The contents of this journal article are original research that have not been published. This article submission has not been submitted elsewhere for consideration.

Corresponding author:

Name: Roy E. Weiss, MD, PhD

Address: University of Miami Miller School of Medicine, 1120 NW 14th St., Room 310F, Miami, FL 33136

Fax: 305-243-2193

Conflicting interests: The authors declare that they have no conflicts of interest.

"Increased prevalence of *TG* and *TPO* mutations in Sudanese children with congenital hypothyroidism"

Ryan J. Bruellman[1]*, Yui Watanabe[1]*, Reham S. Ebrhim[2], Matthew K. Creech[1], Mohamed A. Abdullah[3], Alexandra M. Dumitrescu[4,6], Samuel Refetoff[4,5,7], Roy E. Weiss[1]

[1]Department of Medicine, University of Miami Miller School of Medicine, Miami, Florida.

[2]Department of Paediatrics and Child Health, Faculty of Medicine, University of Almughtaribeen, Khartoum, Sudan

[3]Department of Paediatrics and Child Health, Faculty of Medicine, University of Khartoum, Khartoum, Sudan.

Department of, Medicine[4], and Pediatrics[5], Committee on Molecular Medicine and Nutrition[6] and Committee on Genetics[7] The University of Chicago, Chicago, Illinois.

Short Title: "*TG* and *TPO* in Sudanese Congenital Hypothyroidism"

Keywords: TPO, TG, Congenital Hypothyroidism, Goiter

Corresponding author and reprint requests:

Name: Roy E. Weiss, MD, PhD

Address: University of Miami Miller School of Medicine, 1120 NW 14th St., Room 310F, Miami, FL 33136

Fax: 305-243-2193

Funding: This research was supported by funds from the Esformes Thyroid Research Fund and National Institutes of Health (NIH) grant MD 010722 to R.W., DK 15070 to S.R., and DK 110322 to A.M.D.

Conflicting interests: The authors declare that they have no conflicts of interests.

*Both RJB and YW contributed equally to this paper

Abstract

Context: Congenital hypothyroidism (CH) is due to dyshormonogenesis in 10-15% of subjects worldwide, but accounts for 60% of CH cases in the Sudan.

Objective: To investigate molecular basis of CH in Sudanese families.

Design: Clinical phenotype reporting and serum thyroid hormone measurements. DNA extraction for whole exome sequencing and Sanger sequencing.

Setting: University research center.

Patients: 26 Sudanese families with CH.

Intervention: Clinical evaluation, thyroid function tests, genetic sequencing and analysis. Our samples and information regarding samples from the literature were used to compare *TG* and *TPO* mutation rates in Sudanese against all populations.

Results: Mutations were found in *DUOX1, DUOX2, IYD, SLC26A4, SLC26A7, SLC5A5, TG,* and *TPO* genes. The molecular basis of the CH in 7 families remains unknown. *TG* mutations were significantly higher on average in the Sudanese compared to the average number of *TG* mutations in other populations ($P < 0.05$).

Conclusions: All described mutations occur in domains important for protein structure and function, predicting the CH phenotype. Genotype prediction based on phenotype include low or undetectable thyroglobulin levels for *TG* gene mutations and markedly higher thyroglobulin levels for *TPO* mutations. The reasons for higher incidence of *TG* gene mutations include gene length and possible positive genetic selection due to endemic iodine deficiency.

Précis

Of 14 pathogenic mutations in Sudanese, *TG* and *TPO* genes are the most common. Higher prevalence of *TG* mutation appears to be related to gene length and possible positive selection.

Introduction

Congenital hypothyroidism (CH) is due to abnormal development or architecture of the thyroid gland (dysgenesis) or impaired thyroid hormone (TH) synthesis (dyshormonogenesis). Between 15-20% of worldwide cases with CH are caused by dyshormonogenesis (1). In the Sudanese the rate of dyshormonogenesis is 60% or three times that in the non-Sudanese populations (2). Several factors account for this disproportionate prevalence, including high rates of consanguinity and low levels of iodine intake. There are over 60 known genes that influence thyroid development and TH function (Table 1). *Dual-oxidase 1 (DUOX1)* and *Dual-oxidase 2 (DUOX2)* have homology in their structure with 7 transmembrane domains and both are responsible for H_2O_2 generation, crucial for TH synthesis (3). *Iodotyrosine Deiodinase (IYD)* acts to recycle iodide for future use in TH synthesis with its nitroreductase domain playing a critical role in the recycling (4). Several solute carrier family member genes (*SLC*) including *SLC26A4*, *SLC26A7*, and *SLC5A5* facilitate transport necessary for proper thyroid hormone synthesis. *SLC26A4* and *SLC26A7* are homologous each with over 10 transmembrane domains and both are expressed highly in the thyroid, however, *SLC26A4* is also expressed in the inner ear and thus deleterious mutations present as deafness in the Pendred Syndrome (5,6). *SLC5A5* is known to function as a sodium-iodide symporter, facilitating the uptake of iodide into thyrocytes needed for TH synthesis (4). Two of the most frequent gene defects causing dyshormonogenesis worldwide involve *thyroglobulin* (*TG*) and *thyroid peroxidase* (*TPO*). *TG* encodes thyroglobulin (TG), a large protein which serves in the synthesis of TH, its storage and that of iodine as well (7). The structure of TG consists of three regions and a cholinesterase-like (ChEL) domain. All three regions assist in stabilizing the ChEL domain, which serves as a transporter for TG into areas of TH formation (8). Previously reported mutations of the *TG* gene

showed a variety of phenotypes due to time of diagnosis, location of mutation within the molecule, and iodine intake (9).

TPO oxidizes iodide so that it can be covalently bound to tyrosine residues within TG for TH production (10). TPO consists mainly an alpha-helical conformation with little beta-sheets. Its complex structure has three domains: myeloperoxidase (MPO)-like domain, complement control protein (CCP)-like domain, and epidermal growth factor (EGF)-like domain (10). These domains all play a key role in TPO for TG iodination, using hydrogen peroxide generated mainly by DUOX2, and coupling of iodinated tyrosines (11).

We present 26 Sudanese families with CH in 19 of which deleterious mutations in the functional domains of *TG* and *TPO* as well as other genes related to dyshormonogenic CH. Fifteen of the families are being reported for the first time, with most of the mutations being either novel or have been previously reported without a documented phenotype related to CH. Each of these mutations alter the structure and function of the molecules based on in silico modeling and its location along important functional domains of the protein. The data suggest that gene size, consanguinity and possible positive genetic selection could account for the large proportion of dyshormonogenesis in the Sudanese population.

Materials and Methods

All patients were referred to a pediatric endocrinologist at the University of Khartoum, Sudan presenting with stigmata of CH. Consent from patients or their guardians and their family members was obtained prior to blood sampling. Studies were approved by the University of Miami Institutional Review Board. Thyroid tests done at the time of diagnosis (TSH and FT_4) were done in Sudan and subsequent serum thyroid function tests (TFTs) were completed in Miami, Florida on the Immulite® 1000 (Siemens, Munich, Germany) platform. TFTs included

measuring levels of thyroid stimulating hormone (TSH), total thyroxine (TT_4), total triiodothyronine (TT_3), free thyroxine (FT_4), thyroxine-binding globulin (TBG), thyroglobulin (TG), thyroid peroxidase antibodies (TPOAb), and thyroglobulin antibodies (TGAb). Isolation of genomic DNA from whole blood using the Qiagen QIAamp® DNA Blood Mini Kit (Hilden, Germany) were carried out at the University of Miami. Blood samples were obtained from members of the nuclear family of each individual with CH. This included both parents and all siblings, if available. For each of the families, gDNA from the proband or an affected sibling along with one parent was submitted to whole exome sequencing (WES) (Novogene, Agilent SureSelect Human All Exon V6 Kit). A compilation of thyroid genes linked to thyroid disorders shown on Table 1 was evaluated and possible mutations linked to the phenotype were identified based on predicted functional scores, allele frequency, and zygosity. These mutations were confirmed by Sanger sequencing (Genewiz, Abi 3730xl DNA Analyzer) to verify the WES results and establish the genotype of all sampled family members along with the mode of inheritance. All identified variants were further evaluated by in-silico prediction scores for how detrimental the identified variation, using Sorting Intolerant From Tolerant (SIFT) (12), PolyPhen2_HDIV (13), MutationTaster (14), Combined Annotation Dependent Depletion Score (CADD) (15) and The Human Splicing Factor (16).

The variant prevalence comparison of *TPO* and *TG* genes was done using frequencies of all individuals' whole exome and whole genome sequencing in the gnomAD database against whole exome sequencing data of Sudanese samples collected in our lab along with single-nucleotide polymorphism (SNP) data of Sudanese individuals collected by Hollfelder N, et al. (17). 654 randomly selected SNPs were compared across the *TG* gene, including 152 in exons. 234 randomly selected SNPs were also compared within the *TPO* gene. A paired t-test was

performed to compare the prevalence of mutations in individuals from all ethnic groups versus Sudanese samples. For each of the 26 Sudanese families with CH, each unique causative mutation identified was tallied and the total number of unique causative mutations for each gene was plotted against amino acid length of the gene. Differences in gender were not considered in this analysis.

Results

Gene mutations were found in 19 kindreds (Table 2) with CH (7,18-20). In 7 families we were unable to find the genetic cause of the CH phenotype after valuating all genes on WES with variants which were rare and predicted as deleterious by in-silico prediction tools of the affected individuals. In these 7 families we also evaluated proteins encoded by the genes whether they interacted with the known proteins related to thyroid development, function, serum and cell TH transport and hormone synthesis by using STRING, a database of protein-protein interactions (21). The most frequent mutations identified were in *TG* and *TPO* genes, with 7 and 6 mutations respectively not previously reported (Table 2). For both *TG* and *TPO* mutations the clinical presentations were similar with goiter and/or developmental delay along with characteristic TFTs (Figure 1). In patients with *TG* mutations serum TG levels were low or undetectable (Figure 1) averaging 1.3 ng/mL (reference range 1.7-56 ng/mL) compared to patients with *TPO* mutations, with average TG levels of 165 ng/mL. Comprehensive clinical data and records related to the CH phenotype from many of Sudanese families were not always readily available prior to starting medication, especially if referred from rural clinics.

Mutations were also identified in the *DUOX1* and *DUOX2* genes, with two families (Table 2, Family A and B) having homozygous mutations. Affected members in both Family A and B presented with high TSH and low FT_4 prior to being diagnosed and put on treatment. One

affected from each family also presented with a developmental delay due to a late diagnosis. Another family from the same Sudanese cohort (Family J) also was compound heterozygous for one *DUOX2* and two *TPO* mutations, where the affected with all three mutations all presented with a goiter in addition to abnormal TFTs. One mutation in each of *IYD*, *SLC26A4*, and *SLC26A7* were identified in Family C, D, and E respectively (Table 2). The affected proband in each of these families also presented with high TSH and low FT_4 in addition to other relevant CH phenotypic considerations. All three affected children in Family C including the proband suffered with a development delay and a large goiter along with the abnormal TFTs. The affected proband in Family D presented with abnormal TFTs, a goiter, development delay, and deafness. The proband of Family E presented with abnormal TFTs as well as a developmental delay. Two families had different mutations in the *SLC5A5* gene (Family F and G) with all affected family members originally presenting with high TSH, low FT_4, and a goiter. All affected individuals except Family J were homozygous for each pathogenic variant (Figure 1 and Bruellman R. Data from: Supplemental Figure 3. Figshare. Deposited 27 September 2019. https://doi.org/10.6084/m9.figshare.9916265.v1). All mutations described were linked with the phenotype of CH and were designated as deleterious by in silico prediction tools (Table 2 and Bruellman R. Data from: Supplemental Table 1 - TG Splicing Score. Figshare. Deposited 27 September 2019. https://doi.org/10.6084/m9.figshare.9916253.v1) and had significant alterations in protein structure (Figure 2, Bruellman R. Data from: Supplemental Figure 1. Figshare. Deposited 27 September 2019. https://doi.org/10.6084/m9.figshare.9916256.v1, and Bruellman R. Supplemental Figure 2. Figshare. Deposited 27 September 2019. https://doi.org/10.6084/m9.figshare.9916259.v1.).

The mutations summarized in Table 2 directly impact the CH phenotype due to their location in the functional domains of the respective genes. Figure 2 denotes each mutation's approximate location along its respective gene with missense and splicing mutations occurring in highly conserved domains essential for proper protein function. Each frameshift mutation led to protein truncation.

Prevalence of *TPO* and *TG* Mutations

The average prevalence of 654 SNPs in *TG* and 234 SNPs in *TPO* genes for both the Sudanese population and all other population groups are shown on Table 3. The prevalence of *TG* mutations was significantly higher in the Sudanese populations than all other ethnic groups ($p<0.01$). Analysis of the prevalence of mutations identified in this Sudanese cohort relative to gene length demonstrated a direct correlation of more frequent mutations seen in larger genes; the r^2 value (0.46694) of the trendline calculated in Excel illustrates a positive relationship between gene size and the number of unique mutations (Figure 3).

Discussion

In this report, we have identified new mutations affecting 8 genes that are associated with CH. Mutations reported in *DUOX1*, *DUOX2*, *IYD*, *SLC26A4*, *SLC26A7*, and *SLC5A5* occur along critical domains in the respective genes resulting in a clear CH phenotype in the affected individuals (Figure 2 and Bruellman R. Data from: Supplemental Figure 3. Figshare. Deposited 27 September 2019. https://doi.org/10.6084/m9.figshare.9916265.v1). Only a few mutations have been reported to date in *IYD* (22). While performing a radioactive iodide uptake was not possible during the family visit to the clinician in Sudan, the TFTs and clinical presentation confirm the CH diagnosis. The position of the mutation along the nitroreductase domain of *IYD* would hinder the ability of iodide to be properly recycled for TH synthesis illustrated by TFT

values such as markedly high TG levels (Bruellman R. Data from: Supplemental Figure 3. Figshare. Deposited 27 September 2019. https://doi.org/10.6084/m9.figshare.9916265.v1) similar to TG values in previous families with documented *IYD* mutations (22). The frameshift insertion in *SLC26A7* (Family E) causes an early stop codon in exon 8, shortening the protein significantly from 19 exons in the wild type. Links between *SLC26A7* and a CH phenotype were recently established with 10 reported cases (5,23). The novel *SLC26A7* frameshift with early truncation manifests a severe CH phenotype. The novel *SLC5A5* mutation reported herein occurs along transmembrane domain 7, one of the domains essential for proper coupling and translocation of sodium (24).

Similar to previous studies of CH in different populations, *TPO* and *TG* gene mutations are the most frequent (25-28). Results from consanguineous families with CH in Turkey reported significantly high levels of *TPO* mutations overshadowing all other gene mutations causing CH (25). Non-consanguineous populations have also shown high rates of *TPO* mutations within such populations as the Portuguese, Japanese, and Finnish (26-28). Studies in Korea, China, and Japan also have identified *DUOX2* as a frequent cause of CH (27,29). The Sudanese have a high degree of consanguinity (2) which may have contributed to a preponderance of a particular mutation compared to other populations. Marrying within tribes is a common practice in the Sudan (2), perpetuating a founder's effect within each isolated tribe. However, we cannot ignore other factors that may be responsible for the high prevalence of *TPO* and *TG* gene mutations.

While missense and total mutations in *TPO* were found to be much higher than for other genes of comparable length (Figure 3), the $P = 0.063$ in comparing SNPs between Sudanese and all other populations showed no significant difference (Table 3). The damaging effects being

profound as shown in our Sudanese population can be explained by high prevalence of consanguinity causing high incidence of homozygous rare mutations as well as deficient iodine intake.

Figure 3 illustrates the importance of considering gene length as a reason for higher rates of mutation, as our results show increasing number of mutations with increasing gene amino acid length. The sheer size of the *TG* gene being over 2,700 amino acids spanning 48 exons is one of the reasons for the high incidence of *TG* mutations. However, other studies previously discussed (25-29) also show high rates of mutation in other genes of shorter lengths in consanguineous and non-consanguineous families alike.

Of note, the overall incidence of SNPs in the TG gene was significantly higher in the Sudanese compared to other reported populations (Table 3) and the possibility of positive selection was raised. As we do not have enough genetic data in Sudanese, positive selection in the TG gene cannot be confirmed. However, Bertranpetit and his group performed XP-EHH test on another African population (not Sudanese) and a clear signal of recent selection in the TG gene was found (personal communication). This data supports the possibility of positive selection in the TG gene in the Sudanese population, although direct proof cannot be obtained at this point. Many Sudanese families live in impoverished rural areas that lack the proper intake of iodine (30). This has the potential to aggravate the severity of the clinical manifestation of CH. As previously noted, TG protein devoid of the ChEL domain would result in no TG secretion to iodination sites (31). Further work could potentially determine if TG mutations do, in fact, confer a positive selection advantage due to the potential of the goiter to retain what little iodine is available or to potentially counteract the harmful effects of high TG and thyroid differentiation from high TSH stimulation. It has been reported that signature of positive

selection was observed at some genes involved in growth and metabolism related to thyroid or pituitary function in some African populations, which may contribute to local adaptation to these mutations (32,33). Although differences in diet, climate, and exposure to pathogens among ethnically and geographically diverse African populations are considered to produce distinct selection pressure, the mechanism is unclear. Geography might be another factor of positive selection; however further studies will be necessary to confirm this.

The rate of dyshormonogenesis in the Sudanese population is three times that in other populations. We find increased prevalence of *TG* and *TPO* mutations in Sudanese children with congenital hypothyroidism and our data suggest that gene size, consanguinity and possible positive genetic selection could account for this large proportion of dyshormonogenesis in the Sudanese population.

Conflict of Interest

The authors declare that they have no conflicts of interest.

References

1. Grasberger H. Defects of thyroidal hydrogen peroxide generation in congenital hypothyroidism. *Mol Cell Endocrinol*. 2010;322(1-2):99-106.
2. Mukhtar M, Elshibli EM, Abdelmonein H, Abdullah MA, Elmula IF. Demographic and Clinical Aspects of Congenital Hypothyroidism (Dyshormonogenesis) in Sudan. *Sudan J Med Sci*. 2015;10(3):109-116.
3. Rigutto S, Hoste C, Grasberger H, Milenkovic M, Communi D, Dumont JE, Corvilain B, Miot F, De Deken X. Activation of dual oxidases Duox1 and Duox2: differential regulation mediated by camp-dependent protein kinase and protein kinase C-dependent phosphorylation. *J Biol Chem*. 2009;284(11):6725-6734.
4. Targovnik HM, Citterio CE, Rivolta CM. Iodide handling disorders (NIS, TPO, TG, IYD). *Best Pract Res Clin Endocrinol Metab*. 2017;31(2):195-212.
5. Cangul H, Liao XH, Schoenmakers E, Kero J, Barone S, Srichomkwun P, Iwayama H, Serra EG, Saglam H, Eren E, Tarim O, Nicholas AK, Zvetkova I, Anderson CA, Frankl FEK, Boelaert K, Ojaniemi M, Jaaskelainen J, Patyra K, Lof C, Williams ED, Consortium UK, Soleimani M, Barrett T, Maher ER, Chatterjee VK, Refetoff S, Schoenmakers N. Homozygous loss-of-function mutations in SLC26A7 cause goitrous congenital hypothyroidism. *JCI Insight*. 2018;3(20).
6. Dossena S, Rodighiero S, Vezzoli V, Nofziger C, Salvioni E, Boccazzi M, Grabmayer E, Botta G, Meyer G, Fugazzola L, Beck-Peccoz P, Paulmichl M. Functional characterization of wild-type and mutated pendrin (SLC26A4), the anion transporter involved in Pendred syndrome. *J Mol Endocrinol*. 2009;43(3):93-103.

7. Watanabe Y, Sharwood E, Goodwin B, Creech MK, Hassan HY, Netea MG, Jaeger M, Dumitrescu A, Refetoff S, Huynh T, Weiss RE. A novel mutation in the TG gene (G2322S) causing congenital hypothyroidism in a Sudanese family: a case report. *BMC Med Genet*. 2018;19(1):69.

8. Lee J, Wang X, Di Jeso B, Arvan P. The cholinesterase-like domain, essential in thyroglobulin trafficking for thyroid hormone synthesis, is required for protein dimerization. *J Biol Chem*. 2009;284(19):12752-12761.

9. Rubio IG, Medeiros-Neto G. Mutations of the thyroglobulin gene and its relevance to thyroid disorders. *Curr Opin Endocrinol Diabetes Obes*. 2009;16(5):373-378.

10. Bresson D, Rebuffat SA, Peraldi-Roux S. Localization of the immunodominant region on human thyroid peroxidase in autoimmune thyroid diseases: an update. *J Autoimmune Dis*. 2005;2(1):2.

11. Muzza M, Fugazzola L. Disorders of H2O2 generation. *Best Pract Res Clin Endocrinol Metab*. 2017;31(2):225-240.

12. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40(Web Server issue):W452-457.

13. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249.

14. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11(4):361-362.

15. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894.
16. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37(9):e67.
17. Hollfelder N, Schlebusch CM, Gunther T, Babiker H, Hassan HY, Jakobsson M. Northeast African genomic variation shaped by the continuity of indigenous groups and Eurasian migrations. *PLoS Genet*. 2017;13(8):e1006976.
18. Watanabe Y, Bruellman RJ, Ebrhim RS, Abdullah MA, Dumitrescu AM, Refetoff S, Weiss RE. Congenital Hypothyroidism due to Oligogenic Mutations in Two Sudanese Families. *Thyroid*. 2019;29(2):302-304.
19. Lofrano-Porto A, Barra GB, Nascimento PP, Costa PG, Garcia EC, Vaz RF, Batista AR, Freitas AC, Cherulli BL, Bahmad F, Jr., Figueiredo LG, Neves FA, Casulari LA. Pendred syndrome in a large consanguineous Brazilian family caused by a homozygous mutation in the SLC26A4 gene. *Arq Bras Endocrinol Metabol*. 2008;52(8):1296-1303.
20. Watanabe Y, Ebrhim RS, Abdullah MA, Weiss RE. A Novel Missense Mutation in the SLC5A5 Gene in a Sudanese Family with Congenital Hypothyroidism. *Thyroid*. 2018;28(8):1068-1070.
21. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613.

22. Burniat A, Pirson I, Vilain C, Kulik W, Afink G, Moreno-Reyes R, Corvilain B, Abramowicz M. Iodotyrosine deiodinase defect identified via genome-wide approach. *J Clin Endocrinol Metab*. 2012;97(7):E1276-1283.

23. Ishii J, Suzuki A, Kimura T, Tateyama M, Tanaka T, Yazawa T, Arimasu Y, Chen IS, Aoyama K, Kubo Y, Saitoh S, Mizuno H, Kamma H. Congenital goitrous hypothyroidism is caused by dysfunction of the iodide transporter SLC26A7. *Commun Biol*. 2019;2:270.

24. Portulano C, Paroder-Belenitsky M, Carrasco N. The Na+/I- symporter (NIS): mechanism and medical impact. *Endocr Rev*. 2014;35(1):106-149.

25. Cangul H, Aycan Z, Olivera-Nappa A, Saglam H, Schoenmakers NA, Boelaert K, Cetinkaya S, Tarim O, Bober E, Darendeliler F, Bas V, Demir K, Aydin BK, Kendall M, Cole T, Hogler W, Chatterjee VK, Barrett TG, Maher ER. Thyroid dyshormonogenesis is mainly caused by TPO mutations in consanguineous community. *Clin Endocrinol (Oxf)*. 2013;79(2):275-281.

26. Rodrigues C, Jorge P, Soares JP, Santos I, Salomao R, Madeira M, Osorio RV, Santos R. Mutation screening of the thyroid peroxidase gene in a cohort of 55 Portuguese patients with congenital hypothyroidism. *Eur J Endocrinol*. 2005;152(2):193-198.

27. Park KJ, Park HK, Kim YJ, Lee KR, Park JH, Park JH, Park HD, Lee SY, Kim JW. DUOX2 Mutations Are Frequently Associated With Congenital Hypothyroidism in the Korean Population. *Ann Lab Med*. 2016;36(2):145-153.

28. Lof C, Patyra K, Kuulasmaa T, Vangipurapu J, Undeutsch H, Jaeschke H, Pajunen T, Kero A, Krude H, Biebermann H, Kleinau G, Kuhnen P, Rantakari K, Miettinen P, Kirjavainen T, Pursiheimo JP, Mustila T, Jaaskelainen J, Ojaniemi M, Toppari J, Ignatius

J, Laakso M, Kero J. Detection of Novel Gene Variants Associated with Congenital Hypothyroidism in a Finnish Patient Cohort. *Thyroid*. 2016;26(9):1215-1224.

29. Yu B, Long W, Yang Y, Wang Y, Jiang L, Cai Z, Wang H. Newborn Screening and Molecular Profile of Congenital Hypothyroidism in a Chinese Population. *Front Genet*. 2018;9:509.

30. Ali NI, Elgak SNA, Abdallah YMY, Idriss H. Assessment of iodine supplementation program on thyroid function in Sudan. *Diabetes Metab Syndr*. 2019;13(1):678-680.

31. Lee J, Di Jeso B, Arvan P. Maturation of thyroglobulin protein region I. *J Biol Chem*. 2011;286(38):33045-33052.

32. Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, Bodo JM, Lema G, Fu W, Nyambo TB, Rebbeck TR, Zhang K, Akey JM, Tishkoff SA. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell*. 2012;150(3):457-469.

33. Fan S, Kelly DE, Beltrame MH, Hansen MEB, Mallick S, Ranciaro A, Hirbo J, Thompson S, Beggs W, Nyambo T, Omar SA, Meskel DW, Belay G, Froment A, Patterson N, Reich D, Tishkoff SA. African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biol*. 2019;20(1):82.

Legends for Figures and Tables

Table 1 - List of genes related to thyroid disorders

Table 2 - Variant Type: MS = Missense Mutation; FSD = Frameshift Deletion Mutation; FSI = Frameshift Insertion Mutation; SG = Stop-gain Mutation; SPL = Splice Site Mutation

Table 3 - P-value was calculated by t-test to assess significant difference between the two average frequencies. Significant p-value (<.05) is in bold and underlined.

Figure 1 - Generations are denoted by roman numeral. Each subject is identified by the number just above the corresponding symbol. Laboratory thyroid function tests are aligned below each symbol. Abnormal values are in bold and underlined. Abbreviations: TT_4; total thyroxine, TT_3; total triiodothyronine, FT_4; free thyroxine, TSH; thyroid-stimulating hormone, TBG; thyroxine binding globulin, TG; thyroglobulin, TPO Ab; anti-TPO antibody, TG Ab; anti-thyroglobulin antibody.

Figure 2 - Amino acid numbers are denoted by numbers spanning schematic. Important domains are denoted by roman numerals or by their name for each gene. For the *TPO* gene, the Myeloperoxidase (MPO) domain is denoted by the black box from amino acid position 142 to 737. Catalytic site necessary for proper TPO function is within MPO Domain. The *TPO* Transmembrane domain is also marked along the gene approximately between amino acid 846 and 872. Each of the documented mutations in this report and previously reported by our lab are noted by boxes in their approximate locations. *, a novel mutation.

Figure 3 - Relationship of unique mutations in 19 Sudanese families compared to the gene's amino acid length. Mutations were only counted once in the instance of the same mutation being present in two different families. Trend line is denoted by the dotted line and the r2 is noted next to the line.

Table 1.

Genes Related to thyroid disorder							
AADAT	ALB	ALMS1	ATXN2	CDCA8	DIO1	DIO2	DIO3
DUOX1	DUOXA1	DUOX2	DUOXA2	EXOSC2	FGF8	FOXE1	GLIS3
GNAS	HHEX	HOXA3	IGSF1	IRS4	IYD	JAG1	KDM6A
KMT2D	NCOR2	NKX2-1	NKX2-5	NKX2-6	NTN1	P4HB	PAX8
POU1F1	PROP1	PSMA1	PSMA3	PSMD2	PSMD3	PTH1R	PTRH2
RYR2	SALL1	SECISBP2	SERPINA7	SLC16A2	SLC17A4	SLC26A4	SLC26A7
SLC30A10	SLC5A5	SLCO1C1	STAMBP	TBL1X	TBX1	TG	THRA
THRB	TPO	TRH	TRHR	TRIP11	TRIP12	TSHB	TSHR
TTR	TUBB1	UBR1	VAV3				

Table 2.

Gene Variant Information and Frequencies											
Family	**Gene**	**Variant**	**CH previously reported**	**Variant Type**	**RS or Accession number[1]**	**SIFT[2]**	**Polyphen 2_HDIV[3]**	**Mutation Taster[4]**	**CADD[5]**	**Freq _Afr[6]**	**Freq _All[7]**
A	DUOX1	c.C1525T p.R509W	Yes[8]	MS	rs757808802	0.0, D	1.0, D	1, D	16.38	0.00	0.00
A	DUOX2	c.G3329A p.R1110Q	Yes[8]	MS	rs368488511	0.003, D	0.994, D	1, D	36	0.00	0.00
B	DUOX1	c.C3388G p.H1130D	No	MS	MN115831	0.0, D	1.0, D	1, D	23.1	.	.
B, J	DUOX2	c.1395_1396delCC p.P465fs	Yes[9]	FSD	MH290757	.	.	1, D	.	.	.
C	IYD	c.C835T p.R279C	No	MS	MN115832	0.0, D	1.0, D	1, D	25.3	.	.
D	SLC26A4	c.1197delT p.S399fs	Yes[10]	FSD	rs397516413	.	.	1, D	.	0.00	0.00
E	SLC26A7	c.818_819insTTAT p.C273fs	No	FSI	MN115835	.	.	1, D	.	.	.
F	SLC5A5	c.G749T p.G250V	No	MS	MN115830	0.0, D	1.0, D	1, D	18.58	.	.
G	SLC5A5	c.T1042G p.Y348D	Yes[11]	MS	MH046063	0.001, D	0.999, D	1, D	15.71	.	.
H, I	TPO	c.C1277G p.A426G	No	MS	rs61758082	0.048, D	0.087, B	1, N	13.5	0.03	0.00
J	TPO	c.C1535A p.P512H	Yes[9]	MS	rs150489706	0.0, D	1.0, D	1, D	25.3	0.00	0.00
J	TPO	c.G1759A p.G587R	Yes[9]	MS	rs770562452	0.001, D	1.0, D	1, D	27.4	0.00	0.00
K	TPO	c.G2578A p.G860R	No	MS	rs556552435	0.077, T	0.999, D	1, N	12	0.00	0.00
I	TPO	c.2422del p.C808AfsTer24	No	FSD	rs763662774	.	.	1, D	23.2	0.00	0.00
L	TG	c.5975+401del p.G1992GfsTer75	No	FSD	MN115834	.	.	.	.	.	.
M	TG	c.T6989A p.V2330E	No	MS	MH137705	0.001, D	1.0, D	1, D	28.4	.	.
N	TG	c.G7021A p.G2341S	Yes[12]	MS	NM003235	0.001, D	1.0, D	1, D	34	.	.
O	TG	c.7268del p.V2423VfsTer45	No	FSD	MN115833	.	.	.	.	.	.
P	TG	c.C7655G p.S2552X	No	SG	MH298849	.	.	1, A	46	.	.
Q	TG	c.7909ins p.Y3637Ffs	No	FSI	Pending	.	.	.	.	.	.
Family	**Gene**	**Variant**	**CH previously reported**	**Variant Type**	**RS or Accession number[1]**	**The Human Splicing Finder[13]**					
R, S	TG	c.4816+1G>T	No	SPL	MH137704	-27.7% variation; alteration of wild type donor site, most probably affecting splicing					
J	TPO	c.483-2A>G	Yes[9]	SPL	MH137703	-31.7% variation; alteration of wild type donor site, most probably affecting splicing					

[1]RS number, reference single nucleotide variants number; Accession number, GenBank Accession Number on GenBank sequence database

[2]Sorting Intolerant From Tolerant (SIFT). Scores and predictions are separated by a comma. There are two possible predictions: D (damaging, score ≤0.05); T (tolerated, score >0.05) (12)

[3]Polymorphism Phenotyping v2 HDIV (Polyphen2_HDIV). Scores and predictions are separated by a comma. There are three possible predictions: D (probably damaging, score ≥0.909), P (possibly damaging, 0.447 ≤ score ≤ 0.908), B (benign, score ≤ 0.446) (13)

[4]MutationTaster. Scores and predictions are separated by a comma. The closer a score is to the value 1, the higher the confidence in the prediction. There are four possible predictions: A (disease causing automatic), D (disease causing), N (polymorphism), P (polymorphism automatic) (14)

[5]Combined Annotation Dependent Depletion (CADD) Score. Indicates rarity of variant. Score of 20 indicates variant among top 1% of deleterious variants in human genome. Score of 30 indicates variant among top 0.1% of deleterious variants in human genome (15)

[6]Alternative allele frequency in African population in The Genome Aggregation Database

[7]Alternative allele frequency in all populations in The Genome Aggregation Database

[8]Dual homozygous DUOX1 and DUOX2 Sudanese family previously reported by our lab (18)

[9]Compound heterozygous TPO and DUOX2 Sudanese family previously reported in our lab (18)

[10]Homozygous SLC26A4 mutation found in an affected Sudanese family, previously reported as pathogenic (19)

[11]Homozygous SLC5A5 mutation in Sudanese family previously reported in our lab (20)

[12]Homozygous TG mutation in Sudanese family previously reported in our lab (7)

[13]The Human Splicing Finder. Possible predictions: site broken (variation score < -10%), new site (variation score > 10%) (16)

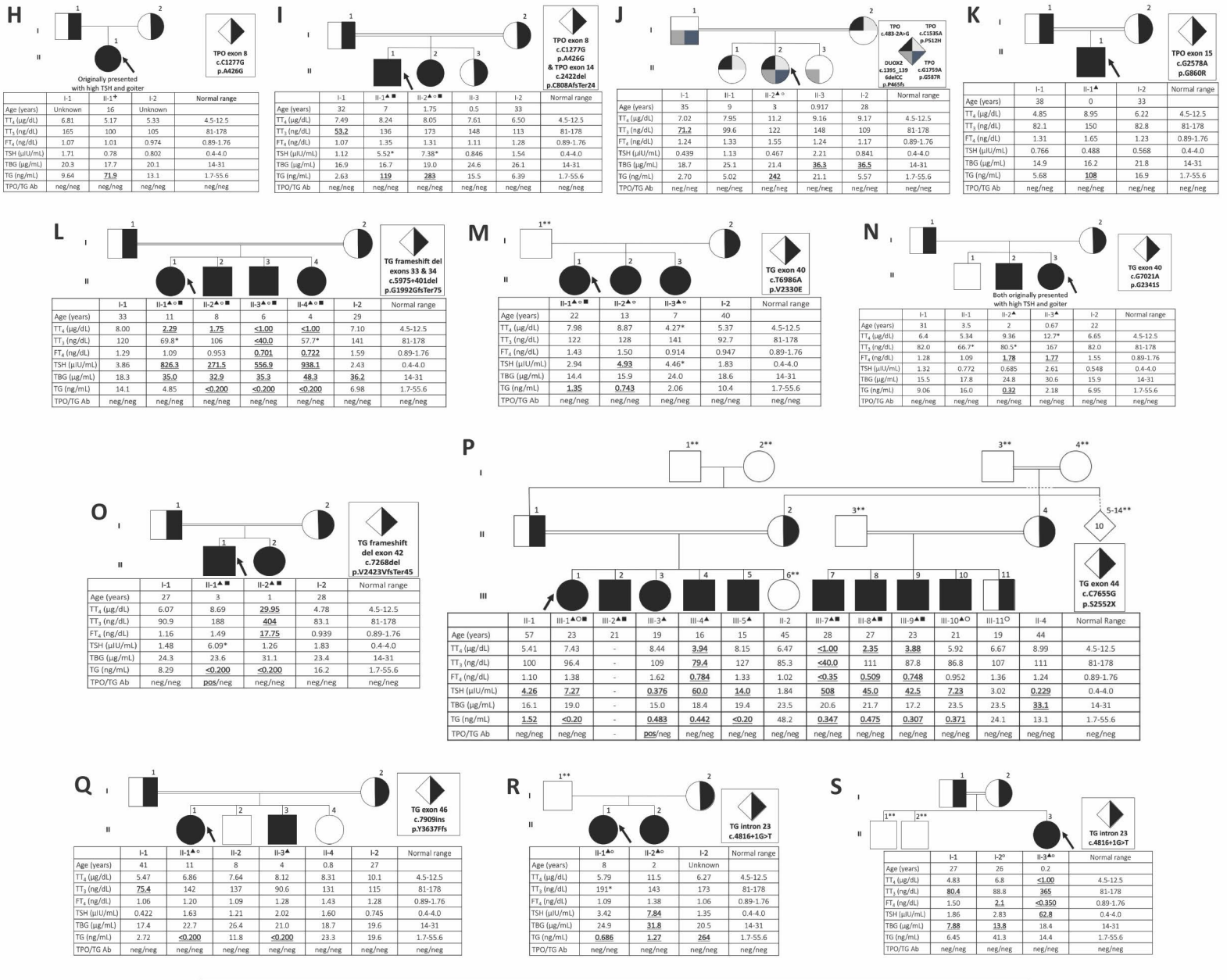
Table 3.

T-Test Results - Sudanese TG and TPO Mutation Rate vs. Other Populations

Gene	Number of SNPs	Frequency of mutations		P-value
		Sudanese (n=234)	All (n=123,136)	
TG	654	0.1512	0.1406	**<0.01**
TPO	234	0.1864	0.175	0.063

Figure 1

TPO / *TG* Family Pedigree and Phenotype Information



H	I-1	II-1✣	I-2	Normal range
Age (years)	Unknown	16	Unknown	
TT_4 (µg/dL)	6.81	5.17	5.33	4.5-12.5
TT_3 (ng/dL)	165	100	105	81-178
FT_4 (ng/dL)	1.07	1.01	0.974	0.89-1.76
TSH (µIU/mL)	1.71	0.78	0.802	0.4-4.0
TBG (µg/mL)	20.3	17.7	20.1	14-31
TG (ng/mL)	9.64	71.9	13.1	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg

I	I-1	II-1▲■	II-2▲○■	II-3	I-2	Normal range
Age (years)	32	7	1.75	0.5	33	
TT_4 (µg/dL)	7.49	8.24	8.05	7.61	6.50	4.5-12.5
TT_3 (ng/dL)	53.2	136	173	148	113	81-178
FT_4 (ng/dL)	1.07	1.35	1.31	1.11	1.28	0.89-1.76
TSH (µIU/mL)	1.12	5.52*	7.38*	0.846	1.54	0.4-4.0
TBG (µg/mL)	16.9	16.7	19.0	24.6	26.1	14-31
TG (ng/mL)	2.63	119	283	15.5	6.39	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg

J	I-1	II-1	II-2▲○	II-3	I-2	Normal range
Age (years)	35	9	3	0.917	28	
TT_4 (µg/dL)	7.02	7.95	11.2	9.16	9.17	4.5-12.5
TT_3 (ng/dL)	71.2	99.6	122	148	109	81-178
FT_4 (ng/dL)	1.24	1.33	1.55	1.24	1.17	0.89-1.76
TSH (µIU/mL)	0.439	1.13	0.467	2.21	0.841	0.4-4.0
TBG (µg/mL)	18.7	25.1	21.4	36.3	36.5	14-31
TG (ng/mL)	2.70	5.02	242	21.1	5.57	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg

K	I-1	II-1▲	I-2	Normal range
Age (years)	38	0	33	
TT_4 (µg/dL)	4.85	8.95	6.22	4.5-12.5
TT_3 (ng/dL)	82.1	150	82.8	81-178
FT_4 (ng/dL)	1.31	1.65	1.23	0.89-1.76
TSH (µIU/mL)	0.766	0.488	0.568	0.4-4.0
TBG (µg/mL)	14.9	16.2	21.8	14-31
TG (ng/mL)	5.68	108	16.9	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg

L	I-1	II-1▲○■	II-2▲○■	II-3▲○■	II-4▲○■	I-2	Normal range
Age (years)	33	11	8	6	4	29	
TT_4 (µg/dL)	8.00	2.29	1.75	<1.00	<1.00	7.10	4.5-12.5
TT_3 (ng/dL)	120	69.8*	106	<40.0	57.7*	141	81-178
FT_4 (ng/dL)	1.29	1.09	0.953	0.701	0.722	1.59	0.89-1.76
TSH (µIU/mL)	3.86	826.3	271.5	556.9	938.1	2.43	0.4-4.0
TBG (µg/mL)	18.3	35.0	32.9	35.3	48.3	36.2	14-31
TG (ng/mL)	14.1	4.85	<0.200	<0.200	<0.200	6.98	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg

M	I-1	II-1▲○■	II-2▲○	II-3▲○	I-2	Normal range
Age (years)	22	13	7	40		
TT_4 (µg/dL)	7.98	8.87	4.27*	5.37	4.5-12.5	
TT_3 (ng/dL)	122	128	141	92.7	81-178	
FT_4 (ng/dL)	1.43	1.50	0.914	0.947	0.89-1.76	
TSH (µIU/mL)	2.94	4.93	4.46*	1.83	0.4-4.0	
TBG (µg/mL)	14.4	15.9	24.0	18.6	14-31	
TG (ng/mL)	1.35	0.743	2.06	10.4	1.7-55.6	
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	

N	I-1	II-1	II-2▲	II-3▲	I-2	Normal range
Age (years)	31	3.5	2	0.67	22	
TT_4 (µg/dL)	6.4	5.34	9.36	12.7*	6.65	4.5-12.5
TT_3 (ng/dL)	82.0	66.7*	80.5*	167	82.0	81-178
FT_4 (ng/dL)	1.28	1.09	1.78	1.77	1.55	0.89-1.76
TSH (µIU/mL)	1.32	0.772	0.685	2.61	0.548	0.4-4.0
TBG (µg/mL)	15.5	17.8	24.8	30.6	15.9	14-31
TG (ng/mL)	9.06	16.0	0.32	2.18	6.95	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg

O	I-1	II-1▲■	II-2▲■	I-2	Normal range
Age (years)	27	3	1	28	
TT_4 (µg/dL)	6.07	8.69	29.95	4.78	4.5-12.5
TT_3 (ng/dL)	90.9	188	404	83.1	81-178
FT_4 (ng/dL)	1.16	1.49	17.75	0.939	0.89-1.76
TSH (µIU/mL)	1.48	6.09*	1.26	1.83	0.4-4.0
TBG (µg/mL)	24.3	23.6	31.1	23.4	14-31
TG (ng/mL)	8.29	<0.200	<0.200	16.2	1.7-55.6
TPO/TG Ab	neg/neg	pos/neg	neg/neg	neg/neg	neg/neg

P	I-1	III-1▲○■	III-2▲■	III-3▲	III-4▲	III-5▲	I-2	III-7▲■	III-8▲■	III-9▲■	III-10▲○	III-11○	II-4	Normal Range
Age (years)	57	23	21	19	16	15	45	28	27	23	21	19	44	
TT_4 (µg/dL)	5.41	7.43	-	8.44	3.94	8.15	6.47	<1.00	2.35	3.88	5.92	6.67	8.99	4.5-12.5
TT_3 (ng/dL)	100	96.4	-	109	79.4	127	85.3	<40.0	111	87.8	86.8	107	111	81-178
FT_4 (ng/dL)	1.10	1.38	-	1.62	0.784	1.33	1.02	<0.35	0.509	0.748	0.952	1.36	1.24	0.89-1.76
TSH (µIU/mL)	4.26	7.27	-	0.376	60.0	14.0	1.84	508	45.0	42.5	7.23	3.02	0.229	0.4-4.0
TBG (µg/mL)	16.1	19.0	-	15.0	18.4	19.4	23.5	20.6	21.7	17.2	23.5	23.5	33.1	14-31
TG (ng/mL)	1.52	<0.20	-	0.483	0.442	<0.20	48.2	0.347	0.475	0.307	0.371	24.1	13.1	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	-	pos/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg

Q	I-1	II-1▲○	II-2	II-3▲	II-4	I-2	Normal range
Age (years)	41	11	8	4	0.8	27	
TT_4 (µg/dL)	5.47	6.86	7.64	8.12	8.31	10.1	4.5-12.5
TT_3 (ng/dL)	75.4	142	137	90.6	131	115	81-178
FT_4 (ng/dL)	1.06	1.20	1.09	1.28	1.43	1.28	0.89-1.76
TSH (µIU/mL)	0.422	1.63	1.21	2.02	1.60	0.745	0.4-4.0
TBG (µg/mL)	17.4	22.7	26.4	21.0	18.7	19.6	14-31
TG (ng/mL)	2.72	<0.200	11.8	<0.200	23.3	19.6	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg	neg/neg

R	II-1▲○	II-2▲○	I-2	Normal range
Age (years)	8	2	Unknown	
TT_4 (µg/dL)	5.79	11.5	6.27	4.5-12.5
TT_3 (ng/dL)	191*	143	173	81-178
FT_4 (ng/dL)	1.09	1.38	1.06	0.89-1.76
TSH (µIU/mL)	3.42	7.84	1.35	0.4-4.0
TBG (µg/mL)	24.9	31.8	20.5	14-31
TG (ng/mL)	0.686	1.27	264	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg

S	I-1	I-2○	II-3▲○	Normal range
Age (years)	27	26	0.2	
TT_4 (µg/dL)	4.83	6.8	<1.00	4.5-12.5
TT_3 (ng/dL)	80.4	88.8	365	81-178
FT_4 (ng/dL)	1.50	2.1	<0.350	0.89-1.76
TSH (µIU/mL)	1.86	2.83	62.8	0.4-4.0
TBG (µg/mL)	7.88	13.8	18.4	14-31
TG (ng/mL)	6.45	41.3	14.4	1.7-55.6
TPO/TG Ab	neg/neg	neg/neg	neg/neg	neg/neg

Figure 2

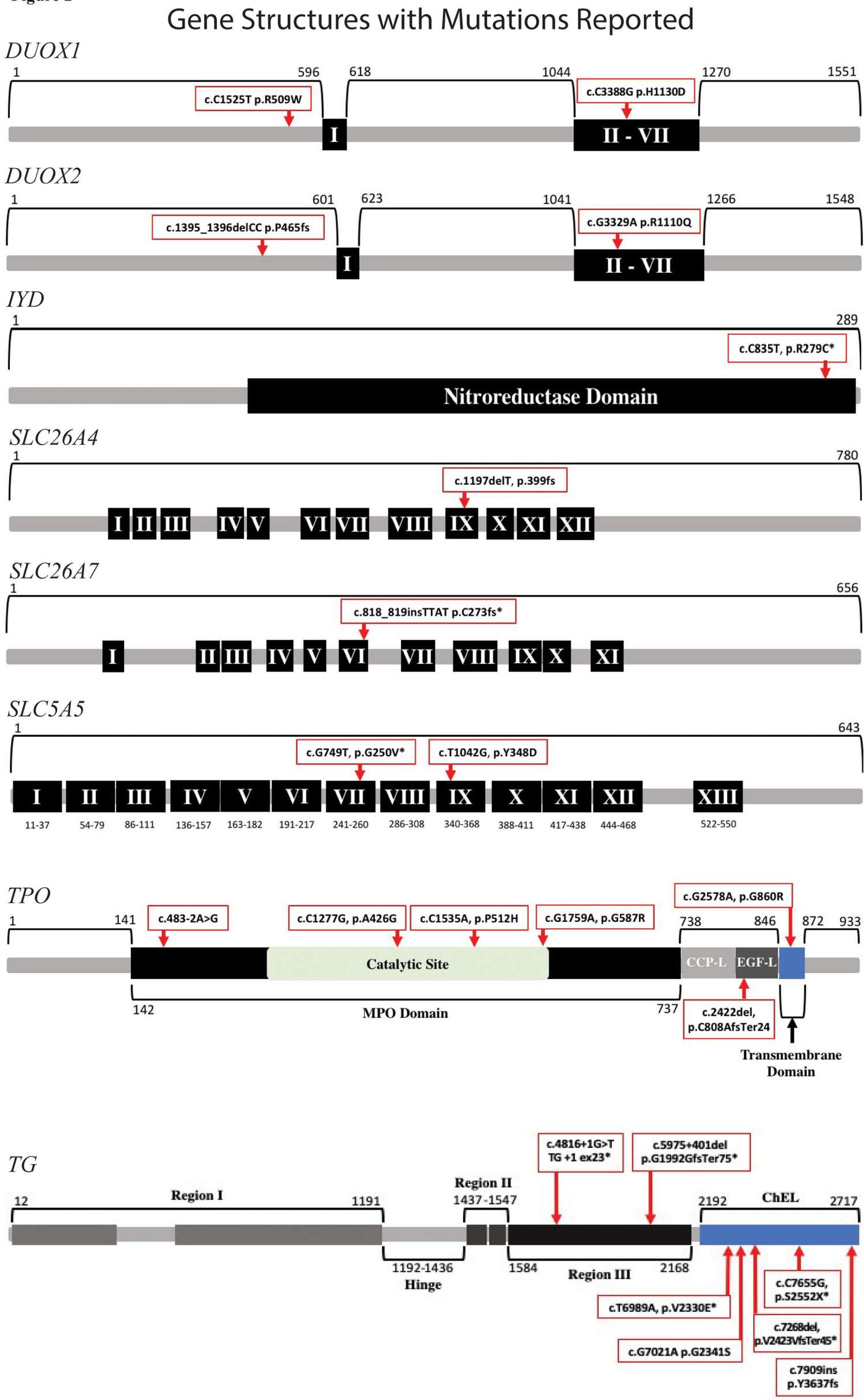
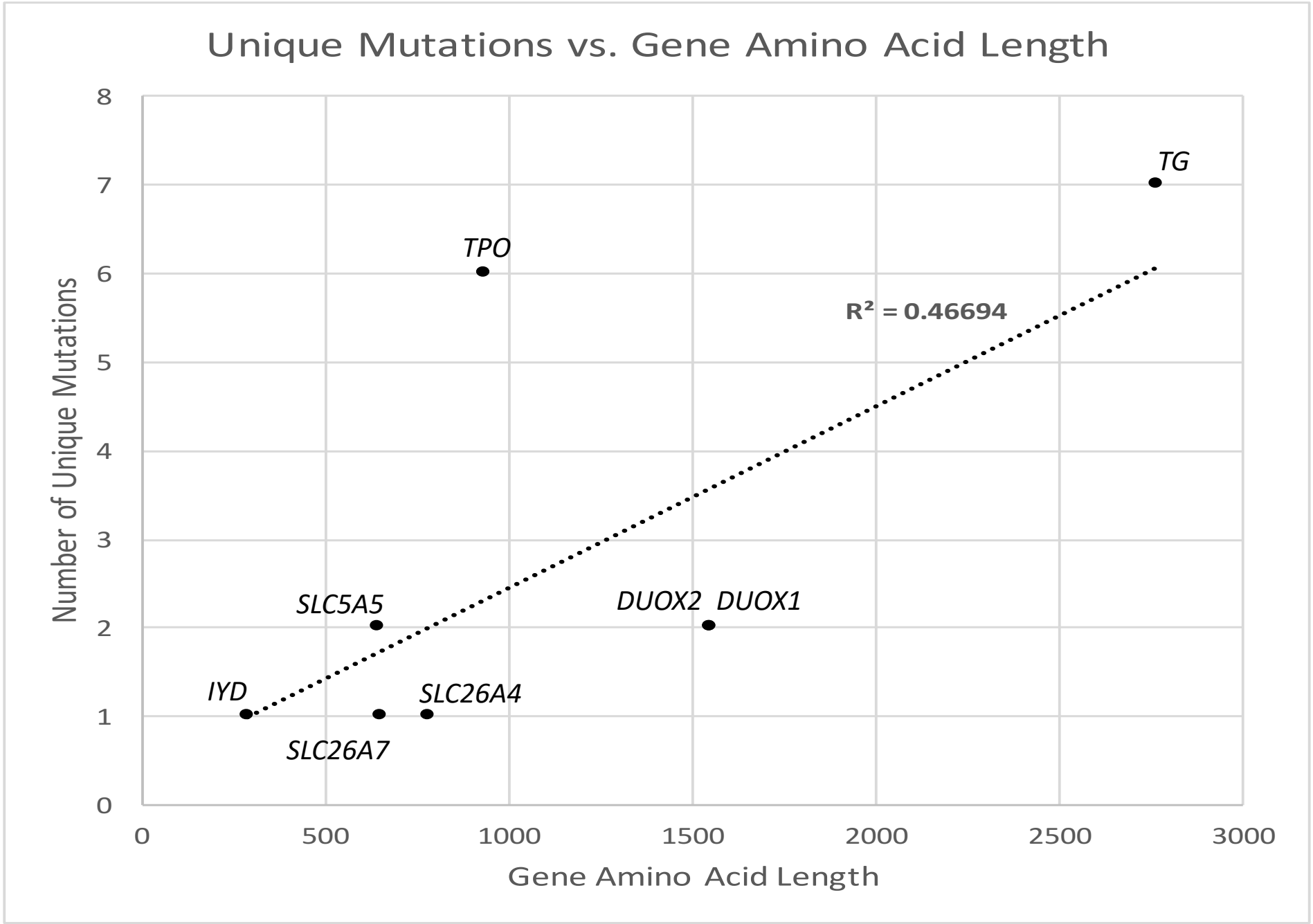

Gene Structures with Mutations Reported
DUOX1
1
596
618
1044
1270
1551
c.C1525T p.R509W
c.C3388G p.H1130D
I
II - VII
DUOX2
1
601
623
1041
1266
1548
c.1395_1396delCC p.P465fs
c.G3329A p.R1110Q
I
II - VII
IYD
1
289
c.C835T, p.R279C*
Nitroreductase Domain
SLC26A4
1
780
c.1197delT, p.399fs
I II III IV V VI VII VIII IX X XI XII
SLC26A7
1
656
c.818_819insTTAT p.C273fs*
I II III IV V VI VII VIII IX X XI
SLC5A5
1
643
c.G749T, p.G250V*
c.T1042G, p.Y348D
I II III IV V VI VII VIII IX X XI XII XIII
11-37 54-79 86-111 136-157 163-182 191-217 241-260 286-308 340-368 388-411 417-438 444-468 522-550
TPO
1
141
c.483-2A>G
c.C1277G, p.A426G
c.C1535A, p.P512H
c.G1759A, p.G587R
c.G2578A, p.G860R
738
846
872
933
Catalytic Site
CCP-L
EGF-L
142
MPO Domain
737
c.2422del, p.C808AfsTer24
Transmembrane Domain
TG
12
Region I
1191
Region II
1437-1547
c.4816+1G>T TG +1 ex23*
c.5975+401del p.G1992GfsTer75*
2192
ChEL
2717
1192-1436
Hinge
1584
Region III
2168
c.C7655G, p.S2552X*
c.T6989A, p.V2330E*
c.7268del, p.V2423VfsTer45*
c.G7021A p.G2341S
c.7909ins p.Y3637fs

Figure 3


Unique Mutations vs. Gene Amino Acid Length
Number of Unique Mutations
Gene Amino Acid Length
0
1
2
3
4
5
6
7
8
0
500
1000
1500
2000
2500
3000
TG
TPO
R² = 0.46694
SLC5A5
DUOX2
DUOX1
IYD
SLC26A4
SLC26A7