Compound hemizygous variants in SERPINA7 gene cause thyroxine-binding globulin deficiency

Yanlan Fang[1], Hong Chen[2], Qingqing Chen[2], Chunlin Wang[3], and Li Liang[3]

[1]Zhejiang University School of Medicine First Affiliated Hospital
[2]Zhejiang University
[3]Zhejiang University School of Medicine

July 25, 2020

Abstract

Thyroxine-binding globulin (TBG) is encoded by SERPINA7 which located on Xq22.2. SERPINA7 variants caused TBG deficiency which does not require treatment, but the decreased thyroxine may be misdiagnosed as hypothyroidism. We discovered some variants of TBG caused by alterations that differ from previously reported. In this study, we enrolled 32 subjects from 10 families and sequenced the SERPINA7 genes of TBG-deficient subjects. Then variants were analyzed to assess their effect on TBG expression and secretion. Bioinformatics database, protein structure and dynamics simulation were used to evaluate the deleterious effects. Finally, we identified 2 novel and 4 known variants, and found 26 of 30 subjects carried the p.L303F. The DynaMut predictions indicated the variants (p.E91K, p.I92T, p.R294C and p.L303F) exhibited decreased stability. Analyses revealed the p.L303F change the protein stability and flexibility, and it had an impact on the function of TBG, but when coexisted with other variants it might change the conformational structure of the protein and aggravate the damage to the protein. We speculated that the existence of a higher number of variants resulted in lower TBG secretion. Keywords: Thyroxine-binding globulin deficiency; SERPINA7 gene; compound variants; single-nucleotide variants

Keywords:

Thyroxine-binding globulin deficiency; *SERPINA7* gene; compound variants; single-nucleotide variants

Introduction

In humans, thyroxine-binding globulin (TBG) is one of the major serum transport proteins of thyroid hormones (THs), and the other major transport proteins include two transthyretins and human serum albumin. Although high amounts of albumin are found in serum, TBG has a greater affinity for thyroxine (T4). Abnormalities in the functionality and amount of TBG cause alterations in the total T4 amount in serum but not the level of free T4. TBG deficiency, including complete TBG deficiency (TBG-CD) and partial TBG deficiency (TBG-PD), manifests as a normal metabolic state but is diagnosed by an abnormal result from a thyroid function test called euthyroid hypothyroidism. Physicians may not recognize these deficiencies and provide an inappropriate prescription, which would lead to side effects.

As far as we know, all inherited TBG-PD variants that have been reported to date are caused by missense variants [1], and this type is the most common form of inherited TBG deficiency, with a frequency of 1 in 4,000 newborns. Since the first description of the human *SERPINA7* gene in 1986, a series of variants and single nucleotide polymorphisms (SNPs) have been identified [2]. The pre-TBG protein is 415 amino acids in length and contains a 20-amino acid signal peptide that directs it into the endoplasmic reticulum. Among these variants, p.L303F, corresponding to p.L283F in the mature protein is the most common SNP. Although we have previously referred to the p.L303F (rs1804495) as p.L283F or TBG-poly, we will refer to

this variant as p.L303F, since this is its standard name in the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/). This SNP is often associated with other variants in the *SERPINA7* gene, whose role is not yet defined[3-7]. Although p.L303F was previously considered to be a "benign" variant, two recent reports have shown that the serum TBG levels were lower in males with p.L303F variant compared to normal controls [8; 9]. The mean TBG concentration in heterozygous females with TBG-PD is usually higher than half the normal level, and this level is intermediate between those found in affected and unaffected males. Affected males present with the classic TBG deficiency, which includes low total T4 and T3 levels, low but detectable TBG expression, and a normal free TH level in serum. The clinical phenotype is often mistaken for hypothyroidism and leads to unnecessary treatment or even drug-induced hyperthyroidism.

In this study, the clinical features and laboratory results from 10 families with TBG deficiency were described. Gene sequencing revealed six variants in the *SERPINA7* gene and five compound hemizygous variants involving these six variants. The effects of TBG-Poly on thyroid function were assessed in individuals by independent T tests. To further confirm the pathogenicity of these *SERPINA7* gene variants, we assessed the expression level and conformational dynamics of these single and compound variants through in vitro assays.

Patients and methods

Subjects

The study was reviewed and approved by the Ethics Committee of The First Affiliated Hospital, Zhejiang University, China (No. 2017-863), and was conducted in agreement with the Declaration of Helsinki Principles. Written informed consent was obtained from the patients' parents.

Clinical evaluations

We evaluated the probands and their parents for clinical manifestations and laboratory thyroid function tests including TT3, TT4, FT3, FT4, TSH and TBG.

Thyroid Function Tests

The concentrations of TBG were determined using a chemiluminescent immunometric assay (Siemens Healthcare Diagnostics Products Limited) with an automated analyzer. Serum thyroid stimulating hormone (TSH) was assayed by chemiluminescence method. Thyroid hormones are measured by automated immunoassay platform. All the samples from each individual were measured in the same assay run.

Genetic diagnosis

Genomic DNA was extracted from the peripheral blood samples of the probands and their parents. The sequence variants detected in the *SERPINA7*gene were described according to the NCBI entry NM_000354.6 (NC_000023.11). The description of variants was according to the Human Genome Variation Society (HGVS) sequence variant nomenclature[10]. The correct nomenclature for variants were checked applying Mutalyzer(https://mutalyzer.nl/).

Analysis of the conservation and pathogenicity of the variants

A variety of vertebrate TBG protein sequences were downloaded from UniProt (https://www.uniprot.org/), and Clustalx 2.1 software was used for sequence alignment. The sequence alignment results were displayed using the online ConSurf Server (http://consurf.tau.ac.il/).

The pathogenicity and conservation of the variants were analyzed using the webserver PREDICT-SNP (https://loschmidt.chemi.muni.cz/predictsnp1/). This tool provides the consensus of several different predictors (SIFT, SNAP, PolyPhen-2, MAPP, and PredictSNP) accompanied by a confidence score for the predictions[11].

The changes in the dynamics and stability of five TBG variants were predicted using the DynaMut server (http://biosig.unimelb.edu.au/dynamut/prediction). Specifically, using this server, a normal mode analysis

(NMA) of wild-type TBG using two different tools, Bio3D[12] and ENCoM[13], and the effects of every single variant on the TBG dynamics and stability induced by changes in vibrational entropy were then assessed. The output was presented in terms of the predicted changes in stability (in kcal/mol) and entropy energy (in kcal/mol/K) between the wild-type and variant structures.

Molecular modeling

The DynaMut server (http://biosig.unimelb.edu.au/dynamut/prediction) was used to predict interactions among amino acid residues, and the results were displayed using PyMOL software. The three-dimensional structure of human TBG (PDB accession code 2RIW and 4X30) was used to study the effect of the identified variants on the protein conformation[14]. The structural representation was generated using the molecular visualization system in the open-source foundation (PyMOL 2.4, https://pymol.org/2/). An image of the variant p.Ile310Phefs* was generated using Swiss-Model (https://swissmodel.expasy.org/interactive).

Construction of SERPINA7 variants

The wild-type (WT) full-length human *SERPINA7* cDNA (GenBank: M14091.1) with EcoR1 and BamH1 restriction enzyme sites was chemically synthesized (Tsingke, China) and cloned into the plasmid vector pcDNA3.1 (+). As described previously, the variants and a 15-20-bp terminal homologous sequence were introduced into the vector. The circular plasmid was complementarily paired (Table S1), and *Escherichia coli* DH5α was used for positive plasmid amplification [15; 16]. Alleles carrying two or more mutations were constructed in the same way.

Transient transfection of SERPINA7 cDNA

We included 15 groups in the analysis: negative control (only pcDNA3.1(+)), positive control (pcDNA3.1(+)-Wild-type), and 13 experimental groups (pcDNA3.1(+)-mut, where mut refers to c.271G>A, c.275T>C, c.631G>A, c.880C>T, c.909G>T, c.927del, c.[271G>A;909G>T], c.[275T>C;909G>T], c.[631G>A;909G>T], c.[880C>T;909G>T], c.[909G>T;927del], c.[271G>A;631G>A] and c.[271G>A; 631G>A;909G>T]). For transient transfection, HEK-293T cells were cultured in 24-well plates until they reached 60%~70% confluence, and the cells were then transfected with 500 ng of plasmid using the Lipofectamine 3000 transfection reagent (Thermo Fisher Scientific).

Quantitative real-time PCR

Forty-eight hours after transfection, total RNA was isolated according to a previously described method[17]. cDNA was obtained via reverse transcription using a reverse transcription kit (TaKaRa, Japan). The sequences of the forward and reverse primers used for PCR amplification were 5'ATCTGGTTCTCTTG-GTACTTG3' and 5'TGGATGACATCTTGTAGAGAGTGGC3', respectively.

Detection of TBG secretion from cells by enzyme-linked immunosorbent assay (ELISA)

Forty-eight hours after transfection, the TBG level in the cell supernatant was determined using an ELISA kit (*Boster* , Wuhan, China), which use the sandwich format and biotin-streptavidin chemistry. The captured antibody was diluted to a final concentration of 1-10 μg/mL in bicarbonate/carbonate antigen-coating buffer, and 100 μL of the diluted antibody was pipetted into each well of a microtiter plate. The plate was covered with adhesive plastic and incubated at 37°C for 30 min. The coating solution was removed, and the plate was washed three times with PBS buffer. Subsequently, 200μL of blocking buffer was pipetted into each well to block residual protein-binding sites. The plate was covered with adhesive plastic and incubated for 1.5 h at 37°C. One hundred microliters of the diluted antibody was then pipetted into the wells with the control, standard, and diluted samples, and 100 μL of the diluted ABC solution was then pipetted into these wells. The absorbance of each well at 450 nm was read using an ELISA reader within 15 min after addition of the stop solution.

Western blot

Forty-eight hours after transfection, the proteins of the cells were extracted using a Whole Cell Extraction Kit (Abcam). The samples were analyzed by 12% SDS-PAGE and immunoblotted as described in the section on immunoblot analyses. Western blotting was performed using a polyclonal anti-TBG antibody (Abcam), and the immunoreactive bands on the membrane were visualized using enhanced chemiluminescence reagents (Super Signal, Pierce) and a CLINX ChemiScope and ECL system (CWBIO)[17].

Visualization of the allele frequency

Data on *SERPINA7* variant allele frequencies were downloaded from the Genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org/). The latest version of gnomAD was downloaded (https://gnomad.broadinstitute.org/downloads), and data on variants of the TBG protein and their corresponding allele frequencies were obtained. In addition, data on the Chinese population were obtained from the Chinese Millionome Database (CMDB) (https://db.cngb.org).

Statistical analysis

GraphPad Prism 7 and Fiji ImageJ for Mac were used for the statistical analyses.

The independent-samples t-test was used for comparisons between two groups. All in vitro data were analyzed using at least three independent replications. The data are presented as the means $\pm$ SDs (n [?] 3).

Results

Clinical evaluations

Our first diagnosis of TBG deficiency was an 11 year old boy and his identical twin brother (family E, II-1 and II-2) who were transferred from a clinic to our hospital due to thyroid dysfunction. Eight months ago, they were admitted to another hospital because they were dissatisfied with their height. Since the thyroid function examination of the proband showed low TT4 and normal TT3 and TSH, it was diagnosed as hypothyroidism at the beginning. Patients were given 75ug/day of sodium thyroxine, which increased to 125ug/day after 8 months until drug-induced hyperthyroidism appeared. Laboratory tests revealed a persistent decrease in TSH. Further tests revealed that they had subnormal serum levels of TBG. After discontinuation of thyroxine therapy, hyperthyroidism disappeared and TSH returned to normal in both patients. These clinical features support the diagnosis of TBG-PD and iatrogenic hyperthyroidism (Table S2).

In the following 2016-2018, we collected 9 more TBG deficient families in clinical practice. At first, these patients came to our hospital because of short stature. On examination their TT4 and/or TT3 levels were found to be abnormal, but FT3, FT4 and TSH were normal. Due to previous misdiagnosis, we further assessed the serum TBG levels in these patients. All of them had lower than normal serum TBG levels. Among them, the serum TBG levels of the four probands (II-1 of family C, II-1 of family D, II-1 of family H and II-1 of family J,) were lower than measurable levels (<3.5μg/mL) and were determined to be TBG-CD. The remaining six probands had lower TBG levels than normal and were identified as TBG-PD (3.5-13.0 μg/mL). And five parents (I-1 of family A,I-2 of family C,; I-1 of family F,; I-1 of family G, and I-2 of family H,) were also identified as TBG-PD. All the probands and their parents had normal FT3, FT4, and TSH levels. The laboratory results of these probands and their parents are summarized in Table 1.

Genetic diagnosis

Sanger sequencing revealed six variants in the *SERPINA7* gene: c.271G>A/p. E91K, c.275T>C/p.I92T, c.631G>A/p.A211T, c.880C>T/p.R294C, c.909G>T/p.L303F, and c.927del/p.I310Ffs*. All the patients, including those with TBG-PD and TBG-CD, carried compound hemizygous variants (c.[275T>C;909G>T], c.[631G>A;909G>T], c.[880C>T;909G>T], c.[927del;909G>T;], and c.[271G>A;631G>A;909G>T]), with the exception of four, who carried only one SNP (c.909G>T). Notably, two novel variants, c.271G>A/p.E91K and c.927del/p.I310Ffs*, were identified (Figure 1). Among the 10 families, 86.7% (26/30) individuals carried p.L303F variant. All patients (18/18) with abnormal thyroid function results had the p.L303F variant.

4

Analysis of the conservation and pathogenicity of the variants

Immature TBG has a putative signal peptide (20-amino acid) at the N-terminus. Figure 2A shows the location of these six variants in the immature TBG. The conservation analysis showed that the wild-type residues at I92 and R294, but not those at G91, A211 and L303, were conserved in vertebrates (Figure 2B).

The novel missense variant p.E91K was predicted to be deleterious by four of five predictors with high confidence scores of 0.61, 0.50, 0.53 and 0.80. And the variants p.I92T, p.R294C, and p.L303F were found to be pathogenic by all five prediction software programs. In contrast, the predictions for variant p.A211T indicated that it is likely benign (Table 2).

To study these changes in the structural stability of TBG induced by the five variants, we utilized the DynaMut webserver, which predicts the variant-induced changes in the dynamics and stability of a protein. The DynaMut predictions showed that the p.E91K, p.I92T and p.L303F variants destabilized the protein and that the p.I92T variant resulted in the lowest stability. In contrast, the p.A211T and p.R294C variants appeared to be relatively stable (Table 3).

Three-dimensional structure modeling

All structural images and amino acid interactions were drawn using PyMOL. The novel frameshift variant c.927del causes early translation termination and thus produces a truncated TBG protein (Figure 3A'). The variant p.E91K, which involves mutation of a glutamic acid residue to lysine, formed a hydrogen bond with S86, T95, Q326, and H327. Due to the size difference between glutamic acid and lysine residues, the residue in the variant is not located in the correct position to produce the same hydrogen bonds as the wild-type residue (Figure 3B, B'). The differences between residues I92 and T92 might interfere with the core structure of the domain: the I92 residue is buried in the core of a domain, whereas the variant residue is markedly smaller, which might create a void in the core of the protein. Moreover, as shown in Figure 3C, C', the p.I92T variant led to a loss of hydrophobic interactions in the protein core, which appears to imply that these variants exhibit more severe pathogenicity. The analysis of the variant p.A211T revealed that the residue in the variant interferes with its interactions with other molecules or parts of the protein because the threonine residue is larger than the alanine residue. In contrast, due to the different hydrophobicity of alanine and threonine, the p.A211T mutation might cause a loss of some of the hydrophobic interactions with other molecules on the protein surface (Figure 3D, D'). Differences in charge and hydrophobicity between arginine and cysteine explain why p.R294C results in a loss of interaction with other molecules. Moreover, residues in the vicinity of the mutated residue are annotated as part of a binding site of TBG. The variant can affect the local structure and thus the TBG binding site (Figure 3E, E'). Further analysis showed that p.L303, which is buried in the core of the domain, is smaller than F303, and the F303 residue might thus interfere with the core structure of the domain (Figure 3F, F').

Decreased levels of TBG were associated with p.L303F variant

To determine whether the p.L303F variant affects serum TBG levels, we compared the serum TBG levels of subjects with p.L3033F hemizygous or homozygous variants to wild-type subjects. As shown in Figure 4A, the TBG levels of p.L303F group(12.1 ± 0.7983, n=8)was significantly lower than that of the wild-type group (21.5 +- 1.831, n=4)(p=0.0002). To test the reliability of our results, we first compared our data with that reported in the literature [6; 8; 9]

PubMed articles on TBG deficiency as of June 2020 are collected. The p.L303F variant is a common SNP locus, which is carried at a higher rate in the population. Moreover, we considered that this variant may affect the level of TBG in individual serum. In addition, the control group should be wild-type individuals with *SERPINA7* gene sequencing. Males with only the p.l303F hemizygous variant and females with only the p.L303F homozygous variant were screened. Meanwhile, serum TBG levels of wild-type individuals with *SERPINA7* gene sequencing were collected as controls. After these articles are included, relevant data are collected from each paper. Nevertheless, there are few studies and reports on TBG levels of individuals with p.L303F variants. Only three previous studies offer eligible individual data. A total of 15 wild-type

individuals and 15 hemizygous males and 2 homozygous females were collected including our study. Raw data of serum TBG level in individuals is show in Table S3. As shown in Figure 4B, the serum TBG level of p.L303F variant (12.79 +- 0.7301, n=17) was significantly lower than wild-type (21.69 +- 1.038, n=15), P<0.0001. There were no significant differences between groups in the level of TT4 (P = 0.664) most likely due to the low sample size of this study (Figure 4C). We pooled *SERPINA7* p.L303F variant and/or wild-type samples along with samples harbouring this variant and/or wild-type from previous research[6; 8; 9]. The TT4 levels of p.L303F group were significantly lower than wild-type group (p=0.0006; Table S3, Figure 4D)

Analysis of secreted TBG levels

The in vitro expression of TBG was analyzed by ELISA and Western blot analysis. Quantitative PCR results revealed no *SERPINA7* gene differences in mRNA levels among groups (Figure S1). The exocrine levels of different variants of TBG were measured by ELISA, and the results showed that the single variants E91K and A211T induced TBG secretion, similar to the results obtained with the wild-type. The levels of TBG with the p.I92T, p.R294C, and p.L303F variants were approximately 54.77%, 67.25%, and 63.13% of the wild-type levels, respectively. The analysis of the compound hemizygous variants showed that the TBG secretion levels obtained with the p.E91K/L303F, p.A211T/L303F, and p.R294C/L303F variants were 50.04%, 63.90%, and 67.26% of the wild-type levels, respectively. In addition, the p.I92T/L303F and p.E91K/R294C/L303F variants resulted in secretion levels that were 33.28% and 32.25% of the wild-type levels, respectively. As expected, TBG secretion could not be detected with the deletion variant (Figure 5A, C).

We then further examined the effects of the variants on the intracellular TBG protein level through a Western blot analysis (Figure 5B, D). The results from the Western blot analysis were consistent with the ELISA results, which indicated that the identified variants affect the stability, but not the extracellular secretion, of TBG.

Expression analysis of cellular TBG

The quantitative PCR results revealed differences in the mRNA levels of the *SERPINA7* among the groups. A Western blot analysis using anti-TBG as the antibody showed that the expression level of TBG protein with the p.E91K variant was similar to the wild-type level. Decreased protein expression levels were obtained for the other variants, particularly the p.I92T/p.L303F and p.E91K/p.A211T/p.L303F variants. The protein band of the p.I310Ffs* variant was not detected for imaging (Figure 5B, D).

Allele frequency analysis

The frequency and distribution of *SERPINA7* variants in the Chinese Millionome database and gnomAD are summarized in Figure 6. The p.L303F variant is the most common in the general population. The p.A211T variant is found at a higher frequency only in the East Asian (non-Chinese) (2.3%) and South Asian (1.4%) populations (Figure 6A). The p.R294C and p.I92T variants, which are found only in the non-Finnish European population, are extremely infrequent and are rare variants. There are some differences in distribution of p.L303F variant in different regions, being lowest in European. The highest frequency is found in the Amish (29.8%), followed by the South Asian (28.6%) and East Asian (non-Chinese) populations (23.9%), whereas the Finnish European (7.8%) and Latino (10%) populations carry relatively low frequencies. This allele has a population frequency of 13.88%, which is the highest frequency of *SERPINA7* gene. There are some differences in distribution of p.L303F variant in different regions (Figure 6B). The highest frequency of p.L303F variant was observed in Amish (29.79%), followed by South Asians (28.62%) and East Asians (non-Chinese) (23.87%). According to the Chinese Millionome database, the mutation frequency of p.L303F variant in Chinese was 19.58%. In addition, only four variants, p.A21V, p.A211T, p.R294H, and p.L303F, were identified among the Chinese population, as detailed in the Millionome database. Similar to the results in other populations, p.L303F is the most common variant in the Chinese population, with a frequency of approximately 19.6%. In contrast, the frequency of the p.A211T allele was found to equal 1.34%, which is similar to that reported in East Asia (1.4%) and less than that found in South Asia (2.26%). No such gene has been reported among the Amish, Ashkenazi Jew, and European populations. The p.I92T and p.L294C alleles have not been reported in the Chinese population (Figure 6C).

6

Discussion

Herein, we present 32 subjects from ten families. Most of these proband came to the hospital for further examination due to their short stature and found the TT4 reduction by accident. Our first case (II-1 and II-2 of family E)was misdiagnosed as hypothyroidism and treated with levothyroxine until iatrogenic hyperthyroidism occurred. TBG is rarely tested in pediatric clinic because it is rarely used as an indicator of disease diagnosis and treatment. Thyroid hormones play a key role in the growth and development of children. Therefore, some children of short stature with decreased serum TT4 are easy to be misdiagnosed as thyroid insufficiency and receive unnecessary treatment.

Among the 10 families, 86.7% (26/30) individuals carried p.L303F variant. Interestingly, all affected individuals (18/18) carried p.L303F variant, also known as TBG-poly. The TBG-poly had been considered as a benign variant in early studies, but our conclusion seems inconsistent with the previous literature[18]. The statistical results of the serum TBG levels in the p.L303F hemizygotes were significantly lower than wild-type individuals (p = 0.0002) in our subjects, which is consistent with two recent reports[8; 9]. The TBG levels of ¿13μg/mL has been described as a cutoff between normal and abnormal. From Figure 4A and 4B, it is clearly shown that all data points are distributed near the critical value line and are significantly different from the TBG level of wild-type individuals. Although TT4 levels in p.L303F hemizygotes (median=55.7 nmol/L) were lower than those in wild-type individuals (median=70.0 nmol/L) in our subjects, there was no significant difference between the two groups due to our insufficient sample size. However, when we combined the subjects' samples with the previous study samples, TT4 levels in the p.L303F group were significantly different from those in the wild-type group (P =0.0006).

We found that most of the patients carry X-linked TBG deficiency-causing variants that are not considered classical variants but are rather present in a new form as compound single-nucleotide variants. Of note, we purport that our analysis involved the use of a variety of predictive methods for determination of the pathogenicity of each variant and their association with TBG deficiency, and the results suggest that these five missense variants appear to be related to TBG deficiency. The predictions obtained with the DynaMut server revealed that the single variants p.E91K, p.I92T and p.L303F reduced the stability of the protein and that the variants p.A211T and p.R294C exerted minor if any effects on the stability. Four probands were diagnosed with TBG-CD in our study, and one of the probands has a frameshift variant in amino acid residue 310 and a stop codon in amino acid residue 320, which results in a truncated protein and can be considered a de novo variant. Another two probands from different families have the same variants, p.I92T and p.L303F. Both algorithms predicted that p.I92T is likely a pathogenic variation. And the comparison of TBG expression among p.I92T, p.L303F and p.I92T/L303F revealed that the lowest level was obtained with the p.I92T/L303F variant, which can explain the sociological results. The proband from family H, which carries all three variants, presented with TBG-CD, whereas the proband from family E, F and G carries p.A211T/L303F and has less severe TBG deficiency. These results showed that the compound variants resulted in a significant decrease in TBG secretion. The presence of a higher number of variants resulted in a greater effect on TBG secretion, which might be a superposition effect. As mentioned above, a western blot analysis showed that the observed alterations in intracellular TBG protein expression observed in the variant and wild-type groups were similar to the serum levels of TBG measured by ELISA, which indicates that these variants alter the stability of TBG in cells without altering the exocrine capacity of this protein.

In summary, we identified 6 variants of *SERPINA7* and 5 compound variants that had not previously been reported as a cause of TBG deficiency. Endocrinologists should better identify inherited TBG deficiency to help prevent misdiagnosis and unnecessary treatment of the disease. More importantly with respect to the etiology of TBG deficiency, the findings identified a de novo mechanism: compound variants in *SERPINA7* might lead to conformational changes in the structure and ultimately influence the stability of the protein. Unlike previous studies, our results indicate that the p.L303F variant changes the function of TBG. And when it coexists with other variations may exacerbate the damage to the function of TBG. The combination of a higher number of variants resulted in decreased TBG secretion.

Some limitation of this study must be acknowledged. First, we did not further examine the T4 binding

7

ability of these variants. Second, despite our best efforts, we have only a handful of individual clinical data. For such a high allele frequency, these data are far from sufficient. In the future, more clinical data and functional evidence will be needed for better analysis.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this article.

WEB RESOURCES

http:// www.hgmd.cf.ac.uk/

https://mutalyzer.nl/

https://www.uniprot.org/

http://consurf.tau.ac.il/

https://loschmidt.chemi.muni.cz/predictsnp1/

https://pymol.org/2/

https://swissmodel.expasy.org/interactive

http://gnomad.broadinstitute.org/

https://gnomad.broadinstitute.org/downloads

https://db.cngb.org

References

1. Pappa T, AM, F., and S., R. (2015). Inherited defects of thyroxine-binding proteins. Best Pract Res Clin Endocrinol Metab 29, 735-747.

2. Irwin L. Flink, Thomas J. Bailey, Thomas A. Gustafson, Bruce E. Markham, and Morkin., E. (1986). Complete amino acid sequence of human thyroxine-binding globulin.pdf. Proc Natl Acad Sci U S A 83, 7708-7712.

3. Janssen OE, Astner ST, Grasberger H, Gunn SK, and S., R. (2000). Identification of Thyroxine-Binding Globulin-San Diego in a Family from Houston and Its Characterization by in Vitro Expression Using Xenopus Oocytes. J Clin Endocrinol Metab 85, 368-372.

4. Mori Y, Takeda K, Charbonneau M, and S., R. (1990). Replacement of Leu227 by Pro in thyroxine-binding globulin (TBG) is associated with complete TBG deficiency in three of eight families with this inherited defect. J Clin Endocrinol Metab 70, 804-809.

5. Su CC, Wu YC, Chiu CY, Won JG, and TS., J. (2003). Two novel mutations in the gene encoding thyroxine-binding globulin (TBG) as a cause of complete TBG deficiency in Taiwan. Clin Endocrinol (Oxf) 58, 409-414.

6. Mannavola, D., Vannucchi, G., Fugazzola, L., Cirello, V., Campi, I., Radetti, G., Persani, L., Refetoff, S., and Beck-Peccoz, P. (2006). TBG deficiency: description of two novel mutations associated with complete TBG deficiency and review of the literature. J Mol Med (Berl) 84, 864-871.

7. Bertenshaw R, Takeda K, and S., R. (1991). Sequencing of the variant thyroxine-binding globulin (TBG)-Quebec reveals two nucleotide substitutions. 1991;48(4):741-744. Am J Hum Genet 48, 741-744.

8. Chen, L.D., Lu, H.J., Gan, Y.L., Pang, S.W., Zheng, Q., Ye, D.M., Huang, X.Y., Qi, H.N., Xu, W.B., Wen, X.Z., et al. (2020). Partial thyroxine-binding globulin deficiency in a family with coding region mutations in the TBG gene. J Endocrinol Invest, published online ahead of print, 2020 Apr 2027.

9. Dang PP, Xiao WW, Shan ZY, Yue Xi, Wang RR, Yu XH, Teng WP, and XC., T. (2019). Novel frameshift mutation causes early termination of the thyroxine-binding globulin protein and complete thyroxine-binding globulin deficiency in a Chinese family: A case report. World J Clin Cases 7, 3887-3894.

10. den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., McGowan-Jordan, J., Roux, A.F., Smith, T., Antonarakis, S.E., and Taschner, P.E. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. Hum Mutat 37, 564-569.

11. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben E.D, Zendulka J, Brezovsky J, and J., D. (2014). PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol 10, e1003440.

12. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, and LS., C. (2006). Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics 22, 2695-2696.

13. Frappier V, and RJ., N. (2014). A coarse-grained elastic network atom contact model and its use in the simulation of protein dynamics and the prediction of the effect of mutations. PLoS Comput Biol 10, e1003569.

14. Qi X, Loiseau F, Chan WL, Yan Y, Wei Z, Milroy LG, Myers RM, Ley SV, Read RJ, Carrell RW, et al. (2011). Allosteric modulation of hormone release from thyroxine and corticosteroid-binding globulins. J Biol Chem 286, 16163-16173.

15. Gibson DG, Smith HO, Hutchison CA, Venter JC, and C., M. (2010). Chemical synthesis of the mouse mitochondrial genome. Nat Methods 7, 901-903.

16. Chen H, Yuan K, Zhang B, Jia Z, Chen C, Zhu Y, Sun Y, Zhou H, Huang W, Liang L, et al. (2019). A Novel Compound Heterozygous CYP17A1 Variant Causes 17alpha-Hydroxylase/17, 20-Lyase Deficiency. Front Genet 10, 996.

17. Liu Z, Song Y, Li D, He X, Li S, Wu B, Wang W, Gu S, Zhu X, Wang X, et al. (2014). The novel mitochondrial 16S rRNA 2336T>C mutation is associated with hypertrophic cardiomyopathy. J Med Genet 51, 176-184.

18. Janssen OE, Bertenshaw R, Takeda K, Weiss R, and S., R. (1992). Molecular basis of inherited thyroxine-binding globulin defects. Trends Endocrinol Metab 3, 49-53.

Table 1: Thyroid function tests and TBG concentration in all probands and their parents

Subject	Family members	Variants in TBG	TT3 (nmol/L)	TT4 (nmol/L)	FT
A	Proband (II-1)	p.L303F Hemizygotes	1.32	**53.4**—	6.1
	Mother (I-2)	p.L303F Heterozygotes	1.82	65.9	5.7
	Father (I-1)	p.L303F Hemizygotes	1.13	**52.4**—	5.8
B	Proband (II-1)	p.L303F Hemizygotes	2.11	62.8	7.4
	Mother (I-2)	p.L303F Heterozygotes	1.98	72.9	4.8
	Father (I-1)	Wild-type	2.56	73.9	5.2
C	Proband (II-1)	p.I92T and p.L303F Hemizygotes	1.04	**43.4**—	6.1
	Mother (I-2)	p.I92T and p.L303F Heterozygotes	2.01	55.6	4.3
	Father (I-1)	p.L303F Hemizygotes	1.89	55.8	5.1
D	Proband (II-1)	p.I92T and p.L303F Hemizygotes	1.35	**30.6**—	7.3

Subject	Family members	Variants in TBG	TT3 (nmol/L)	TT4 (nmol/L)	FT
	Mother (I-2)	ND	1.89	69.0	3.4
	Father (I-1)	Wild-type	2.02	67.5	4.5
E	Proband (II-1)	p.A211T and p.L303F Hemizygotes	1.27	**41.5—**	5.6
	Twin brother (II-2)	p.A211T and p.L303F Hemizygotes	1.12	**35.4—**	5.6
	Litter brother (II-3)	p.L303F Hemizygotes	1.20	100.8	4.6
	Mother (I-2)	p.L303F Heterozygotes	1.18	105.2	4.4
	Father (I-1)	p.L303F Hemizygotes	1.38	91.1	4.8
F	Proband (II-1)	p.A211T and p.L303F Hemizygotes	1.32	**46.7—**	5.9
	Mother (I-2)	p.A211T Heterozygotes and p.L303F Homozygotes	1.45	60.3	5.2
	Father (I-1)	L303F Hemizygotes	1.10	55.6	4.1
G	Proband (II-1)	p.A211T Heterozygotes and p.L303F Homozygotes	1.79	**55.2—**	6.2
	Mother (I-2)	p.L303F Heterozygotes	1.33	89.7	5.2
	Father (I-1)	p.A211T and p.L303F Hemizygotes	1.24	57.7	4.3
H	Proband (II-1)	p.E91K, p.A211T and p.L303F Hemizygotes	1.10	**36.9—**	6.1
	Mother (I-2)	p.E91K, p.A211T and p.L303F Heterozygotes	1.03	**52.6—**	3.1
	Father (I-1)	Wild-type	1.50	67.9	5.5
I	Proband (II-1)	p.R294C and p.L303F Hemizygotes	1.24	**49.9—**	5.7
	Mother (I-2)	p.R294C Heterozygotes and p.L303F Homozygotes	1.13	**52.8—**	5.7
	Father (I-1)	p.L303F Hemizygotes	1.19	**53.1—**	4.3
J	Proband (II-1)	p.I310Ffs* and p.L303F Hemizygotes	**0.72—**	**31.7—**	4.8
	Mother (I-2)	ND	1.84	66.7	4.2
	Father (I-1)	Wild-type	1.97	70.8	3.8

TT3: total triiodothyronine (1.02-2.96 nmol/L); TT4: total thyroxine (55.4-161.3 nmol/L); FT3: free tri-iodothyronine (2.77-6.31 pmol/L); FT4: free thyroxine (10.5-24.4 pmol/L); TSH: thyroid stimulating hormone (0.38-4.34 mIU/L); TBG: thyroxine-binding globulin (13.0-39.0 μg/mL)

A-I: Nine families; ND: Not done; The abnormal values are in bold.

Table2. Predictions and confidence scores for the TBG variants obtained with the PREDICT-SNP server

Variants	Reference ID	PredictSNP	PredictSNP	MAPP	MAPP	PolyPhen-2	PolyPhen-2	S
		Prediction	Accuracy	Prediction	Confidence score	Prediction	Confidence score	F
p.E91K	-	Deleterious	61%	Neutral	64%	Deleterious	50%	I
p.I92T	-	Deleterious	79%	Deleterious	81%	Deleterious	55%	I
p.A211T	rs2234036	Neutral	83%	Neutral	63%	Neutral	68%	N
p.R294C	rs182605024	Deleterious	55%	Deleterious	46%	Deleterious	59%	I
p.L303F	rs1804495	Deleterious	72%	Deleterious	57%	Deleterious	43%	I

Table 3: Predicted results obtained with the DynaMut server after the amino acid change(s) in the TBG variants

Variants				vib
	DynaMut	ENCoM	DUET	ENCoM
p.E91K	-0.082 kcal/mol (Destabilizing)	-0.067 kcal/mol (Destabilizing)	-0.327 kcal/mol (Destabilizing)	0.084 kcal.m
p.I92T	-2.727 kcal/mol (Destabilizing)	-0.603 kcal/mol (Destabilizing)	-3.066 kcal/mol (Destabilizing)	0.754 kcal.m
p.A211T	0.618 kcal/mol (Stabilizing)	0.177 kcal/mol (Destabilizing)	-0.637 kcal/mol (Destabilizing)	-0.221 kcal.
p.R294C	0.152 kcal/mol (Stabilizing)	0.032 kcal/mol (Destabilizing)	-0.135 kcal/mol (Destabilizing)	-0.040 kcal.

10

Variants				vib
p.L303F	-0.716 kcal/mol (Destabilizing)	0.584 kcal/mol (Stabilizing)	-1.750 kcal/mol (Destabilizing)	-0.730 kcal.

Table S1. Mutagenic primers

Variants	Sequence (5'-3')
c.271G>A	cctgctgcagcacccaaactAagattgtggaga
c.275T>C	tgctgcagcacccaaactgagaCtgtggagaccttg
c.631G>A	attcactttaaagcccagtggAcaaatccttttgatc
c.880C>T	aaaacactgaagaagtggaactgcttactaca
c.909G>T	gaagggatgggttgacttTtttgttccaaag
c.927del	tgtttgttccaaagttttcatttctgccacatatgacc

Table S2 misdiagnose and error therapy of probands from family E

Patients	Times	Levothyroxine (μg)	TT3 (nmol/L)	TT4 (nmol/L)	FT3 (pmol/L)	FT4 (pmol/L)	TSH (mU/L)	TBG (μg/mL
II-1 (Family E)	Initial diagnosis	75	1.40	**36.0—**	ND	ND	1.45	ND
	Eight months later	125	1.90	64.2	6.00	16.10	**0.00—**	ND
	Dosage reduction	50	1.42	**53.1—**	**6.92—**	17.03	**0.01—**	**11.0—**
	Drug withdrawal	0	1.27	**41.5—**	5.64	17.27	**1.31—**	ND
II-2 (Family E)	Initial diagnosis	50	1.20	**40.1—**	ND	ND	1.01	ND
	Eight months later	100	1.70	65.8	**6.88—**	20.97	**0.00—**	ND
	Dosage reduction	50	**0.90—**	**52.4—**	5.60	22.90	**0.02—**	**10.4—**
	Drug withdrawal	0	1.12	**35.4—**	5.69	15.41	0.96	ND

TT3: total triiodothyronine (1.02-2.96 nmol/L); TT4: total thyroxine (55.4-161.3 nmol/L); FT3: free triiodothyronine (2.77-6.31 pmol/L); FT4: free thyroxine (10.5-24.4 pmol/L); TSH: thyroid stimulating hormone (0.38-4.34 mIU/L); TBG: thyroxine-binding globulin (13.0-39.0 μg/mL); ND: Not done; The abnormal values are in bold.

Table S3 Serum TBG levels in individuals with p.L303F hemizygotes or p.L303F homozygotes and wild-type genotype controls.

Family members	Sex	Variants in TBG	TT4 nmol/L	TBG (μg/mL)	Rese
II-1	Female	p.L303F Homozygotes	86.2	17.30	A

Family members	Sex	Variants in TBG	TT4 nmol/L	TBG (µg/mL)	Rese
II-4	Male	p.L303F Hemizygotes	81.1	17.20	
III-1	Male	p.L303F Hemizygotes	84.8	17.60	
III-2	Male	p.L303F Hemizygotes	73.3	10.70	
III-4	Male	p.L303F Hemizygotes	75.0	15.30	
III-5	Male	p.L303F Hemizygotes	86.3	10.40	
IV-8	Male	p.L303F Hemizygotes	71.1	8.90	
IV-9	Female	p.L303F Homozygotes	90.4	13.60	
II-2	Male	p.L303F Hemizygotes	62.09	9.6	B
I-1 (A)	Male	p.L303F Hemizygotes	52.4	8.8	Present study
II-1 (A)	Male	p.L303F Hemizygotes	53.4	9.8	
II-1 (B)	Male	p.L303F Hemizygotes	62.8	10.1	
I-1 (C)	Male	p.L303F Hemizygotes	55.8	14.6	
I-1 (E)	Male	p.L303F Hemizygotes	91.1	13.8	
II-3 (E)	Male	p.L303F Hemizygotes	100.8	13.0	
I-1 (F)	Male	p.L303F Hemizygotes	55.6	12.2	
I-1 (I)	Male	p.L303F Hemizygotes	53.1	14.5	
II-2	Male	Wild-type	132.7	21.30	A
II-5	Female	Wild-type	107.6	23.50	
II-8	Male	Wild-type	113.6	33.00	
III-9	Male	Wild-type	94.5	20.90	
IV-2	Male	Wild-type	104.8	21.50	
IV-6	Male	Wild-type	114.3	21.80	
IV-10	Male	Wild-type	99.9	16.40	
IV-12	Male	Wild-type	93.8	18.70	
IV-13	Male	Wild-type	116.9	23.00	
IV-14	Female	Wild-type	103.9	21.50	
I-2	Male	Wild-type	90.96	17.8	B
AII/1	Male	Wild-type	-	30.0	C
I-I (B)	Male	Wild-type	73.9	16.5	Present study
I-1 (D)	Male	Wild-type	67.5	25.3	
I-1 (H)	Male	Wild-type	67.9	21.9	
I-1 (J)	Male	Wild-type	70.8	22.3	

Figure 1: Ten pedigrees with TBG deficiency and Sanger sequencing chromatograms. I: A-J indicates ten pedigrees with TBG deficiency. The arrowheads denote the probands. The affected individuals with partial TBG deficiency are shown as pink dots in the middle of the box. The affected individuals with complete TBG deficiency are shown as red dots in the middle of the box. The orange square shows that the clinically affected subject harbors the p.E91K variant; the green square shows that the clinically affected subject harbors the p.I92T variant; the purple square shows that the clinically affected subject harbors the p.A211T variant; the yellow square shows that the clinically affected subject harbors the p.R294C variant; the blue square shows that the clinically affected subject harbors the p.L303F variant; the gray square shows that the clinically affected subject harbors the p.I310Ffs* variant; and the unfilled symbols indicate that that the subjects are clinically unaffected. The asterisks indicate the individuals who are not available for gene sequencing. Line filling indicates homozygosity. # indicates individuals with reduced total thyroxine levels. II: Sanger sequencing chromatograms. The variants are indicated by an arrow.
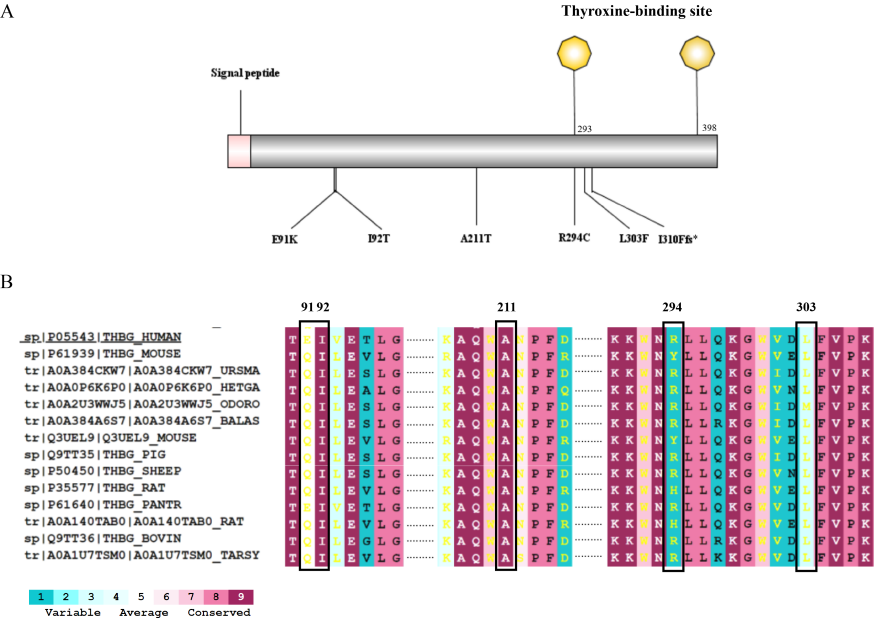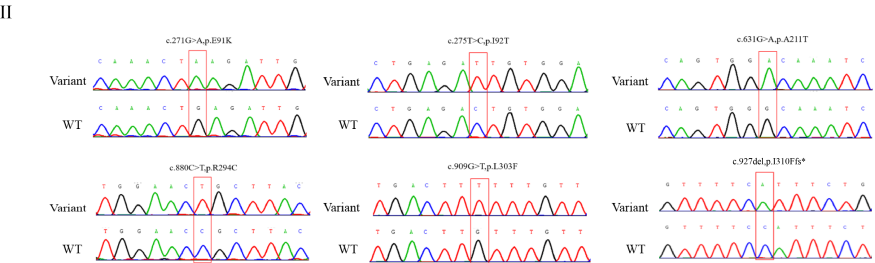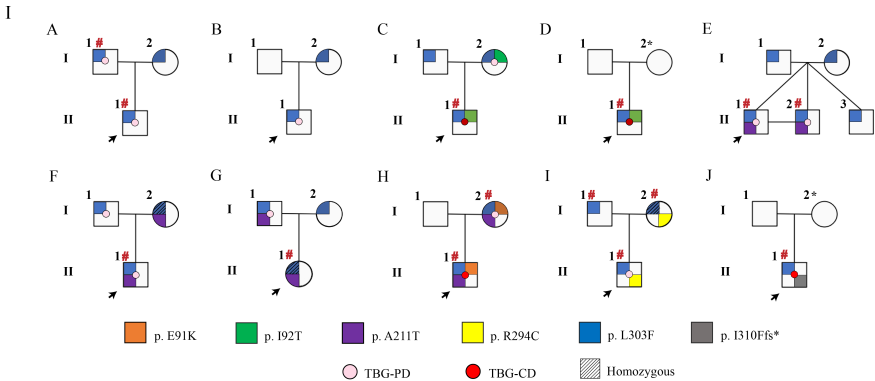
Figure2. Analyze distribution and conservation of TBG variants. (A) Schematic of human preproTBG protein showing the location of 6 different TBG missense variants, and the resulting amino acid changes. (B): Conservation of p.G91, p.A211, p.R294, and p.L303 in TBG proteins from 14 different vertebrates.
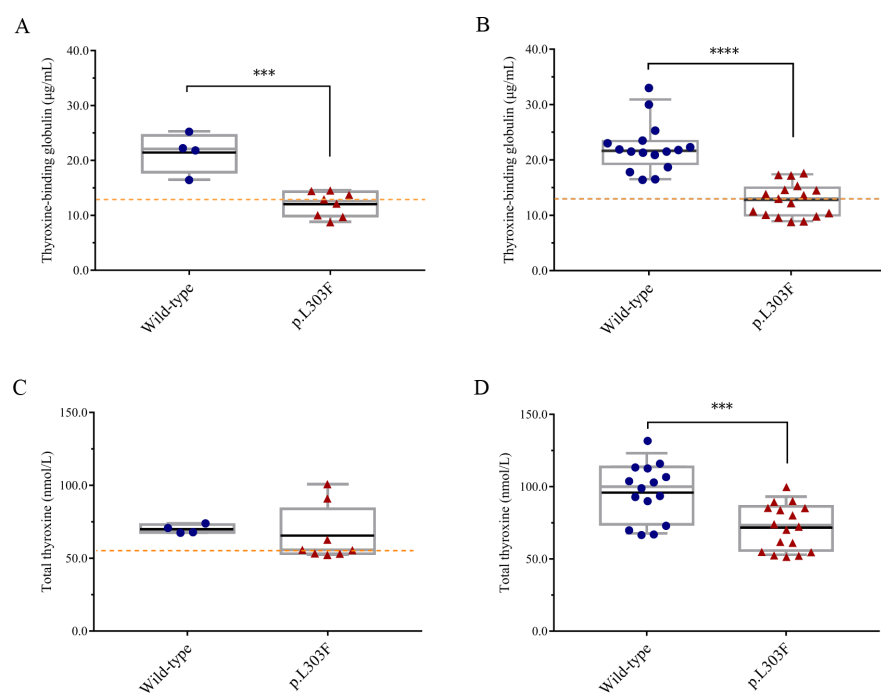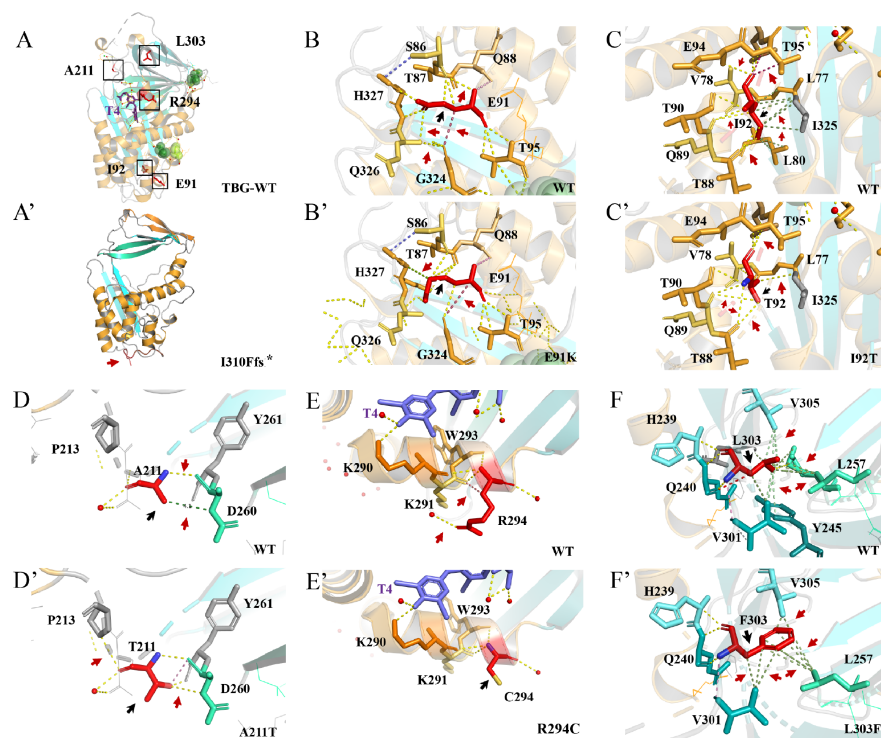
Figure 3: Prediction of interactions between amino acid residues. The wild-type and variant residues are colored in red and are represented as sticks alongside the surrounding residues that are involved in any type of interactions. The red ball indicates H_2O. The red arrows point to interactions between the altered amino acids, and the black arrows point to wild-type or variant amino acid residues. The colors of the contacts are based on the following: yellow, hydrogen bonds; red, ionic interactions; green, hydrophobic contacts; and pink, carbonyl contacts. A: Three-dimensional structure of the wild-type TBG. Black boxes mark the locations of the five variants. A': Three-dimensional structure of the p.I310Ffs* variant. A red arrow points to the truncated protein terminus. B, B': Interaction between the local amino acids of the wild-type and the p.E91K variant. C, C': Interaction between the local amino acids of the wild-type and the p.I92T variant. D, D': Interaction between the local amino acids of the wild-type and the p.A211T variant. E, E': Interaction between the local amino acids of the wild-type and the p.R294C variant. F, F': Interaction between the local amino acids of the wild-type and the p.L303F variant.
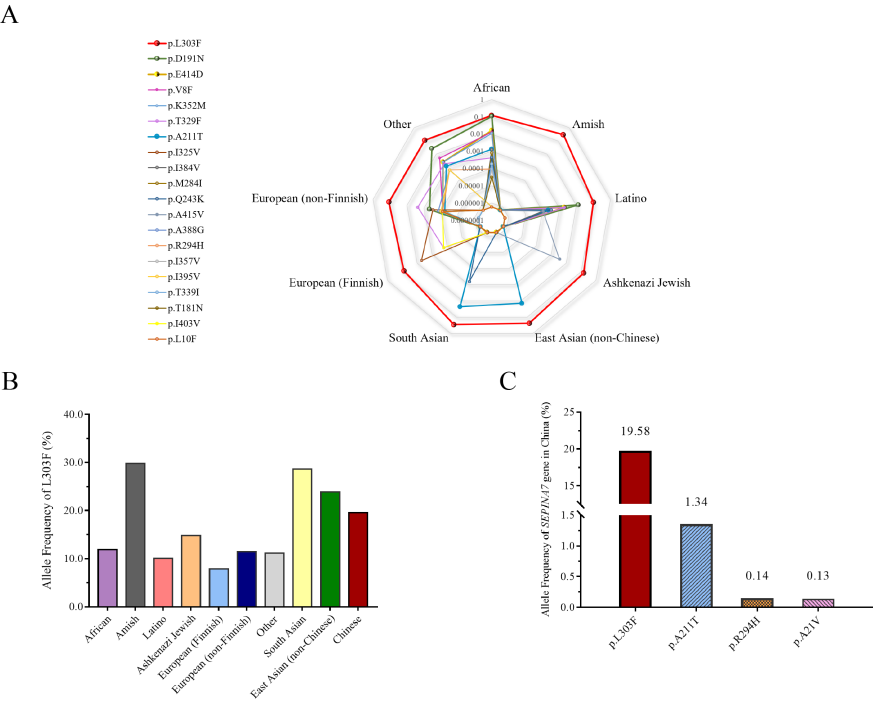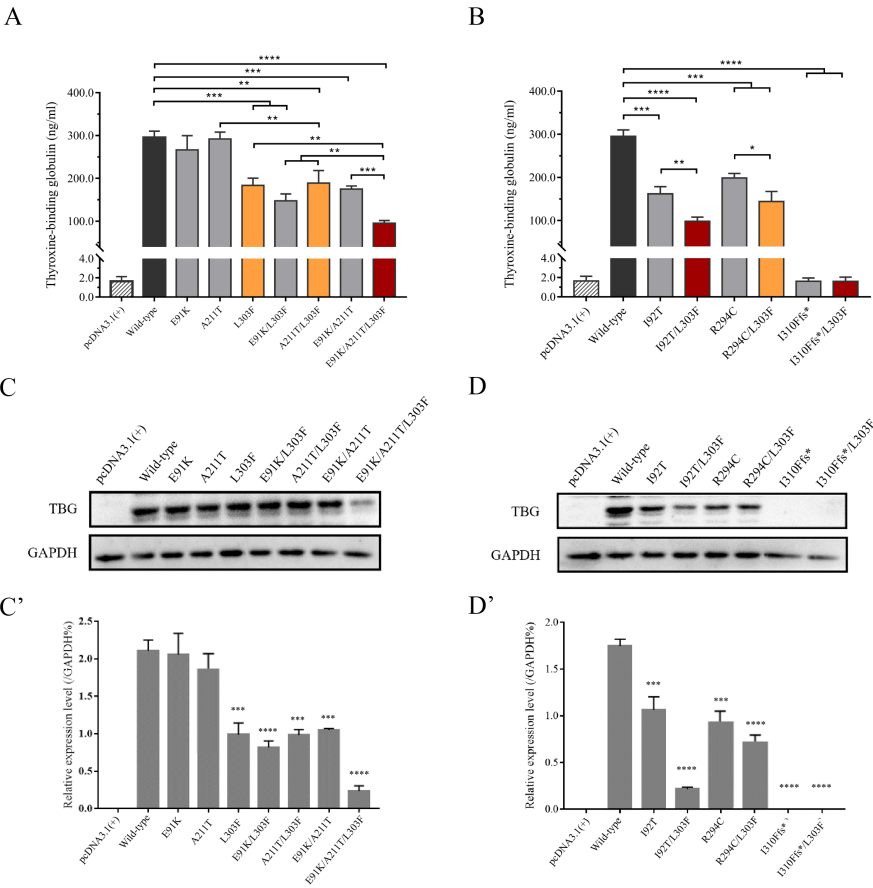
Figure 4. Box plots of pairwise comparisons for TBG and TT4 concentrations. The black horizontal lines represent the mean, and boxes represent the quartiles of data range including the median. The scatter points show all data points. (A): The comparison of TBG levels between two groups in this study. (B): Our research data incorporates previous research data for two sets of analysis. Group Wild-type: wild-type individuals with *SERPINA7* gene sequencing. Group p.L303F: individuals with p.L303F hemizygotes or p.L3033F homozygotes. The orange dotted line indicates threshold of the normal TBG level and TBG deficiency clinically (13.0μg/mL). ***p < 0.001 and ****p < 0.0001.

Figure 5: Extracellular and intracellular TBG levels. A, B: Detection of extracellular TBG secretion by ELISA. The genotypes of the TBG-CD patients are filled with red, the genotypes of the TBG-PD patients are filled with orange, and the wild-type is filled with black color. C-D: The protein expression level was measured by Western blot analysis. The data are presented as the means ± SDs, *p<0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001.

Figure 6: Frequency distribution of the *SERPINA7* gene variants in the African, Amish, Latino, Ashkenazi Jewish, East Asian, Finnish, non-Finnish European, South Asian, China and other populations. (A) Frequency distribution of the top 20 variants in *the SERPINA7* gene among different populations detailed in the Genome Aggregation Database. (B) Frequency distribution of the p.L303F variants among different populations. (C) Allele frequency of the p.A21V, p.A211T, p.R294C and p.L303F variants in China.

13

I

p. E91K   p. I92T   p. A211T   p. R294C   p. L303F   p. I310Ffs*

TBG-PD   TBG-CD   Homozygous

II

c.271G>A,p.E91K   c.275T>C,p.I92T   c.631G>A,p.A211T

c.880C>T,p.R294C   c.909G>T,p.L303F   c.927del,p.I310Ffs*

A

Thyroxine-binding site

Signal peptide

293   398

E91K   I92T   A211T   R294C   L303F   I310Ffs*

B

91 92   211   294   303

sp|P05543|THBG_HUMAN
sp|P61939|THBG_MOUSE
tr|A0A384CKW7|A0A384CKW7_URSMA
tr|A0A0P6K6P0|A0A0P6K6P0_HETGA
tr|A0A2U3WWJ5|A0A2U3WWJ5_ODORO
tr|A0A384A6S7|A0A384A6S7_BALAS
tr|Q3UEL9|Q3UEL9_MOUSE
sp|Q9TT35|THBG_PIG
sp|P50450|THBG_SHEEP
sp|P35577|THBG_RAT
sp|P61640|THBG_PANTR
tr|A0A140TAB0|A0A140TAB0_RAT
sp|Q9TT36|THBG_BOVIN
tr|A0A1U7TSM0|A0A1U7TSM0_TARSY

1 2 3 4 5 6 7 8 9
Variable   Average   Conserved

14

A

B

C

D

C'

D'

A

- p.L303F
- p.D191N
- p.E414D
- p.V8F
- p.K352M
- p.T329F
- p.A211T
- p.I325V
- p.I384V
- p.M284I
- p.Q243K
- p.A415V
- p.A388G
- p.R294H
- p.I357V
- p.I395V
- p.T339I
- p.T181N
- p.I403V
- p.L10F

African
Amish
Latino
Ashkenazi Jewish
East Asian (non-Chinese)
South Asian
European (Finnish)
European (non-Finnish)
Other

B

C

16

A

B