Human Molecular Genetics, 2014, Vol. 23, No. 20 **5505–5517**
doi:10.1093/hmg/ddu250
Advance Access published on May 22, 2014

Genome-wide association study identifies a novel susceptibility gene for serum TSH levels in Chinese populations

Ming Zhan[1,2,†], Gang Chen[3,†], Chun-Ming Pan[1,†], Zhao-Hui Gu[4,†], Shuang-Xia Zhao[1,2,†], Wei Liu[1,2,†], Hai-Ning Wang[1], Xiao-Ping Ye[1], Hui-Jun Xie[1], Sha-Sha Yu[1,2], Jun Liang[5], Guan-Qi Gao[6], Guo-Yue Yuan[7], Xiao-Mei Zhang[8], Chun-Lin Zuo[9], Bin Su[10], Wei Huang[11], Guang Ning[1,2], Sai-Juan Chen[1,‡], Jia-Lun Chen[2,‡] and Huai-Dong Song[1,2,‡,*] for The China Consortium for the Genetics of Autoimmune Thyroid Disease[¶]

[1]State Key Laboratory of Medical Genomics and [2]Shanghai Institute of Endocrinology and Metabolism, Department of Endocrinology, Ruijin Hospital Affiliated to SJTU School of Medicine, Shanghai 200025, China, [3]Department of Endocrinology, Fujian Provincial Hospital, Fujian Medical University, Fuzhou 350001, China, [4]Shanghai Center for Systems Biomedicine, SJTU, Shanghai 200240, China, [5]Department of Endocrinology, The Central Hospital of Xuzhou Affiliated to Xuzhou Medical College, Xuzhou, Jiangsu Province 221109, China, [6]Department of Endocrinology, Linyi People's Hospital, Linyi, Shandong Province 276003, China, [7]Department of Endocrinology, The Hospital Affiliated to Jiangsu University, Zhenjiang, Jiangsu Province 213001, China, [8]Department of Endocrinology, The First Hospital Affiliated to Bengbu Medical College, Bengbu, Anhui Province 233004, China, [9]Department of Endocrinology, The First Hospital Affiliated to Anhui Medical University, Hefei, Anhui Province 230022, China, [10]Department of Endocrinology, The Tenth People's Hospital Affiliated to Tongji University, Shanghai 200072, China and [11]Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center, Shanghai 201303, China

Received October 28, 2013; Revised April 15, 2014; Accepted May 19, 2014

Thyroid-stimulating hormone (TSH) is a sensitive indicator of thyroid function. High and low TSH levels reflect hypothyroidism and hyperthyroidism, respectively. Even within the normal range, small differences in TSH levels, on the order of 0.5–1.0 mU/l, are associated with significant differences in blood pressure, BMI, dyslipidemia, risk of atrial fibrillation and atherosclerosis. Most of the variance in TSH levels is thought to be genetically influenced. We conducted a genome-wide association study of TSH levels in 1346 Chinese Han individuals. In the replication study, we genotyped four candidate SNPs with the top association signals in an independent isolated Chinese She cohort ($n = 3235$). We identified a novel serum TSH susceptibility locus within *XKR4* at 8q12.1 (rs2622590, $P_{combined} = 2.21 \times 10^{-10}$), and we confirmed two previously reported TSH susceptibility loci near *FOXE1* at 9q22.33 and near *CAPZB* at 1p36.13, respectively. The rs2622590_T allele at *XKR4* and the rs925489_C allele near *FOXE1* were correlated with low TSH levels and were found to be nominally associated to patients with papillary thyroid carcinoma (PTC) (OR = 1.41, $P = 0.014$ for rs2622590_T, and OR = 1.61, $P = 0.030$ for rs925489_C). The rs2622590 and rs925489 genotypes were also correlated with the expression levels of *FOXE1* and *XKR4*, respectively, in PTC tissues ($P = 2.41 \times 10^{-4}$ and $P = 0.02$). Our findings suggest that the SNPs in *XKR4* and near *FOXE1* are involved in the regulation of TSH levels.

*To whom correspondence should be addressed. Tel: +8621 64370045-610808; Fax: +8621 64743206; Email: huaidong_s1966@163.com
†Joint first authors.
‡Joint senior authors.
¶A full list of The China Consortium for the Genetics of Autoimmune Thyroid Disease members and affiliations appears at the end of this manuscript.

INTRODUCTION

Impinging on virtually every tissue of the body, thyroid hormones affect a variety of metabolic and developmental processes in humans (1). They are vital to brain development and skeletal maturation in fetuses and infants, and they influence protein synthesis, fat and carbohydrate metabolism, bone deposition, metabolic rate, temperature regulation and blood pressure in adults (2–4).

Thyroid function is regulated by a homeostatic negative-feedback loop involving the pituitary–thyroid axis and is evaluated by measuring circulating concentrations of thyroid-stimulating hormone (TSH), free thyroxine (FT4), and free triiodothyronine (FT3) (5). Secreted by the pituitary gland, TSH interacts with the TSH receptor on thyroid cells to control the release of T4 and T3 through a series of signal transduction pathways (6). TSH and FT4 have an inverse, log-linear relationship such that small changes in FT4 levels result in dramatic changes in TSH secretion. Hence, the serum TSH concentration is a very sensitive indicator of the function of the thyroid gland (7).

The variability of serum TSH concentration is more dramatic among healthy individuals than within healthy individuals (measured multiple times in the same individual) (8), suggesting that the regulation of the thyroid-hormone axis is exact and varies among individuals. Evidence from studies of healthy Danish twins has shown that genetic factors can account for 64% (57–70%) of the variance in circulating TSH levels (9). Another heritability study of Mexican Americans showed that non-genetic covariates only accounted for 2–18% of the total phenotypic variation, whereas genetic factors accounted for 26–64% of the total phenotypic variation (10).

Previous studies have identified many putative genetic susceptibility variants for serum TSH levels, but unequivocal replication has been limited to a few loci (such as *PDE8B* at 5q13.3, *CAPZB* at 1p36.13, *MAF* at 16p23.2 and *NR3C2* at 4q31.23) (11–17). Moreover, more recent GWAS studies in large samples recruited from Caucasian by Gudmundssonn *et al*. (15) and Porcu *et al*. (17), identified 22 novel susceptibility loci for TSH level. All the 26 TSH risk loci only explained about 4.3–5.6% of the inter-individual variability in serum TSH levels, suggesting that as-yet-unknown variants may also be important.

To identify novel serum TSH susceptibility loci, we conducted a GWAS of 1346 healthy Chinese Han individuals and an independent replication study of 3235 healthy Chinese She individuals. We identified a new susceptibility locus (rs2622590, near the *XKR4* region at 8q12.1; $P_{\text{combined}} = 2.21 \times 10^{-10}$) and confirmed two previously reported loci (near *FOXE1* at 9q22.33 and near *CAPZB* at 1p36.13).

RESULTS

Genome-wide association study of serum TSH levels

The clinical and demographic characteristics of the study participants are shown in Table 1. After stringent QC filters, a total of 483 947 genotyped SNPs and 8 019 905 imputed SNPs were analyzed for associations with serum TSH levels among 1346 individuals whose serum TSH levels ranged from 0.35 and 4.94 mU/l (Fig. 1). The results for the typed and imputed SNPs ($P < 10^{-2}$) (http://www.dropbox.com/s/q32p5yrd1accd3a/Data%20for%20TSH.zip), organized by chromosome location in a Manhattan plot, are shown in Figure 2. A quantile–quantile plot of TSH levels is shown in the Supplementary Material (Supplementary Material, Fig. S1). The estimated inflation factor was modest ($\lambda = 1.004$), and thus the distribution of *P*-values for the association tests shows no evidence of systematic bias. A principal component analysis (PCA) and a multidimensional scaling (MDS) analysis, both described in our previous study (18), showed that all of the subjects clustered around Chinese and Japanese lines of descent.

In our initial GWAS, 10 genotyped SNPs located in five different chromosomal regions had *P*-values $<1 \times 10^{-5}$ (Table 2). From those 10 SNPs, we identified the four SNPs, located in three different chromosomal regions, with the strongest ($P < 10^{-6}$) serum TSH associations: rs1348271 at 11q22.1 ($P = 1.09 \times 10^{-7}$), rs925489 and rs7850258 at 9q22.3 ($P = 3.19 \times 10^{-7}$ and $P = 3.19 \times 10^{-7}$, respectively), and rs2622590 at 8p12.1 ($P = 6.49 \times 10^{-7}$). The minor allele frequencies (MAFs) for the four SNPs ranged from 0.05 to 0.46.

Analyses of replication and combined data

Given that the sample size of the initial GWAS data set in the current study is underpowered to detect common variants associated with TSH levels, the three SNPs (rs1348271, rs925489 and rs2622590) with the strongest serum TSH associations, and one SNP (rs6683419, $P_{\text{GWAS}} = 1.33 \times 10^{-5}$) in 1p36.13 region previously reported to be associated with serum TSH levels (13), were selected and genotyped in an independent sample of the Southern Chinese She population, who is one of the important ethnic minorities, amounts to 0.7 million people in China (in the 2010 census). They mainly work in agriculture, forestry, animal husbandry, fishing and water industry, and have their own living customs. The marriage between Han and She were prohibited 60 years ago. The replication sample was recruited from the She population residing in Ningde city of Fujian Province and consisted of 3235 individuals with serum TSH levels between 0.35 and 4.94 mU/l (19). Three of the four selected SNPs (rs925489, $P = 9.13 \times 10^{-9}$; rs2622590, $P = 5.62 \times 10^{-6}$; and rs6683419, $P = 5.86 \times 10^{-4}$) showed evidence of replication in the tested data set (Table 3).

In the combined samples, two SNPs (rs925489 at 9q22.33, $P = 1.02 \times 10^{-13}$; and rs2622590 at 8q12.1, $P = 2.21 \times 10^{-10}$) showed unequivocal evidence of association with serum TSH levels, with a genome-wide significance threshold of $P = 5 \times 10^{-8}$; and one more SNP (rs6683419 at 1p36.13) nearly met the genome-wide significance level for association with serum TSH levels ($P = 2.90 \times 10^{-7}$) (Table 3). These three SNPs included one new serum TSH susceptibility locus (the *XKR4* region at 8q12.1) and two previously reported TSH susceptibility loci (Table 3).

Notably, we identified a novel serum TSH susceptibility locus, rs2622590 ($P = 2.21 \times 10^{-10}$) in the 8p12.1 region. In the initial GWAS, we analyzed 2966 imputed SNPs and 164 genotyped SNPs in a ~900 kb linkage disequilibrium (LD) block containing *XKR4* and *SBF1P1*, *TMEM68*, *TGS1* and *LYN* located at 8p12.1 (Fig. 3A). We found that all of the SNPs with $P < 10^{-5}$ within the LD block were located within a

Table 1. Characteristics of samples used in the GWAS and replication analyses

Study population	*n*	Female, *n* (%)	Age (years) (Mean ± SD)	BMI (kg/m^2) (Mean ± SD)	TSH (mU/l) (Mean ± SD)	Log_{10} TSH (Mean ± SD)
GWAS						
Xuzhou	713	545 (76.44)	47.71 ± 7.70	22.27 ± 2.19	1.81 ± 0.84	0.21 ± 0.20
Linyi	633	482 (76.15)	50.33 ± 9.90	NA	1.50 ± 0.73	0.13 ± 0.21
Replication						
Ningde	3235	1,767 (54.62)	50.89 ± 12.70	22.74 ± 3.11	1.61 ± 0.84	0.15 ± 0.23

Values are means and standard deviations (SD) unless otherwise specified.
BMI, body mass index; TSH, thyroid-stimulating hormone; NA, not assessed, SD, standard deviation.

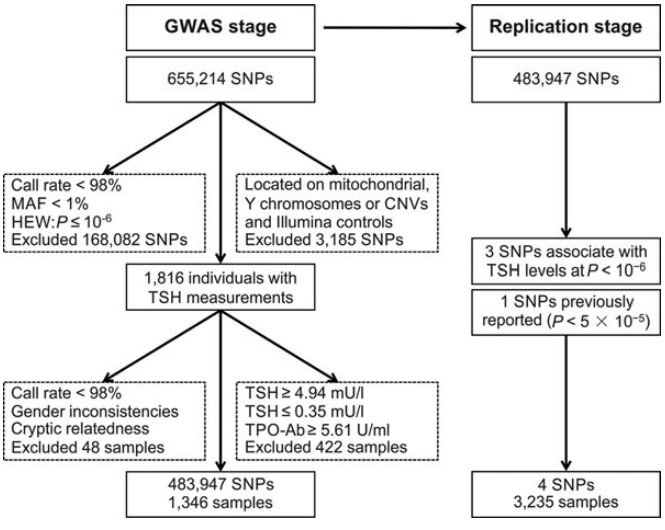

Figure 1. Flowchart for quality filtering in the two-stage GWAS for serum TSH levels. In the initial GWAS stage, quality filtering was performed on SNPs and samples before analysis to ensure robust association tests. Of the 655 214 markers assayed, the Y-chromosome and mitochondrial SNPs, CNV-related markers, and Illumina controls were excluded ($n = 3185$), leaving 652 029 SNPs for further analysis. We further excluded SNPs with a low call rate (< 98%), an MAF < 1% within the population, or a significant deviation from Hardy-Weinberg equilibrium ($P \leq 10^{-6}$). Samples were excluded if they had a high missing call rate (<98%), gender inconsistencies, cryptic relatedness, abnormal TSH (TSH ≥4.94 or ≤0.35 mU/l) or high TPO-Ab (TPO-Ab ≥5.61 U/ml). In the replication study, three SNPs associated with TSH ($P < 10^{-6}$) were selected as candidate loci for replication. Additionally, one SNP that was previously reported to be associated with TSH was also chosen for genotyping in the replication samples.

35 kb region of intron 2 of *XKR4* (Fig. 3A; Supplementary Material, Tables S1 and S2). After conditioning on rs2622590 ($P_{GWAS} = 6.49 \times 10^{-7}$; Table 3), none of the other SNPs at 8p12.1 remained significantly ($P < 0.01$) associated with serum TSH levels, suggesting that rs2622590 might be an independent SNP associated with serum TSH levels. Therefore, we selected rs2622590 for further genotyping in the replication sample. The replication study confirmed the serum TSH association of rs2622590 ($P_{replication} = 5.62 \times 10^{-6}$; Table 3), and the association reached the genome-wide significance level in the combined analysis ($P_{combined} = 2.21 \times 10^{-10}$; Table 3). After adjustment for the relevant covariates, the serum TSH levels among individuals with the rs2622590 CC genotype were higher than those among individuals with the homozygous-susceptible TT genotype in the GWAS cohort (1.79 ± 0.82 for the CC genotype, 1.52 ± 0.75 for the TT genotype; $P = 6.31 \times 10^{-6}$; Supplementary Material, Table S3) and in the replication cohort (1.67 ± 0.85 for the CC genotype, 1.47 ± 0.79 for the TT genotype; $P = 3.44 \times 10^{-6}$; Supplementary Material, Table S3). Rs2622590 is located in intron 2 of *XKR4*, 3.5 kb downstream of the pseudogene *SBF1P1* (Fig. 3A). *XKR4* is expressed abundantly in the tissues of the brain, esophagus and stomach and relatively sparsely in the tissues of the thyroid, kidney and pancreas (Fig. 4C). To further test whether the TSH associated SNPs regulate the expression of *XKR4*, we inspected three *cis*-gene expression quantitative trait loci (*cis*-eQTL) databases from European Caucasian population (20–22) and found that out of the 327 TSH associated SNPs ($P < 0.05$) in a ~900 kb region on 8q12.1 (Supplementary Material, Table S2), no SNPs were correlated with the expression of *XKR4*. Furthermore, a ~40 kb region surrounding the rs2622590 on 8q12.1 was analyzed to identify regulatory elements by inspection the UCSC Genome Browser. Interestingly, a ~5 kb fragment near to rs2622590, which contained four TSH strongly associated SNPs (rs2975987, rs2939632, rs2975986 and rs2975985), was conserved from zebrafish to human (Fig. 5). These four SNPs were in complete LD with rs2622590 ($r^2 > 0.97$ in the 1000 Genomes Asian (ASN) samples). The ~5 kb fragment near to rs2622590 was a DNase I hypersensitive site based on searching the ENCODE database (http://genome.ucsc.edu/ENCODE/) (23), suggesting that the fragment containing TSH associated SNPs might be a transcript regulatory element (Fig. 5). Moreover, non-synonymous SNPs in *XKR4* were in complete LD with the strong associated SNPs in the TSH risk locus on 8q12.1.

Rs965513 is located at 9q22.33 and was associated with serum TSH levels and thyroid cancer in previous studies (12,15). The 9q22.33 region contains one thyroid-specific transcription factor, *FOXE1* (or thyroid transcription factor 2, *TTF2*). In our initial GWAS, we obtained the genotypes of 2944 imputed SNPs and 186 genotyped SNPs in an ~800 kb LD block at 9q22.33. We found that the SNPs with the strongest serum TSH associations ($P < 1 \times 10^{-6}$) within the ~30 kb region were in a complete LD block ($r^2 > 0.8$ in the 1000 Genomes ASN samples) located between *XPA* and *FOXE1* (Fig. 3B, Supplementary Material, Table S4). In our replication study, we analyzed a previously genotyped SNP, rs925489, which was in complete LD with rs965513 ($r^2 = 0.97$ in our Chinese Han samples). We confirmed the serum TSH association of rs925489 in the replication study, and we found that rs925489 reached the genome-wide significance level in the combined cohorts ($P_{GWAS} = 3.19 \times 10^{-7}$; $P_{replication} = 9.13 \times 10^{-9}$; $P_{combined} = 1.02 \times 10^{-13}$; Table 3). Rs925489 is located about

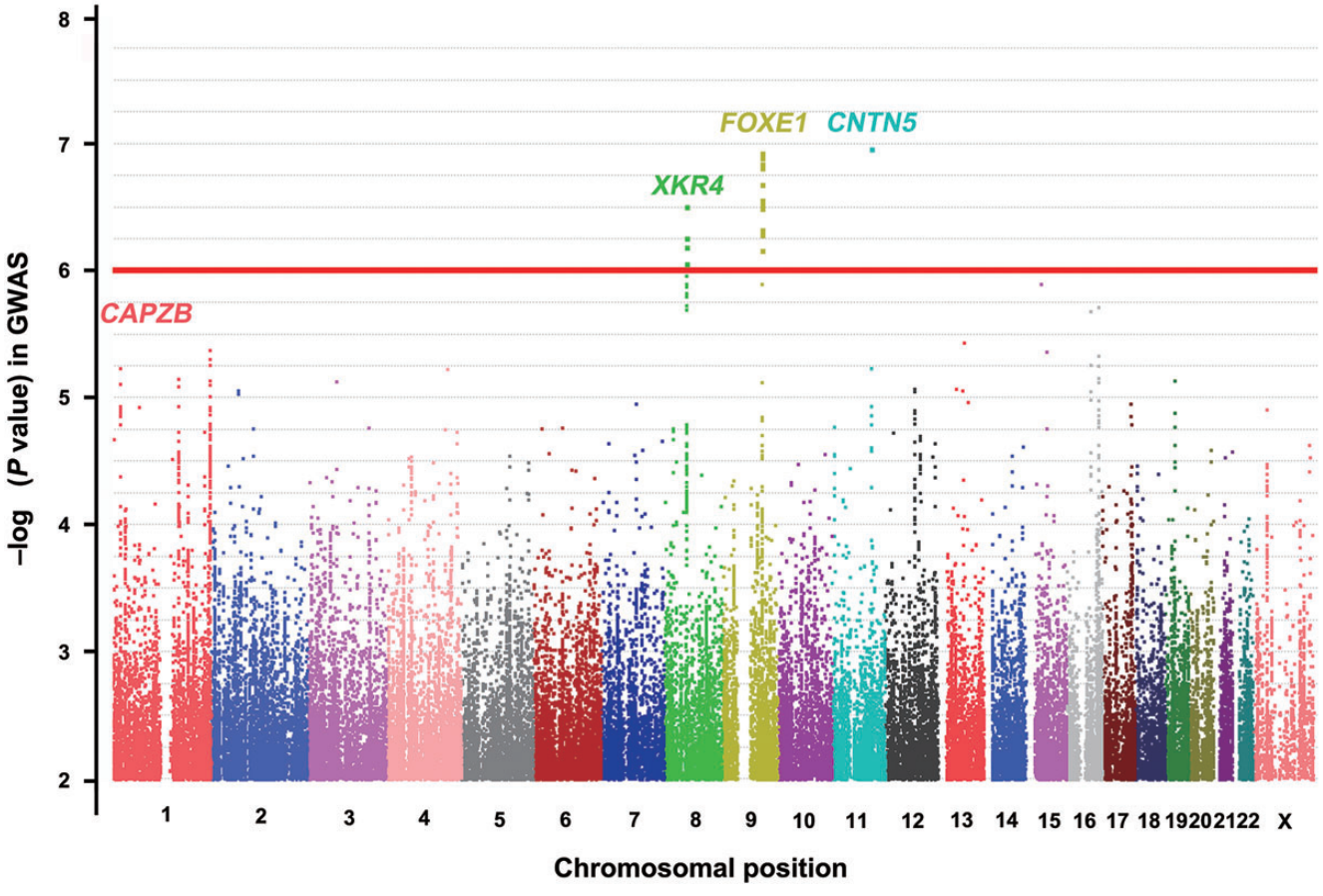

Figure 2. Manhattan plots of the SNPs associated with serum TSH levels in the GWAS data. Manhattan plots of associations between SNPs and TSH levels. Typed and imputed SNPs ($P < 10^{-2}$) are represented on the *X*-axis, organized by chromosome. On the *Y*-axis, statistical significance is expressed as -$\log_{10}$ *P* values. The red horizontal line represents the discovery cohorts (1346 subjects) threshold *P*-value of 10^{-6}.

69 kb upstream of *FOXE1*, and 87 kb upstream of *XPA* (Fig. 3B). Notably, *FOXE1* is specifically expressed in human thyroid tissues (Fig. 4C).

Rs6683419, located in the 1p36.13 region harboring *CAPZB, MINOS1, NBL1* and other genes, was previously reported to be associated with serum TSH levels in European population (13,15–17). Our initial GWAS provided the genotypes of 1300 imputed SNPs and 86 genotyped SNPs in a ~400 kb LD block in the 1p36.13 region. Among all genotyped SNPs, rs6683419 had the most significant association with TSH levels ($P_{GWAS} = 1.33 \times 10^{-5}$) (Supplementary Material, Table S5). We therefore analyzed rs6683419 in the replication study (Fig. 3C and Table 3). After the results of the discovery and replication studies were combined, rs6683419 showed a suggestive serum TSH association that nearly met the genome-wide significance level ($P_{replication} = 5.86 \times 10^{-4}$; $P_{combined} = 2.90 \times 10^{-7}$; Table 3). Rs10917469, rs10799824 and rs10917477 in the 1p36.13 region, which were previously reported as the best SNPs associated with serum TSH levels in European cohort, respectively (13,15–17), were further analyzed in our GWAS. Although rs10917469 and rs10799824 were only nominally associated with TSH levels (both $P_{GWAS} = 0.015$), we did find rs10917477 was strongly associated with TSH levels in our Chinese Han population ($P_{GWAS} = 4.19 \times 10^{-4}$) (Supplementary Material, Table S5). Moreover, we found that rs10917469 and rs10799824 were not in complete LD with rs6683419 (both $r^2 = 0.056$ in the 1000 Genomes ASN samples), otherwise rs10917477 was strongly correlated with rs6683419 ($r^2 = 0.744$ in the 1000 Genomes ASN samples) (Table 4).

Though our GWAS implicated *XKR4, FOXE1* and *CAPZB* gene with a significant effect on variation of serum TSH levels, other variants also contribute to this quantitative trait. Therefore, we further analyzed the 45 SNPs in the 26 loci, which were reported to be associated with TSH levels in previous studies, and found that out of the 45 SNPs, 27 in 15 risk loci were associated with TSH levels in our GWAS data ($P < 0.05$) (Table 4). Interestingly, there were 36 loci harbored more than one SNPs associated with TSH level at the $P < 1 \times 10^{-4}$ in our GWAS data, which included four loci reported to be associated with TSH in previous studies. However, only nine out of 280 loci associated with TSH level at $P < 1 \times 10^{-3}$ in our GWAS data , were confirmed to be the susceptibility loci for TSH in previous studies (Table 4) (11–17).

Association of the TSH candidate SNPs with thyroid diseases and expression patterns in target tissues

Serum TSH levels were previously reported to be correlated with the risk of thyroid cancer and could be a marker to predict thyroid cancer relapse (24–26). Hence, we investigated whether the four

Table 2. Most significant associations for TSH levels from the GWAS cohorts

Chr.	SNP	Position	A1/A2	MAF	β^b	t	P-value[a]	Annotated gene	Distance
1p36.13	rs6683419	19827780	G/A	0.257	0.039	4.371	1.33×10^{-5}	*CAPZB*	15.7 kb
1q44	rs12033048	246429681	T/C	0.436	−0.034	−4.347	1.49×10^{-5}	*SMYD3*	Intron
2p14	rs11681944	68322579	T/C	0.437	−0.035	−4.451	9.24×10^{-6}	*C1D*	32.4 kb
7q21.12	rs43123	87876090	G/A	0.171	0.051	4.412	1.11×10^{-5}	*SRI*	19.8 kb
8q12.1	rs2622590	56358274	A/G	0.406	−0.039	−4.999	6.54×10^{-7}	*XKR4*	Intron
8q12.1	rs2929029	56387105	G/A	0.247	−0.037	−4.823	1.58×10^{-6}	*XKR4*	Intron
9q22.33	rs925489	100546600	C/T	0.461	−0.074	−5.138	3.19×10^{-7}	*FOXE1*	68.9 kb
9q22.33	rs7850258	100549013	G/A	0.125	−0.074	−5.138	3.19×10^{-7}	*FOXE1*	66.5 kb
9q22.33	rs965513	100556109	C/T	0.077	−0.070	−4.868	1.26×10^{-6}	*FOXE1*	59.4 kb
9q22.33	rs10759944	100556972	A/G	0.077	−0.070	−4.868	1.26×10^{-6}	*FOXE1*	58.6 kb
9q22.33	rs7870926	100796945	A/G	0.076	0.041	4.497	7.49×10^{-6}	*ANP32B*	18.7 kb
11q22.1	rs11223396	100118859	A/G	0.076	−0.074	−4.366	1.36×10^{-5}	*CNTN5*	Intron
11q22.1	rs1348271	100137346	T/C	0.053	−0.087	−5.014	6.04×10^{-7}	*CNTN5*	Intron
16q21	rs2582597	60125785	G/T	0.051	−0.046	−4.565	5.44×10^{-6}	NA	Intergenic
16q23.2	rs11866972	79738425	A/G	0.233	0.034	4.423	1.05×10^{-5}	*MAF*	103.8 kb
17q25.1	rs4789661	72425159	T/C	0.397	−0.039	−4.411	1.11×10^{-5}	*GPR142*	2.5 kb

SNP, single-nucleotide polymorphism; Chr., chromosome; A1/A2, Min-allele/Max-allele; MAF, minor allele frequency; NA, not assessed.
[a]*P*-values are from linear regression models adjusted for age, gender, and geographic region.
[b]β-values estimate the difference in TSH level and are obtained from linear regression models in PLINK.

serum-TSH candidate SNPs identified in our initial GWAS were associated with papillary thyroid carcinoma (PTC) in a sample of 108 patients with PTC and 1490 control individuals from the Chinese Han population. Two of the SNPs, rs2622590 and rs925489, were nominally associated with PTC (rs2622590, $P = 0.014$, OR = 1.41, 95% CI = 1.07–1.87; rs925489, $P = 0.030$, OR = 1.61, 95% CI = 1.04–2.47; Table 5).

Although we apply a Bonferroni correction for multiple testing of association between SNPs (rs2622590, rs1348271, rs6683419 and rs925489) and PTC (P 0.05/4 = 0.0125), rs965513 near *FOXE1* on 9q22.33 ($P = 0.030$) and rs2622590 in *XKR4* on 8p12.1 ($P = 0.014$) were no longer associated with PTC. However, we found that rs965513 in *FOXE1* was significantly associated with PTC in Chinese population in a previous report ($P = 1.18 \times 10^{-4}$, OR = 1.53, 95% CI = 1.23–1.90; Supplementary Material, Table S6) (27). Because rs965513 was in complete LD with rs925489 ($r^2 = 0.97$ in our Chinese Han samples), we performed a meta-analysis and found that rs925489 or rs965513 was significantly associated with PTC ($P_{\text{Meta}} = 1.28 \times 10^{-5}$, OR = 1.54, 95% CI = 1.27–1.88) in combined Chinese PTC population (Supplementary Material, Table S6). We have neither more PTC samples nor reported data about the genotypes of SNPs in *XKR4* region to further confirmed the association of rs2622590 with PTC. Therefore, the relation of the *XKR4* SNP to PTC remains to be determined, and further studies based on larger PTC samples to confirm these two findings were needed.

The PTC risk alleles of the two SNPs (the TT genotype for rs2622590 and the CC genotype for rs925489) were also associated with low serum TSH levels (Table 5 and Supplementary Material, Table S3). None of the four candidate SNPs were associated with Graves' disease (GD), however, in a sample of 1442 patients with GD and 1468 control individuals collected from the Chinese Han population in our previous GWAS (18) (Table 5).

We used quantitative real-time PCR to measure the expression of *XKR4*, *FOXE1* and *CAPZB* in 66 PTC tissues and 40 normal thyroid tissues adjacent to PTC tissues. We found that the expression levels of *XKR4, FOXE1* and *CAPZB* in the PTC tissues were significantly higher than those in the normal thyroid tissues adjacent to PTC tissues ($P = 3.85 \times 10^{-7}$, 3.24×10^{-4}, and 2.67×10^{-15}, respectively; Fig. 4A). Moreover, the rs2622590_T and rs925489_C alleles, which were the risk alleles for PTC and were correlated with lower serum TSH levels, were also correlated with high expression levels of *XKR4* and *FOXE1*, respectively, in the thyroid tissues of patients with PTC ($P_{\text{ANOVA}} = 2.41 \times 10^{-4}$ for rs2622590; $P_{\text{ANOVA}} = 0.02$ for rs925489; Fig. 4B). The rs6683419 genotypes, however, were not correlated with the expression of *CAPZB* in the PTC tumor tissues ($P_{\text{ANOVA}} = 0.65$; Fig. 4B).

DISCUSSION

The identification of susceptibility loci and genes related to serum TSH levels may provide important insight into the regulation of thyroid hormones, and may also be valuable for designing new preventative and therapeutic strategies for the disorders influenced by TSH, such as hypothyroidism, dyslipidemia, cardiovascular disease, myocardial infarction and atrial fibrillation (28–34). Through a two-stage GWAS of circulating TSH levels in individuals from the Chinese populations, we identified a novel susceptibility locus in *XKR4* at 8p12.1, and we confirmed two previously reported susceptibility loci near *FOXE1* at 9q22.33 and near *CAPZB* at 1p36.13 (12–17).

This study is the first to associate rs2622590, located in intron 2 of *XKR4* at 8p12.1, with serum TSH levels. The *XKR4* gene encodes a member of the XK, Kell blood group complex subunit-related protein family. Kell and XK are two distinct red blood cell-membrane proteins that form the Kell blood-group complex. XK is considered to be a membrane transport protein, and the absence of XK can lead to McLeod syndrome, a rare X-linked hereditary disease characterized by red blood cell acanthocytosis and late-onset central nervous system and neuromuscular abnormalities (35). More recently, a few SNPs

Table 3. Association results for the selected SNPs in the GWAS and replication phases

SNP	Chr.	Position	A1/A2	Nearest gene	GWAS ($n = 1346$)			Replication ($n = 3235$)				Combined ($n = 4581$)			
					MAF	β^a	*P*-value[b]	*n*	MAF	β^a	*P*-value[b]	*n*	MAF	β^a	*P*-value[b]
Novel loci															
rs2622590	8q12.1	56358274	A/G	*XKR4*	0.436	−0.039	6.49×10^{-7}	3,077	0.449	−0.026	5.62×10^{-6}	4,423	0.445	−0.030	2.21×10^{-10}
rs1348271	11q22.1	100137346	T/C	*CNTN5*	0.051	−0.092	1.09×10^{-7}	3,120	0.027	−0.012	4.89×10^{-1}	4,466	0.034	−0.046	4.14×10^{-4}
Known loci															
rs6683419	1p36.13	19827780	G/A	*CAPZB*	0.257	0.039	1.33×10^{-5}	3,159	0.268	0.022	5.86×10^{-4}	4,505	0.265	0.027	2.90×10^{-7}
rs925489	9q22.33	100546600	C/T	*FOXE1*	0.077	−0.074	3.19×10^{-7}	3,027	0.122	−0.053	9.13×10^{-9}	4,373	0.109	−0.058	1.02×10^{-13}

SNP, single-nucleotide polymorphism; Chr, chromosome; A1/A2, Min-allele/Max-allele; MAF, minor allele frequency.
[a]β-values estimate the difference in TSH level and are obtained from linear regression models in PLINK.
[b]*P*-values are from linear regression models adjusted for age, gender, geographic region and race.

near *XKR4* at the 8p12.1 locus have been reported to be associated with attention deficit/hyperactivity disorders, responses to iloperidone and risperidone, and paclitaxel-induced peripheral sensory neuropathy (36–39). *XKR4* is abundantly expressed in the tissues of the brain, esophagus and stomach, but is relatively weakly expressed in the tissues of the thyroid, kidney and pancreas. We found that the expression level of *XKR4* was significantly higher in the tumor tissues of patients with PTC than in the adjacent normal thyroid tissues. The rs2622590 genotypes were correlated with the expression of *XKR4* in the tumor tissues of patients with PTC. Determining whether the rs2622590 risk allele mediates the regulation of serum TSH levels by *XKR4*, however, will require further study.

We found that the rs2622590_T and rs925489_C alleles, which were associated with lower serum TSH levels, might be risk alleles for patients with PTC. These alleles were significantly correlated with the enhanced expression of *XKR4* and *FOXE1*, respectively, in the PTC tumor tissues. *FOXE1*, located about 69 kb downstream of rs925489, is important for both pituitary-gland development and thyroid-gland development (40–42). *FOXE1* is also necessary for the synthesis of the thyroid hormones T3 and T4, because it regulates the transcription of the *TG* (thyroglobulin) and *TPO* (thyroperoxidase) genes by binding to response elements in the promoter regions (43,44).

It is well known that the product of *TG* is the precursor of T3 and T4, the synthesis of which is catalyzed by *TPO*. Therefore, it is tempting to presume that the C allele of rs925489 increases the expression of *FOXE1* in thyroid tissues, up-regulating *TG* and *TPO* transcription and enhancing the synthesis of T3 and T4; the increased synthesis of thyroid hormones acts in a negative-feedback loop, further decreasing the serum concentration of TSH, leading to the association of the C allele of rs925489 with lower TSH levels. Indeed, all the four PTC risk alleles except rs965513 near to *FOXE1* were associated with the lower TSH levels and trended to increased FT4 levels. Moreover, for rs966423 on 2q35 and rs116909374 on 14q13.3, the lower-TSH alleles were significantly associated with increase FT4 at the $P < 0.05$ level (15). Although the PTC risk allele of rs965513 in *FOXE1* was associated with lower TSH and FT4 levels, the risk allele appeared to be associated with the increase FT3 level ($P < 0.05$) (15). Therefore, all the data suggested that the PTC risk allele trended to lower the TSH and increase the physiological function of thyroid as well as supported our hypothesis. With regard to the SNPs near to *FOXE1* at 9q22.33, ours and previous reports supported that rs925489, a proxy rs965513 ($r^2 = 0.97$ in our Chinese Han samples), was associated with TSH concentration at the GWAS significant level (12,15). However, the more recent study from Porcu *et al.* (17) found that rs7045138 and rs965513 near *FOXE1* were associated with FT4 level ($P = 1.5 \times 10^{-11}$ and 3.45×10^{-8}, respectively), but not associated with the TSH level. Interestingly, the rs965513 was weakly linked with rs7045138 in Caucasian, but not linked with rs7045138 in Chinese Han population ($r^2 =$ 0.409 in the 1000 Genomes CEU samples and 0.001 in the 1000 Genomes ASN samples, respectively). In our GWAS data, rs965513, but not rs7045138 was significantly associated with TSH level ($P = 1.26 \times 10^{-6}$ and 0.70, respectively). These data suggested that two SNPs near *FOXE1* region were independently associated with the TSH or FT4 levels, respectively. Regretfully, the FT4 levels have not been collected in our current

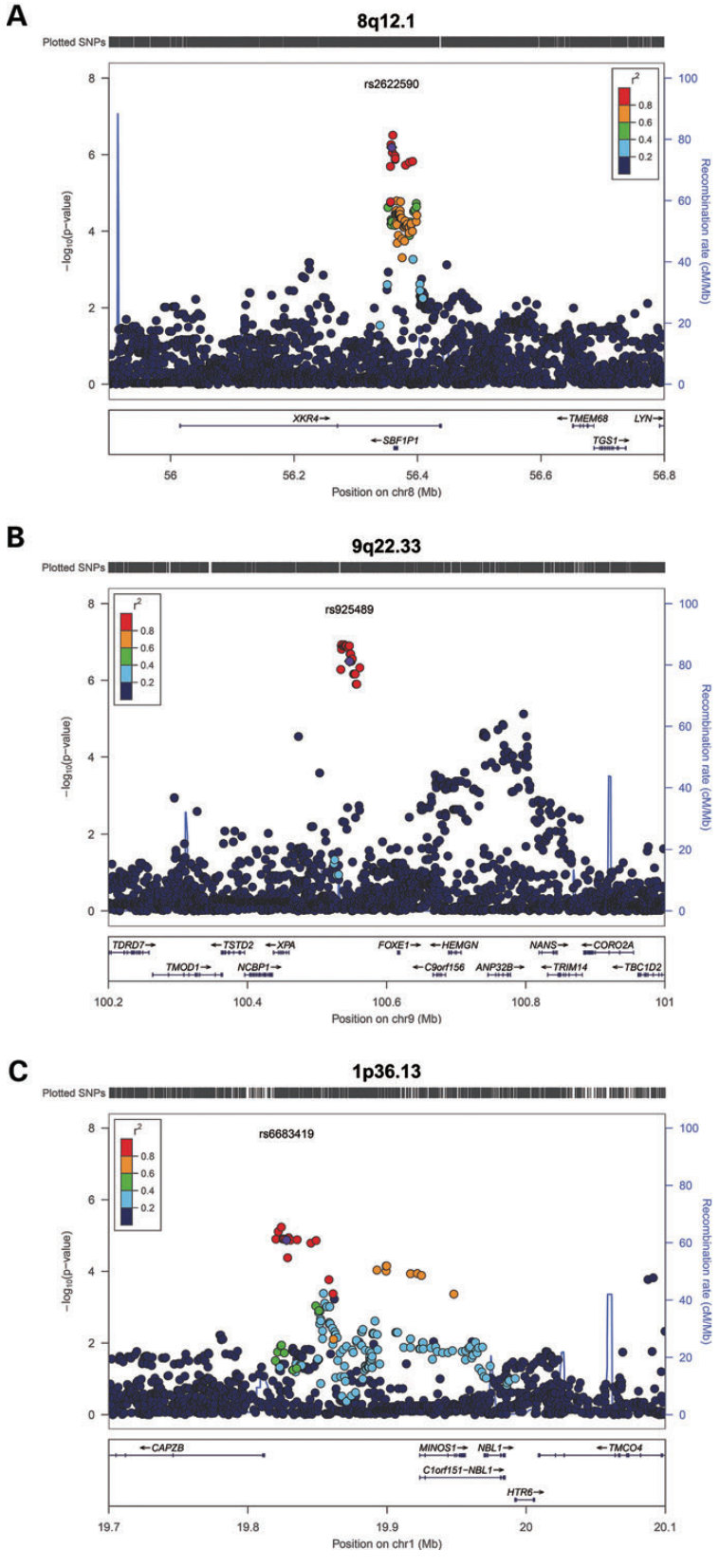

Figure 3. Regional plots of association results at 8p12.1, 9q22.33 and 1p36.13. Regional plots of association results in the GWAS samples at 8p12.1 (**A**), 9q22.33 (**B**) and 1p36.13 (**C**). The color of each SNP spot, ranging from red to blue, reflects the magnitude of the r^2 between the SNP spot and the candidate SNP (purple circle) within each candidate locus for serum TSH levels. Genetic recombination rates, estimated using the 1000 Genomes ASN samples, are plotted in the background in cyan. The vertical axis represents the -$\log_{10}$ *P*-value. Physical positions are based on NCBI build 36/hg19 of the human genome.

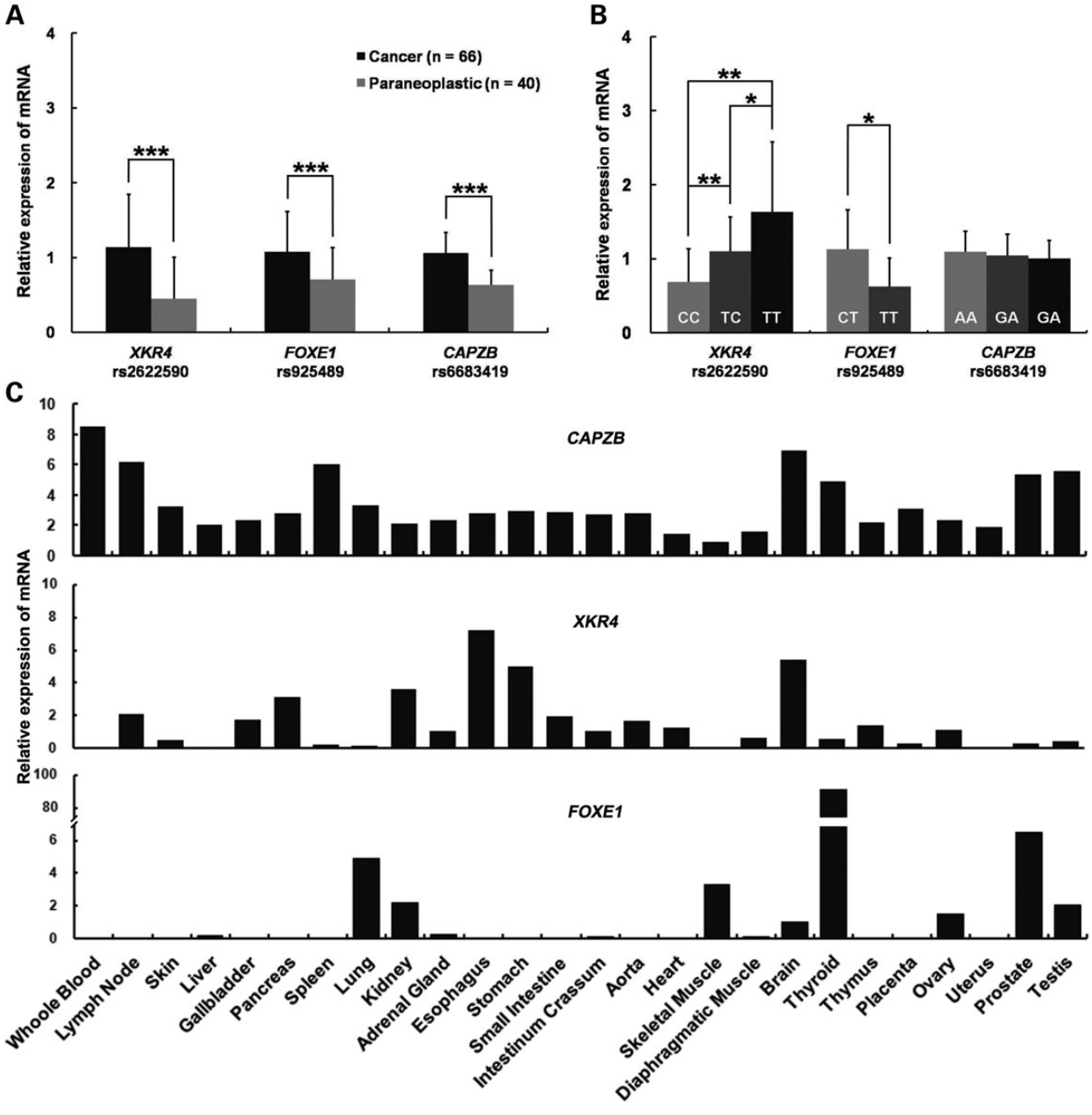

Figure 4. The expression analyses of the candidate genes *XKR4*, *FOXE1* and *CAPZB* at 1p36.13, 8p12.1 and 9q22.33. (**A**) The mRNA levels of *XKR4*, *FOXE1* and *CAPZB* in the tumor tissues ($n = 66$) and adjacent normal thyroid tissues ($n = 40$) of patients with PTC. Data are shown as means ± SD. The mRNA levels of *XKR4*, *FOXE1* and *CAPZB* are normalized to that of *GAPDH*. (**B**) The correlation of *XKR4, FOXE1* and *CAPZB* mRNA levels in PTC tumor tissues with the genotypes of rs2622590 in *XKR4* (CC, $n = 14$; TC, $n = 34$; TT, $n = 18$), rs925489 near *FOXE1* (TT, $n = 59$; CT, $n = 7$) and rs6683419 near *CAPZB* (AA, $n = 28$; GA, $n = 30$; GG, $n = 8$), respectively (normalized to *GAPDH*). (**C**) The expression patterns of *CAPZB*, *XKR4* and *FOXE1* in human tissues resolved by real-time PCR. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. Error bars, ± SD.

study, the association study of the SNPs in *FOXE1* region with FT4 and TSH levels in the Chinese population in further to unlock the perplexed was expected.

The rs2622590_T allele, which was associated with lower serum TSH levels, was also correlated with the increased expression of *XKR4* in thyroid tumor tissues; however, the relationship between serum TSH levels and *XKR4* expression in thyroid tissues remains unknown. Consistent with the results of our study, the alleles of four other SNPs associated with lower serum TSH levels were found to increase the risk of PTC in previous studies (12,15). All of the available data suggest that low serum TSH levels are correlated with an increased risk of PTC.

In summary, we identified a new serum TSH susceptibility locus in the *XKR4* region at 8q12.1 and confirmed two previously reported susceptibility loci near *FOXE1* at 9q22.33 and *CAPZB* at 1p36.13, respectively. The rs2622590_T allele in *XKR4* and

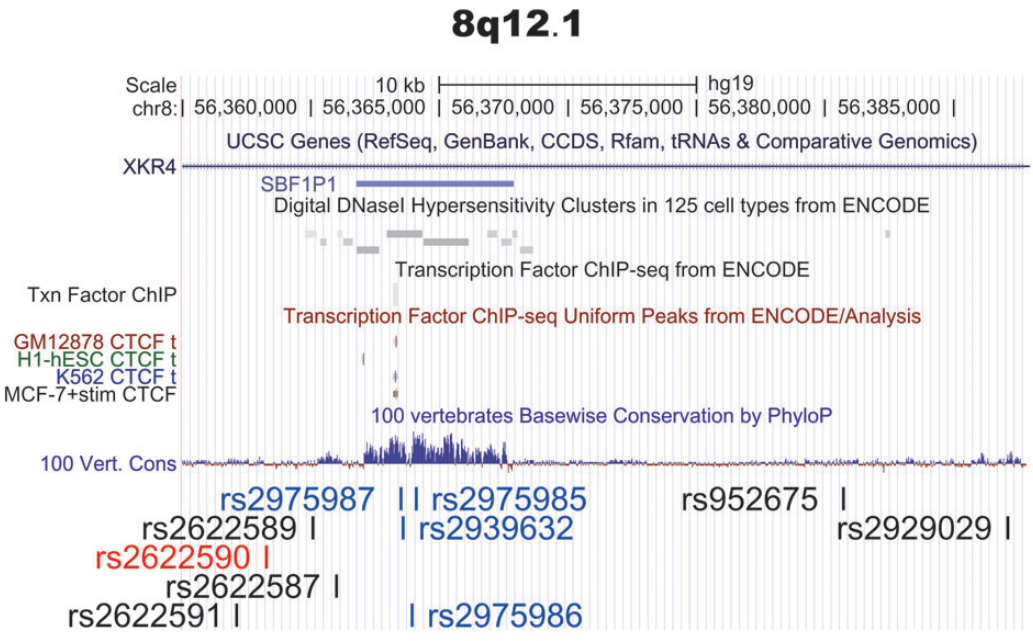

Figure 5. The epigenetic analysis of *XKR4* at 8q12.1 from the ENCODE database. Rs2622590 shows no DNase I hypersensitivity and does not bind any transcription factors. However, a ~5 kb fragment near to rs2622590, which contained four TSH strongly associated SNPs (rs2975987, rs2939632, rs2975986 and rs2975985), was conserved from zebrafish to human. The data from ENCODE manifested a ~35 kb LD block in the 8q12.1 region containing rs2975987, rs2939632, rs2975986 and rs29759854 (blue SNPs) were located in a DNaseI hypersensitivity clusters, which were in complete LD with rs2622590 ($r^2 > 0.97$ in the 1000 Genomes ASN samples). SNPs in red font are candidate SNPs in our study. SNPs in black font are in high LD with candidate SNPs ($r^2 > 0.8$ in the 1000 Genomes ASN samples).

the rs925489_C allele near *FOXE1*, both associated with lower serum TSH levels, were correlated with the increased expression of *XKR4* and *FOXE1*, respectively, in the thyroid tumor tissues, and possibly conferred a predisposition to PTC. These findings provided new insights into the mechanisms regulating serum TSH levels and suggested that low serum TSH levels were correlated with susceptibility to PTC.

MATERIALS AND METHODS

GWAS and replication cohorts

The GWAS cohorts included in this study were collected from the Chinese Han population through collaborations with two hospitals in northern China (Xuzhou and Linyi) (18,45,46). The replication cohort was collected from the Chinese She population, who is one of the important ethnic minorities, residing in the Ningde City, Fujian Province in southern China. The marriage between Han and She were prohibited 60 years ago, suggesting the She ethnic is an independent isolated population in China (19). Subjects were excluded if they had thyroid nodules, autoimmune thyroid disease (AITD) or a family history of AITD or other autoimmune disorders.

During a physical examination, 5 ml blood samples were collected from each participant and stored at −80°C for DNA extraction and measurements of serum TSH sensitivity and TPO-Ab levels. TSH and TPO-Ab were measured using a chemiluminescence immunoassay (CLIA) according to the manufacturer's instructions in our laboratory (Siemens ADVIA Centaur CP). To exclude subjects with clinical or subclinical AITD and subjects taking medications likely to influence serum thyroid function, we measured the serum levels of TPO-Ab among the 1816 GWAS subjects and the 4233 replication subjects. Based on those measurements, we excluded 1420 individuals with TPO-Ab levels <5.60 U/ml and TSH levels <4.93 mU/l or <0.36 mU/l. Ultimately, we included 1394 and 3235 subjects for the GWAS and replication studies, respectively (Fig. 1). The median TSH level was 1.62 mU/l and the interquartile range was 1.02–2.06 mU/l. We also included 108 patients with PTC recruited from the First hospital affiliated to Bengbu Medical College, in northern China to evaluate whether selected loci were associated with PTC. The sample sizes, sex distributions, mean ages, BMIs and serum TSH levels of the GWAS and replication participants are summarized in Table 1.

All of the participating cohorts were granted approval by the local Ethics Committee from Ruijin Hospital, the Central Hospital of Xuzhou, Linyi People's Hospital, the Fujian Province Hospital Affiliated to Fujian Medical University and the First Hospital Affiliated to Bengbu Medical College. All of the subjects in this study provided written informed consent using protocols approved by the local Ethics Committees.

GWAS genotyping and initial quality control

DNA was extracted from the blood samples using a FUJIFILM QuickGene-610L system. Genome-wide genotyping was performed using the Illumina Human Omni-Express BeadChip platform on the Illumina Human660-Quad BeadChips at the Chinese National Human Genome Center in Shanghai, China (18). Genotype clustering was conducted using Illumina GenomeStudio V2011.1 software based on the 660W-Quad_v1_H manifest files, which converted the fluorescence intensities into SNP genotypes. The mean call rate across all 1394

Table 4. The association of the top SNPs in 26 previously reported TSH risk loci with TSH in our GWAS data

Chr.	Marker name	Annotated Gene	Tested SNP	*P*-value[a]	r^{2}[b]	Top SNP[c]	*P*-value[a]
1p36.13	rs10799824	*CAPZB*	rs10799824	1.51×10^{-2}	0.056	rs2314146	5.80×10^{-6}
	rs10917469	*CAPZB*	rs10917469	1.49×10^{-2}	0.056		
	rs10917477	*CAPZB*	rs10917477	4.19×10^{-4}	0.744		
1p31.3	rs334725	*NFIA*	rs334725	1.78×10^{-2}	0.005	rs334707	2.86×10^{-2}
	rs334699	*NFIA*	rs334699	2.75×10^{-2}	0.004		
1p13.3	rs17020124	*VAV3*	rs17020124	3.41×10^{-3}	0.734	rs9787296	4.66×10^{-4}
2p25.3	rs11694732	*TPO*	rs11694732	7.10×10^{-1}	0.098	rs4927648	2.74×10^{-2}
2q35	rs737308	*IGFBP5*	rs737308	4.30×10^{-2}	0.702	rs2712172	1.25×10^{-2}
	rs13015993	*IGFBP5*	rs13015993	3.21×10^{-2}	0.727		
4q31.23	rs10028213	*NR3C2*	rs10028213	7.48×10^{-3}	0.266	rs9968300	1.77×10^{-5}
	rs10030849	*NR3C2*	rs10030849	7.48×10^{-3}	N/A		
	rs10032216	*NR3C2*	rs10032216	7.48×10^{-3}	N/A		
5q13.3	rs4704397	*PDE8B*	rs4704397	2.14×10^{-2}	0.001	rs10055027	1.36×10^{-3}
	rs6885099	*PDE8B*	rs6885099	7.55×10^{-3}	0		
	rs2046045	*PDE8B*	rs2046045	8.13×10^{-3}	0		
6p21.1	rs729761	*VEGFA*	NA	NA	N/A	rs10948107	9.57×10^{-4}
	rs9472138	*VEGFA*	rs9472138	3.39×10^{-2}	0		
	rs6923866	*VEGFA*	rs6923866	1.85×10^{-3}	0.915		
	rs11755845	*VEGFA*	rs11755845	1.60×10^{-3}	0.831		
6q24.3	rs9497965	*SASH1*	rs9497965	1.75×10^{-1}	0.031	rs58488869	8.03×10^{-3}
6q27	rs753760	*PDE10A*	rs753760	7.39×10^{-2}	0.825	rs58885758	7.15×10^{-3}
	rs3008043	*PDE10A*	rs3008043	2.42×10^{-2}	0.866		
8p12	rs7825175	*NRG1*	rs7825175	1.68×10^{-2}	0.685	rs11991469	6.56×10^{-3}
	rs2439302	*NRG1*	rs2439302	6.89×10^{-3}	0.947		
9p24.2	rs1571583	*GLIS3*	rs1571583	4.17×10^{-2}	0.064	rs11999791	2.62×10^{-2}
9q22.33	rs965513	*FOXE1*	rs965513	1.26×10^{-6}	0.777	rs10818050	1.17×10^{-7}
9q34.2	rs657152	*ABO*	rs657152	8.81×10^{-1}	0.091	rs10901263	6.60×10^{-4}
10q24.2	rs7913135	*NKX2-3*	rs7913135	2.27×10^{-1}	1.000	rs7913750	1.97×10^{-1}
11p11.2	rs17723470	*PRDM11*	rs17723470	2.66×10^{-2}	N/A	rs72907072	2.57×10^{-4}
	rs7128207	*PRDM11*	rs7128207	2.83×10^{-1}	N/A		
12q23.1	rs61938844	*ELK3*	NA	NA	N/A	rs12814360	2.16×10^{-2}
14q13.3	rs1537424	*MBIP*	rs1537424	2.48×10^{-1}	0.217	rs2780310	5.64×10^{-4}
	rs116909374	*MBIP*	NA	NA	N/A		
14q32.12	rs34269820	*ITPK1*	rs34269820	3.79×10^{-2}	0.211	rs1263428	5.62×10^{-3}
	rs11624776	*ITPK1*	rs11624776	1.04×10^{-1}	0.166		
14q32.33	rs73362602	*SIVA1*	rs73362602	5.47×10^{-1}	N/A	rs12896579	1.28×10^{-2}
15q21.2	rs73398284	*FGF7*	rs73398284	8.00×10^{-2}	N/A	rs74406736	5.55×10^{-3}
	rs10519227	*FGF7*	rs10519227	1.76×10^{-1}	N/A		
15q26.1	rs17776563	*MIR1179*	rs8030854 ($r^{2} = 0.94$)	2.49×10^{-1}	0.007	rs74027996	6.89×10^{-3}
16p23.2	rs7190187	*MAF*	rs7190187	1.18×10^{-4}	0.71	rs4575545	1.91×10^{-6}
	rs3813582	*MAF*	rs3813582	1.50×10^{-5}	0.955		
17q24.3	rs9915657	*SOX9*	rs9915657	2.54×10^{-1}	0.021	rs16977126	1.20×10^{-2}
19p13.2	rs4804416	*INSR*	rs4804416	3.74×10^{-2}	0.108	rs4804433	1.47×10^{-2}
	rs10420008	*INSR*	rs10420008	6.52×10^{-2}	0.332		
20p11.21	rs6082762	*FOXA2*	rs6082762	2.91×10^{-1}	0.609	rs4499503	3.26×10^{-2}

Chr., chromosome; SNP, single-nucleotide polymorphism; NA, not assessed.
[a]*P*-values were from linear regression models adjusted for age, gender and geographic region.
[b]r^{2} in the 1000 Genomes ASN samples.
[c]Top SNP; the top association signal in each TSH risk locus in our data (recombination rate < 0.3 in the 1000 Genomes ASN samples).

genotyped samples was 99.8%. No individuals were excluded from the GWAS analysis. Quality control was performed on the SNPs and samples before analysis using the PLINK software to ensure robust association tests (47).

Of the 655 214 markers assayed, 3185 were removed because they were from the Y or mitochondrial chromosomes or were CNV-related. We discarded 168 082 markers because of deviation from Hardy–Weinberg equilibrium $P \leq 10^{-6}$, genotype call rates below 98%, or MAFs < 0.01, leaving 483 947 SNPs for subsequent analyses. Next, we excluded 48 samples with call rates $<98\%$, gender inconsistencies or cryptic relatedness. Ultimately, 1346 samples were available for the GWAS analysis (Fig. 1).

Genotype imputation

To infer the genotypes of SNPs in the GWAS cohort that were not previously genotyped, we used the IMPUTE2 software (48) with the 1000G phase-1 interim impute data (March 2012) as a reference. We used an estimated probability >0.90 to call imputed genotypes. Taking imputation uncertainty into account, the association analysis was carried out using the SNPTESTv2 software (frequentist) association tests with score method (49).

SNP selection and genotyping in the replication study

We selected three SNPs (rs1348271, rs925489 and rs2622590) that were strongly associated with serum TSH levels ($P < 1 \times 10^{-6}$)

Table 5. Association of the selected SNPs with PTC risk and GD

SNP	PTC association analysis (108 versus 1490)					GD association analysis (1442 versus 1468)				
	Risk allele	Case MAF	Control MAF	OR (95% CI)	*P*-value	Risk allele	Case MAF	Control MAF	OR (95% CI)	*P*-value
rs2622590	T	0.519	0.432	1.41 (1.07–1.87)	0.014	C	0.426	0.439	1.03 (0.93–1.14)	0.567
rs1348271	G	0.046	0.051	1.11 (0.58–2.14)	0.757	T	0.046	0.050	1.10 (0.86–1.40)	0.445
rs6683419	G	0.315	0.262	1.29 (0.96–1.74)	0.092	G	0.265	0.264	1.00 (0.89–1.12)	0.935
rs925489	C	0.120	0.079	1.61 (1.04–2.47)	0.030	C	0.089	0.078	1.16 (0.72–1.04)	0.125

SNP, single-nucleotide polymorphism; PTC, papillary thyroid carcinoma; GD, Graves' disease; OR, odds ratio for the risk allele; CI, confidence interval.

as candidate loci for the replication study (rs7850258 was in complete LD with rs925489, $r^2 = 1.00$ in our Chinese Han samples). We also selected SNP rs6683419 ($P = 1.33 \times 10^{-5}$) at 1p36.13, which was previously reported (13) to be associated with serum TSH levels. In the replication study, the DNA concentrations of all samples were standardized to 50 ng/μl in 96-well plates, and one negative control (DNase-free and RNase-free water) was included in one of the 96 wells at random. Four SNPs were genotyped using the ABI ViiA^{TM} 7 Real-Time PCR System according to the manufacturer's protocol by technicians who were blinded to the sample status. The average call rate was >95%.

Statistical analysis

To reduce potential bias caused by deviation of TSH levels from a normal distribution, we applied the natural-log transformation to normalize the distributions of all phenotypic data before testing for associations. A standard linear regression model was used to analyze the association of each SNP with quantitative traits, assuming an additive genetic model, which was implemented in the PLINK software package (47). For typed or imputed genotypes, the allelic dosage at each SNP was the independent variable, adjusted for primary covariates of age, gender, geographic region and race. A *P*-value of 5×10^{-8} was used as the cutoff for genome-wide significance. The effects of selected SNPs on PTC risk were further evaluated for the case–control samples using logistic regression models.

PCA and MDS analysis were implemented using the EIGENSTRAT (50) and PLINK (47) software, respectively, to evaluate population stratification in the GWAS samples. The Haploview software version 4.2 was used to generate the genome-wide plot of *P*-values, and regional plots were generated using the LocusZoom Version 1.1 software (51). Quantile–quantile plots were constructed by plotting the observed distribution of the *P*-values for the given SNPs against the theoretical distribution of the expected *P*-values. The calculations and the plots were implemented using the PLINK and R statistical packages.

mRNA expression analyses

In the mRNA expression analysis, 66 PTC samples and 40 paraneoplastic tissue samples were surgically collected from patients. Total RNA was isolated using Trizol reagent (Invitrogen) according to the manufacturer's instructions and then treated with DNase I at room temperature for 10 min to degrade possible contaminating genomic DNA. cDNAs were made from 1 μg RNA templates using reverse transcriptase and oligo(dT) primer (Takara). Quantitative real-time PCR for a series of genes was performed in triplicate using the SYBR Green and ABI ViiA^{TM} 7 Real-Time PCR System. Data were analyzed and presented relative to the expression of the GAPDH housekeeping gene. The primer sequences used for real-time PCR are shown in the Supplementary Material, Table S7. We performed statistical analysis of the expression data using an ANOVA and an unpaired Student's *t*-test (the two-tail *P*-values are indicated on the figures).

URLs

PLINK v1.07, http://pngu.mgh.harvard.edu/~purcell/plink/

R statistical environment version 2.10.0, http://www.r-project.org/

LocusZoom Version 1.1, http://csg.sph.umich.edu/locuszoom/

IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html/

SNPTEST v2 software, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html/

ENCODE database, http://genome.ucsc.edu/ENCODE/

AUTHOR CONTRIBUTIONS

H.-D.S. was responsible for the coordination of the project. H.-D.S., J.-L.C, S.-J.C, G.N., C.-M.P., S.-X.Z. and M.Z. contributed to study design. The writing team consisted of H.-D.S., M.Z. and S.-X.Z. H.-D.S., M.Z., C.-M.P. and S.-X.Z. contributed to the project management. G.C., H.-D.S., L.J., G.-Q.G., X.-M.Z., C.-L.Z., B.S., C.-M.P., G.-Y.Y., W.L., S.-X.Z. and M.Z. took part in the collection of clinical samples, DNA and sample QC. W.H. contributed to whole genome genotyping. C.-M.P., M.Z., X.-P.Y., S.-S.Y. and H.-J.X. took part in replication genotyping. C.-M.P., M.Z. and X.-P.Y. contributed to *CAPZB*, *XKR4* and *FOXE1* real-time RT-PCR. Z.-H.G., S.-X.Z., H.-N.W. and M.Z. performed imputation analysis and took part in statistical analysis.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are indebted to the staff, patients and all other individuals involved in this study for their dedication and contributions.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by National Basic Research Program of China (973) (2010CB529204 and 2012CB517604), National Natural Science Foundation of China (81370888, 81100553, 31171127, 81270863, 81200568, 81370965, 81270864, 81272748 and 81101444), Program for Graves′ Disease Innovative Research Team of Shanghai Municipal Education Commission, National Natural Science Foundation of Jiangsu Province (BK2009208), National Natural Science Foundation of Shandong Province (ZH2011HM067), Natural Science Foundation of Anhui Province (1208085MH140) and Dr Innovation Fund of SJTU School of Medicine (BXJ201208).

REFERENCES

1. Porterfield, S.P. and Hendrich, C.E. (1993) The role of thyroid hormones in prenatal and neonatal neurological development--current perspectives. *Endocr. Rev.*, **14**, 94–106.
2. Asvold, B.O., Bjoro, T., Nilsen, T.I. and Vatten, L.J. (2007) Association between blood pressure and serum thyroid-stimulating hormone concentration within the reference range: a population-based study. *J. Clin. Endocrinol. Metab.*, **92**, 841–845.
3. Asvold, B.O., Vatten, L.J., Nilsen, T.I. and Bjoro, T. (2007) The association between TSH within the reference range and serum lipid concentrations in a population-based study. *The HUNT Study.* Eur. J. Endocrinol., **156**, 181–186.
4. Oppenheimer, J.H., Schwartz, H.L., Mariash, C.N., Kinlaw, W.B., Wong, N.C. and Freake, H.C. (1987) Advances in our understanding of thyroid hormone action at the cellular level. *Endocr. Rev.*, **8**, 288–308.
5. Larsen, P.R. (1982) Thyroid-pituitary interaction: feedback regulation of thyrotropin secretion by thyroid hormones. *N. Engl. J. Med.*, **306**, 23–32.
6. Vassart, G. and Dumont, J.E. (1992) The thyrotropin receptor and the regulation of thyrocyte function and growth. *Endocr. Rev.*, **13**, 596–611.
7. Keffer, J.H. (1996) Preanalytical considerations in testing thyroid function. *Clin. Chem.*, **42**, 125–134.
8. Andersen, S., Pedersen, K.M., Bruun, N.H. and Laurberg, P. (2002) Narrow individual variations in serum T(4) and T(3) in normal subjects: a clue to the understanding of subclinical thyroid disease. *J. Clin. Endocrinol. Metab.*, **87**, 1068–1072.
9. Hansen, P.S., Brix, T.H., Sorensen, T.I., Kyvik, K.O. and Hegedus, L. (2004) Major genetic influence on the regulation of the pituitary-thyroid axis: a study of healthy Danish twins. *J. Clin. Endocrinol. Metab.*, **89**, 1181–1187.
10. Samollow, P.B., Perez, G., Kammerer, C.M., Finegold, D., Zwartjes, P.W., Havill, L.M., Comuzzie, A.G., Mahaney, M.C., Goring, H.H., Blangero, J. *et al.* (2004) Genetic and environmental influences on thyroid hormone variation in Mexican Americans. *J. Clin. Endocrinol. Metab.*, **89**, 3276–3284.
11. Arnaud-Lopez, L., Usala, G., Ceresini, G., Mitchell, B.D., Pilia, M.G., Piras, M.G., Sestu, N., Maschio, A., Busonero, F., Albai, G. *et al.* (2008) Phosphodiesterase 8B gene variants are associated with serum TSH levels and thyroid function. *Am. J. Hum. Genet.*, **82**, 1270–1280.
12. 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.
13. Panicker, V., Wilson, S.G., Walsh, J.P., Richards, J.B., Brown, S.J., Beilby, J.P., Bremner, A.P., Surdulescu, G.L., Qweitin, E., Gillham-Nasenya, I. *et al.* (2010) A locus on chromosome 1p36 is associated with thyrotropin and thyroid function as identified by genome-wide association study. *Am. J. Hum. Genet.*, **87**, 430–435.
14. Medici, M., van der Deure, W.M., Verbiest, M., Vermeulen, S.H., Hansen, P.S., Kiemeney, L.A., Hermus, A.R., Breteler, M.M., Hofman, A., Hegedus, L. *et al.* (2011) A large-scale association analysis of 68 thyroid hormone pathway genes with serum TSH and FT4 levels. *Eur. J. Endocrinol.*, **164**, 781–788.
15. Gudmundsson, J., Sulem, P., Gudbjartsson, D.F., Jonasson, J.G., Masson, G., He, H., Jonasdottir, A., Sigurdsson, A., Stacey, S.N., Johannsdottir, H. *et al.* (2012) Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat. Genet.*, **44**, 319–322.
16. Rawal, R., Teumer, A., Volzke, H., Wallaschofski, H., Ittermann, T., Asvold, B.O., Bjoro, T., Greiser, K.H., Tiller, D., Werdan, K. *et al.* (2012) Meta-analysis of two genome-wide association studies identifies four genetic loci associated with thyroid function. *Hum. Mol. Genet.*, **21**, 3275–3282.
17. Porcu, E., Medici, M., Pistis, G., Volpato, C.B., Wilson, S.G., Cappola, A.R., Bos, S.D., Deelen, J., den Heijer, M., Freathy, R.M. *et al.* (2013) A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. *PloS Genet.*, **9**, e1003266.
18. Chu, X., Pan, C.M., Zhao, S.X., Liang, J., Gao, G.Q., Zhang, X.M., Yuan, G.Y., Li, C.G., Xue, L.Q., Shen, M. *et al.* (2011) A genome-wide association study identifies two new risk loci for Graves' disease. *Nat. Genet.*, **43**, 897–901.
19. Lin, Y., Lai, X., Chen, B., Xu, Y., Huang, B., Chen, Z., Zhu, S., Yao, J., Jiang, Q., Huang, H. *et al.* (2011) Genetic variations in CYP17A1, CACNB2 and PLEKHA7 are associated with blood pressure and/or hypertension in She ethnic minority of China. *Atherosclerosis*, **219**, 709–714.
20. Dixon, A.L., Liang, L., Moffatt, M.F., Chen, W., Heath, S., Wong, K.C., Taylor, J., Burnett, E., Gut, I., Farrall, M. *et al.* (2007) A genome-wide association study of global gene expression. *Nat. Genet.*, **39**, 1202–1207.
21. Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., Maouche, S., Germain, M., Lackner, K., Rossmann, H. *et al.* (2010) Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One*, **5**, e10693.
22. Fairfax, B.P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., Ellis, P., Langford, C., Vannberg, F.O. and Knight, J.C. (2012) Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.*, **44**, 502–510.
23. Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
24. Boelaert, K., Horacek, J., Holder, R.L., Watkinson, J.C., Sheppard, M.C. and Franklyn, J.A. (2006) Serum thyrotropin concentration as a novel predictor of malignancy in thyroid nodules investigated by fine-needle aspiration. *J. Clin. Endocrinol. Metab.*, **91**, 4295–4301.
25. Haymart, M.R., Repplinger, D.J., Leverson, G.E., Elson, D.F., Sippel, R.S., Jaume, J.C. and Chen, H. (2008) Higher serum thyroid stimulating hormone level in thyroid nodule patients is associated with greater risks of differentiated thyroid cancer and advanced tumor stage. *J. Clin. Endocrinol. Metab.*, **93**, 809–814.
26. McLeod, D.S., Watters, K.F., Carpenter, A.D., Ladenson, P.W., Cooper, D.S. and Ding, E.L. (2012) Thyrotropin and thyroid cancer diagnosis: a systematic review and dose-response meta-analysis. *J. Clin. Endocrinol. Metab.*, **97**, 2682–2692.
27. Wang, Y.L., Feng, S.H., Guo, S.C., Wei, W.J., Li, D.S., Wang, Y., Wang, X., Wang, Z.Y., Ma, Y.Y., Jin, L. *et al.* (2013) 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.*, **50**, 689–695.
28. Persani, L. (2012) Clinical review: central hypothyroidism: pathogenic, diagnostic, and therapeutic challenges. *J. Clin. Endocrinol. Metab.*, **97**, 3068–3078.
29. Asvold, B.O., Bjoro, T. and Vatten, L.J. (2013) Associations of TSH levels within the reference range with future blood pressure and lipid concentrations: 11-year follow-up of the HUNT study. *Eur. J. Endocrinol.*, **169**, 73–82.
30. Pasqualetti, G., Tognini, S., Polini, A., Caraccio, N. and Monzani, F. (2013) Is subclinical hypothyroidism a cardiovascular risk factor in the elderly? *J. Clin. Endocrinol. Metab.*, **98**, 2256–2266.
31. Nanchen, D., Gussekloo, J., Westendorp, R.G., Stott, D.J., Jukema, J.W., Trompet, S., Ford, I., Welsh, P., Sattar, N., Macfarlane, P.W. *et al.* (2012) Subclinical thyroid dysfunction and the risk of heart failure in older persons at high cardiovascular risk. *J. Clin. Endocrinol. Metab.*, **97**, 852–861.
32. Hak, A.E., Pols, H.A., Visser, T.J., Drexhage, H.A., Hofman, A. and Witteman, J.C. (2000) Subclinical hypothyroidism is an independent risk factor for atherosclerosis and myocardial infarction in elderly women: the Rotterdam Study. *Ann. Intern. Med.*, **132**, 270–278.
33. Bielecka-Dabrowa, A., Mikhailidis, D.P., Rysz, J. and Banach, M. (2009) The mechanisms of atrial fibrillation in hyperthyroidism. *Thyroid Res.*, **2**, 4.

34. Sawin, C.T., Geller, A., Wolf, P.A., Belanger, A.J., Baker, E., Bacharach, P., Wilson, P.W., Benjamin, E.J. and D'Agostino, R.B. (1994) Low serum thyrotropin concentrations as a risk factor for atrial fibrillation in older persons. *N. Engl. J. Med.*, **331**, 1249–1252.
35. Calenda, G., Peng, J., Redman, C.M., Sha, Q., Wu, X. and Lee, S. (2006) Identification of two new members, XPLAC and XTES, of the XK family. *Gene*, **370**, 6–16.
36. Lantieri, F., Glessner, J.T., Hakonarson, H., Elia, J. and Devoto, M. (2010) Analysis of GWAS top hits in ADHD suggests association to two polymorphisms located in genes expressed in the cerebellum. *Am. J. Med. Genet. B Neuropsychiatr. Genet.,* **153B**, 1127–1133.
37. Fijal, B.A., Stauffer, V.L., Kinon, B.J., Conley, R.R., Hoffmann, V.P., Witte, M.M., Zhao, F. and Houston, J.P. (2012) Analysis of gene variants previously associated with iloperidone response in patients with schizophrenia who are treated with risperidone. *J. Clin. Psychiatry*, **73**, 367–371.
38. Lavedan, C., Licamele, L., Volpi, S., Hamilton, J., Heaton, C., Mack, K., Lannan, R., Thompson, A., Wolfgang, C.D. and Polymeropoulos, M.H. (2009) Association of the NPAS3 gene and five other loci with response to the antipsychotic iloperidone identified in a whole genome association study. *Mol. Psychiatry*, **14**, 804–819.
39. Leandro-Garcia, L.J., Inglada-Perez, L., Pita, G., Hjerpe, E., Leskela, S., Jara, C., Mielgo, X., Gonzalez-Neira, A., Robledo, M., Avall-Lundqvist, E. *et al.* (2013) Genome-wide association study identifies ephrin type A receptors implicated in paclitaxel induced peripheral sensory neuropathy. *J. Med. Genet.*, **50**, 599–605.
40. De Felice, M., Ovitt, C., Biffali, E., Rodriguez-Mallon, A., Arra, C., Anastassiadis, K., Macchia, P.E., Mattei, M.G., Mariano, A., Scholer, H. *et al.* (1998) A mouse model for hereditary thyroid dysgenesis and cleft palate. *Nat. Genet.*, **19**, 395–398.
41. Fagman, H. and Nilsson, M. (2011) Morphogenetics of early thyroid development. *J. Mol. Endocrinol.*, **46**, R33–R42.
42. Parlato, R., Rosica, A., Rodriguez-Mallon, A., Affuso, A., Postiglione, M.P., Arra, C., Mansouri, A., Kimura, S., Di Lauro, R. and De Felice, M. (2004) An integrated regulatory network controlling survival and migration in thyroid organogenesis. *Dev. Biol.*, **276**, 464–475.
43. 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.
44. Ortiz, L., Aza-Blanc, P., Zannini, M., Cato, A.C. and Santisteban, P. (1999) The interaction between the forkhead thyroid transcription factor TTF-2 and the constitutive factor CTF/NF-1 is required for efficient hormonal regulation of the thyroperoxidase gene transcription. *J. Biol. Chem.*, **274**, 15213–15221.
45. Song, H.D., Liang, J., Shi, J.Y., Zhao, S.X., Liu, Z., Zhao, J.J., Peng, Y.D., Gao, G.Q., Tao, J., Pan, C.M. *et al.* (2009) Functional SNPs in the SCGB3A2 promoter are associated with susceptibility to Graves' disease. *Hum. Mol. Genet.*, **18**, 1156–1170.
46. Zhao, S.X., Xue, L.Q., Liu, W., Gu, Z.H., Pan, C.M., Yang, S.Y., Zhan, M., Wang, H.N., Liang, J., Gao, G.Q. *et al.* (2013) Robust evidence for five new Graves' disease risk loci from a staged genome-wide association analysis. *Hum. Mol. Genet.*, **22**, 3347–3362.
47. 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.
48. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*., **5**, e1000529.
49. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
50. 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.
51. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.

APPENDIX

The china consortium for the genetics of autoimmune thyroid disease members

Huai-Dong Song [State Key Laboratory of Medical Genomics, Molecular Medicine Center, Shanghai Institute of Endocrinology and Metabolism, Ruijin Hospital Affiliated to Shanghai Jiaotong University (SJTU) School of Medicine, Shanghai, China], Shuang-Xia Zhao (State Key Laboratory of Medical Genomics, Molecular Medicine Center, Shanghai Institute of Endocrinology and Metabolism, Ruijin Hospital Affiliated to SJTU School of Medicine, Shanghai, China), Chun-Ming Pan (State Key Laboratory of Medical Genomics, Ruijin Hospital Affiliated to SJTU School of Medicine, Shanghai, China), Jun Liang (Department of Endocrinology, The Central Hospital of Xuzhou Affiliated to Xuzhou Medical College, Xuzhou, Jiangsu Province, China), Xiao-Mei Zhang (Department of Endocrinology, The First Hospital Affiliated to Bengbu Medical College, Bengbu, Anhui Province, China), Guo-Yue Yuan (Department of Endocrinology, The Hospital Affiliated to Jiangsu University, Zhenjiang, Jiangsu Province, China), Chang-Gui Li (Department of Endocrinology and Gout Laboratory, Medical School Hospital of Qingdao University, Qingdao, Shandong Province, China), Jia-Lun Chen (Shanghai Institute of Endocrinology and Metabolism, Department of Endocrinology, Ruijin Hospital Affiliated to SJTU School of Medicine, Shanghai, China), Guan-Qi Gao (Department of Endocrinology, Linyi People's Hospital, Linyi, Shandong Province, China), Li-Bin Liu (Department of Endocrinology, Xiehe Hospital Affiliated to Fujian Medical University, Fuzhou, Fujian Province, China), Gang Chen (Department of Endocrinology, Fujian Province Hospital, Fuzhou, Fujian Province, China), Qing Su (Department of Endocrinology, Xin-Hua Hospital Affiliated to SJTU School of Medicine, Shanghai 200092, China), Yong-De Peng (Department of Endocrinology, The First People's Hospital Affiliated to SJTU School of Medicine, Shanghai, China), Jia-Jun Zhao (Department of Endocrinology, Shandong Province Hospital, Shandong University, Jinan, China).