ORIGINAL ARTICLE

Combined linkage analysis and exome sequencing identifies novel genes for familial goiter

Junxia Yan[1], Tsutomu Takahashi[2], Toshihiro Ohura[3], Hiroyuki Adachi[2], Ikuko Takahashi[2], Eishin Ogawa[4], Hiroko Okuda[1], Hatasu Kobayashi[1], Toshiaki Hitomi[1], Wanyang Liu[1], Kouji H Harada[1] and Akio Koizumi[1]

Familial goiter is a genetic disease showing heterogeneous expression. To identify causative genes, we investigated three multigenerational goiter families with an autosomal dominant inheritance pattern. We performed genome-wide linkage analysis on all the families, combined with whole-exome sequencing in two affected individuals from each family. For linkage analysis, we considered loci with logarithm of odds (LOD) scores >1.5 as candidate regions for identification of rare variants. In one of the families, we found two rare heterozygous missense variants, p.V56M in *RGS12* and p.G37D in *GRPEL1,* which segregate with goiter and are both located within the same haplotype on 4p16. This haplotype was not observed in 150 controls. In the other two families, we identified two additional rare missense variants segregating with goiter, p.A551T in *CLIC6* on 21q22.12 and p.V412A in *WFS1* on 4p16. In controls, the minor allele frequency (MAF) of p.V412A in WFS1 was 0.017 while p.A551T in CLIC6 was not detected. All identified genes (*RGS12*, *GRPEL1*, *CLIC6* and *WFS1*) show expression in the human thyroid gland, suggesting that they may play a role in thyroid gland function. Moreover, these four genes are novel with regard to their involvement in familial goiter, supporting genetic heterogeneity of this disease.

INTRODUCTION

Nontoxic thyroid goiter is a common disorder characterized by a diffuse or nodular enlargement of the thyroid gland. It is caused by chronic stimulation of thyroid-stimulating hormone (TSH), and is not the result of inflammatory or neoplastic proliferation.[1] Both nongenetic and/or genetic factors contribute to the development of goiter, and with regard to nongenetic factors, it is well established that the incidence of goiter is largely dependent on iodine intake.[2] In addition, goitrogens (for example, thiocyanates and isoflavones), cigarette smoking, gender, age and increased body mass index are also important risk factors.[3–5] Moreover, family and twin studies clearly demonstrate a genetic predisposition to goiter development,[4,6] and it is likely that interactions between environmental factors and an underlying genetic predisposition ultimately determine goiter occurrence.

In familial cases of goiter, the majority of families show an autosomal dominant inheritance pattern with locus heterogeneity.[7–10] To date, linkage studies have identified many different loci including *MNG-1* on 14q31,[7] a locus between *TSHR* and *MNG-1* on 14q31,[8] a locus on Xp22,[9] and four other novel candidate loci on chromosomes 2q, 3p, 7q and 8p.[10] Except for *MNG-1*, the

candidate genes responsible for the other linkage signals are so far unidentified.

The aim of the present study was to identify causative genes in three Japanese families with goiter. The probands of these three families were found through the neonatal mass screening program. To identify causative genes, we performed whole-exome sequencing in each family in parallel with linkage analysis. We anticipated that the combination of linkage analysis and exome sequencing would significantly enhance the capability of positional cloning. As all probands were also diagnosed with congenital hypothyroidism (CH), we also systematically investigated all known candidate genes related to CH.

MATERIALS AND METHODS

The study involved three Japanese multigenerational goiter families from the Tohoku area, one family was from Miyagi Prefecture and the other two from Akita Prefecture. In addition, 150 controls from each area were used. This study was approved by the Institutional Review Board and Ethics Committee of Kyoto University School of Medicine, Japan. Written informed consent was obtained from all participants.

[1]Department of Health and Environmental Sciences, Graduate School of Medicine, Kyoto University, Kyoto, Japan; [2]Department of Pediatrics, Graduate School of Medicine, Akita University, Akita, Japan; [3]Department of Pediatrics, Sendai City Hospital, Sendai, Japan and [4]Department of Pediatrics, School of Medicine, Teikyo University, Tokyo, Japan
Correspondence: Professor A Koizumi, Department of Health and Environmental Sciences, Graduate School of Medicine, Kyoto University, Konoe-cho, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan.
E-mail: koizumi.akio.5v@kyoto-u.ac.jp

Diagnostic criteria

A goiter-affected status was determined by the following criteria: (1) clinical assessment, by palpation and ultrasonographic examination, indicating diffuse or nodular enlargement of the thyroid gland; or (2) diagnosed with hypothyroidism due to high TSH levels ($\geqslant 10 \mu IU ml^{-1}$) in the neonatal mass screening program with follow-up by clinicians; or (3) having undergone a thyroidectomy because of goiter; or (4) having a history of diagnosed or an obviously observed goiter phenotype of deceased individuals.

Pedigrees

The MYG family included 15 individuals, of which 7 were affected with goiter (Figure 1). The proband was a 9-year-old girl (IV-3) who was born to non-consanguineous parents with a normal delivery (41 weeks gestation, birth weight 3416 g). Based on her high TSH level ($68.4 \mu IU ml^{-1}$) in the neonatal mass screening program, L-thyroxine therapy was started at 2 weeks of age. A follow-up thyroid function test revealed that her TSH level was slightly increased ($6.25 \mu IU ml^{-1}$; normal range 0.29–5.11 $\mu IU ml^{-1}$). Ultrasonographic examination indicated the thyroid gland had a slightly diffuse enlargement bilaterally. As for IV-3, her younger sister (IV-4), a 7-year-old girl, and her first-cousin's younger sister (IV-2), an 8-year-old girl, also immediately received hormone therapy based on high TSH levels (39.4 and $62.1 \mu IU ml^{-1}$, respectively) in the neonatal mass screening program. Antibody examination showed all the three patients were negative for anti-thyroid peroxidase, anti-thyroglobulin and anti-thyroid stimulating hormone receptor antibodies. Since initiation of hormone replacement therapy in the patients, TSH levels were within the normal range and thyroid function was normal. Growth, development and intelligence were also assessed as normal. IV-1 was classified as normal, with no goiter or other thyroid disease detected over the 10-year study period: IV-1, IV-2, IV-3 and IV-4 have been followed up from 2002 to 2013. Within this family, III-4, III-3 and II-3 have mild diffuse goiter. TSH levels were slightly increased in III-4 ($7.2 \mu IU ml^{-1}$) and thyroglobulin was increased in III-3 and II-3 (73 and $90 ng ml^{-1}$, respectively; normal level $< 30 ng ml^{-1}$). Free triiodothyronine (FT3) and thyroxine (FT4) levels were in the normal range. The deceased great-grandmother (I-2) had goiter and underwent a thyroidectomy. Thyroid functions were normal with no goitrous thyroids in other family members.

A clinical description of the THS and THM families has been described previously.[11] In brief, these two families contain four multinodular goiter-affected individuals in four generations. The probands were two young girls (<10 years) who visited the hospital periodically based on high TSH levels identified in the neonatal mass screening program.

The THS family included 10 members. The proband (IV-1), a 3-year-old girl, showed multinodular goiter bilaterally despite receiving thyroid hormone therapy since she was 1 month old. Thyroid function testing revealed increased TSH levels ($19.5 \mu IU ml^{-1}$) and ^{123}I-uptake (3 h 11.4%; normal rate 5.4–12.0%), with normal FT3 and FT4 ($3.5 pg ml^{-1}$ and $1.03 ng dl^{-1}$, respectively), and thyroglobulin ($30 ng ml^{-1}$) levels. Microsome and $KClO_4$ discharge tests were both negative. Within this family, I-2, II-2 and III-4 had goiter, with thyroidectomies performed in II-2 and III-4. The pathology of the surgically removed thyroids revealed adenomatous goiter without neoplastic cell proliferation or lymphocyte infiltration. The remaining family members did not have goiter, as determined by palpation and ultrasonographic examination.

The THM family included 13 members. The proband (IV-3), an 8-year-old girl, had mild and diffuse goiter, with abnormal TSH levels ($16.6 \mu IU ml^{-1}$) and ^{123}I-uptake (3 h 15.9%). Microsome and $KClO_4$ discharge tests were both negative. She had suspected hypothyroidism (TSH $38.9 \mu IU ml^{-1}$) in the neonatal mass screening program, but replacement therapy was not started because of normal T3 and T4 levels. Her father (III-4) was diagnosed with goiter at 30 years of age. Thyroid function tests identified increased TSH ($5.1 \mu IU ml^{-1}$; normal range 0.34–3.5 $\mu IU ml^{-1}$) and TG ($300 ng ml^{-1}$) levels, FT3 and FT4 within the normal range ($3.2 pg ml^{-1}$ and $1.03 ng dl^{-1}$, respectively), and a negative microsome test. A CT scan showed bilateral swelling of the thyroid glands and a nodular lesion in the left lobe. Aspiration biopsy did not show lymphocyte infiltration or malignancy in the nodular lesion. The great-grandmother (I-2) also had goiter, identified as multinodular goiter by ultrasonographic examination, but TSH, FT3 and FT4 were all normal, with a negative microsome test. The deceased grandfather (II-2) also had goiter and had undergone a thyroidectomy. Pathological examination revealed multinodular goiter without malignant cell proliferation and no lymphocyte infiltration.

Controls

We randomly selected 150 healthy controls from Miyagi and Akita Prefectures to determine the minor allele frequency (MAF) for the identified candidate variants. Blood samples were collected from healthy individuals as part of their annual health checkup in 1999 (Akita) and 2005 (Miyagi), and stored at $-20 °C$ in the Kyoto University Specimen Banks.[12] Control individuals did not have thyroid disease (self-reported). The 150 Miyagi controls were 21–65 years old (mean age 47.1 ± 11.2 years; male:female, 123:27). The 150 Akita controls were 30–82 years old (mean age 58.2 ± 10.7 years; male:female, 37:113).

Genome-wide linkage analysis and simulations

Genomic DNA was extracted from peripheral blood lymphocytes using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). Genome-wide linkage analysis was performed in each family (MYG without II-1, 14 members; THS, 10 members; THM without III-1, IV-1 and IV-2, 10 members) using the ABI Prism Linkage Mapping Set (Version2; Applied Biosystems, Foster City, CA, USA), with 382 markers, 10 cm apart, covering all 22 autosomes. Polymerase chain reaction (PCR) amplification was performed and fluorescence-labeled DNA products mixed and electrophoresed on an ABI Prism 3100 Genetic Analyzer. Alleles were collected and analyzed using 3130 Data Collection (Version 3.0) and Genemapper (Version 4.0) software. The disease allele frequency was set at 0.01 and population allele frequencies assigned as equal portions of individual alleles. A multipoint parametric linkage analysis was performed using GeneHunter (Ver2.1_r6) and logarithm of the odds (LOD) scores obtained.[13] Haplotypes were constructed (and segregated haplotypes identified to determine susceptibility loci), and then combined with exome data to search for candidate causative variants in each family.

Because of relatively small family sizes and genetic heterogeneity, linkage analysis failed to map the disease locus to a single genomic site and the maximum LOD score did not reach the accepted LOD score of 3. Therefore, to identify the linkage locus, we introduced alternative approaches to compensate for the low statistical power. The exome approach, which identifies any potential coding variant or splicing site variant within families, is prone to detect false-negative loci as a result of the low statistical power. Therefore, to determine the genome-wide false-positive rate of LOD scores, we introduced a new threshold LOD score of 1.5, and randomly simulated linkage analysis in our families, with 100 simulations performed in each family. The average genome-wide false-positive rate of linkage analysis was then calculated in each family.

Whole-exome sequencing

Considering genetic heterogeneity, we selected two affected individuals in each family for exome sequencing, IV-2 and IV-3 in the MYG family, III-4 and IV-1 in the THS family, and III-4 and IV-3 in the THM family (Figure 1). We used the HiSeq 2000 platform (Illumina, San Diego, CA, USA) for exome sequencing. The target regions (CCDS exonic regions and flanking intronic regions, totally ~50 Mb of genomic DNA) were captured using the SureSelect Human All Exon 50 Mb Kit (Agilent) according to the manufacturer's instructions. Briefly, genomic DNAs were extracted from peripheral blood and randomly fragmented by acoustic fragmentation (Covaris, Woburn, MA, USA), then purified using a QIAquick PCR Purification kit (Qiagen). Adapters were ligated to each end of the fragments, and the resulting DNA library purified (using QIAquick PCR Purification kit), amplified by ligation-mediated PCR and captured by hybridization using the SureSelect Biotinylated RNA library 'baits' (Agilent) for enrichment. The magnitude of enrichment of captured ligation-mediated PCR products was determined using the Agilent 2100 Bioanalyzer. Next, each captured library was loaded onto a HiSeq 2000 platform, and paired-end sequencing performed with read lengths of 101 bp, using two channels. Sequence reads were mapped to the reference human genome (Ghr37/hg19; UCSC Genome Browser hg19) using Burrows-Wheeler
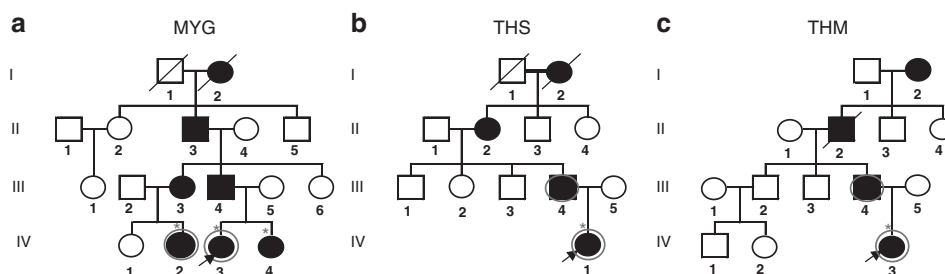
Figure 1 Pedigrees of the three Japanese goiter families. The MYG family is from Miyagi Prefecture (**a**), and the THS and THM families are from Akita Prefecture (**b** & **c**), in the Tohoku area of Japan. Filled and unfilled symbols indicate goiter affected and unaffected individuals, respectively. Squares and circles represent male and females, respectively. Arrows indicate probands and a slash indicates deceased individuals. Asterisks show goiter-affected individuals, identified through the neonatal mass screening program. Red circles show individuals on whom whole-exome sequencing was performed. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

Aligner 0.5.9 software (http://bio-bwa.sourceforge.net/index.shtml). Single-nucleotide variants (SNVs) and small insertion/deletions (Indels) were detected using the Genome Analysis Toolkit (GATK) (http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit).

Sequence data were filtered against multiple databases, namely, dbSNP135 (ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/ASN1_flat), 1000 Genomes Project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/) and NHLBI ESP6500 (http://evs.gs.washington.edu/EVS/), and five in-house control exome databases, using various filtering strategy. Registration was searched in the dbSNP database (Build 135; NCBI, www.ncbi.nlm.nih.gov), and reference SNP number was provided if the variant had been registered in the current dbSNP database. For the analysis of exome data, non-coding and synonymous variants were filtered out, and only missense, nonsense, read-through and Indels were used for further analysis. Annotation of variants or markers to their physical positions was based on the Human Genome overview (Build GRCh37.p5).

Sanger sequencing
The candidate variants were directly sequenced using the Sanger method. Forward and reverse PCR primers were designed for each candidate variant (Supplementary Tables S1 and S4). PCR products were run on 2% agarose gels, and target bands excised and purified using QIAquick Gel Extraction kit (Qiagen). PCR amplification and sequencing was carried out using the GeneAmp PCR System 9700. Sequencing products were purified using Centri-Sep columns (Princeton Separations, Adelphia, NJ, USA), and sequences determined directly using BigDye Terminator v 1.1 Cycle Sequencing Kit (Applied Biosystems). PCR products were directly sequenced on an ABI PRISM3100 Genetic Analyzer, and sequence analysis performed using Sequencing Analysis v 5.3.1.

Prediction of function and homology alignment
The effect of rare or novel non-synonymous SNPs was assessed using Polyphen2 (Prediction of functional effects of human nsSNPs;[14] http://genetics.bwh.harvard.edu/pph2) and SIFT (Sorting Tolerant From Intolerant) algorithm;[15] (http://sift.bii.a-star.edu.sg/), which predict damage to protein function or structure based on amino-acid conservation and structural data. Homology of the candidate variants was determined using protein BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Restriction enzyme assays
We used restriction fragment length polymorphisms to determine the MAF in controls. Appropriate restriction enzymes for identifying genotypes of target variants were determined using NEBcutter (http://tools.neb.com/NEBcutter2/). Primers and restriction enzymes are described in Supplementary Table S2.

Reverse transcription (RT)-PCR
For candidate gene expression analysis in human thyroid gland tissue, we purchased the Human Total RNA Master Panel II (Clontech, Takara, Japan). Using designed target primers for each gene (Supplementary Table S3), RT-PCR was performed and products run in 1.5% agarose gels for 15 min.

RESULTS
Linkage analysis
In all three families, goiter was transmitted in an autosomal dominant inheritance pattern (Figure 1). Multipoint parametric linkage analysis did not identify a shared susceptibility locus among the families. However, in the MYG and THM families, we identified a susceptibility locus with a maximum LOD score of 3.8, from markers D4S412 to D4S419 at 4p16.3–4p15.3. In the THS and THM families, candidate loci partially overlapped at D3S1262, as described previously.[11] Our initial assumption had been that the two families either shared a variant, or had different variants within the same gene (allelic heterogeneity). However, despite exome sequencing, we were unable to find a shared variant or gene in the corresponding linkage region. Thus, as the disease shows genetic heterogeneity, we decided to combine our linkage analysis with exome sequencing, to search for causative variants in each family separately.

We simulated linkage analysis 100 times in each of the three families separately, and found that the genome-wide average false-positive rate for LOD scores >1.5 were all less than 1% (MYG, 0.56% (95% confidence interval (CI): 0–1.3); THS, 0.52% (95% CI: 0–1.3); THM, 0.94% (95% CI: 0–2.1)). Therefore, for this study, we considered loci with LOD scores >1.5 to be potential candidate regions.

Based on the given criteria, the following loci within each family were considered candidate regions for gene analysis. In the MYG family, a single locus was identified with a maximum LOD score of 2.41, and covering a 10.4-Mb genomic region, flanked by markers D4S412–D4S403 (Chr4: 3 380 694–Chr4: 13 750 946) on 4p16 (Figure 2a). In the THS family, three loci, D3S1262–D3S1311 at 3q27–3q29 (Chr3: 186 223 479–Chr3: 197 018 138; 10.8 Mb genomic region), D12S86–D12S1659 at 12q24.2–12q24.33 (Chr12: 119 170 322–Chr12: 129 416 408; 10.3 Mb genomic region) and D21S1914–D21S266 at 21q21–21q22 (Chr21: 25 622 401–Chr21: 42 684 564; 17.1 Mb genomic region), with maximum LOD scores of 1.81, 1.51 and 1.81, respectively, were identified (Figure 2b). In the THM family, we identified four susceptibility loci, D3S1262–D3S1580 at 3q27–3q28 (Chr3: 186 233 479–Chr3: 188 542 981; 2.3 Mb genomic region), D4S412–D4S391 at 4p16.3–4p16.14 (Chr4: 3 380 694–Chr4:
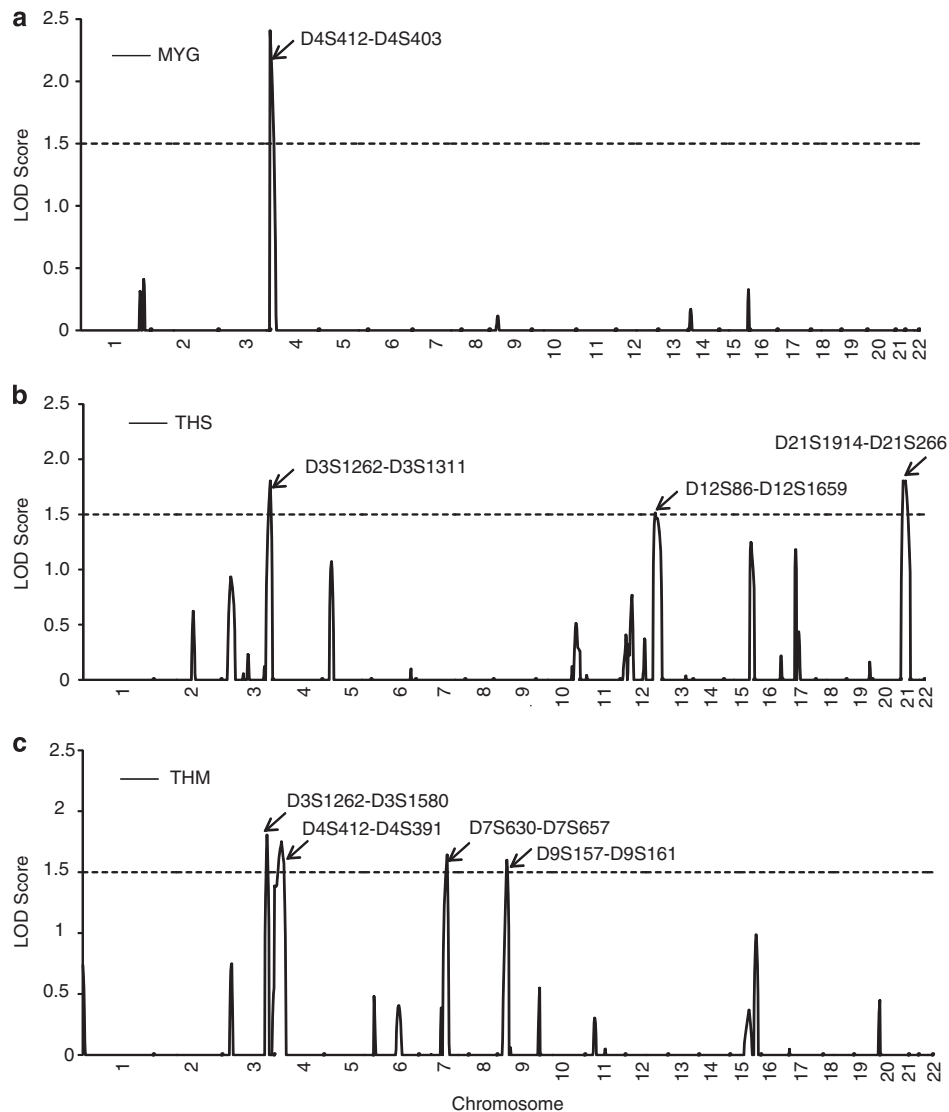
Figure 2 Genome-wide linkage analysis on the three Japanese goiter families. (**a**) A single genomic region between markers D4S412 and D4S403 (Chr4: 3 380 694–Chr4: 13 750 946), spanning 10 Mb of genomic region, and with a maximum logarithm of odds (LOD) score of 2.41 was identified in the MYG family. (**b**) Three susceptibility regions between markers D3S1262–D3S1311 (Chr3: 186 223 479–Chr3: 197 018 138; 10.8 Mb), D12S86–D12S1659 (Chr12: 119 170 322–Chr12: 129 416 408; 10.3 Mb) and D21S1914–D21S266 (Chr21: 25 622 401–Chr21: 42 684 564; 17.1 Mb) with maximum LOD scores of 1.81, 1.51 and 1.81, respectively, were identified in the THS family. (**c**) Four susceptibility genomic regions between markers D3S1262–D3S1580 (Chr3: 186 233 479–Chr3: 188 542 981; 2.3 Mb), D4S412–D4S391 (Chr4: 3 380 694–Chr4: 27 612 448; 24.3 Mb), D7S630–D7S657 (Chr7: 88 443 684–Chr7: 92 806 229; 4.4 Mb) and D9S157–D9S161 (Chr9: 17 628 302–Chr9: 27 632 327; 10.0 Mb) with maximum LOD scores of 1.81, 1.75, 1.64 and 1.60, respectively, were identified in the THM family. The physical position of markers is based on GRCh37.p5 primary assembly.

27 612 448; 24.3 Mb genomic region), D7S630–D7S657 at 7q21.13–7q21.21 (Chr7: 88 443 684–Chr7: 92 806 229; 4.4 Mb genomic region) and D9S157–D9S161 at 9p22–9p21.2 (Chr9: 17 628 302–Chr9: 27 632 327; 10.0 Mb genomic region), with maximum LOD scores of 1.81, 1.75, 1.64 and 1.60, respectively (Figure 2c). We focused on these candidate regions to search for causative variants using exome sequencing.

Exome-sequencing analysis
We performed exome capture using the Agilent SureSelect Human All Exon 50 Mb Kit, combined with massively parallel sequencing, generating 5.8–10.6 billion bases for six affected individuals. After mapping to the human reference genome (UCSC Genome Browser

hg19), we obtained 4.1–7.9 Gb of targeted exome sequence suitable for mapping, with a mean sequencing depth between 54- and 102-fold. On average, >88.1% of exomes were covered to a read depth of at least 10-fold (Table 1), which has been deemed sufficient from previous studies.[16–19]

After variant annotation, we focused our analysis primarily on nonsynonymous variants (missense, nonsense and read-through), splice acceptor and donor site variants, and coding Indels, anticipating that synonymous variants were less likely to be pathogenic. Using the public databases, dbSNP135, 1000 Genomes Project and NHLBI ESP6500, and five in-house control exome databases, we filtered our exome data. Given that familial goiter with an autosomal dominant inheritance pattern is rare, we assumed that the variants are also rare.

Table 1 Summary of exome sequencing results

Data	MYG		THS		THM	
	IV-2	*IV-3*	*III-4*	*IV-1*	*III-4*	*IV-3*
Sequencing and mapping data						
Raw data yield (Gb)	10.6	9.5	8.9	9.5	5.9	5.8
No. of effective bases (Gb) mapped to genome	7.9	6.3	5.8	6.4	4.1	4.5
Exome capture						
Effective bases on target region (Mb)[a]	5274	4172	3826	4238	2782	3009
Average sequencing depth on target region	102.32	80.94	74.24	82.23	53.97	58.38
Coverage of target region (%)	99.08	98.87	99.13	98.9	98.7	98.47
Fraction of target region covered with at least 10 X (%)	93.51	92.11	91.73	92.16	88.17	89.21
SNV and indel annotation						
No. of total SNVs	224 607	195 851	209 111	214 609	157 337	167 222
No. of total indels	20 772	20 258	20 172	20 094	18 705	19 093
Filtering criteria						
MS/NS/SS/RT/Indels in CCDS coding region	7952	7493	7963	8009	7952	8093
Less common variants (MAF<0.05)[b]	754	731	799	807	747	792
Heterozygous shared by two affected individuals	125		324		287	
Not presented in 5 in-house exome database	125		324		287	
Read depth ⩾10	125		287		253	
In the linkage regions	3		4		3	
	RGS12, c.166G>A		MASP1, c.64G>A		C4orf6, c.58_59insT	
	GRPEL1, c.110G>A		SYNJ1, c.4215_4216insAATACT		WFS1, c.1235T>C	
	CC2D2A, c.2882T>C		CLIC6, c.1651G>A		KIAA1797, c.2890C>G	
			SH3BGR, c.356_357insAGA			

Abbreviations: CCDS, consensus coding sequence; MAF, minor allele frequency; MS/NS/SS/RT, missense, nonsense, splicing site and read through variants; MYG, Miyagi Prefecture; SNVs, single-nucleotide variants; indels, small insertions and deletions.
[a]Target region, CCDS exonic regions and flanking intronic regions totally ∼50 Mb of genomic DNA.
[b]MAF, minor allele frequency in NHLBI ESP6500 and 1000 Genomes Project Phase I databases.

As it is also well known that both genetic and environmental factors play critical roles in goiter development,[1,2] we assumed low penetrance of the trait to minimize false-negative rates. Taken together, we focused on less common variants (MAF<0.05 in the 1000 Genomes Project phase I and NHLBI ESP6500 databases), filtering the data accordingly (Table 1). Overall, we detected 125, 287 and 253 SNVs and Indels in three families. Combined with linkage analysis, three or four genes with candidate variants were identified in each family.

In the MYG family we identified *RGS12* (c.166G>A, p.V56M; NM_198229.2), *GRPEL1* (c.110G>A, p.G37D; NM_025196.2) and *CC2D2A* (c.2882T>C, p.I961T; NM_001080522.2). In the THS family, *MASP1* (c.64G>A, p.V22M; NM_139125.3), *SYNJ1* (c.4215_4216insAATACT, p.1405_1406insNT; NM_003895.3), *CLIC6* (c.1651G>A, p.A551T; NM_053277.1) and *SH3BGR* (c.356_357 insAGA, p.119delinsGE; NM_007341.2) were identified. In the THM family *C4orf6* (c.58_59insT, p.120fs; NM_005750.2), *WFS1* (c.1235T>C, p.V412A; NM_006005.3) and *KIAA179* (c.2890C>G, p.L964V; NM_017794.3) were identified (Table 1). For SNVs, we used Polyphen2 and SIFT to predict the functional impact, thereby eliminating *CC2D2A* as a candidate, as p.I961T was predicted benign by both methods.

Confirmation of exome sequencing by direct sequencing and segregation analysis
Next, we performed Sanger sequencing to confirm the exome findings and determine segregation of the candidate variants with the goiter phenotype (Table 2). We found p.V56M in *RGS12* and p.G37D in *GRPEL1* completely segregated with goiter in the MYG family. In addition, p.A551T in *CLIC6* and p.V412A in *WFS1* also completely segregated with goiter in the THS and THM families, respectively (Figure 3). The other variants identified by exome sequencing were validated but did not segregate with the goiter phenotype.

Minor allele frequency in controls
We determined the MAF of candidate genes in population-relevant controls (Table 3, Supplementary Figure S1). The MAF of p.V56M (*RGS12*) and p.G37D (*GRPEL1*) was 12 (0.04) and 2 (0.007), respectively, in 150 controls from Miyagi Prefecture. In the MYG family, these two variants were located on the same chromosomal haplotype (Figure 3a). Carriers of these two rare variants were not observed in controls, suggesting the haplotype frequency is <1%. The MAF of p.A551T (*CLIC6*) and p.V412A (*WFS1*) was 0 (0) and 5 (0.017), respectively, in 150 controls from Akita Prefecture.

Species homology and gene expression in the thyroid gland
We compared homology of the candidate variants in mammals, and found conservation of the target amino acids in all candidate proteins (Figure 4).

Using RT-PCR, we determined expression of the candidate genes in the human thyroid gland, and found that all genes were expressed in human thyroid tissue (Figure 5).

Table 2 Exome-sequencing results from linkage regions

Pedigree	Linkage region	Maximum LOD score	Exome findings in linkage regions							Function prediction		
			Gene	Position[a]	Transcript	dbSNP[b]	SNV	Protein	MAF[c]	Polyphen2	SIFT	Segregation
MYG	D4S412–D4S403	2.41	RGS12	chr4: 3318063	NM_198229.2	rs80169463	c.166G>A	p.V56M	0.0055/0.0449	Probably damaging	Tolerate	Yes
			GRPEL1	chr4: 7065903	NM_025196.2	rs140117073	c.110G>A	p.G37D	0.0032/0.0169	Possibly damaging	Damaging	Yes
			CC2D2A	chr4: 15560840	NM_001080522.2	rs76626268	c.2882T>C	p.I961T	0.0023/0.0281	Benign	Tolerate	—
THS	D3S1262–D3S1311	1.81	MASP1	chr3: 187003786	NM_139125.3	rs77189011	c.64G>A	p.V22M	0.0032/0.0169	Probably damaging	Tolerate	No
	D12S86–D12S1659	1.51	/									
	D21S1914–D21S266	1.81	SYNJ1	chr21: 34003928	NM_003895.3	rs71640263	c.4215_4216insAATACT	p.1405_1406insNT	NA	—	—	No
			CLIC6	chr21: 36081038	NM_053277.1	rs145427334	c.1651G>A	p.A551T	0.0005/0	Probably damaging	Damaging	Yes
			SH3BGR	chr21: 40883671	NM_007341.2	rs75027626	c.356_357insAGA	p.119delinsGE	NA	—	—	No
THM	D3S1262–D3S1580	1.81	/									
	D4S412–D4S391	1.75	C4orf6	chr4: 5527115	NM_005750.2	rs11426893	c.58_59insT	p.I20fs	NA	—	—	No
			WFS1	chr4: 6302757	NM_006005.3	rs144951440	c.1235T>C	p.V412A	0.0037/0.0112	Probably damaging	Tolerate	Yes
	D7S630–D7S657	1.64	/									
	D9S157–D9S161	1.60	KIAA1797	chr9: 20923696	NM_017794.3	rs117748510	c.2890C>G	p.L964V	0.0069/0.0169	Possibly damaging	Tolerate	No

Abbreviations: delins, deletion-insertion; fs, frame shift; MYG, Miyagi perfecture; SNV, single nucleotide variant; NA, not available; —, not checked.
[a]GRCh37.p5—primary assembly; /, no sequence variants;
[b]SNP annotation was based on NCBI dbSNP135 database. MAFs for these SNPs are cited from 1000 Genome Phase I database.
[c]MAF, minor allele frequency in 1000 Genomes Phase I database (1092 individuals worldwide / 89 Japanese individuals); none of these variants were found in the NHLBI ESP6500 database.
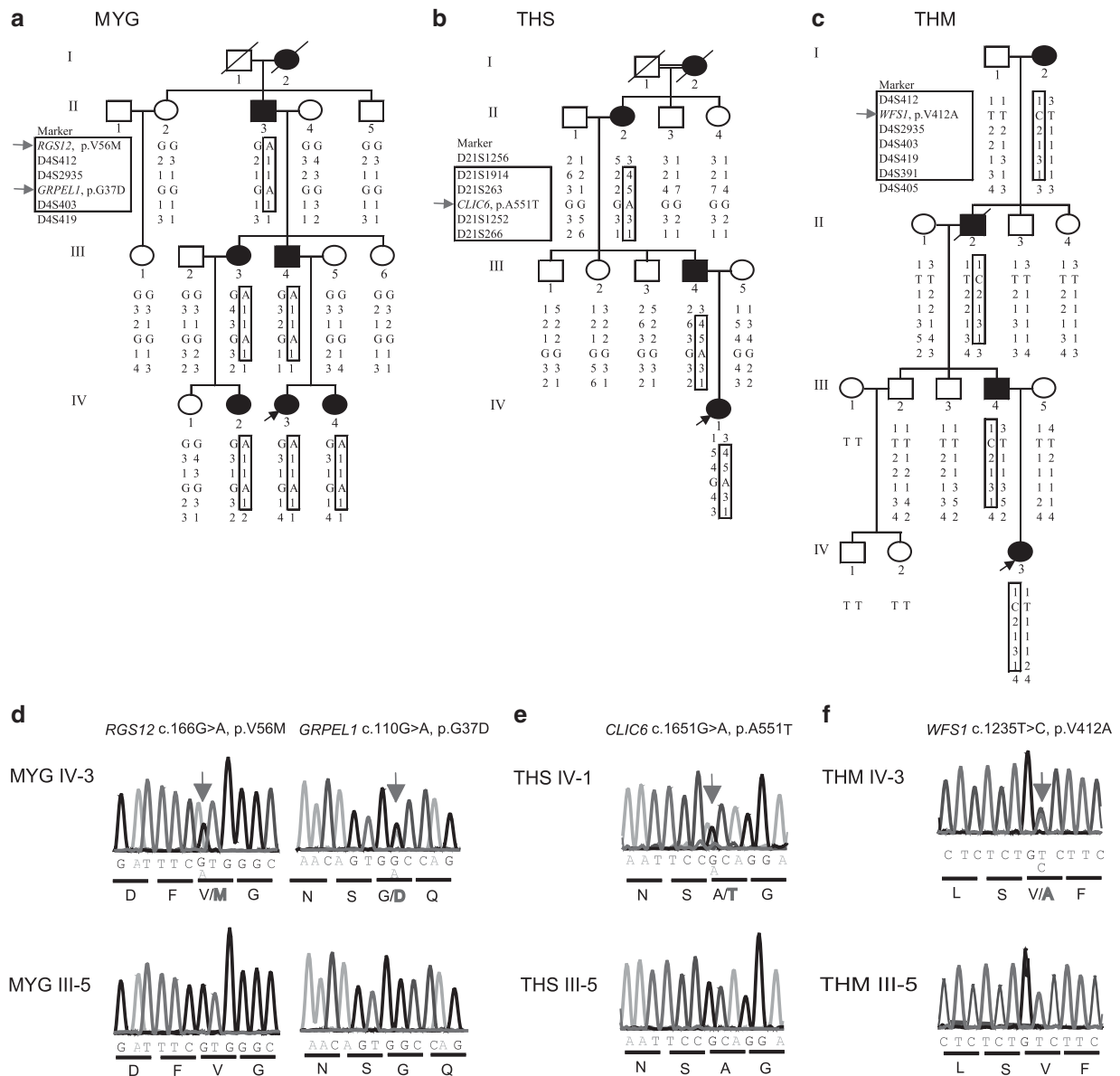
Figure 3 Risk haplotypes associated with goiter and the segregating variants identified in each family. (a–c) Risk haplotypes and segregating variants identified in each family. Markers and segregating variants are shown on the left side. The genotype for each individual is illustrated. Haplotypes were estimated by Genehunter (Ver2.1_r6). Risk haplotypes are boxed and arrowheads show segregating variants. In the MYG family (a), two rare heterozygous missense variants (c.166G>A, p.V56M in *RGS12* and c.110G>A, p.G37D in *GRPEL1*) segregate with the goiter phenotype. In the THS (b) and THM (c) families, rare missense variants c.1651G>A, p.A551T in *CLIC6* (b) and c.1235T>C, p.V412A in *WFS1* (c) segregate with the goiter phenotype in each family, respectively. (d–f) Representative Sanger sequence chromatograms of each variant are shown for affected (upper panels) and unaffected (bottom panels) individuals. Predicted amino acid changes and surrounding amino acids are indicated below the sequence. Mutated nucleotides are indicated using red arrowheads. V, valine; M, methionine; G, glycine; D, aspartic acid; A, alanine; T, threonine. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

Known genes related to CH

All probands from the goiter families in this study were diagnosed with CH based on high TSH levels in the neonatal mass screening program. Thus, using our exome-sequencing data, we also systematically investigated all known candidate genes related to CH, namely *TG, TSHR, TSHB, SLC5A5, SLC26A4, IYD, TPO, DUOX2, DUOXA2, NKX2-1, FOXE1, PAX8* and *NKX2-5*.[2,20] We found several rare or novel missense or splicing site variants, p.S1222L, p.G1479R,

IVS47 + 1C > T and p.N2616I in *TG*, p.S305R and p.C636F in *TSHR*, and p.G322S in *NKX2-1* (Table 4). We confirmed these variants by Sanger sequencing (Figure 6).

In the MYG family, the rare missense variants, p.S1222L in *TG* and p.C636F in *TSHR*, are present in non-affected, as well as affected subjects. In *TG*, the p.S1222L variant is shared by all the goiter-affected and four unaffected individuals (II-2, II-5, III-6 and IV-1). II-2, II-5 and III-6 did not show any goiter, and IV-1 did not show

a *RGS12* p.V56M

Homo sapiens	42	C V L S C V M R G S P A D F [V] G L R A G D Q I L A V N E I N 71
Pan paniscus	42	C V L S C V M R G S P A D F [V] G L R A G D Q I L A V N E I N 71
Nomascus leucogenys	42	C V L S C V M R G S P A D F [V] G L R A G D Q I L A V N E I N 71
Canis lupus familiaris	42	C V L S C V M R G S P A D F [V] G L R A G D Q I L A V N E I N 71
Ailuropoda melanoleuca	42	C V L S C V M R G S P A D F [V] G L R A G D Q I L A V N E I N 71
Equus caballus	42	C V L S C V M R G S P A D F [V] G L R A G D Q I L A V N E I N 71

b *GRPEL1* p.G37D

Homo sapiens	22	P S P R L L C T A T K Q K N S [G] Q N L E E D M G Q S E Q K A 51
Macaca mulatta	22	P S P R L L C T A T K Q K N S [G] Q N L E E D M G Q S E Q K A 51
Callithrix jacchus	22	P S P R L L C T A T K Q K N S [G] Q N L E E D M G Q S E Q K A 51
Saimiri boliviensis	22	P S P R L L C T A T K Q K N S [G] Q N L E E D M G Q S E Q K A 51

c *CLIC6* p.A551T

Homo sapiens	538	Y P K L G T Q H P E S N S [A] G N D V F A K F S A F I K N T K 567
Mus musculus	448	Y P K L G T Q H P E S N S [A] G N D V F A K F S A F I K N T K 477
Sus scrofa	492	Y P K L G T Q H P E S N S [A] G N D V F A K F S A F I K N T K 521
Papio anubis	498	Y P K L G T Q H P E S N S [A] G N D V F A K F S A F I K N T K 527
Bos taurus	507	Y P K L G T Q H P E S N S [A] G N D V F A K F S A F I K N T K 536
Pan troglodytes	508	Y P K L G T Q H P E S N S [A] G N D V F A K F S A F I K N T K 537
Oryctolagus cuniculus	489	Y P K L G T Q H P E S N S [A] G N D V F A K F S A F I K N T K 518

d *WFS1* p.V412A

Homo sapiens	398	G W N H L E P Y A H F L L S [V] F F V I F S F P L A S K D C I 427
Nomascus leucogenys	398	G W N H L E P Y A H F L L S [V] F F V I F S F P L A S K D C I 427
Canis lupus familiaris	392	G W N H L E P Y V H F L L S [V] F F V I F S F P L A S K D C I 421
Saimiri boliviensis	398	G W N H L E P Y A H F L L S [V] F F V I F S F P L A S K D C I 427
Gallus gallus	383	G W N H L E P Y I Y F L L S [V] F F V I F S F P L A S K D C I 412
Taeniopygia guttata	390	G W N H L E P Y I Y F L L S [V] F F V I F S F P L A S K D C I 419

Figure 4 Homology of candidate variants within mammalian species. (**a–d**) In mammals, BLAST alignments (http://blast.ncbi.nlm.nih.gov/Blast.cgi) identified conservation of valine at position 56 of *RGS12*, glycine at position 37 of *GRPEL1*, alanine at position 551 of *CLIC6* and valine at position 412 of *WFS1*. Arrowheads indicate the position of conserved amino acids.

Table 3 Minor allele frequency in controls

			Genotype			
Gene	*SNV*	*Protein*	*Wild type*	*Hetero-zygous*	*Homo-zygous*	*MAF*
Miyagi controls						
RGS12	c.166G>A	p.V56M	140	8	2	0.041
GRPEL1	c.110G>A	p.G37D	148	2	0	0.007
Haplotype[a]	—	—	150	0	0	0
Akita controls						
CLIC6	c.1651G>A	p.A551T	150	0	0	0
WFS1	c.1235T>C	p.V412A	145	5	0	0.017

Abbreviations: MAF, minor allele frequency; SNV, single-nucleotide variant.
[a]Haplotype of c.166G>A in *RGS12* and c.110G>A in *GRPEL1* in controls.

a *RGS12*

Exon15 | Exon16
cDNA-F   cDNA-R

GRPEL1

Exon2 | Exon3 | Exon4
cDNA-F   cDNA-R

CLIC6

Exon4 | Exon5
cDNA-F cDNA-R

WFS1

Exon3 | Exon4
cDNA-F cDNA-R

b

Size (bp) Blank RGS12 GRPEL1 CLIC6 WFS1
200
100

Figure 5 Gene expression in the human thyroid gland. (**a**) Exonic location of cDNA primers used for RT-PCR amplification of *RGS12*, *GRPEL1*, *CLIC6* and *WFS1*. Primer sequences are provided in Supplementary Table S3. (**b**) Electrophoresis of RT-PCR products. The expected sizes for each product are 132, 161, 161 and 164 bp, respectively. All four genes show expression in the human thyroid gland.

goiter in the 10-year follow-up. Thus, we excluded p.S1222L in *TG* as a candidate causal variant. The p.C636F variant in *TSHR* did not segregate with the goiter phenotype, with three affected individuals (III-3, IV-2 and IV-4) who did not carry this variant. Another missense variant, p.G1479R in *TG*, was transmitted from the healthy individual (III-2).

In the THS family, the rare missense variants, p.N2616I in *TG* and p.S305R in *TSHR* were identified. Segregation analysis revealed that these two variants were transmitted from a healthy family member (II-1).
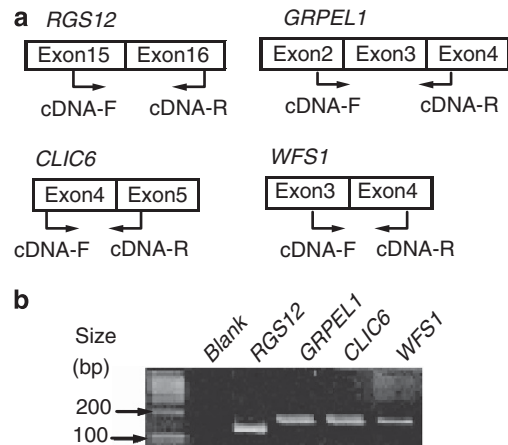
In the THM family, the splicing site variant, IVS47 + 1C>T in *TG*, and a novel missense variant, p.G322S in *NKX2-1*, were identified. The IVS47 + 1C>T variant was not carried by the affected proband (IV-3) and did not segregate within this family. The p.G322S variant in *NKX2-1* was transmitted from a healthy family member (I-1).

Table 4 Genes known to be involved in congenital hypothyroidism or familial mutinodular goiter

Gene	Location	Exome findings							THS		THM			MYG		Function prediction	
		Position[a]	SNV	Transcript	Exon	Protein	dbSNP[b]	MAF	III-4	IV-1	III-4	IV-4	IV-3	IV-2	IV-3	Polyphen2	SIFT
TG	8q24	chr8: 133918963	c.3665C>T	NM_003235	17	p.S1222L	rs12549018	0.0096/0.0169	0/0	0/0	0/0	0/0	0/0	0/1	0/1	Possibly damaging	Damaging
		chr8: 133931677	c.4435G>A		21	p.G1479R	rs114211101	0.0014/0.0112	0/0	0/0	0/1	0/1	0/0	0/1	0/0	Probably damaging	Damaging
		chr8: 134125846	IVS47+1C>T		Splicing site	—	—	—	0/1	0/1	0/1	0/1	0/1	0/0	0/0	—	Damaging
		chr8: 134128945	c.7847A>T		48	p.N2616I	—	0.0769/0	0/1	0/1	0/1	0/1	0/0	0/0	0/0	Benign	Damaging
TSHR	14q31	chr14: 81609317	c.915T>A	NM_000369	11	p.S305R	rs142122217	0.0018/0.0112	0/0	0/0	0/0	0/0	0/0	0/1	0/1	Possibly damaging	Tolerant
		chr14: 81610309	c.1907G>T		11	p.C636F	—	—	0/0	0/0	0/0	0/1	0/1	0/1	0/1	Probably damaging	Damaging
		chr14: 369866635	c.964G>A	NM_003317	2	p.G322S	—	0.0023/0.0056	0/0	0/0	0/0	0/1	0/1	0/0	0/0	Benign	Tolerant
NKX2-1	14q13	No sequence variants															
TSHB	1p13	No sequence variants															
SLC5A5	19p13.11	No sequence variants															
SLC26A4	7q31	No sequence variants															
IYD	6q25.1	No rare sequence variants															
TPO	2p25	No rare sequence variants															
DUOX2	15q15.3	No rare sequence variants															
DUOXA2	15q15.1	No rare sequence variants															
FOXE1	9q22	No sequence variants															
PAX8	2q13	No sequence variants															
NKX2-5	5q34	No sequence variants															
MNG1	14q	No sequence variants															

Abbreviations: DUOX2, dual oxidase 2; DUOXA2, dual oxidase maturation factor 2; FOXE1, forkhead box E1; IYD, iodotyrosine deiodinase; MAF, minor allele frequency; MNG1, multinodular goiter 1; NKX2-1, NK2 homeobox 1; NKX2-5, NK2 transcription factor related 5; PAX8, paired box8; SLC26A4, solute carrier family 26, member 4, also known as PDS, Pendred syndrome; SNV, single nucleotide variant; TG, thyroglobulin; TPO, thyroid peroxidase; TSHR, thyroid-stimulating hormone receptor; TSHB, thyroid stimulating hormone, beta; SLC5A5, solute carrier family 5 (sodium iodide symporter), member 5.
Rare sequence variants, with a minor allele frequency <0.05 in 1000 Genomes (phase I) and NHLBI ESP6500 databases.
MAF, minor allele frequency in 1000 Genomes (phase I) database (1092 individuals worldwide/89 Japanese individuals).
0/0, wild type; 0/1, heterozygous; 0 and 1 represent major and minor alleles, respectively.
aGRCh37, p5—primary assembly.
bSNP annotation based on NCBI dbSNP135 database.

Thus, as both variants were either transmitted from a healthy individual or not carried by obvious affected individuals, we can exclude them as causative in our families.

DISCUSSION

Linkage analysis is an effective method to detect susceptibility loci with a large effect size.[21] Traditionally, a combination of linkage analysis and Sanger sequencing has been used to search for causative variants for disease.[22] However, because of locus and disease heterogeneity, the conventional approach has been difficult to perform, as the causative variants may be present in any number of candidate genes. The recent development of massively parallel DNA-sequencing technologies provides a powerful way to identify causative variants responsible for Mendelian or common disorders.[23] Even though exomes only constitute about 1% of the human genome, they are estimated to be the major source of causative variants, constituting 85% of disease-causing mutations.[24] Thus, by combining traditional linkage analysis and whole-exome sequencing, we have maximized our chances of identifying a causative variant.

Using this approach, we identified rare missense variants in four genes (RSG12, GRPEL1, CLIC6 and WFS1) that segregate with familial goiter. These genes have not previously been linked to goiter, and in this regard are novel. All variants are predicted by Polyphen2 or SIFT to be detrimental to the protein, are rare in controls and are expressed in thyroid, suggesting that they likely play an important role in familial goiter development.

RGS12, located at 4p16.3, is a member of the regulator of G protein signaling (RGS) gene family. The RGS family modulates function of G proteins by activating the intrinsic guanosine triphosphatase activity of the α-subunits.[25] RGS12 encodes a protein that may function as a transcriptional repressor in addition to its role as a guanosine triphosphatase-activating protein. In hot and cold thyroid nodules, mRNA transcripts of RGS12 are significantly downregulated compared with normal tissue, implying that loss-of-function variants of RGS transcripts in thyroid nodules may contribute to goiter development.[26] In the MYG family, we found a missense variant (p.V56M) predicted by Polyphen to be detrimental to RGS12 function, segregating with the goiter phenotype, in accordance with its probable role in goiter pathology. However, it is unlikely that this variant acts independently, as its prevalence in controls was high (MAF, 0.041). Interestingly, in the same haplotype from this family, we found another rare missense variant, p.G37D, in GRPEL1. GRPEL1 is a GrpE-like 1, mitochondrial (Escherichia coli) gene that encodes an essential component of the presequence-associated motor complex.[27] GRPEL1 is necessary for translocation of transit peptide-containing proteins from the inner mitochondrial membrane to the mitochondrial matrix, in an ATP-dependent manner, controlling nucleotide-dependent binding of mitochondrial HSP70 to substrate proteins.[27] In the local population, the GRPEL1 MAF was 0.7%; therefore, as for RGS12, relatively high prevalence does not support GRPEL1 as an independent gene for familial goiter within this family. However, if we assume an interaction between these two genes in the thyroid gland, the identified variants may potentially impair this interaction; therefore the haplotype carrying these two variants may be a susceptibility locus. Whether or how these two genes interact is still unknown; yet, as the haplotype carrying these two variants was not observed in 150 controls, nor in any of the publically available databases we examined, we are tempted to speculate that the combination of these two rare variants leads to goiter predisposition in this family.
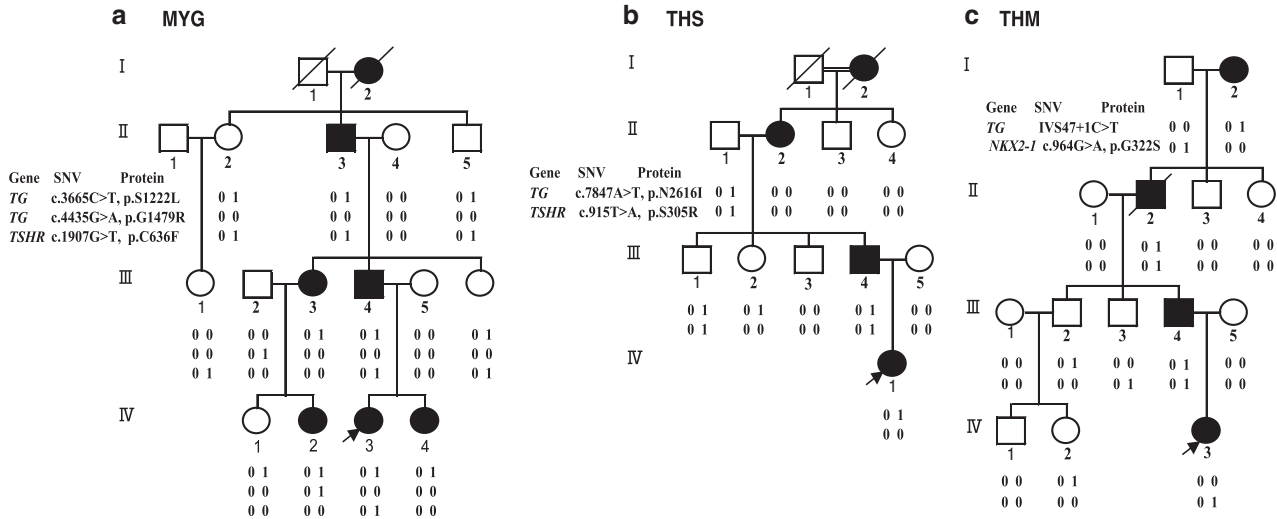
Figure 6 Genotyping and segregation analysis of known candidate genes related to congenital hypothyroidism (CH). Gene and candidate variants identified from exome data are shown on the left side. Corresponding genotypes are shown below each individual. 0 1, heterozygous; 0 0, wild type. 0 and 1 represent major and minor alleles, respectively. (a) In the MYG family, rare (p.S1222L in *TG*) and novel (p.G1479R in *TG* and p.C636F in *TSHR*) missense variants were identified from exome sequencing, and confirmed by Sanger sequencing, but did not segregate with the goiter phenotype. (b) In the THS family, exome sequencing identified novel (p.N2616I in *TG*) and rare (p.S305R in *TSHR*) missense variants. Segregation analysis revealed these two variants were transmitted from the healthy individual, II-1, and showed no segregation with the goiter phenotype. (c) In the THM family, exome sequencing identified rare splicing site (IVS47 + 1C>T in *TG*), and novel (p.G322S in *NKX2-1*) variants. Segregation analysis revealed that p.G322S in *NKX2-1* was transmitted from the healthy individual, I-1. Both variants showed no segregation with the goiter phenotype.

In the THS family, the proband (IV-1, a 3-year-old girl) developed multinodular goiter bilaterally, despite receiving thyroid hormone therapy since she was 1 month old. Two affected individuals (II-2 and III-4) have undergone thyroidectomies because of huge and compressed goiters. In this family, we identified a rare missense variant (p.A551T) in *CLIC6*, which completely segregated with the goiter phenotype. We did not detect this variant in 150 controls. Additionally, this variant is not detected in the NHLBI ESP6500 database and is very rare in the 1000 Genomes Project (MAF, 0.0005 in 1092 worldwide individuals, and not detected in 89 Tokyo Japanese individuals). *CLIC6*, located at 21q22.12, encodes a member of the chloride intracellular channel family of proteins.[28] It is expressed predominantly in the stomach, pituitary and brain, interacting with D2-like dopamine receptors directly and through scaffolding proteins.[29] *CLIC6* may also be involved in secretion regulation, potentially through chloride ion transport regulation. The rare variant found in the THS family is located at the C-terminal, alpha helical domain of *CLIC6*, a highly conserved region. Thus, it is feasible that p.A551T in *CLIC6* is the causative mutation within the THS family. Further functional analysis of *CLIC6* is needed to ascertain its exact role in thyroid function and goiter development.

In the THM family, we unexpectedly found that a rare missense variant (p.V412A) in *WFS1* completely segregates with the goiter phenotype. *WFS1* is known to be responsible for *WFS1*-related disorders, including Wolfram syndrome[30] and *WFS1*-related low-frequency sensory hearing loss (also known as *DFNA6/14/38* low-frequency sensorineural hearing loss[31,32]), which are inherited in an autosomal recessive and dominant pattern, respectively. In the various clinical manifestations of *WFS1*-related disorders, hypothyroidism is reported but its frequency is unknown.[33] In the THM family, the affected individuals have goiter, but thyroid functions are normal and they do not exhibit hearing loss. The proband (IV-3) was diagnosed with mild and diffuse goiter at 8 years of age, her father (III-4) with euthyroid goiter at 30 years of age, and her great-grandmother (I-2)

had goiter with normal thyroid function. The MAF of the *WFS1* variant in controls was 0.017, showing relatively high prevalence of this variant in the local population. However, we observed that this variant is rare in the 1000 Genomes Project database (MAF, 0.0037) and not present in the NHLBI ESP6500 database. Thus, it is likely that a combination of triggering environmental factors and a genetic predisposition results in euthyroid familial goiter development, as reported previously.[1–4] Moreover, we speculate that based on our results, P.V412A in *WFS1* may be a contributing genetic factor, at least in this family.

Because of their known and important role in hormone synthesis and thyroid physiology, *TG, TSHR, TSHB, SLC5A5, SLC26A4, IYD, TPO, DUOX2, DUOXA2, NKX2-1, FOXE1, PAX8* and *NKX2-5* are potential candidate genes for familial goiter.[2,20] Therefore, using our exome-sequencing data, we systematically investigated all of these first-line candidate genes. We found several rare or novel missense or splicing site variants in *TG, TSHR* and *NKX2-1*, but can exclude them as causative in our families, as most of the variants are either transmitted from healthy individuals or not carried by all the affected individuals. The one exception is the rare heterozygous missense variant (p.S1222L) in *TG*, which is shared by all the goiter-affected individuals in the MYG family. Given that goiter penetrance is low and the variant causes various thyroid diseases, this variant was reasonable to be the causative one. However, in the MYG family, we found that the goiter has obvious early age onset. There were three girls diagnosed with CH, who developed goiter in the early age stage. Another young girl (IV-1), a carrier of this variant, was normal throughout the 10-year follow-up period. As this variant also presented in four unaffected individuals, we considered it to be less pathogenic within the family. Moreover, further evidence also suggested that this variant was unlikely to be the major causative one. First, the pattern of goiter inheritance contradicts the mode of disease inheritance typically observed with *TG* mutations. To date, 50 mutations in *TG* have been identified in CH patients, and they are

predominantly inherited in an autosomal recessive manner, with affected individuals either homozygous or compound heterozygous for mutations.[34] In contrast, in the MYG family, the goiter is inherited in an autosomal dominant pattern. Second, there is a phenotypic discrepancy from traditional phenotypes attributed to *TG* mutations. Although clinical manifestations of CH display wide phenotype variation, from euthyroidism to severe hypothyroidism, generally CH patients with *TG* mutations exhibit low serum TG levels and elevated TSH with simultaneous low serum FT4.[34] However, the affected individuals in the MYG family show elevated TG levels (II-3, $90\,ng\,ml^{-1}$; III-3, $73\,ng\,ml^{-1}$; normal level, $<30\,ng\,ml^{-1}$). Thus, although p.S1222L in *TG* is common in the MYG family, it is unlikely to contribute to the goiter phenotype expressed in this family. Additionally, although two rare heterozygous missense variants (p.C636F and p.S305R) in *TSHR* were identified in these goiter families, we excluded them as causative because they were either transmitted from a healthy individual or not carried by all affected individuals. This is consistent with reports that monoallelic mutations in *TSHR* may be less pathogenic because of their estimated high prevalence.[35]

Our current study has several limitations. First, the family sizes are relatively small. As familial goiter is rare (prevalence of CH with a family history of goiter is estimated at 1/60 000 in Akita Prefecture[11]), we have only identified three families during the past decades, with the family sizes not large enough to provide sufficient information for linkage analysis. Inevitably, several false-positive or negative loci would be identified by linkage analysis. To compensate for this low statistical power and to minimize false-positive locus identification, we combined exome sequencing and genome-wide linkage analysis, with a LOD score of 1.5 as a threshold for candidate locus identification. Second, even though exome-sequencing analysis is an effective way to detect coding sequencing variants, it cannot reveal mutations (that is, SNPs and structural variants) in noncoding regulatory regions, such as promoters or enhancers. Third, in our goiter families, the detection of goiter phenotypes is highly dependent on the mass screening program. Certain phenotypes may be undetected in the years before the introduction of mass screening. Consequently, we were unable to determine a consistent phenotype and the results may be confounded by bias. Fourth, our analysis is based on genetic homogeneity within the families, and we did not consider genetic heterogeneity. Finally, we did not conduct functional analyses to evaluate the effects of the variants on thyroid function, as complicated biological processes in the thyroid gland hampered our approach.

Familial goiter is a heterogeneous disease, at the level of both the phenotype and the genotype. In our three goiter families we identified four genes, *RGS12, GRPEL1, CLIC6* and *WFS1*, that are novel in their association with the goiter phenotype. We excluded all first-line candidate genes as candidates. Our findings are consistent with the general agreement that additional proteins are likely to be involved in the mechanisms of goiter development, as the goiter biochemical pathway has not yet been fully elucidated. However, without further functional analysis, we cannot definitely conclude they are causative biologically. Further functional research is needed to provide a comprehensive understanding of the biological mechanisms involved in familial goiter.

1   Bottcher, Y., Eszlinger, M., Tonjes, A. & Paschke, R. The genetics of euthyroid familial goiter. *Trends. Endocrinol. Metab.* **16,** 314–319 (2005).
2   Krohn, K., Führer, D., Bayer, Y., Eszlinger, M., Brauer, V., Neumann, S. *et al.* Molecular pathogenesis of euthyroid and toxic multinodular goiter. *Endocr. Rev.* **26,** 504–524 (2005).
3   Knudsen, N., Laurberg, P., Perrild, H., Bülow, I., Ovesen, L. & Jørgensen, T. Risk factors for goiter and thyroid nodules. *Thyroid.* **12,** 879–888 (2002).
4   Brix, T. H., Kyvik, K. O. & Hegedüs, L. Major role of genes in the etiology of simple goiter in females: a population-based twin study. *J. Clin. Endocrinol. Metab.* **84,** 3071–3075 (1999).
5   Hansen, P. S., Brix, T. H., Bennedbaek, F. N., Bonnema, S. J., Kyvik, K. O. & Hegedüs, L. Genetic and environmental causes of individual differences in thyroid size: a study of healthy Danish twins. *J. Clin. Endocrinol. Metab.* **89,** 2071–2077 (2004).
6   Malamos, B., Koutras, D. A., Kostamis, P., Rigopoulos, G. A., Zerefos, N. S. & Yataganas, X. A. Endemic goiter in Greece: a study of 379 twin pairs. *J. Med. Genet.* **4,** 16–18 (1967).
7   Bignell, G. R., Canzian, F., Shayeghi, M., Stark, M., Shugart, Y. Y., Biggs, P. *et al.* Familial nontoxic multinodular thyroid goiter locus maps to chromosome 14q but does not account for familial nonmedullary thyroid cancer. *Am. J. Hum. Genet.* **61,** 1123–1130 (1997).
8   Neumann, S., Willgerodt, H., Ackermann, F., Reske, A., Jung, M., Reis, A. *et al.* Linkage of familial euthyroid goiter to the multinodular goiter-1 locus and exclusion of the candidate genes thyroglobulin, thyroperoxidase, and Na$^+$/I$^-$ symporter. *J. Clin. Endocrinol. Metab.* **84,** 3750–3756 (1999).
9   Capon, F., Tacconelli, A., Giardina, E., Sciacchitano, S., Bruno, R., Tassi, V. *et al.* Mapping a dominant form of multinodular goiter to chromosome Xp22. *Am. J. Hum. Genet.* **67,** 1004–1007 (2000).
10 Bayer, Y., Neumann, S., Meyer, B., Rüschendorf, F., Reske, A., Brix, T. *et al.* Genome-wide linkage analysis reveals evidence for four new susceptibility loci for familial euthyroid goiter. *J. Clin. Endocrinol. Metab.* **89,** 4044–4052 (2004).
11 Takahashi, T., Nozaki, J., Komatsu, M., Wada, Y., Utsunomiya, M., Inoue, K. *et al.* A new locus for a dominant form of multinodular goiter on 3q26.1-q26.3. *Biochem. Biophys. Res. Commun.* **284,** 650–654 (2001).
12 Koizumi, A., Harada, K. H., Inoue, K., Hitomi, T., Yang, H. R., Moon, C. S. *et al.* Past, present and future of environmental specimen banks. *Environ. Health. Prev. Med.* **14,** 307–318 (2009).
13 Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. & Lander, E. S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58,** 1347–1363 (1996).
14 Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods.* **7,** 248–249 (2010).
15 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4,** 1073–1081 (2009).
16 Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461,** 272–276 (2009).
17 Bell, C. J., Dinwiddie, D. L., Miller, N. A., Hateley, S. L., Ganusova, E. E., Mudge, J. *et al.* Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med.* **3,** 65ra4 (2011).
18 Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P. *et al.* Genetic diagnosis by whole-exome capture and massively parallel DNA sequencing. *Proc. Natl. Acad. Sci. USA.* **106,** 19096–19101 (2009).
19 Chang, H., Jackson, D. G., Kayne, P. S., Ross-Macdonald, P. B., Ryseck, R. P. & Siemers, N. O. Exome sequencing reveals comprehensive genomic alterations across eight cancer cell lines. *PLoS One* **6,** e21097 (2011).
20 Park, S. M. & Chatterjee, V. K. Genetics of congenital hypothyroidism. *J. Med. Genet.* **42,** 379–389 (2005).
21 Bailey-Wilson, J. E. & Wilson, A. F. Linkage Analysis in the Next-Generation Sequencing Era. *Hum. Hered.* **72,** 228–236 (2011).
22 Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nat. Genet.* **33** (suppl) 228–237 (2003).
23 Mardis, E. R. & Next-generation, D. N. A. sequencing methods. *Annu. Rev. Genomics. Hum. Genet.* **9,** 387–402 (2008).
24 Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12,** 745–755 (2011).
25 Watson, N., Linder, M. E., Druey, K. M., Kehrl, J. H. & Blumer, K. J. RGS family members: TPase-activating proteins for heterotrimeric Gprotein-subunits. *Nature* **383,** 172–175 (1996).
26 Tönjes, A., Miedlich, S., Holzapfel, H. P., Eszlinger, M., Arkenau, C. & Paschke, R. Expression of regulators of g protein signaling mRNA is differentially regulated in hot and cold thyroid nodules. *Thyroid.* **14,** 896–901 (2004).
27 Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T. & Pfanner, N. Importing Mitochondrial Proteins: Machineries and Mechanisms. *Cell.* **138,** 628–644 (2009).
28 Hattori, M., Fujiyama, A., Taylor, T. D., Watanabe, H., Yada, T., Park, H. S. *et al.* The DNA sequence of human chromosome 21. *Nature* **405,** 311–319 (2000).
29 Griffon, N., Jeanneteau, F., Prieur, F., Diaz, J. & Sokoloff, P. CLIC6, a member of the intracellular chloride channel family, interacts with dopamine D(2)-like receptors. *Brain. Res. Mol. Brain. Res.* **117,** 47–57 (2003).

30 Strom, T. M., Hörtnagel, K., Hofmann, S., Gekeler, F., Scharfe, C., Rabl, W. *et al.* Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (Wolframin) coding for a predicted transmembrane protein. *Hum Mol. Genet.* **7,** 2021–2028 (1998).

31 Komatsu, K., Nakamura, N., Ghadami, M., Matsumoto, N., Kishino, T., Ohta, T. *et al.* Confirmation of genetic homogeneity of nonsyndromic low-frequency sensorineural hearing loss by linkage analysis and a DFNA6/14 mutation in a Japanese family. *Hum. Genet.* **47,** 395–399 (2002).

32 Fukuoka, H., Kanda, Y., Ohta, S. & Usami, S. Mutations in the WFS1 gene are a frequent cause of autosomal dominant nonsyndromic low-frequency hearing loss in Japanese. *J. Hum. Genet.* **52,** 510–515 (2007).

33 Tranebjaerg, L., Barrett, T. & Rendtorff, N. D. WFS1-Related Disorders. Updated 2009 Jun 2. Available from: http://www.ncbi.nlm.nih.gov/books/NBK4144/

34 Targovnik, H. M., Esperante, S. A. & Rivolta, C. M. Genetics and phenomics of hypothyroidism and goiter due to thyroglobulin mutations. *Mol. Cell. Endocrinol.* **322,** 44–55 (2010).

35 Narumi, S., Muroya, K., Abe, Y., Yasui, M., Asakura, Y., Adachi, M. *et al.* TSHR mutations as a cause of congenital hypothyroidism in Japan: a population-based genetic epidemiology study. *J. Clin. Endocrinol. Metab.* **94,** 1317–1323 (2009).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)