Genetic Polymorphism Predisposing to Differentiated Thyroid Cancer: A Review of Major Findings of the Genome-Wide Association Studies

Vladimir A. Saenko[1], Tatiana I. Rogounovitch[2]

Departments of [1]Radiation Molecular Epidemiology, [2]Global Health, Medicine and Welfare, Atomic Bomb Disease Institute, Nagasaki University, Nagasaki, Japan

Thyroid cancer has one of the highest hereditary component among human malignancies as seen in medical epidemiology investigations, suggesting the potential meaningfulness of genetic studies. Here we review researches into genetic variations that influence the chance of developing non-familial differentiated thyroid cancer (DTC), focusing on the major findings of the genome-wide association studies (GWASs) of common single-nucleotide polymorphisms (SNPs). To date, eight GWAS have been performed, and the association of a number of SNPs have been reproduced in dozens of replication investigations across different ethnicities, including Korea and Japan. Despite the cumulative effect of the strongest SNPs demonstrates gradual increase in the risk for cancer and their association signals are statistically quite significant, the overall prediction ability for DTC appears to be very limited. Thus, genotyping of common SNPs only would be insufficient for evidence-based counseling in clinical setting at present. Further studies to include less significant and rare SNPs, non-SNP genetic information, gene-gene interactions, ethnicity, non-genetic and environmental factors, and development of more advanced computational algorithms are warranted to approach to personalized disease risk prediction and prognostication.

Keywords: Thyroid neoplasms; Genetic loci; Genome-wide association study; Polymorphism, single nucleotide; Genetic predisposition to disease; Genetic testing

INTRODUCTION

Thyroid cancer displays a rising trend during the last decades worldwide. The increase in disease rate is almost completely attributed to the papillary thyroid carcinoma (PTC), which accounts for over 80% of all thyroid cancer in patients below the age of 70. Prevalence of the second most frequent histological type of thyroid malignancy originating, similarly to PTC, from follicular epithelial cells, the follicular thyroid carcinoma (FTC),

is about 5%. The two types of thyroid cancer, PTC and FTC, are referred to as differentiated thyroid cancer (DTC) to contrast them from poorly differentiated and anaplastic carcinomas, which also originate from follicular cells, and from medullary thyroid carcinoma, originating from parafollicular C cells.

So far, several factors have been established to increase the risk for DTC. From the point of view of their nature, risk factors can be broadly subdivided into modifiable (those that could be changed) and non-modifiable (cannot be changed). Examples of

modifiable factors include the lifestyle (in particular, obesity and diabetes in women), low-iodine diet (a risk factor for FTC, not characteristic for iodine-rich countries such as Korea and Japan), environmental agents (ionizing radiation, especially if exposure takes place in childhood, as seen in A-bomb hibakusha of Hiroshima and Nagasaki, in Chernobyl, and from medical exposures), and enhanced medical surveillance (such as e.g., ultrasound thyroid screening, which actually does not increase the risk for cancer biologically but sharply elevates the detection rate in usually asymptomatic individuals thus affecting thyroid cancer morbidity on the medical statistics metrics). Among nonmodifiable factors, the following are important: female gender (women have thyroid cancer about three times more frequent than men, this holds true for Korea and Japan [1-3]), age range from 40 to 70 years old (the highest rate), a family history of benign or malignant thyroid disease, ethnicity (people of Asian and European ancestry appear to have higher risk than other populations), hereditary syndromes (e.g., familial adenomatous polyposis, Carney complex type I, Cowden disease, Werner syndrome, acromegaly, etc.), and genetic predisposition.

Genetic predisposition to DTC should be distinguished from hereditary syndromes because its basis is the genetic polymorphism inherent in humans rather than mutations of certain genes which underlie syndromic diseases and display Mendelian form of inheritance. While causative mutations disrupting intracellular pathways and eventually resulting in tumor development are highly penetrant, evidence from different types of human cancers and multifactorial or complex diseases demonstrates that the source of genetic predisposition includes a large number of low-penetrance alleles with rather small effect sizes [4,5]. In this article, we review the state of the field of researches into genetic variations influencing the chance of developing non-familial DTC in humans focusing on genetic association studies of common single-nucleotide polymorphisms (SNPs).

SINGLE-NUCLEOTIDE POLYMORPHISMS IN HUMAN GENOME AND THE GENETIC COMPONENT OF NON-SYNDROMIC THYROID CANCER

SNP is defined as a change at a single position in a DNA sequence among individuals. The current build 151 of National Center for Biotechnology Information (NCBI) dbSNP database includes 660.8 million genetic variants of which 113.9 million are validated (compare with about 62.7 million with 44.3 million validated in the build 138 of April 2013, 5 years ago) [6].

Given the size of human genome of 3.2 billion base pairs, a variation may be expected to occur once in each 5th to 10th nucleotide on average, which is not really correct. It should be born in mind that dbSNP contains all available information on short genetic variations, not only SNPs, that some variants may be redundant or display a very low frequency so they are not necessarily found in every human, or may be ethnicity-specific. Actually, SNPs are estimated to occur once in 100 to 300 bases throughout the individual genome.

In the database, SNPs are designated "rs1234567890," where "rs" stands for "reference SNP" and "1234567890" may be any number 2-digit or longer. SNPs are usually termed "common" if their frequency is >1% in a population, while those less frequent are called "rare." Sometimes it is possible to find in the literature definitions for the "common" SNP frequency of ≥5%, "rare" ranging from ≥1% to <5%, and "subpolymorphic" for ≤1% [7]. SNPs are the most frequent and an easy to detect type of genetic variation, which makes them a facile target for genetic association studies.

Thyroid cancers, similarly to many other complex diseases, are recognized to be caused by both environmental and hereditary factors. The extent of genetic component in a malignancy can be measured based on the information on familial cancer aggregation if a family-cancer database is available in a country or a region. One of such databases with high-quality information is located in Sweden. According to it, the proportion of susceptibility to thyroid cancer accounted for by the genetic factors was 53%, the highest among 15 types of human malignancies studied [8]. Another study from the United States reported the population attributable risk for thyroid cancer due to familial factors of 28%, which was the 8th top of 40 cancer types [9]. Estimate of the relative risk of having thyroid cancer in the first-degree relatives was 6.9, which is comparable to the value of 8.5 for the overall familial relative risk in the same cohort [10]. This data attest to the potential meaningfulness of genetic studies in thyroid cancer.

TYPES AND METHODOLOGY OF GENETIC ASSOCIATION STUDIES

The aim of a genetic association study is the identification of genetic variants correlating with quantitative traits (e.g., human height, blood pressure) or the presence or absence of a certain phenotype (e.g., cancer, including DTC, or other conditions). Genetic variants such as SNPs may point at key genes involved in the mechanism of disease or may have their own function.

Genetic research into complex diseases is expected to provide insights into the biological processes underlying pathological state, and may lead to clinical advances by determining potential therapeutic targets and biomarkers, which can be useful for disease prevention or for tailored approaches to disease diagnosis and optimization of treatment ultimately improving the prognosis.

The total number of published genetic studies of thyroid cancer since 2000 is around 500, and the progress in the identification of genetic variants affecting susceptibility to DTC has been extensive. In the historical perspective, genetic association studies have passed through several generations. First studies were analyzing one to two SNPs in a sample of about a hundred, followed by more SNPs in several hundred of samples, then with a number of samples approaching to a thousand and increasing number of SNPs conforming with epidemiological design. Latest studies are consortium-based investigations enrolling several or even tens or hundreds thousand of participants. Increasing sample size is critical to achieve sufficient statistical power to detect moderate and weak association signals.

In general, genetic association studies can be subdivided into those hypothesis-based and hypothesis-free. The first type of investigations is called "target gene approach," which focuses on a limited set of genes combined in a pathway or a functional or ontological category. Because the number of SNPs in such studies ranges from a few to several hundred, genotyping is performed using low to medium throughput polymerase chain reaction based technologies. Examples of such studies in thyroid cancer are the analyses of three SNPs in the *RET* gene [11] one SNP in tumor protein p53 (*TP53*) [12], nine SNPs in six DNA damage response genes [13], two SNPs in ataxia telangiectasia mutated (*ATM*) and forkhead box E1 (*FOXE1*) [14] and 141 SNPs in 43 DNA repair genes in patients with Chernobyl thyroid cancer [15], or 768 tag and putative functional SNPs in 97 genes in genes involved in thyroid cell differentiation and proliferation, and in genes differentially expressed in thyroid cancer tissues [16].

The hypothesis-free study does not make assumption regarding possible genes or pathways involved in a disease. Genotyping is performed on a genome scale. Such works are known as genome-wide association studies (GWASs), and the number of analyzed SNPs in them is limited only by the technologies employed. Using modern human SNP microarrays, about a million SNPs can be interrogated simultaneously with Affymetrix chips and about 4 million with an Illumina platform. Using next-generation sequencing (NGS), all (i.e., millions) SNPs can theoreti-

cally be detected in a DNA sample.

Both target gene and GWA studies typically have case-control design. Studies aiming at discovery of disease-related SNPs usually have two or more stages. In the first stage, which is the "discovery phase," a full set of SNPs is genotyped (using e.g., microarrays) in a relatively small sample. The purpose is to determine the set of SNPs displaying an association signal at liberal P value. In the second stage, called "validation," these SNPs are analyzed using single-track assays in an extended independent sample. Only those SNPs that display the same direction of the effect (e.g., a particular "risk allele" of a given SNP is more frequent in patients than in controls similarly to the discovery phase findings) and with appropriate statistical significance are considered for further analysis. Validation step may be repeated again in another independent sample of cases and controls in order to further narrow down the number of SNPs, although this is not mandatory.

Statistical analysis is performed at each stage of the study. Usually this may be a chi-square test for trend or a logistic regression analysis, in which different models of inheritance can be tested. Joint (meta-) analysis of findings from different stages is run at the end to report the results typically as an odds ratio (OR) with its confidence intervals. OR represents an estimate of risk per allele if the analysis is performed for the multiplicative (log-additive) model of inheritance, which is most common for genetic association studies. Also reported are P values for the SNPs analyzed in the work. Statistical methods for genetic association study result analysis are now well established and are analytically rigorous, including controlling for multiple comparisons, population stratification, relatedness, and technical quality. With regard to correction for multiple testing, a $P = 5E-08$ is currently considered to be a reasonable threshold for standard GWAS [17].

MAJOR GWAS FINDINGS IN DIFFERENTIATED THYROID CANCER
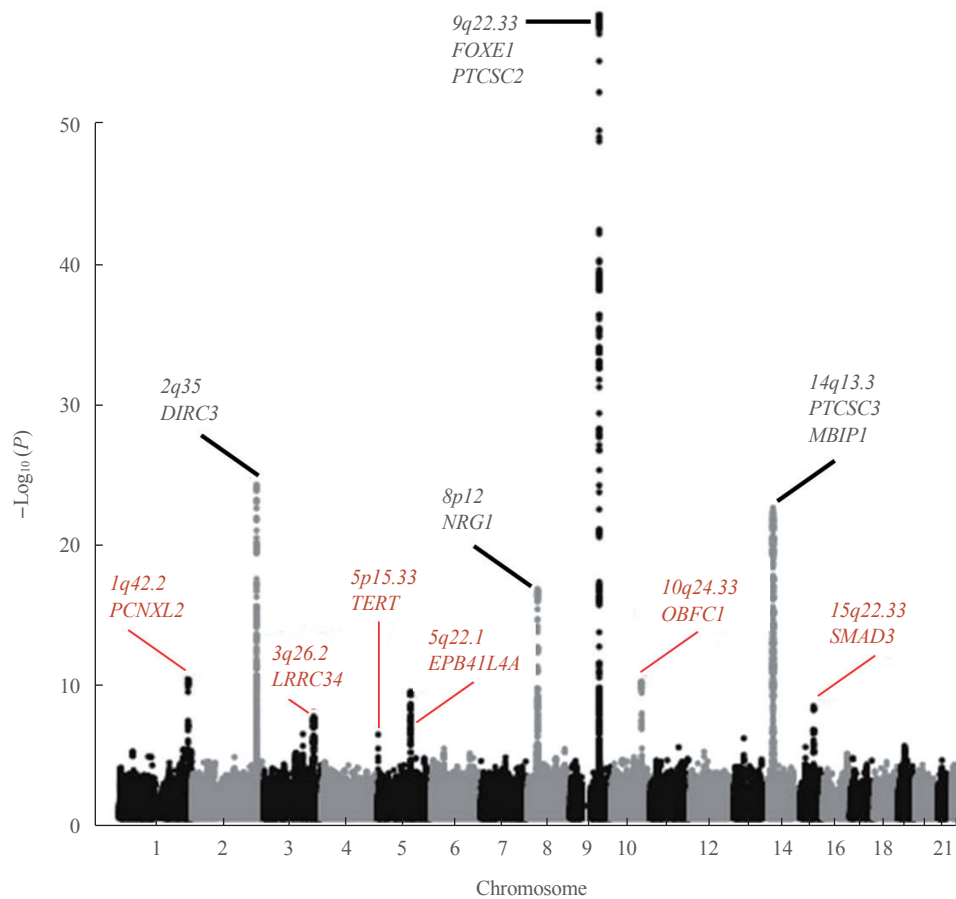
Populations of European ancestry

Most target gene studies have been effective in identifying SNPs nominally associated ($P < 0.05$) with DTC. However, association signals were rather weak, the results were mixed, poorly reproducible and generally lacked causative evidence. The most successful large-scale target gene study was performed by Landa et al. [16], which not only discovered a SNP (rs1867277, chromosome 9q22.33) in the 5' untranslated region (5'UTR) of the *FOXE1* gene (also known as thyroid transcrip-

tion factor 2) but also established that it regulates *FOXE1* expression by differentially recruiting upstream transcription factor 1/2 (USF1/USF2).

Results of GWAS were by far more meaningful. To date, seven GWAS of DTC, principally PTC, have been performed [18-24]. The number of participants (cases and controls cumulatively) in the studies ranged from about 2,000 to some 280,000 in the most recent, and reported effect sizes in terms of OR ranging from 1.2 to 2.1. Fig. 1 presents the results of the latest GWAS in the populations of European origin [24], which has confirmed the previous findings and reported novel SNPs associated with

risk for DTC; corresponding information from GWAS is shown in Table 1. Several SNPs, including rs11693806 (disrupted in renal carcinoma 3 [*DIRC3*]), rs2439302 (neuregulin 1 [*NRG1*]), rs965513 and rs1867277 (*FOXE1* and/or papillary thyroid carcinoma susceptibility candidate 2 [*PTCSC2*]), and rs944289 (papillary thyroid carcinoma susceptibility candidate 3 [*PTCSC3*] and/or NK2 homeobox 1 [*NKX2-1*]) were replicated in the Japanese population [25-28]. Rs944289, rs965513, rs966423, and rs2439302 were also replicated in a Chinese study [29].

The variant rs966423 on chromosome 2q35 is located in the *DIRC3* gene whose function is not established but it is thought

Fig. 1. A Manhattan plot of the combined thyroid cancer genome-wide association study (GWAS) results in the populations of European ancestry according to [24]. The vertical axis is the negative log₁₀-transformed *P* values for association signals of single-nucleotide polymorphisms across 22 autosomal chromosomes (horizontal axis). Annotated in black color are the loci discovered in earlier studies and replicated by different groups. Red color correspond to six novel loci associated with thyroid cancer detected by meta-analysis of GWAS data from 3,001 patients and 287,550 controls of the European descent. Note that statistical significance of association of the newly detected loci is generally lower than of those discovered earlier. The figure is derived from the open access article [24] according to a Creative Commons Attribution 4.0 International License and is presented here with minor modifications. *FOXE1*, forkhead box E1; *PTCSC2*, papillary thyroid carcinoma susceptibility candidate 2; *PCNXL2*, pecanex homolog 2; *DIRC3*, disrupted in renal carcinoma 3; *LRRC34*, leucine rich repeat containing 34; *TERT*, telomerase reverse transcriptase; *EPB41L4A*, erythrocyte membrane protein band 4.1 like 4A; *NRG1*, neuregulin 1; *OBFC1*, STN1, CST complex subunit; *PCNXL3*, pecanex homolog 3; *MBIP1*, MAP3K12 binding inhibitory protein 1; *SMAD3*, SMAD family member 3.

to be a putative tumor suppressor.

The SNP rs2439302, at 8p12, is located in the first intron of the *NRG1* gene encoding neuregulin 1, which is a human epidermal growth factor receptor 3 (HER3) ligand. *NRG1* can activate proliferative and survival mitogen activated protein kinase (MAPK) and AKT signaling pathways under conditions causing HER2/HER3 dimer induction in thyroid cancer cells [30].

Localized about 60 kB upstream and centromeric to *FOXE1*, rs966513 was the first and the strongest SNP consistently reported as a genetic determinant of susceptibility to thyroid cancer [18], but its functional relevance was established only recently [31]. The lead rs966513 and several other SNPs on chromosome 9q22.33 that are in linkage disequilibrium with rs966513, were shown to modify the activities of long-range enhancers involved in the transcriptional regulation of *FOXE1* and *PTCSC2*, a newly discovered thyroid-specific long intergenic noncoding RNA gene whose chromosomal position partly overlaps with that of the *FOXE1* promoter [31]. Both rs965513 and rs1867277 are functionally involved in the transcriptional regulation of *FOXE1* and *PTCSC2*. The risk allele of rs965513 was associated with decreased expression of *FOXE1*, unspliced *PTCSC2* and thyroid stimulating hormone receptor (*TSHR*) in normal thyroid tissue [31]. Interestingly, our recent study demonstrated that overexpression of FOXE1 in the thyroids of transgenic mice restrained the proliferation of follicular cells [32], in support of the functional effect of rs965513. Thus, besides of its role of a thyroid-specific transcription factor, *FOXE1* may also have a tumor suppressor effect.

In a functional study, rs944289 at 14q13.3 was shown to regulate expression of *PTCSC3*, a lincRNA gene with tumor suppressor properties in thyroid cancer cell lines [33]. Restoration of *PTCSC3* expression in cell lines inhibited cell growth and affected the expression of genes corresponding to (1) DNA replication, recombination and repair, gene expression, amino acid metabolism; (2) cellular movement, tumor morphology, cell death; and (3) cellular assembly and organization, cellular function and tissue morphology networks. *PTCSC3* expression was significantly downregulated in PTC as compared to normal thyroid in our study of PTC from Japan [27], in line with the mentioned work. Of note, rs944289 (*PTCSC3*) and rs2439302 (*NRG1*) were associated not only with thyroid cancer, but also with follicular adenoma [27]. This indicates that the mechanisms mediated by *PTCSC3* and NRG1 are likely to play roles not only in carcinogenesis but more broadly in thyroid tumorigenesis.

The 14q13.3 chromosome locus is likely to contain more than one genetic variant predisposing to thyroid cancer. The SNP rs116909374 reported in association with serum thyroid stimulating hormone level and thyroid malignancy [20] may point at the MAP3K12 binding inhibitory protein 1 (*MBIP*) gene whose product regulates JNK pathway involved in intracellular signaling of many types of human cancers including thyroid malignancy.

The recent discovery of six additional SNPs was possible due to a very large sample size in a consortium-based study (Table 1, lower part) [24]. The lead SNP at 1q42.2, rs12129938, is localized in the first intron of pecanex homolog 2 (*PCNXL2*; *Drosophila*). The role of highly conserved *PCNXL2* gene is not established, but it contains coding mononucleotide repeats that are associated with tumors of high microsatellite instability.

The lead variant, rs6793295 at 3q26.2, results in a missense mutation (p.Ser249Gly) in leucine rich repeat containing 34 (LRRC34), which is a protein suggested to play a role in ribosome biogenesis in pluripotent stem cells [34]. The role of this gene in the thyroid is unknown and it is possible that other genes in this chromosomal region should be considered.

The SNP rs10069690 at 5p15.33 is localized in the fourth intron of telomerase reverse transcriptase (*TERT*) which is an essential component of telomere length maintenance complex. *TERT* is weakly expressed in normal thyroid tissue but is reactivated in many cancers, including thyroid cancer [35] through the transcriptional regulation. Mutations in the *TERT* promoter in thyroid cancer have a prognostic potential when coexist with the $BRAF^{V600E}$ mutation [36].

The closest gene to a variant at chromosome 5q22.1, rs-73227498, is erythrocyte membrane protein band 4.1 like 4A (*EPB41L4A*), which mediate interactions between the cytoskeleton and plasma membrane, is thought to be involved in the β-catenin signaling pathway. β-Catenin is a marker of epithelial-to-mesenchymal transition in cancer and the aberrant activation of Wnt/β-catenin signaling was implicated in thyroid carcinogenesis [37].

The SNP rs7902587 is located at 10q24.33 approximately16 Kb upstream of STN1, CST complex subunit (*OBFC1*), whose product is involved in the stimulation of DNA polymerase-alpha-primase activity important for DNA replication, and also a member of a telomere-associated complex. Relationship of *OBFC1* to thyroid cancer is not established. Interestingly, another genetic variant in this region was previously found to confer risk to melanoma [38].

Rs2289261 at 15q22.33 is located in SMAD family member 3 (*SMAD3*) intron, which encodes a transcriptional mediator of

Table 1. Major Results of Genome-Wide Association Study of Differentiated Thyroid Cancer in the Population of European Ancestry[a]

Chromosome	SNP (RA/OA)	Annotation	SNP function	Nearest gene	Risk allele frequency	Allelic OR[b]	Replicated[c]
2q35	rs966423 (C/T)	Intron	NE	*DIRC3*	0.41	1.28	No
	rs11693806 (C/G)	Exon, non-coding	NE		0.32	1.43	No
8p12	rs2439302 (G/C)	Intron 1	Risk allele decreases *NRG1* expression	*NRG1*	0.48	1.32	Yes
9q22.33	rs965513 (A/G)	Intergenic	Enhancer element involved in the regulation of *PTCSC2* and *FOXE1* expression	*FOXE1*, *PTCSC2*	0.40	1.70	Yes
	rs1867277 (A/G)[d]	5'UTR	Risk allele regulates FOXE1 transcription through the differential recruitment of USF1/USF2 transcription factors	*FOXE1*	0.52	1.49	Yes
14q13.3	rs944289 (T/C)	Non-coding, promoter	Risk allele decreases promoter activation by weakening the binding affinity of the p42 C/EBPα and β	*PTCSC3*, *NKX2-1*	0.60	1.35	Yes
	rs116909374 (T/C)	Intergenic	NE	*MBIP1*	0.04	1.71	Absent[e]
Recently discovered SNPs							
1q42.2	rs12129938 (A/G)	Intron 1	NE	*PCNXL2*	0.80	1.32	No
3q26.2	rs6793295 (T/C)	Missense variant (p.Ser249Gly)	NE	*LRRC34*	0.76	1.23	No
5p15.33	rs10069690 (T/C)	Intron 4	NE	*TERT*	0.28	1.20	No
5q22.1	rs73227498 (A/T)	Intergenic	NE	*EPB41L4A*	0.87	1.37	No
10q24.33	rs7902587 (T/C)	Intergenic	NE	*OBFC1*	0.11	1.41	Absent
15q22.33	rs2289261 (C/G)	Intron	NE	*SMAD3*	0.68	1.23	No

SNP, single-nucleotide polymorphism; RA, risk allele; OA, other allele; OR, odds ratio; NE, not established; *DIRC3*, disrupted in renal carcinoma 3; *NRG1*, neuregulin 1; *PTCSC2*, papillary thyroid carcinoma susceptibility candidate 2; *FOXE1*, forkhead box E1; 5'UTR, 5' untranslated region; USF1/USF2, upstream transcription factor 1/2; C/EBP, CCAAT (cytosine-cytosine-adenosine-adenosine-thymidine)/enhancer binding protein; *PTCSC3*, papillary thyroid carcinoma susceptibility candidate 3; *NKX2-1*, NK2 homeobox 1; *MBIP1*, MAP3K12 binding inhibitory protein 1; *PCNXL2*, pecanex homolog 2; *LRRC34*, leucine rich repeat containing 34; *TERT*, telomerase reverse transcriptase; *EPB41L4A*, erythrocyte membrane protein band 4.1 like 4A; *OBFC1*, STN1, CST complex subunit; *SMAD3*, SMAD family member 3.
[a]Data according to the results from [24]; [b]Allelic effect size in terms of odds ratio; [c]Replication studies in Asian population(s); [d]Data from the target gene study ref [16]; [e]Zero allelic frequency in the Asian populations.

transforming growth factor β (TGFβ) signaling. TGFβ is abundantly expressed in thyroid cancer epithelium in contrast no normal follicular cells [39] and is thought to be involved in epithelial-to-mesenchymal transition. It is likely, that rs2289261 points at the importance of TGFβ pathway in thyroid carcinogenesis.

The Korean GWAS study
The only large-scale GWAS of DTC in Asian populations was recently performed in Korea [40]. Although SNPs reported in the study were not always the same as in European series, in part due to different genotyping platforms, the findings have unambiguously confirmed the association of the *DIRC3*, *NRG1*, *FOXE1*, and *NKX2-1* loci with DTC (Table 2, Fig. 2). In addition, seven novel significant association signals were found: rs11175834 (methionine sulfoxide reductase B3 [*MSRB3*], chromosome 12q14.3), rs4915076 (vav guanine nucleotide exchange factor 3 [*VAV3*], chromosome 1p13.3), rs1874564

(septin 11 [*SEPT11*], chromosome 4q21.1), rs9858271 (fragile histidine triad [*FHIT*], chromosome 3p14.2), rs7248104 (insulin receptor [*INSR*], chromosome 19p13.2), rs16934253 (solute carrier family 8 member B1 [*SLC24A6*], chromosome 12q24.13), and rs4649295 (*PCNXL2*, chromosome 1q42.2). Note that the latter locus was first reported as conferring risk for DTC in the populations of European ancestry [24] just a few months before the Korean study that in fact has independently replicated the association.

The product of *MSRB3*, a methionine sulfoxide reductase, is a selenoprotein. Selenocystein is essential for deiodinases involved in thyroid hormone biosynthesis enzymes; it has been also implicated in thyroid autoimmunity [41]. Therefore, MSRB3 may be relevant to thyroid cancer.

VAV3 is a guanine nucleotide exchange factor for the Rho family of GTPases, mediating actin cytoskeletal rearrangements and regulation of gene expression. *VAV3* expression has been

Table 2. Results of Genome-Wide Association Study of Differentiated Thyroid Cancer in the Korean Population[a]

Chromosome	SNP (RA/OA)	Annotation	Nearest gene	Risk allele frequency	Allelic OR[b]
Markers in the loci common with the populations of European ancestry					
2q35	rs12990503 (G/C)	Intron	*DIRC3*	0.69	1.34
8p12	rs6996585 (G/A)	Intron	*NRG1*	0.29	1.39
	rs12542743 (C/T)	Intron		0.32	1.36
	rs2439302 (G/C)[c]	Intron 1		0.27	1.37
9q22.33	rs72753537 (C/T)	Intron	*FOXE1*	0.10	1.41
14q13.3	rs34081947 (T/C)	Intron	*NKX2-1*	0.47	1.27
	rs944289 (T/C)[c]	Non-coding, promoter		0.51	1.25
1q42.2	rs4649295 (T/C)	Intron	*PCNXL2*	0.68	1.23
Markers specific for the Korean population					
12q14.3	rs11175834 (T/C)	Intron	*MSRB3*	0.20	1.37
1p13.3	rs4915076 (T/C)	Intron	*VAV3*	0.76	1.33
4q21.1	rs1874564 (G/A)	Intron	*SEPT11*	0.75	1.31
3p14.2	rs9858271 (G/A)	Intron	*FHIT*	0.48	1.26
19p13.2	rs7248104 (A/G)	3'UTR	*INSR*	0.41	1.22
12q24.13	rs16934253 (A/C/G)[d]	3'UTR	*SLC24A6 (SLC8B1)*	0.03	1.51

SNP, single-nucleotide polymorphism; RA, risk allele; OA, other allele; OR, odds ratio; *DIRC3*, disrupted in renal carcinoma 3; *NRG1*, neuregulin 1; *FOXE1*, forkhead box E1; *NKX2-1*, NK2 homeobox 1; *PCNXL2*, pecanex homolog 2; *MSRB3*, methionine sulfoxide reductase B3; *VAV3*, vav guanine nucleotide exchange factor 3; *SEPT11*, septin 11; *FHIT*, fragile histidine triad; *INSR*, insulin receptor; 3'UTR, 3' untranslated region; *SLC24A6*, solute carrier family 8 member B1; *SLC8B1*, solute carrier family 8 member B1.

[a]Data according to the results from [40]; [b]Allelic effect size in terms of odds ratio; [c]Same SNPs genotyped in the European populations, see Table 1; [d]This SNP displayed the strongest association with FTC: allelic frequency 0.07, OR=3.32.
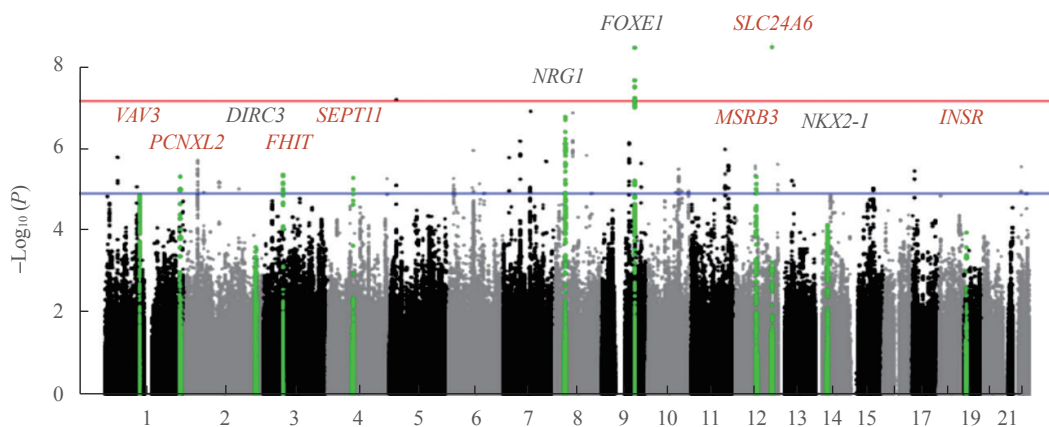
Fig. 2. A Manhattan plot of the genome-wide association study of differentiated thyroid cancer (DTC) in the Korean population according to [40]. The vertical axis is the negative log$_{10}$-transformed *P* values for association signals of single-nucleotide polymorphisms across 22 autosomal chromosomes (horizontal axis). Annotated in black color are the loci described earlier in the populations of European ancestry. Red color correspond to seven novel loci associated with DTC and specific to the Korean population based on gene scans of 470 patients and 8,279 controls. The red horizontal line represents the genome-wide significance threshold *P*=5E-08, and the blue horizontal line represents the genome-wide suggestiveness threshold *P*=1E-05. The figure is derived from the open access article [40] according to a Creative Commons Attribution 4.0 International License and is presented here with minor modifications. *VAV3*, vav guanine nucleotide exchange factor 3; *PCNXL2*, pecanex homolog 2; *DIRC3*, disrupted in renal carcinoma 3; *FHIT*, fragile histidine triad; *SEPT11*, septin 11; *NRG1*, neuregulin 1; *FOXE1*, forkhead box E1; *SLC24A6*, solute carrier family 8 member B1; *MSRB3*, methionine sulfoxide reductase B3; *NKX2-1*, NK2 homeobox 1; *INSR*, insulin receptor.

shown to correlate with the presence of *RET/PTC* rearrangements and point mutations of the RAS family genes in thyroid cancer cells [42] implicating it in thyroid carcinogenesis.

SEPT11 encodes a component of a filament-forming cytoskeletal GTPase potentially related to cytokinesis [43]. The presence of risk allele of the candidate SNP in the *SEPT11* locus correlated with extrathyroidal extension in the Korean series, thus pointing at possible SEPT11 involvement in thyroid cancer progression.

FHIT is a P1-P3-bis(5'-adenosyl) triphosphate hydrolase participating in purine metabolism. *FHIT* is likely to be a tumor suppressor gene in human lung [44], digestive tract [45], and gallbladder cancers [46]. In thyroid neoplastic tissues, *FHIT* abnormalities were reported [47]. Furthermore, homozygous deletion of *FHIT* exons 5 and 8, and changes in protein expression due to promoter methylation were implicated in DTC [48].

With regard to *SLC24A6*, which encodes a mitochondrial sodium and calcium ion exchanger, whose relationship to DTC pathogenesis is not established yet, an interesting finding was the stronger genetic variant association with FTC rather than PTC. This is a novel finding suggesting that at least some genetic markers of susceptibility to DTC are different between thyroid cancer types.

It is also worth noting that several SNPs in the *NRG1*, *DIRC3*, *NKX2-1*, and *FOXE1* loci displayed stronger association signals than in the European series. Therefore, together with SNPs in *MSRB3*, *VAV3*, *SEPT11*, *FHIT*, and *INSR* loci, these comprise DTC susceptibility markers specific for the Korean population. In contrast, SNPs in inner mitochondrial membrane peptidase subunit 2 (*IMMP2L*), DEAH-box helicase 35 (*DHX35*), arylsulfatase B (*ARSB*), and WDR11 antisense RNA 1 (*WDR11-AS1*) loci, which were significant in some European studies, showed no association signal in the Korean population; these are likely to be specific for some Caucasian ethnic groups.

CUMULATIVE EFFECT AND PREDICTION ABILITY OF RISK-CONFERRING SNPs

Cumulative effect of the major SNPs were estimated in different populations. When five strongest SNPs, i.e., rs966423 (*DIRC3*), rs2439302 (*NRG1*), rs965513 (*PTCSC2/FOXE1*), rs944289 (*PTCSC3/NKX2-1*), and rs116909374 (*MBIP1*), were considered, each additional risk allele increased the odds of thyroid cancer by 1.51 in Ohio and 1.35 in the Polish cohort, and the presence of 7 or more risk alleles conferred the ORs as high as 13.38 and 6.16, respectively [49]. In a study of Italian DTC, in

which 12 SNP were tested, odds of cancer was close to eight in the carriers of 14 or more risk alleles as compared to those with seven or less [23]. All differences and trends were highly significant. This is an additional fact in support of SNP relevance to DTC.

Unexpectedly, however, highly significant SNPs appear to have very limited predictive ability for DTC. When receiver operating characteristic analysis was applied to Ohio and Polish cohorts, the area under curve (AUC) were 71% and 61%, respectively [49]. In the Chinese study, AUC estimates ranged from 54% to 60% [50]. Furthermore, the five SNPs explained only about 11% of the familial risk of DTC in Ohio and about 6% of the familial risk in the Polish cohort. In the Chinese population, the overall familial relative risk based on SNP data was 5.98%, which nearly coincides with the results in the Polish group. In the Italian study, 12 SNPs explained about 4% of variation in DTC risk, and the *FOXE1* SNP alone explained less than 1%. These findings are in drastic contrast with the estimates of the impact of hereditary factors on thyroid cancer risk obtained in medical epidemiology investigations in familial series, which ranged >20% to 50% [8,9].

CONCLUSIONS

Remarkable progress in the genetics analysis of susceptibility to thyroid cancer has led to the identification of several chromosomal loci displaying significant association signals. Of them rs966423 (*DIRC3*), rs2439302 (*NRG1*), rs965513 (*PTCSC2/FOXE1*), rs944289 (*PTCSC3/NKX2-1*), and rs116909374 (*MBIP1*) are the most reproducible. Independent studies continue to discover new markers, some of which may be ethnicity-specific. Although exact biological mechanisms of all genetic variants' effects are not fully understood, the high statistical significance of association signals unambiguously implicates their involvement in thyroid carcinogenesis. There is little doubt that many more SNPs that may contribute to genetic predisposition to thyroid cancer exist, but their effect would be difficult to detect statistically unless extremely large cohorts are enrolled.

Cumulatively, SNP studies results are consistent with a multifactorial and polygenic model of susceptibility to DTC, and so far demonstrate that much of the heritability is still missing. In clinical setting, the goal of genotyping is to provide predictions that are significant enough to be used in counseling or even to inform treatment particularities and prognosis of thyroid cancer patients, as well as other individuals being prospectively evaluated for thyroid cancer risk. Apparently, this point has not been

reached so far with the available markers thereby restricting the usefulness of genetic data for DTC in the clinical practice at a moment. On the other hand, this demonstrates a need in research, which would help to narrow the gap between heredity estimates from the medical epidemiology and genetic epidemiology investigations.

Further studies would be anticipated in two major directions. First, the inclusion in analysis less significant and rare SNPs, non-SNP genetic information, ethnicity and environmental factors. Advancements in NGS technologies and its improving cost-effectiveness hold promise in this regard. However, profusion of high-throughput data would unlikely improve our understanding of the genetic component in complex diseases without parallel progress in study design and computational means. Therefore, the second major research avenue would include better definition of disease phenotype (i.e., not just simply DTC, but e.g., cancer morphology and invasive characteristics) and pathway- and biological network-oriented computations. These are expected to be able to estimate the effects of multiple small associations, gene-environment and gene-gene interactions which may function synergistically. There is still much remains to be discovered on the way to better personalized disease risk prediction and prognostication.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

ORCID

Vladimir A. Saenko *https://orcid.org/0000-0003-2844-3121*

REFERENCES

1. Statistics Korea. Cancer incident cases and incidence rates by site (24 items) and sex [Internet]. Daejeon: Statistics Korea; c2014 [cited 2018 May 3]. Available from: http://kosis.kr/eng/statisticsList/statisticsList_01List.jsp?vwcd=MT_ETITLE&parentId=D#SubCont.

2. Center for Cancer Control and Information Services, National Cancer Center. Projected cancer statistics, 2017 [Internet]. Tokyo: National Cancer Center; 2017 [updated 2017 Oct 27; cited 2018 May 3]. Available from: https://ganjoho.jp/en/public/statistics/short_pred.html.

3. Saika K, Matsuda T, Sobue T. Incidence rate of thyroid cancer by histological type in Japan. Jpn J Clin Oncol 2014;44:1131-2.

4. Sueta A, Ito H, Kawase T, Hirose K, Hosono S, Yatabe Y, et al. A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. Breast Cancer Res Treat 2012;132:711-21.

5. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med 2008;359:2208-19.

6. National Center for Biotechnology Information. NCBI dbSNP Build 151 [Internet]. Bethesda: NCBI; 2018 [updated 2018 Mar 22; cited 2018 May 3]. Available from: https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi?view+summary=view+summary&build_id=151.

7. Carvajal-Carmona LG. Challenges in the identification and use of rare disease-associated predisposition variants. Curr Opin Genet Dev 2010;20:277-81.

8. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. Int J Cancer 2002;99:260-6.

9. Kerber RA, O'Brien E. A cohort study of cancer risk in relation to family histories of cancer in the Utah population database. Cancer 2005;103:1906-15.

10. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. J Natl Cancer Inst 1994;86:1600-8.

11. Stephens LA, Powell NG, Grubb J, Jeremiah SJ, Bethel JA, Demidchik EP, et al. Investigation of loss of heterozygosity and SNP frequencies in the RET gene in papillary thyroid carcinoma. Thyroid 2005;15:100-4.

12. Rogounovitch TI, Saenko VA, Ashizawa K, Sedliarou IA, Namba H, Abrosimov AY, et al. TP53 codon 72 polymorphism in radiation-associated human papillary thyroid cancer. Oncol Rep 2006;15:949-56.

13. Akulevich NM, Saenko VA, Rogounovitch TI, Drozd VM, Lushnikov EF, Ivanov VK, et al. Polymorphisms of DNA damage response genes in radiation-related and sporadic papillary thyroid carcinoma. Endocr Relat Cancer 2009;16: 491-503.

14. Damiola F, Byrnes G, Moissonnier M, Pertesi M, Deltour I, Fillon A, et al. Contribution of ATM and FOXE1 (TTF2) to risk of papillary thyroid carcinoma in Belarusian children exposed to radiation. Int J Cancer 2014;134:1659-68.

15. Lonjou C, Damiola F, Moissonnier M, Durand G, Malakhova I, Masyakin V, et al. Investigation of DNA repair-related SNPs underlying susceptibility to papillary thyroid carcinoma reveals MGMT as a novel candidate gene in Belarusian children exposed to radiation. BMC Cancer 2017;17:328.

16. Landa I, Ruiz-Llorente S, Montero-Conde C, Inglada-Perez L, Schiavi F, Leskela S, et al. The variant rs1867277 in FOXE1 gene confers thyroid cancer susceptibility through the recruitment of USF1/USF2 transcription factors. PLoS Genet 2009;5:e1000637.

17. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol 2008;32:227-34.

18. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Sigurdsson A, Bergthorsson JT, et al. Common variants on 9q22.33 and 14q13.3 predispose to thyroid cancer in European populations. Nat Genet 2009;41:460-4.

19. Takahashi M, Saenko VA, Rogounovitch TI, Kawaguchi T, Drozd VM, Takigawa-Imamura H, et al. The FOXE1 locus is a major genetic determinant for radiation-related thyroid carcinoma in Chernobyl. Hum Mol Genet 2010;19:2516-23.

20. Gudmundsson J, Sulem P, Gudbjartsson DF, Jonasson JG, Masson G, He H, et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. Nat Genet 2012;44:319-22.

21. Kohler A, Chen B, Gemignani F, Elisei R, Romei C, Figlioli G, et al. Genome-wide association study on differentiated thyroid cancer. J Clin Endocrinol Metab 2013;98:E1674-81.

22. Mancikova V, Cruz R, Inglada-Perez L, Fernandez-Rozadilla C, Landa I, Cameselle-Teijeiro J, et al. Thyroid cancer GWAS identifies 10q26.12 and 6q14.1 as novel susceptibility loci and reveals genetic heterogeneity among populations. Int J Cancer 2015;137:1870-8.

23. Figlioli G, Chen B, Elisei R, Romei C, Campo C, Cipollini M, et al. Novel genetic variants in differentiated thyroid cancer and assessment of the cumulative risk. Sci Rep 2015; 5:8922.

24. Gudmundsson J, Thorleifsson G, Sigurdsson JK, Stefansdottir L, Jonasson JG, Gudjonsson SA, et al. A genome-wide association study yields five novel thyroid cancer risk loci. Nat Commun 2017;8:14517.

25. Bychkov A, Saenko V, Nakashima M, Mitsutake N, Rogounovitch T, Nikitski A, et al. Patterns of FOXE1 expression in papillary thyroid carcinoma by immunohistochemistry. Thyroid 2013;23:817-28.

26. Matsuse M, Takahashi M, Mitsutake N, Nishihara E, Hirokawa M, Kawaguchi T, et al. The FOXE1 and NKX2-1 loci are associated with susceptibility to papillary thyroid carcinoma in the Japanese population. J Med Genet 2011;48:645-8.

27. Rogounovitch TI, Bychkov A, Takahashi M, Mitsutake N, Nakashima M, Nikitski AV, et al. The common genetic variant rs944289 on chromosome 14q13.3 associates with risk of both malignant and benign thyroid tumors in the Japanese population. Thyroid 2015;25:333-40.

28. Nikitski AV, Rogounovitch TI, Bychkov A, Takahashi M, Yoshiura KI, Mitsutake N, et al. Genotype analyses in the Japanese and Belarusian populations reveal independent effects of rs965513 and rs1867277 but do not support the role of FOXE1 polyalanine tract length in conferring risk for papillary thyroid carcinoma. Thyroid 2017;27:224-35.

29. Wang YL, Feng SH, Guo SC, Wei WJ, Li DS, Wang Y, et al. Confirmation of papillary thyroid cancer susceptibility loci identified by genome-wide association studies of chromosomes 14q13, 9q22, 2q35 and 8p12 in a Chinese population. J Med Genet 2013;50:689-95.

30. Montero-Conde C, Ruiz-Llorente S, Dominguez JM, Knauf JA, Viale A, Sherman EJ, et al. Relief of feedback inhibition of HER3 transcription by RAF and MEK inhibitors attenuates their antitumor effects in BRAF-mutant thyroid carcinomas. Cancer Discov 2013;3:520-33.

31. He H, Li W, Liyanarachchi S, Jendrzejewski J, Srinivas M, Davuluri RV, et al. Genetic predisposition to papillary thyroid carcinoma: involvement of FOXE1, TSHR, and a novel lincRNA gene, PTCSC2. J Clin Endocrinol Metab 2015; 100:E164-72.

32. Nikitski A, Saenko V, Shimamura M, Nakashima M, Matsuse M, Suzuki K, et al. Targeted Foxe1 overexpression in mouse thyroid causes the development of multinodular goiter but does not promote carcinogenesis. Endocrinology 2016;157:2182-95.

33. Jendrzejewski J, He H, Radomska HS, Li W, Tomsic J, Liyanarachchi S, et al. The polymorphism rs944289 predis-

poses to papillary thyroid carcinoma through a large inter-genic noncoding RNA gene of tumor suppressor type. Proc Natl Acad Sci U S A 2012;109:8646-51.

34. Luhrig S, Siamishi I, Tesmer-Wolf M, Zechner U, Engel W, Nolte J. Lrrc34, a novel nucleolar protein, interacts with npm1 and ncl and has an impact on pluripotent stem cells. Stem Cells Dev 2014;23:2862-74.

35. Liu X, Bishop J, Shan Y, Pai S, Liu D, Murugan AK, et al. Highly prevalent TERT promoter mutations in aggressive thyroid cancers. Endocr Relat Cancer 2013;20:603-10.

36. Matsuse M, Yabuta T, Saenko V, Hirokawa M, Nishihara E, Suzuki K, et al. TERT promoter mutations and Ki-67 label-ing index as a prognostic marker of papillary thyroid carci-nomas: combination of two independent factors. Sci Rep 2017;7:41752.

37. Ishigaki K, Namba H, Nakashima M, Nakayama T, Mit-sutake N, Hayashi T, et al. Aberrant localization of beta-catenin correlates with overexpression of its target gene in human papillary thyroid cancer. J Clin Endocrinol Metab 2002;87:3433-40.

38. Law MH, Bishop DT, Lee JE, Brossard M, Martin NG, Mo-ses EK, et al. Genome-wide meta-analysis identifies five new susceptibility loci for cutaneous malignant melanoma. Nat Genet 2015;47:987-95.

39. Kimura ET, Kopp P, Zbaeren J, Asmis LM, Ruchti C, Ma-ciel RM, et al. Expression of transforming growth factor beta1, beta2, and beta3 in multinodular goiters and differen-tiated thyroid carcinomas: a comparative study. Thyroid 1999;9:119-25.

40. Son HY, Hwangbo Y, Yoo SK, Im SW, Yang SD, Kwak SJ, et al. Genome-wide association and expression quantitative trait loci studies identify multiple susceptibility loci for thy-roid cancer. Nat Commun 2017;8:15966.

41. Negro R. Selenium and thyroid autoimmunity. Biologics 2008;2:265-73.

42. Riesco-Eizaguirre G, Santisteban P. New insights in thyroid follicular cell biology and its impact in thyroid cancer therapy. Endocr Relat Cancer 2007;14:957-77.

43. Hanai N, Nagata K, Kawajiri A, Shiromizu T, Saitoh N, Hasegawa Y, et al. Biochemical and cell biological charac-terization of a mammalian septin, Sept11. FEBS Lett 2004; 568:83-8.

44. Joannes A, Bonnomet A, Bindels S, Polette M, Gilles C, Burlet H, et al. Fhit regulates invasion of lung tumor cells. Oncogene 2010;29:1203-13.

45. Ohta M, Inoue H, Cotticelli MG, Kastury K, Baffa R, Palaz-zo J, et al. The FHIT gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. Cell 1996; 84:587-97.

46. Priya TP, Kapoor VK, Krishnani N, Agrawal V, Agarwal S. Fragile histidine triad (FHIT) gene and its association with p53 protein expression in the progression of gall bladder cancer. Cancer Invest 2009;27:764-73.

47. Zou M, Shi Y, Farid NR, al-Sedairy ST, Paterson MC. FHIT gene abnormalities in both benign and malignant thyroid tu-mours. Eur J Cancer 1999;35:467-72.

48. Yin DT, Wang L, Sun J, Yin F, Yan Q, Shen RL, et al. Ho-mozygous deletion but not mutation of exons 5 and 8 of the fragile histidine triad (FHIT) gene is associated with features of differentiated thyroid carcinoma. Ann Clin Lab Sci 2010; 40:267-72.

49. Liyanarachchi S, Wojcicka A, Li W, Czetwertynska M, Stachlewska E, Nagy R, et al. Cumulative risk impact of five genetic variants associated with papillary thyroid carci-noma. Thyroid 2013;23:1532-40.

50. Guo S, Wang YL, Li Y, Jin L, Xiong M, Ji QH, et al. Signifi-cant SNPs have limited prediction ability for thyroid cancer. Cancer Med 2014;3:731-5.