these genes reside in BlockB. Seven out of the nine markers showing suggestive association signals are located at either 5' or 3' flanking region of the *FOXE1* gene in BlockC. In addition, an imputation analysis was performed for SNP markers in Blocks A, B and C using genotypes of International HapMap Project as reference.

9

We identified three additional SNPs in BlockB, namely rs7030280, rs10983700 and rs1588635, located approximately 9~11kb centromeric to rs925489, showing similar levels of association (imputed p=2.8 x 10^{-9} for rs7030280 and rs10983700, imputed p=3.7 x 10^{-9} for rs1588635) (Supplementary Table 2).   No other SNP markers in BlockA or BlockC 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 x 10^{-9}) in the Icelandic sporadic PTC patients, rs965513 was strongest (p=6.8 x 10^{-20}, OR=1.77) (Table 2).   We therefore selected rs965513 located approximately 57 kb upstream to *FOXE1* for further genotyping by Taqman using an independent sample set (termed as Study3) of 259 cases and 648 controls (Table 1).   The strong association (p=2.0 x 10^{-4}) was reproduced, and was further reinforced when the genotypes of the three studies were combined for meta-analysis (*mhp*=4.8 x 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 x 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 x 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 x 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 Study1, p=0.43 in

Study2 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.   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 ten 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 x 10^{-9}, OR=1.97, 95%CI: 1.58-2.47) than the late-onset cases (p=6.0 x 10^{-8},

OR=1.52, 95%CI: 1.31-1.77) when compared with 1,268 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 x 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 a 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 BlockC containing *FOXE1* in both the Icelandic and Belarusian studies may indicate the

existence of DNA sequences in BlockB 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

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 x 10^{-6} were identified, of which one was on chromosome 9p and the other

on chromosome 12p.   Since neither of these chromosomal loci has 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, ten 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 - 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 ~ 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 ± 1.57 µU/ml for thyrotropin (normal range 0.5 - 5.0 µU/ml) and 1.17 ± 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 Study2, 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) according to the manufacturer's protocol. DNA concentration and purity were measured with a Nanodrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA). The samples were stored at -80°C until use.

Genome-wide association 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 (Study1 and Study2), respectively. Validation of genome scan results (Study3) was performed by Taqman analysis using a third independent sample set consisting of 259 cases and 648 controls.

Study1 - Genome scan

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).   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.

Study2 – Genome scan

567,512 autosomal SNPs were genotyped in 214 thyroid cancer patients (mean age ± SD: 5.8 ± 5.2 years) using the same SNP arrays as those used in Study1. Genotype calls of 448 Russian DNA samples were used as population-based controls. The same exclusion criteria as Study1 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 Study1, in 52 cases and 235 controls for males, and in 161 cases and 213 controls for females.

Study3 - 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) according to the manufacturer's guidelines. A pre-designed 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 Analyses

A case-control association in each study was examined using trend Chi-squared 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 Study1 and Study2 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 Study1 to Study3. The LD structure was derived using the genotypes of Study1 and Study2 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 425kb-region between rs4742698 and rs4618817 were examined for association. In

the process of imputation, 50 Markov chain iterations were implemented.

Acknowledgments

Conflicts of Interest

We declare that we have no competing financial interests.

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., Carr, Z. (2006) In *Report of the UN Chernobyl Forum Expert Group "Health"*, 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.

Legends to Figures

Figure 1. Manhattan plot of the combined GWAS results for Studies 1 and 2.

P-values calculated by the Trend Chi-squared 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 x 10^{-8}.

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-squared 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 Study1 to Study3.   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 pair-wise LD

estimates of the SNPs located within the region using the genome scan results of Study1 and

Study2.

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
Study1	PTC1	Cases	187	0~17	3.0±3.8	3~20	10.0±4.1
	CTR1	Controls	172	0~17	1.5±2.8	-	-
Study2	PTC2	Cases	214	0~17	5.8±5.2	2~22	13.9±5.5
	CTR2	Controls*	448	-	-	-	-
Study3	PTC3	Cases	259	0~18	6.8±5.5	3~22	16.5±4.4
	CTR3	Controls	648	0~26	6.2±5.9	-	-

* Russian population controls from other genetic studies

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

Marker	Allele[1] ref.	var	Chr	Position	Study	Freq var. Case	Control	Trend p[2]	OR (95% C.I.)[3]	mhp[4]	OR (95% C.I.)[3]	mhp[4]	OR (95% C.I.)[3]	Freq var. Case	Control	p[5]	OR (95% C.I.)[3]
										Study1+2		Study1+2+3		Gudmundsson *et al*			
rs1512261	G*	T	9	99562351	1	0.500	0.419	0.029	1.39 (1.03, 1.87)	6.9×10^{-6}	1.53 (1.27, 1.84)						
					2	0.509	0.391	5.2×10^{-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)	5.2×10^{-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	3.3×10^{-5}	1.89 (1.40, 2.55)	1.7×10^{-9}	1.79 (1.48, 2.16)						
					2	0.481	0.349	6.0×10^{-6}	1.73 (1.37, 2.18)								
rs7850258	A*	G	9	99588834	1	0.487	0.334	3.3×10^{-5}	1.89 (1.40, 2.55)	1.7×10^{-9}	1.79 (1.48, 2.16)						
					2	0.481	0.349	6.0×10^{-6}	1.73 (1.37, 2.18)								
rs965513	A*	G	9	99595930	1	0.487	0.334	3.3×10^{-5}	1.89 (1.40, 2.55)	4.9×10^{-9}	1.76 (1.45, 2.12)	4.8×10^{-12}	1.65 (1.43, 1.91)	0.490	0.352	6.8×10^{-20}	1.77 (1.57, 2.00)
					2	0.476	0.352	1.7×10^{-5}	1.68 (1.33, 2.12)								
					3	0.462	0.367	2.0×10^{-4}	1.48 (1.20,1.83)								
rs10759944	A*	G	9	99596793	1	0.487	0.334	3.3×10^{-5}	1.89 (1.40, 2.55)	3.5×10^{-9}	1.77 (1.46, 2.14)			0.490	0.352	1.7×10^{-19}	1.77 (1.57, 2.01)
					2	0.479	0.352	1.2×10^{-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)	2.5×10^{-6}	1.56 (1.29, 1.87)						
					2	0.502	0.393	2.0×10^{-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)	1.4×10^{-6}	1.63 (1.33, 1.98)			0.387	0.285	1.9×10^{-12}	1.58 (1.39, 1.80)
					2	0.397	0.273	4.0×10^{-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)	2.6×10^{-4}	1.41 (1.17, 1.70)			0.488	0.385	2.8×10^{-9}	1.52 (1.32, 1.74)
					2	0.477	0.398	0.0070	1.37 (1.09, 1.73)								

	ref[1]	var[1]	Chr	Position	Block	freq	freq	p[2]	OR (95% C.I.)[3]	p[4]	OR (95% C.I.)[3]		freq	freq	p[5]	OR (95% C.I.)[3]
rs907580	T*	C	9	99662418	1	0.374	0.314	0.10	1.30 (0.96, 1.78)	5.7×10^{-6}	1.58 (1.30, 1.92)		0.395	0.281	1.1×10^{-14}	1.66 (1.46, 1.89)
					2	0.396	0.273	4.2×10^{-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)	4.5×10^{-5}	1.47 (1.22, 1.77)		0.472	0.359	2.6×10^{-13}	1.60 (1.41, 1.81)
					2	0.465	0.369	9.2×10^{-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)	4.6×10^{-5}	1.47 (1.22, 1.77)		0.472	0.359	2.2×10^{-13}	1.59 (1.41, 1.81)
					2	0.465	0.369	9.5×10^{-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)	5.2×10^{-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	2.5×10^{-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 Table 1.

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

[2] The p-values using Trend χ^2 test are shown.

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

[4] The Trend χ^2 Mantel-Haenszel *p*-values are shown.

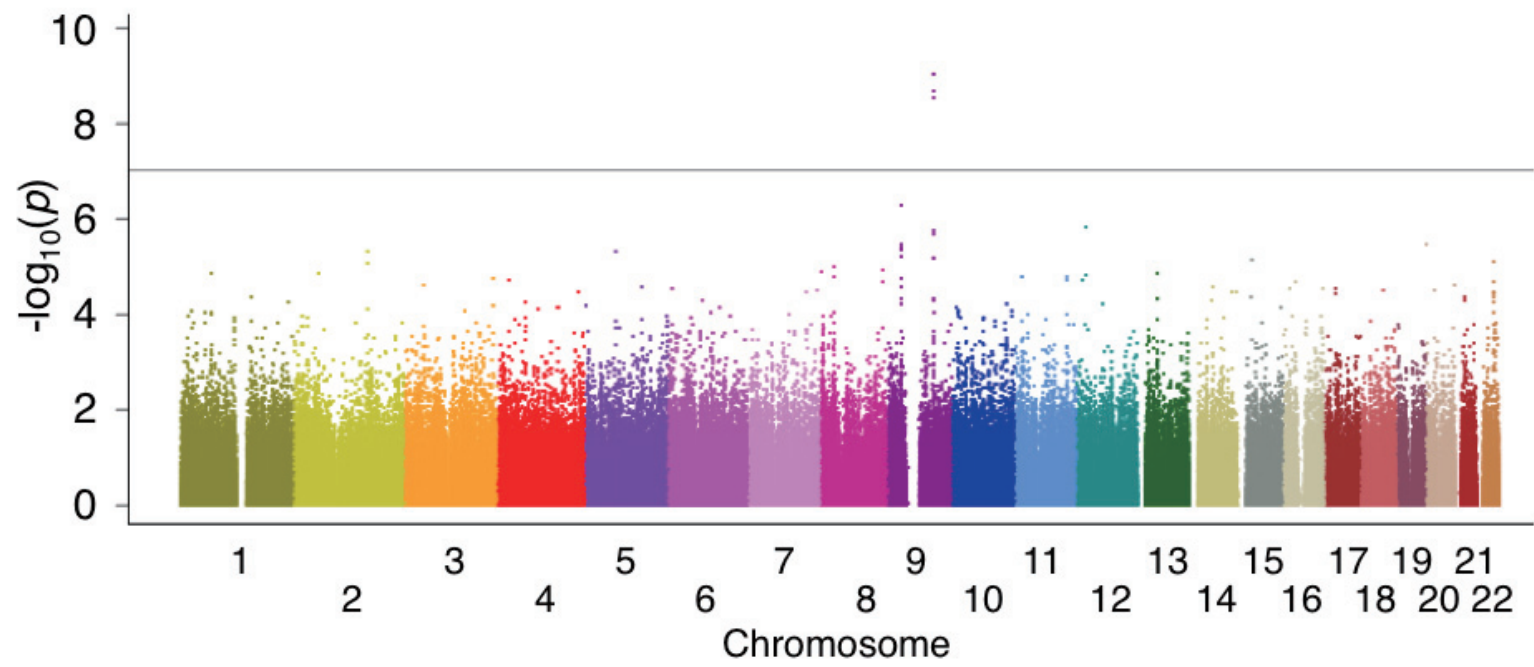[5] The *p*-values using a standard likelihood ratio χ^2 statistic are shown.

Figure 1

Figure 2