Genome-wide Association Study Identifies Four Genetic Loci Associated with Thyroid Volume and Goiter Risk

Alexander Teumer,[1,10] Rajesh Rawal,[2,10] Georg Homuth,[1,10] Florian Ernst,[1] Margit Heier,[9] Matthias Evert,[3] Frank Dombrowski,[3] Uwe Völker,[1] Matthias Nauck,[4] Dörte Radke,[5] Till Ittermann,[5] Reiner Biffar,[6] Angela Döring,[2] Christian Gieger,[2] Norman Klopp,[2] H.-Erich Wichmann,[2,7,8] Henri Wallaschofski,[4] Christa Meisinger,[9,10] and Henry Völzke[5,10,*]

Thyroid disorders such as goiters represent important diseases, especially in iodine-deficient areas. Sibling studies have demonstrated that genetic factors substantially contribute to the interindividual variation of thyroid volume. We performed a genome-wide association study of this phenotype by analyzing a discovery cohort consisting of 3620 participants of the Study of Health in Pomerania (SHIP). Four genetic loci were associated with thyroid volume on a genome-wide level of significance. Of these, two independent loci are located upstream of and within *CAPZB*, which encodes the β subunit of the barbed-end F-actin binding protein that modulates actin polymerization, a process crucial in the colloid engulfment during thyroglobulin mobilization in the thyroid. The third locus marks *FGF7*, which encodes fibroblast growth factor 7. Members of this protein family have been discussed as putative signal molecules involved in the regulation of thyroid development. The fourth locus represents a "gene desert" on chromosome 16q23, located directly downstream of the predicted coding sequence *LOC440389*, which, however, had already been removed from the NCBI database as a result of the standard genome annotation processing at the time that this study was initiated. Experimental proof of the formerly predicted mature mRNA, however, demonstrates that LOC440389 indeed represents a real gene. All four associations were replicated in an independent sample of 1290 participants of the KORA study. These results increase the knowledge about genetic factors and physiological mechanisms influencing thyroid volume.

From the clinical and the public-health point of view, thyroid disorders such as nontoxic and toxic goiter are relevant diseases in previously and currently iodine-deficient areas. Whereas goiters are highly prevalent in iodine-deficient regions, it is less commonly present in iodine-replete areas.[1,2] The effect of iodine deficiency on goiter risk is pronounced by cigarette smoking, whereas this association is not present in regions with optimal iodine supply.[3–5] Additional environmental factors include gender, age, and body mass index.[3,5] There is no doubt that genetic factors also play a substantial role in the etiology of simple goiter.[6,7] Sibling studies from Denmark, a region with previously mild to moderate iodine deficiency, demonstrated a higher intraclass correlation for thyroid volume in monozygotic twins as compared to dizygotic twins, suggesting that genetic factors account for approximately 61%–78% of the interindividual variation of the thyroid volume.[7] Whereas genetic loci associated with clinically overt euthyroid multinodular goiter were already mapped in linkage analyses, genome-wide association studies (GWAS) investigating genetic factors with regard to thyroid enlargement have not been conducted so far. Thus, we have performed a GWAS on thyroid volume in Germany, a previously iodine-deficient area with moderate iodine deficiency in the northeast and moderate to severe iodine deficiency in the south.[8,9] A voluntary iodine fortification program was introduced in Germany during the 1980s. In December 1993, improved legislations concerning the iodization of table salt became effective, which contributed to an increase in the use of iodized salt for food production, resulting in a stable iodine supply during the past 15 years.

In the discovery-stage GWAS, 3620 individuals, aged 20–79 years, from the baseline examinations of the Study of Health in Pomerania (SHIP-0[10]) in West Pomerania (northeast Germany) were analyzed for associations of SNPs with the phenotypes "thyroid volume" and "goiter." The lead SNPs of the four identified loci that exhibited genome-wide significant associations for "thyroid volume" and the corresponding SNPs of these loci that showed the strongest associations to "goiter" were replicated in 1290 individuals, aged 30–79, years from the Kooperative Gesundheitsforschung in der Region Augsburg (KORA F4, southern Germany[11]). Finally, we performed a combined GWAS analysis using data from both studies.

[1]Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University, 17487 Greifswald, Germany; [2]Institute of Epidemiology, Helmholtz Zentrum München, 85764 Neuherberg, Germany; [3]Institute for Pathology, University Medicine Greifswald, 17475 Greifswald, Germany; [4]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, 17475 Greifswald, Germany; [5]Institute for Community Medicine, University Medicine Greifswald, 17475 Greifswald, Germany; [6]Department of Prosthetic Dentistry, Gerostomatology and Dental Materials, University Medicine Greifswald, 17475 Greifswald, Germany; [7]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, 80539 Munich, Germany; [8]Klinikum Großhadern, 81377 Munich, Germany; [9]Institute of Epidemiology II, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[10]These authors contributed equally to this work.
*Correspondence: voelzke@uni-greifswald.de
DOI 10.1016/j.ajhg.2011.04.015.

Table 1. Cohort Characteristics

	Study of Health in Pomerania (SHIP)	Cooperative Health Research in the Region of Augsburg, Survey 4 (KORA F4)
Study design	population-based	population-based
Sample size	3620	1290
Age in years (range)	49 (20–81)	60 (32–79)
Females (%)	1716 (47.4)	541 (41.9)
Current smokers (%)	1143 (31.57)	210 (16.29)
BSA in m^2 (SD)	1.89 (0.21)	1.90 (0.21)
Thyroid volume in ml (SD)	21.28 (11.43)	21.55 (10.9)
Presence of goiter (%)	1325 (36.6)	467 (36.2)
Thyroid measurement	Ultrasound VST-Gateway 5 MHz linear array transducer (Diasonics)	SONOLINE G50 5 MHz linear array transducer (Siemens Medical)
Thyroid volume calculation	length × width × depth × 0.479 [ml] for each lobe[34]	
Thyroid measurement QC	Intra- and interobserver reliabilities within and between both studies were assessed before the start of each study and afterwards annually during the studies; analyses were performed according to Bland and Altman.[35] All measurements of the thyroid volume for within and between study comparisons showed Spearman correlation coefficients of > 0.85 and mean differences (+ 2 SD) of the mean bias of $< 5\%$ (<25%).	
Sample exclusions by phenotype	Individuals taking thyroid medication or reporting thyroid disorders, women pregnant at the time of thyroid measurement	

All participants were of European ancestry. Approval was obtained by local ethic committees, and informed consent was given by all participants. Goiter was defined as a thyroid volume of > 18 ml in women and of > 25 ml in men.[12] Subjects with known thyroid disease or those with previous or current antithyroid treatment were excluded from the analyses, because potentially relevant treatment effects on thyroid volume cannot be quantified.[13,14] The detailed characteristics of the study populations, exclusion criteria, and quality control procedures are described in Table 1. Genotyping information and GWAS details are specified in Table S1 (available online).

All SNPs with a minor allele frequency < 0.01 were excluded. Since all X-linked SNPs were excluded from imputation,[15] 27.399 directly genotyped SNPs of chromosome X and 46 directly genotyped SNPs from mtDNA (SNP call rates $\geq 80\%$, $p_{HWE} > 0.001$, MAF $> 1\%$) were tested in an additional analysis in the discovery stage. None of these SNPs showed genome-wide significant associations with "thyroid volume" or "goiter." Associations were tested with the use of a linear additive model on natural log-transformed thyroid volume (ml) for the "thyroid volume" phenotype and a logistic regression analysis for the "goiter" phenotype, respectively. Adjustment for age, gender, current smoking state (yes or no), and body surface area (BSA) was performed for all analyses. All p values of the discovery GWAS and the results of the meta-analysis were corrected for genomic control. Only SNPs, for which association data from both studies were available, were included in the meta-analysis. Associations were considered to have genome-wide significance below a p value of 5×10^{-8}.[16] Genomic control was applied both for the individual cohorts and for the combined results. The estimated genomic control was low for "thyroid volume" and "goiter" for both the individual-cohort analysis (Table S1) and the combined analyses ($\lambda_{GC} = 1.058$ and $\lambda_{GC} = 1.021$, respectively), suggesting little residual confounding due to population stratification (Figure S1). All SNPs found to be associated with one of the two phenotypes of interest in the discovery GWAS, the replication stage, or the combined analysis were in Hardy-Weinberg equilibrium ($p > 0.001$) in both studies. To identify independently associated loci, SNPs were clumped with the use of the PLINK[17] clumping algorithm of ($r^2 > 0.1$, 1 Mb distance) based on genotype data of 4105 SHIP participants. To validate the independence of the four loci for "thyroid volume" in the discovery stage, the lead SNPs of these loci were analyzed together in a multivariate linear regression model, in which the associations remained significant and mostly unchanged, indicating statistical independence of the four SNPs from each other (Table 2).

The discovery analysis identified four loci associated with "thyroid volume" at a genome-wide significance level. Two of these loci were also significantly associated with "goiter," whereas the other two missed genome-wide significance in the discovery stage (see Table 3). The strongest associations were found for the *CAPZB* region on chromosome *1p36*. Within this region, two independent loci were significantly associated with both "thyroid volume" and "goiter": at the locus upstream of *CAPZB*, rs12138950 represented the lead SNP for both phenotypes. Within *CAPZB*, rs12091047 represented the lead SNP for "thyroid volume," whereas rs12033437 was the lead SNP for "goiter." For the third locus at 15q21, the lead SNPs

Table 2. Results from the Analysis of Independence and the Explained Variance of the Lead SNPs of the Four Loci Associated with the "Thyroid Volume" Phenotype in the Discovery-Stage GWAS

Discovery-Stage GWAS		SHIP				KORA			
Locus	Lead SNP	p Value	Effect	SE	Variance Explained	p Value	Effect	SE	Variance Explained
1	rs12138950	4.700×10^{-13}	0.093	0.013	3.33%	3.988×10^{-6}	0.103	0.022	3.18%
2	rs1354920	2.530×10^{-9}	0.060	0.010		6.619×10^{-3}	0.048	0.018	
3	rs17767491	7.390×10^{-11}	0.063	0.010		4.437×10^{-4}	0.060	0.017	
4	rs12091047	3.250×10^{-8}	−0.054	0.010		2.775×10^{-4}	−0.059	0.016	

The values were calculated by analyzing all SNPs adjusted for sex, age, smoking status, and body surface area in a linear regression model. Compared to the results in Table 3, the p values have changed only marginally, indicating statistical independence of the four SNPs from each other.

for both phenotypes (rs1354920 for "thyroid volume" and rs1023683 for "goiter") were located within *C15orf33* in close vicinity of *FGF7*.

The fourth locus at 16q23 was located within a 110 kb distance of the next annotated gene (*MAF*) and was not in linkage disequilibrium with it, and it was therefore initially designated as a "gene desert" (see Figure 2). The lead SNPs of this locus were rs17767491 for "thyroid volume" and rs3813579 for "goiter."

All four loci were positively replicated in the second stage in *KORA-F4* (Table 3). Finally, a GWAS using the SNP data from both studies was performed, including a study population of n = 4910. This combined analysis did not reveal additional, yet-unidentified genome-wide significant associations with the two phenotypes. However, all four loci detected in the primary stages were confirmed and exhibited even stronger associations than those in the discovery analysis (Table 4, Figure 1). In several cases, the combined analysis yielded lead SNPs that were different from those of the discovery GWAS but were consistently found to be in distinct linkage disequilibrium with the former. The lead SNPs of the loci at 16q23 ("gene desert") and at 1p36 within *CAPZB* now exhibited clear genome-wide significant associations for "goiter" (Table 4, Figure 1). We tested for interactions between the lead SNPs of the four "thyroid volume" loci in SHIP, but we did not observe any significant results ($p > 0.05$). For each individual, the number of alleles of all four loci increasing the thyroid volume was counted, and the mean increment of the log thyroid volume per allele was estimated. The results are shown in Table 5. Sensitivity analyses included anti-TPO antibody status and serum concentrations of TSH, free T3, and free T4 as potential confounders without substantially affecting the key results (Table S2).

The two *CAPZB* loci associated with both phenotypes are clearly independent from each other, as their lead SNPs are not in significant linkage disequilibrium ($r^2 = 0.004$). Furthermore, the effects of their minor alleles on "thyroid volume" and "goiter" are inverse. *CAPZB* encodes the two β subunit isoforms of the capping protein (CP), also known as the barbed-end actin binding protein. CP represents a heterodimeric protein composed of α and β subunits. The α1 and α2 subunit isoforms of CP are encoded by *CAPZA1* and *CAPZA2*, respectively. Both β subunit isoforms encoded by *CAPZB* are specified by differential

Table 3. SNPs within the Four Loci that Show the Strongest Association with the "Thyroid Volume" and "Goiter" Phenotypes in the Discovery-Stage GWAS

Locus	Chr.	Position	Lead SNP SHIP	p_{GC} SHIP	p KORA	p_{GC} META	Allele 1	Allele 2	Freq1	ImpQual	R^2
Thyroid Volume											
1	1	19711702	rs12138950	9.340×10^{-14}	2.333×10^{-06}	2.876×10^{-18}	C	A	0.154	0.989	same SNP
2	15	47612815	rs1354920	1.130×10^{-08}	7.656×10^{-03}	4.923×10^{-10}	T	C	0.302	0.970	0.60
3	16	78302988	rs17767491	1.240×10^{-10}	8.571×10^{-04}	8.071×10^{-13}	G	A	0.322	1.000	1.00
4	1	19644512	rs12091047	2.530×10^{-09}	8.400×10^{-05}	1.777×10^{-12}	T	C	0.343	0.993	0.97
Goiter											
1	1	19711702	rs12138950	6.666×10^{-11}	7.111×10^{-05}	3.641×10^{-14}	C	A	0.154	0.989	0.60
2	15	47494739	rs1023683	1.651×10^{-08}	9.199×10^{-07}	3.864×10^{-13}	T	A	0.758	0.991	0.83
3	16	78306777	rs3813579	8.138×10^{-08}	8.457×10^{-04}	3.868×10^{-10}	G	A	0.482	0.997	same SNP
4	1	19622481	rs12033437	1.499×10^{-07}	2.844×10^{-05}	4.055×10^{-11}	T	C	0.334	1.000	0.97

Allele 1, effect allele; Allele 2, other allele; Freq1, frequency of allele 1; ImpQual, imputation quality measurement (observed by expected variance ratio); pGC, p value of the association after genomic control has been applied; R^2, linkage disequilibrium value for the respective lead SNP with the lead SNP of the meta-analysis GWAS of the same locus.

Table 4. SNPs within the Four Loci with Strongest Association with the "Thyroid Volume" and "Goiter" Phenotypes in the Combined Meta-Analysis

Locus	Chr.	Position	Lead SNP	p_{GC}	Effect	SE	Allele 1	Allele 2	Freq1
Thyroid Volume									
1	1	19711702	rs12138950	2.876×10^{-18}	−0.102	0.012	A	C	0.847
2	15	47522589	rs4338740	1.441×10^{-12}	−0.067	0.009	T	C	0.738
3	16	78302049	rs17767419	9.418×10^{-15}	0.068	0.009	T	C	0.321
4	1	19638105	rs12045440	3.237×10^{-14}	0.067	0.009	T	G	0.664
Goiter									
1	1	19715941	rs10917468	1.114×10^{-14}	0.66	0.59–0.73	T	C	0.783
2	15	47522589	rs4338740	2.843×10^{-13}	0.69	0.63–0.76	T	C	0.739
3	16	78306777	rs3813579	3.868×10^{-10}	1.32	1.21–1.44	A	G	0.518
4	1	19638105	rs12045440	1.649×10^{-11}	1.38	1.26–1.51	T	G	0.664

Chr, chromosome; Allele 1, effect allele; Allele 2, other allele; Freq1, frequency of allele 1; pGC, p value of the association after genomic control has been applied.

pre-mRNA splicing.[18] In spite of pronounced amino acid sequence differences between the carboxy termini of β1 and β2, both subunit isoforms exhibit comparable actin-binding activities.[19] Therefore, the β subunit carboxy termini, besides binding actin, may interact with different target proteins that might regulate their activity.[20] This is substantiated by the fact that the two β isoforms exhibit tissue-specific expression: whereas β2 represents the prevailing isoform of nonmuscle tissues, β1 predominates in muscle tissues.[18]

In the thyroid, TSH-induced engulfment of the colloid by extension of microvilli and filopodia protruding in the thyroid follicular lumen from the surface of thyrocytes represents a key step for thyroglobulin mobilization. The resulting endocytotic vesicles fuse with lysosomes, and proteolysis of thyroglobulin releases mono-, di-, tri-, and tetraiodthyronine (T_1, T_2, T_3, and T_4, respectively). Polymerization of actin is crucial in the formation and extension of microvilli and filopodia. CP represents a major antagonist of filopodia formation.[21,22] The elongation status of the barbed end of actin filaments can be regarded as the net result of the interplay between capping and anti-capping activities, with CP as the major barbed-end terminator.[21] On the other hand, the barbed-end actin-binding activity of CP is modulated by additional regulatory proteins that are, in turn, able to bind to and sequester the former. Known examples of this are the CARMIL and V-1/myotrophin proteins that can bind to CP and inhibit its binding to the barbed end of the actin filament, i.e., its capping function.[22] Interestingly, the CP-binding activity of V-1/myotrophin is regulated by cAMP, the most important second messenger involved in the TSH signal transduction.[23] The lead SNP showing the strongest association with both phenotypes in our combined analysis, rs12138950, is located in the upstream region of *CAPZB* (Figure 2). Therefore, it can be speculated that the causative sequence variant underlying the association of this locus with "thyroid volume" and "goiter" might influence the activity of the *CAPZB* promoter, thereby modulating the expression level of the gene. As the minor allele of this locus is associated with increased thyroid volume and risk of goiter, the postulated causative polymorphism most probably represents a promoter-up variant, causing an increased β2 subunit amount in the thyrocyte. Given that both subunits are unstable in the absence of the other subunit but stabilized in its presence,[24] the total amount of the active CP heterodimer might also be increased. According to this model, alleviated reception of the incoming TSH/cAMP signal as a result of attenuated uncapping activity would result in reduced thyroglobulin engulfment by filopodia, decreased T3/T4 release, and, in turn, compensatory thyroid hyperplasia and increased thyroid volume.

In the second *CAPZB* locus, the lead SNP for both phenotypes in the combined analysis, rs12045440, is located within the first of the nine introns of *CAPZB* (Figure 2). Because this locus is marked by a genic lead SNP, one may hypothesize that at least one, yet-unidentified, sequence polymorphism in linkage disequilibrium with the associated SNPs represents a nonsynonymous SNP causing an amino acid exchange in the encoded β2 subunit. Because the minor allele of this locus is associated with decreased thyroid volumes, the β2 variant specified by this allele might exhibit a more sensitive response to the TSH/cAMP signal, resulting in enhanced uncapping compared to the major allele. This could be caused by an improved interaction of CP with at least one protein that negatively regulates the capping activity, such as cAMP-activated V-1/myotrophin. Accordingly, improved reception of the incoming TSH/cAMP signal would result in accelerated thyroglobulin engulfment by filopodia, increased T3/T4 release, and, in turn, compensatory thyroid hypoplasia and decreased thyroid volume.

Recently, Panicker et al. described an association between the SNP rs10917469 and TSH serum concentration.[25] This SNP is located upstream of *CAPZB* and in strong linkage disequilibrium with the lead SNPs of our

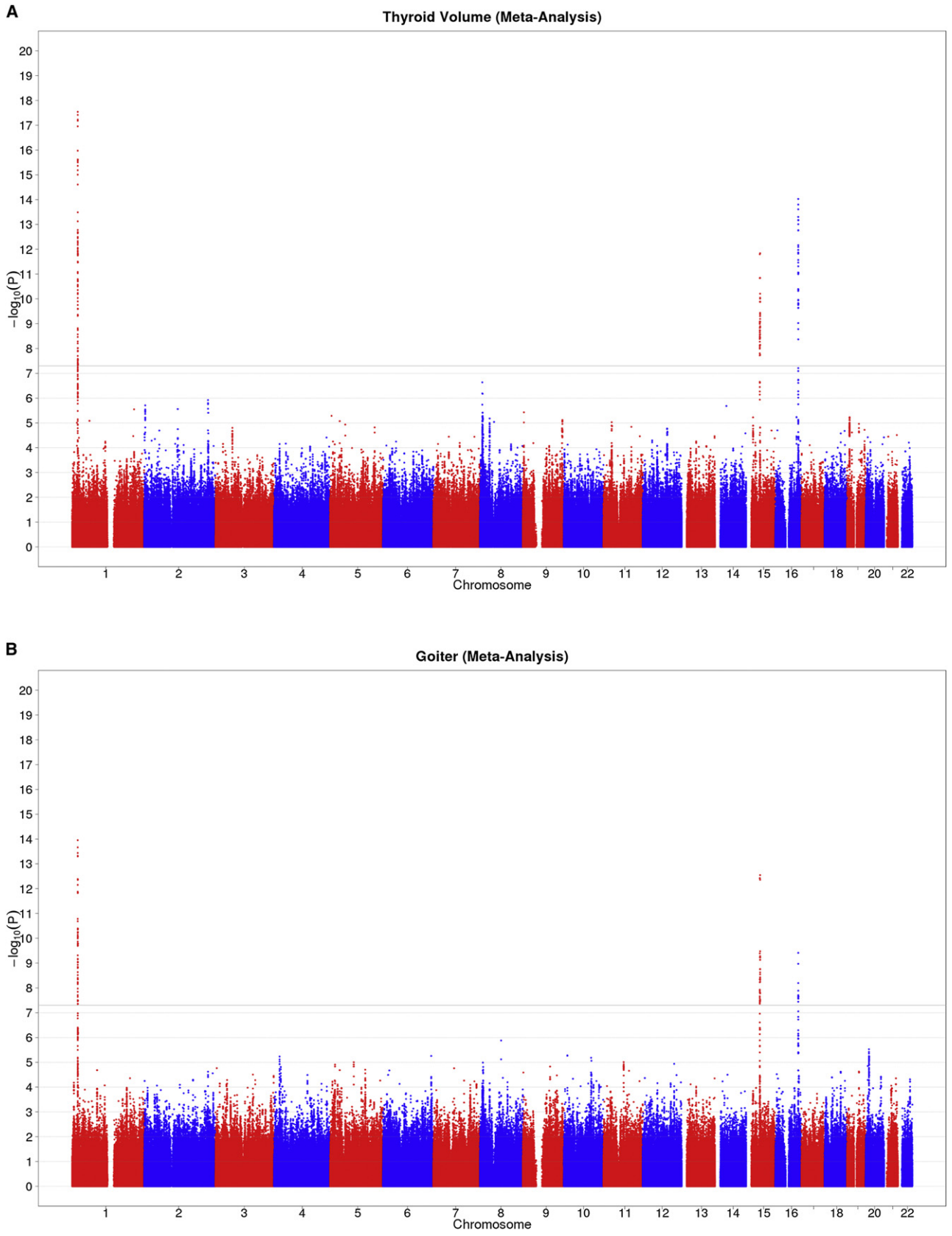

Figure 1. Manhattan Plots Showing the Significance of Association of All SNPs in the Meta-Analysis with the Phenotypes
(A and B) "Thyroid volume" (A) and "goiter" (B). SNPs are plotted on the *x* axis according to their position on each chromosome against association with the respective phenotype on the *y* axis (shown as −log10 p value). SNPs were filtered by minor allele frequency of 1%. The black horizontal line indicates the threshold for genome-wide significance.

Table 5. Mean of Log Thyroid Volumes of Individuals Carrying the Specified Number of "Thyroid Volume"-Increasing Alleles of the Four Lead SNPs in the Combined Analysis

	SHIP				KORA			
No. of Risk Alleles	**Mean**	**95% CI**	**Freq.**	**Freq. %**	**Mean**	**95% CI**	**Freq.**	**Freq. %**
0	2.77	[2.68,2.85]	76	2.10	2.75	[2.60,2.89]	29	2.26
1	2.84	[2.8,2.87]	459	12.69	2.81	[2.75,2.87]	160	12.45
2	2.90	[2.87,2.92]	962	26.59	2.91	[2.87,2.95]	347	27.00
3	2.97	[2.94,2.99]	1077	29.77	2.95	[2.91,2.99]	412	32.06
4	3.01	[2.99,3.04]	726	20.07	3.06	[3.01,3.11]	249	19.38
5	3.10	[3.05,3.15]	253	6.99	3.25	[3.17,3.34]	74	5.76
6	3.16	[3.06,3.26]	61	1.69	3.26	[3.04,3.46]	13	1.01
7	3.32	[2.94,3.7]	4	0.11	3.46	[2.70,4.23]	1	0.08
Total			3618	100.00			1285	100.00

The number of risk alleles ranges from 0 to 7. There were no individuals homozygous for the thyroid increasing allele at all four loci. Freq denotes frequency; the number of individuals carrying the corresponding number of risk alleles. Freq %, percentage of individuals carrying the corresponding number of risk alleles.

combined analysis for "thyroid volume" and "goiter" (rs12138950: $r^2 = 1.00$; rs10917468: $r^2 = 0.60$, respectively). The minor G allele of rs10917469 (MAF = 0.16) is associated with lower TSH serum concentrations than the major A allele. Because the minor alleles of rs12138950 (MAF = 0.15) and rs10917468 (MAF = 0.22) are in linkage disequilibrium with the minor allele of rs10917469, they obviously belong to one common haplotype that is associated not only with TSH serum concentrations, but also with thyroid volume and goiter.

The results of Panicker et al.[25] can be integrated in the model described above: the increased activity of the *CAPZB* promoter mediated by the associated haplotype will cause attenuated uncapping activity in response to the incoming TSH/cAMP signal in thyrocytes and, finally, compensation by increased thyroid volumes. After thyroid growth finally ceases, this compensatory hyperplasia results in the production of amounts of T3 and T4 that are even above the physiological threshold values. As the result of negative feedback regulation, the increased amounts of T3 and T4 would trigger reduced production and secretion of TSH until the concentrations of the thyroid hormones are again in the physiological range. In the new equilibrium, an increased thyroid volume would then be associated with decreased TSH serum concentrations, whereas free T3 and T4 concentrations would be inconspicuous. Supportive for this extended model, in the study of Panicker et al. as well as in ours, serum-free T4 and T3 concentrations did not differ between the genotype groups.

The 15q21 lead SNP, rs4338740, which was identical for "thyroid volume" and "goiter," is located within the second intron of *FGF7* (see Figure 2). Because the minor allele of rs4338740 is associated with increased thyroid volume and goiter risk and the *FGF7* locus is marked by a genic lead SNP, the linked causative sequence variant might cause an amino acid exchange in the encoded FGF7 protein or a modified mRNA splicing site; however, neither nonsynonymous SNPs nor sequence variants predicted to influence splicing sites were identified in linkage disequilibrium, according to the HapMap release 22 CEU data set. Therefore, the causative sequence variation obviously has not yet been identified. FGF7 is a potent epithelial-cell-specific growth factor, whose mitogenic activity is predominantly exhibited in keratinocytes, and it is therefore also named keratinocyte growth factor (KGF). The members of the fibroblast growth factor (FGF) protein family exhibit several mitogenic and cell-survival activities and are involved in various biological processes, including cell growth and morphogenesis.[26] The corresponding FGF receptors are encoded by a tyrosine kinase gene family encompassing at least four members. The main FGF7 receptor is the isoform IIIb of the FGF receptor 2 (FGFR2). Transgenic mice deficient for the murine counterpart of this receptor (FgfR2-IIIb) exhibit, among other phenotypes, thyroid agenesis.[27] Similarly, broad midgestational expression of a kinase-deficient variant of this receptor isoform in mice causes athyreosis.[28] Hence, there are FGF signals essential for the development of the thyroid gland. As putative candidates, FGF1, FGF3, FGF7, and FGF10 were already discussed.[29] Indeed, *FGF10* knockout mice exhibited diverse phenotypes closely related to those for FgfR2-IIIb-deficient animals, including athyreosis.[30] The FGF10 function might be related to the maintenance of the thyroid primordium or the regulation of mitotic activity of the thyroid gland rather than to the induction of thyroid development.[31] Therefore, it can be predicted that the causative sequence variant underlying the observed association causes an enhanced FGF7 signal, thus mediating a more pronounced proliferation of thyroid cells in risk-allele carriers, resulting in thyroid hyperplasia and increased thyroid volume.

An open question concerned the origin of the FGF7 protein obviously influencing the thyroid volume. The FGF protein family members involved in thyroid

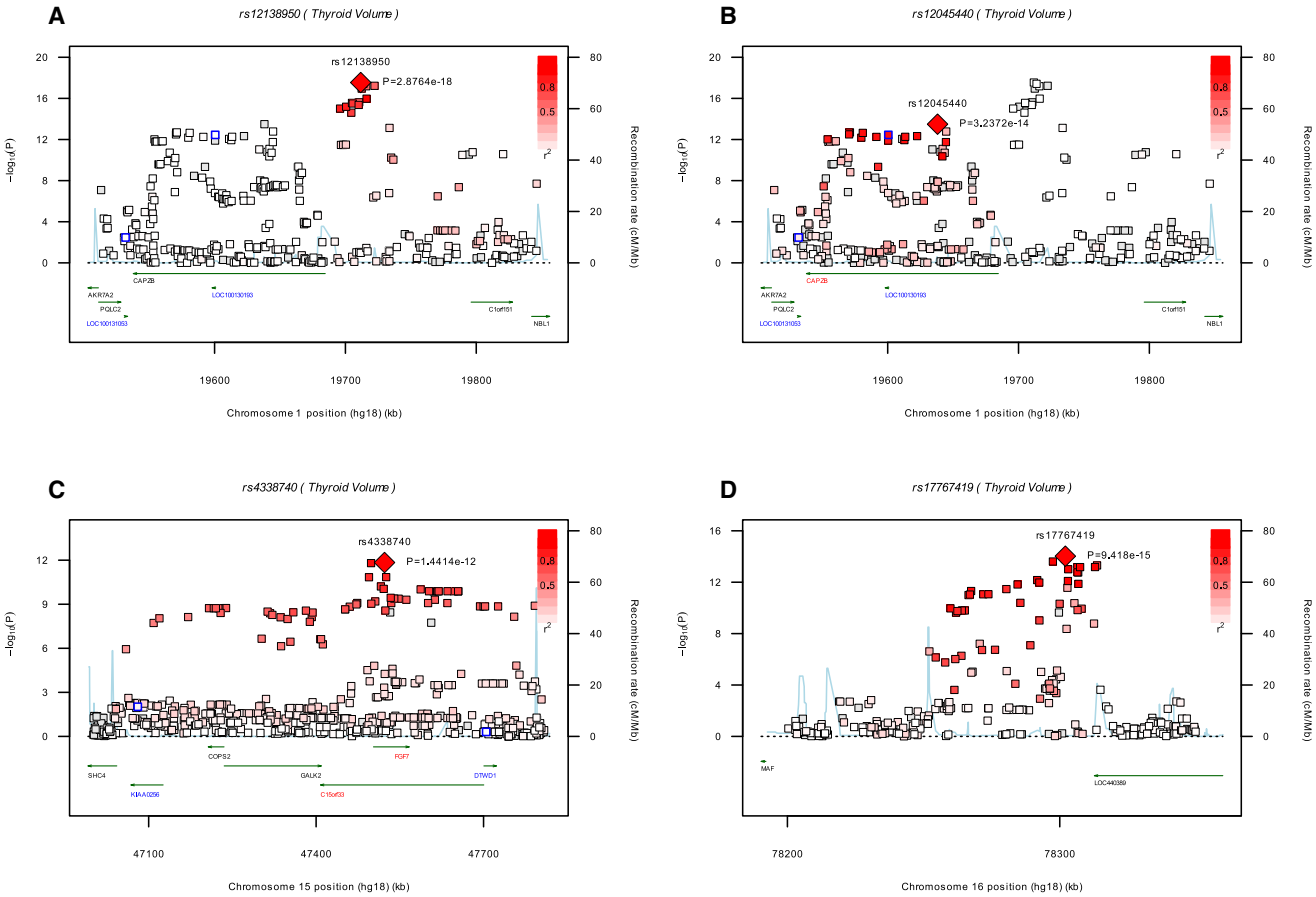

Figure 2. Regional Association Plots Showing the Association Signals in the Regions of the Four Loci Associated with the "Thyroid Volume" Phenotype on the −log10 Scale as a Function of Chromosome Position in the Meta-Analysis
(A–D) The region upstream of *CAPZB* on chromosome 1 (A), the genic region within *CAPZB* on chromosome 1 (B), the *FGF7* region on chromosome 15 (C), and the "gene desert" on chromosome 16 (D). Large diamonds in red indicate the lead SNPs exhibiting the lowest p values for association with the "thyroid volume" phenotype. The correlations (r^2) between each of the surrounding SNPs and the respective lead SNP are indicated by red shading. Genes and nonsynonymous SNPs are labeled in blue. Lead SNPs that are located within genes are indicated by red letters. The left *y* axis indicates the p values for the association with "thyroid volume," and the right *y* axis indicates the estimated recombination rates (HapMap Phase III), shown in light blue. Genes and the direction of transcription (NCBI) are displayed by green bars.

development might be produced and secreted by the mesenchyme surrounding the thyroid gland.[29] Alternatively, FGF7 could be produced by thyroid cells themselves in an autocrine fashion. To differentiate between these possibilities, thyroid tissue, surgically removed from ten patients with nodular goiter, was selected from the archive of the Institute of Pathology, University Medicine Greifswald, and immunohistochemical analyses using anti-FGF7 primary antibodies (KGF; Abcam, catalog number ab9598) were performed. Staining was found in each sample, strongest in leiomyocytes of arterial walls and endothelial cells. The staining in follicle epithelial cells was observed in each case, usually of moderate intensity (Figure 3) but varying from negatively to strongly stained regions. This result demonstrates that FGF7 involvement in the modulation of thyroid volumes is present throughout the cytoplasm of thyroid follicle epithelial cells and therefore at least partially acts in an autocrine manner.

Both lead SNPs of the "gene desert" on 16q23 are located directly downstream of a predicted coding sequence (gene symbol: *LOC440389*), which, however, had already been removed from the NCBI database as a result of the standard genome annotation processing at the time that this study was initiated. However, the predicted amino acid sequence of the putatively encoded protein exhibits 100% identity with a predicted *Pan troglodytes* protein (encoded by *LOC454261*). Furthermore, the local genomic context of this region is comparable between both species: the distance between *MAF* and *LOC440389* in *Homo sapiens* and the distance between *MAF* and *LOC454261* in *Pan troglodytes* both amount to 121 kb, and in both species, both genes are transcribed in the same direction. The two lead SNPs are located within a region of elevated linkage disequilibrium that encompasses the furthermost 3′- end of the putative gene and the region immediately downstream of it. Therefore, we hypothesized that the putative causative sequence variant(s) affecting the thyroid size

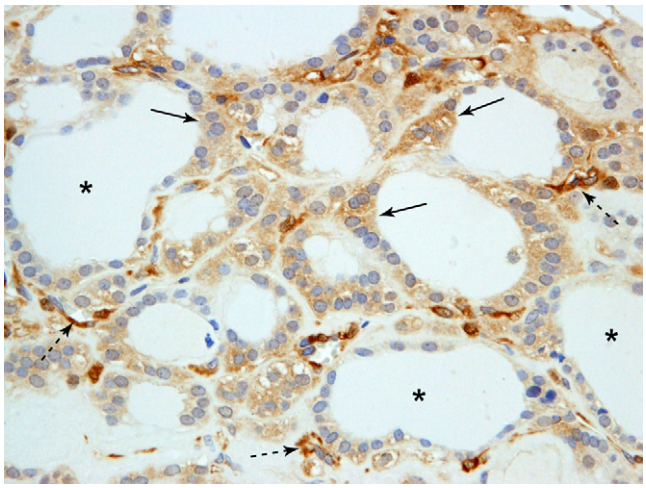

Figure 3. Immunohistochemical Detection of FGF7
Immunohistochemical detection of FGF7 in thyroid follicle epithelial cells (arrows). Colloid within the follicles is unstained (asterisks). Endothelial cells also react positively (dotted arrows). The lower edge of the panel measures 250 μm.

might influence the *LOC440389* expression via modulation of its 3′- trailer sequence. Indeed, it is known that 3′- untranslated regions (3′- UTRs) of human protein-coding genes are rich in miRNA target sites, and it has been proposed that miRNA regulation may be affected by polymorphisms in 3′-UTRs. [32]

Because these considerations were based on the assumption that *LOC440389* indeed represents a "real" gene, an experiment aimed at detecting a specific, mature *LOC440389* mRNA was designed. Because the *LOC440389* locus influences the thyroid volume, it was speculated that expression of the postulated gene occurs with increased likelihood in thyroid tissue. Therefore, a tissue-specific cDNA library was generated by reverse transcription of human thyroid total RNA (Ambion). This library served as a template in a PCR reaction using a primer pair designed for amplification of the predicted cDNA derived from the postulated *LOC440389* mRNA (see Supplemental Material and Methods). The complete amplified *LOC440389* coding region spanning all four exon-intron boundaries was calculated to encompass 258 base pairs. Given that the 3′-primer was extended by a tail containing a T7 promoter sequence, the total size of the PCR product should amount to 282 base pairs. Indeed, a specific PCR product of the predicted size was obtained. Sequencing of both strands of this PCR product (LGC Genomics) proved that the cDNA sequence exactly corresponded to the published *LOC440389* mRNA sequence (see Supplemental Material and Methods) expected from the database entry (XM_498648.3). This clearly demonstrated that *LOC440389* indeed represents a real gene and that the withdrawal of its annotation from the NCBI database was incorrect. Furthermore, tissue-specific expression of this gene in the thyroid was proven. To determine the size of the *LOC440389* transcript, we performed an RNA hybridization analysis using a digoxigenin-labeled RNA probe complementary to the complete predicted *LOC440389* mRNA and total RNA prepared from thyroid tissue and, as a control, from skeletal muscle, as described previously.[33] The cheminoluminescence signals obtained clearly demonstrated the presence of a distinct mRNA in the predicted size range of 700 nucleotides (Figure 4). Densitometrical analysis of the bands revealed that the transcript was threefold more abundant in thyroid than in skeletal muscle. Additional analyses are necessary for determining the physiological function of the protein encoded by *LOC440389*, especially in the context of its relationship to thyroid volume. The finding of a stronger expression of the gene in thyroid tissue than in skeletal muscle tissue might serve as a first hint of a thyroid-specific function of the protein.

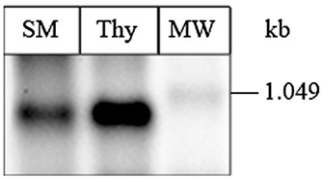

Figure 4. Northern Analysis Demonstrating the Presence of an *LOC440389*-Specific mRNA of around 0.8 kb in Skeletal Muscle Tissue and Thyroid Tissue
SM, skeletal muscle tissue; Thy, thyroid tissue. The first two lanes (SM and Thy) contained 5 μg of total tissue-specific RNA each. The right lane (MW) represents the RNA molecular weight standard. Total RNA was separated by electrophoresis in a 1.2% gel, and, after blotting, the nylon membrane was hybridized with a *LOC440389*-specific RNA probe.

Finally, it has to be emphasized that iodine supply of a population represents the major environmental risk factor for goiter. Given that both populations investigated herein were recruited from areas with highly comparable iodine supply, the present results are representative for formerly and, probably, currently iodine-deficient regions. The representativeness of our findings for iodine-replete regions, however, has to be investigated.

Supplemental Data

Supplemental Data include Supplemental Material and Methods, one table, and two figures and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

SHIP is part of the Community Medicine Research net of the Ernst-Moritz-Arndt-University Greifswald, Germany, which is funded by the Federal Ministry of Education and Research, the Ministry of Cultural Affairs, and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare (Erlangen, Germany) and the Federal State of Mecklenburg-West Pomerania. The SHIP authors are grateful to the contribution of Anja Wiechert and Astrid Petersmann in generating the SNP

data and to Marc Schaffer for his assistance in the RNA analysis. The University of Greifswald is a member of the "Center of Knowledge Interchange" program of Siemens AG. Data analyses were further supported by the German Research Foundation (DFG Vo 955/10-1) and the Federal Ministry of Nutrition, Agriculture and Comsumer's Safety.

The KORA research platform was initiated and financed by the Helmholtz Center Munich, by the German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research, and by the State of Bavaria. The work of KORA is supported by the German Federal Ministry of Education and Research (BMBF) in the context of the German National Genome Research Network (NGFN-2 and NGFN-plus). Our research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. Thyroid examinations were funded by Sanofi-Aventis within the framework of the Papillon Initiative.

Received: December 23, 2010
Revised: March 28, 2011
Accepted: April 21, 2011
Published online: May 12, 2011

Web Resources

The URLs for data presented herein are as follows:

HapMap, http://www.hapmap.org
IMPUTE software, http://www.stats.ox.ac.uk/~marchini/software/gwas/impute
METAL package, http://www.sph.umich.edu/csg/abecasis/metal
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
QUICKTEST, http://toby.freeshell.org/software/quicktest.shtml
R 2.4.1, http://www.R-project.org
SNAP, http://www.broadinstitute.org/mpg/snap/
SNPTEST, http://www.stats.ox.ac.uk/~marchini/software/gwas/snptest.html
UCSC Genome Browser, http://genome.ucsc.edu

References

1. Laurberg, P., Pedersen, K.M., Hreidarsson, A., Sigfusson, N., Iversen, E., and Knudsen, P.R. (1998). Iodine intake and the pattern of thyroid disorders: a comparative epidemiological study of thyroid abnormalities in the elderly in Iceland and in Jutland, Denmark. J. Clin. Endocrinol. Metab. *83*, 765–769.
2. Teng, W., Shan, Z., Teng, X., Guan, H., Li, Y., Teng, D., Jin, Y., Yu, X., Fan, C., Chong, W., et al. (2006). Effect of iodine intake on thyroid diseases in China. N. Engl. J. Med. *354*, 2783–2793.
3. Völzke, H., Schwahn, C., Kohlmann, T., Kramer, A., Robinson, D.M., John, U., and Meng, W. (2005). Risk factors for goiter in a previously iodine-deficient region. Exp. Clin. Endocrinol. Diabetes *113*, 507–515.
4. Ittermann, T., Schmidt, C.O., Kramer, A., Below, H., John, U., Thamm, M., Wallaschofski, H., and Völzke, H. (2008). Smoking as a risk factor for thyroid volume progression and incident goiter in a region with improved iodine supply. Eur. J. Endocrinol. *159*, 761–766.
5. Knudsen, N., Laurberg, P., Perrild, H., Bülow, I., Ovesen, L., and Jørgensen, T. (2002). Risk factors for goiter and thyroid nodules. Thyroid *12*, 879–888.
6. Brix, T.H., Kyvik, K.O., and Hegedüs, L. (1999). Major role of genes in the etiology of simple goiter in females: a population-based twin study. J. Clin. Endocrinol. Metab. *84*, 3071–3075.
7. Hansen, P.S., Brix, T.H., Bennedbaek, F.N., Bonnema, S.J., Kyvik, K.O., and Hegedüs, L. (2004). Genetic and environmental causes of individual differences in thyroid size: a study of healthy Danish twins. J. Clin. Endocrinol. Metab. *89*, 2071–2077.
8. Meng, W., Schindler, A., Horack, S., Lux, E., and Muche, A. (1998). Renal iodine excretion by students in East Germany. A prospective study 1989 to 1996. Med. Klin. (Munich) *93*, 347–351.
9. Meng, W.S.A. (1998). Iodine Supply in Germany. In Elimination of Iodine Deficiency Disorders (IDD) in Central and Eastern Europe, the Commonwealth of Independent States and the Baltic States, F. Delange, A. Robertson, E. McLoughney, and G. Gerasimov, eds. (Munich, Germany: World Health Organization), pp. 21–27.
10. Volzke, H., Alte, D., Schmidt, C.O., Radke, D., Lorbeer, R., Friedrich, N., Aumann, N., Lau, K., Piontek, M., Born, G., et al. (2010). Cohort Profile: The Study of Health in Pomerania. Int. J. Epidemiol. *40*, 294–307.
11. Wichmann, H.E., Gieger, C., and Illig, T.; MONICA/KORA Study Group. (2005). KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen *67* (*Suppl 1*), S26–S30.
12. Gutekunst, R., Becker, W., Hehrmann, R., Olbricht, T., and Pfannenstiel, P. (1988). Ultrasonic diagnosis of the thyroid gland. Dtsch. Med. Wochenschr. *113*, 1109–1112.
13. Brix, T.H., Kyvik, K.O., and Hegedüs, L. (2001). Validity of self-reported hyperthyroidism and hypothyroidism: comparison of self-reported questionnaire data with medical record review. Thyroid *11*, 769–773.
14. Bergmann, M.M., Jacobs, E.J., Hoffmann, K., and Boeing, H. (2004). Agreement of self-reported medical history: comparison of an in-person interview with a self-administered questionnaire. Eur. J. Epidemiol. *19*, 411–416.
15. Soranzo, N., Spector, T.D., Mangino, M., Kühnel, B., Rendon, A., Teumer, A., Willenborg, C., Wright, B., Chen, L., Li, M., et al. (2009). A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat. Genet. *41*, 1182–1190.
16. Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet. Epidemiol. *32*, 381–385.
17. 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., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
18. Schafer, D.A., Korshunova, Y.O., Schroer, T.A., and Cooper, J.A. (1994). Differential localization and sequence analysis of capping protein beta-subunit isoforms of vertebrates. J. Cell Biol. *127*, 453–465.
19. Schafer, D.A., Jennings, P.B., and Cooper, J.A. (1996). Dynamics of capping protein and actin assembly in vitro: uncapping barbed ends by polyphosphoinositides. J. Cell Biol. *135*, 169–179.
20. Yamashita, A., Maeda, K., and Maéda, Y. (2003). Crystal structure of CapZ: structural basis for actin filament barbed end capping. EMBO J. *22*, 1529–1538.

21. Mejillano, M.R., Kojima, S., Applewhite, D.A., Gertler, F.B., Svitkina, T.M., and Borisy, G.G. (2004). Lamellipodial versus filopodial mode of the actin nanomachinery: pivotal role of the filament barbed end. Cell *118*, 363–373.
22. Cooper, J.A., and Sept, D. (2008). New insights into mechanism and regulation of actin capping protein. Int. Rev. Cell Mol. Biol. *267*, 183–206.
23. Kitazawa, M., Yamakuni, T., Song, S.Y., Kato, C., Tsuchiya, R., Ishida, M., Suzuki, N., Adachi, E., Iwashita, S., Ueno, S., et al. (2005). Intracellular cAMP controls a physical association of V-1 with CapZ in cultured mammalian endocrine cells. Biochem. Biophys. Res. Commun. *331*, 181–186.
24. Amatruda, J.F., Gattermeir, D.J., Karpova, T.S., and Cooper, J.A. (1992). Effects of null mutations and overexpression of capping protein on morphogenesis, actin distribution and polarized secretion in yeast. J. Cell Biol. *119*, 1151–1162.
25. 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.
26. Ornitz, D.M., and Itoh, N. (2001). Fibroblast growth factors. Genome Biol. *2*, REVIEWS3005.
27. Revest, J.M., Spencer-Dene, B., Kerr, K., De Moerlooze, L., Rosewell, I., and Dickson, C. (2001). Fibroblast growth factor receptor 2-IIIb acts upstream of Shh and Fgf4 and is required for limb bud maintenance but not for the induction of Fgf8, Fgf10, Msx1, or Bmp4. Dev. Biol. *231*, 47–62.
28. Celli, G., LaRochelle, W.J., Mackem, S., Sharp, R., and Merlino, G. (1998). Soluble dominant-negative receptor uncovers essential roles for fibroblast growth factors in multi-organ induction and patterning. EMBO J. *17*, 1642–1655.
29. De Moerlooze, L., Spencer-Dene, B., Revest, J.M., Hajihosseini, M., Rosewell, I., and Dickson, C. (2000). An important role for the IIIb isoform of fibroblast growth factor receptor 2 (FGFR2) in mesenchymal-epithelial signalling during mouse organogenesis. Development *127*, 483–492.
30. Ohuchi, H., Hori, Y., Yamasaki, M., Harada, H., Sekine, K., Kato, S., and Itoh, N. (2000). FGF10 acts as a major ligand for FGF receptor 2 IIIb in mouse multi-organ development. Biochem. Biophys. Res. Commun. *277*, 643–649.
31. De Felice, M., and Di Lauro, R. (2004). Thyroid development and its disorders: genetics and molecular mechanisms. Endocr. Rev. *25*, 722–746.
32. Landi, D., Gemignani, F., Naccarati, A., Pardini, B., Vodicka, P., Vodickova, L., Novotny, J., Försti, A., Hemminki, K., Canzian, F., and Landi, S. (2008). Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer. Carcinogenesis *29*, 579–584.
33. Homuth, G., Masuda, S., Mogk, A., Kobayashi, Y., and Schumann, W. (1997). The dnaK operon of Bacillus subtilis is heptacistronic. J. Bacteriol. *179*, 1153–1164.
34. Brunn, J., Block, U., Ruf, G., Bos, I., Kunze, W.P., and Scriba, P.C. (1981). Volumetric analysis of thyroid lobes by real-time ultrasound (author's transl). Dtsch. Med. Wochenschr. *106*, 1338–1340.
35. Bland, J.M., and Altman, D.J. (1986). Regression analysis. Lancet *1*, 908–909.