Analysis of worldwide carrier frequency and predicted genetic prevalence of congenital hypothyroidism based on a general population database

Kyung Sun Park[1]

[1]Kyung Hee University School of Medicine

August 20, 2020

Abstract

Background: To assess how genomic information of the general population reflects probabilities of developing diseases and the differences in those probabilities among ethnic groups, a general population database was analyzed with an example of congenital hypothyroidism. Methods: Ten candidate genes that follow an autosomal recessive inheritance pattern in congenital hypothyroidism (SLC5A5, TPO, TG, IYD, DUOXA2, DUOX2, TSHR, TSHB, TRHR, FOXE1) in the gnomAD database (v2.1.1) were analyzed. The carrier frequency (CF) and predicted genetic prevalence (pGP) were estimated. Results: The total CF in the overall population was 2.9%. DUOX2 showed the highest CF (1.46%), followed by TG (0.47%), TPO (0.42%), TSHR (0.16%), DUOXA2 (0.15%), IYD (0.08%), TRHR (0.07%), SLC5A5 (0.07%), TSHB (0.04%), and FOXE1 (0%). The pGP in the overall population was 6.57 individuals per 100.000 births (1:15,216). The highest pGP was in the East Asian group at 44.65 per 100,000 births (1:2,240), followed by Finnish (14.21), other (5.99), Non-Finnish European (5.96), African (3.80), South Asian (3.07), Latino (2.99), and Ashkenazi Jewish (1.52) groups. Conclusion: The feasibility of genetic screening for congenital hypothyroidism may be determined by ethnicity. Comparing the pGP with the real prevalence of congenital hypothyroidism indicates that genetic screening in East Asian may be feasible.

Introduction

Genetic screening is a type of genetic testing that is designed to identify a specified population at a higher risk of having or developing a disease with the aim of prevention or early treatment (Andermann & Blancquaert, 2010). Generally, genetic screening is performed as targeted testing for known hotspot variations. The development of next-generation sequencing (NGS) techniques has introduced a new genomic era by producing massive genomic data and reducing costs. Recently, the Genome Aggregation Database (gnomAD, https://gnomad.broadinstitute.org/) has been constructed as a very large database that contains genomic information of the general population worldwide (Karczewski et al., 2020; Lek et al., 2016). Several companies have launched proactive genetic testing for generally healthy individuals without a personal or family history using NGS techniques for identifying particular genes or performing whole exome/genome sequencing that is not confined to hotspot variations.

At present, major questions are how genomic information of a generally healthy population reflects probabilities of developing diseases and the differences in those probabilities among ethnic groups. Additional questions are the use of genetic testing to provide useful and crucial information and eventually prevent diseases in healthy individuals. To answer these questions, genomic data associated with congenital hypothyroidism, which is one of the major achievements of preventive medicine (Buyukgebiz, 2013), were analyzed based on the general population database and their carrier frequency and genetic prevalence were estimated by ethnicity.

Methods

Database analysis

The gnomAD data (v2.1.1) was obtained from https://gnomad.broadinstitute.org/. A total of 10 candidate genes that follow an autosomal recessive inheritance pattern in congenital hypothyroidism (*SLC5A5* , *TPO* , *TG* , *IYD* ,*DUOXA2* , *DUOX2* , *TSHR* , *TSHB* , *TRHR* ,*FOXE1*) were analyzed. The gnomAD database (v2.1.1) contains genetic variants from 125,748 exomes and 15,708 genomes and eight populations, including African/African American, Latino Ashkenazi Jewish, East Asian, Finnish, Non-Finnish European, South Asian, and other groups.

Genetic variant classification

All genetic variants in the 10 candidate genes reported in the gnomAD database (v.2.1.1) were classified into five categories, benign, likely benign, uncertain significance, likely pathogenic, and pathogenic, following the 2015 American College of Medical Genetics and Genomics (ACMG)/Association for Molecular Pathology (AMP) standards and guidelines (Richards et al., 2015). Loss-of-function variants of these 10 candidate genes were presumed to be responsible for a congenital hypothyroidism mechanism. The GRCh37/hg19 genomic build was used for all position descriptions. All variants were described according to HGVS variant nomenclature standards (http://varnomen.hgvs.org/) (den Dunnen et al., 2016) and analyzed based on the transcript selected by matched annotation from NCBI and EMBL-EBI (MANE) (https://www.ncbi.nlm.nih.gov/refseq/MANE/) using a Mutalyzer program (https://mutalyzer.nl/). If the genetic variants were not known pathogenic or likely pathogenic variants (PLPV), the null variants (stop-gain, splice site disrupting, or frameshift variants) with flags of low-confidence predicted loss-of-function (pLoF) or pLof flag by loss-of-function transcript effect estimator (LOFTEE, https://github.com/konradjk/loftee) were filtered. For the prediction of variant pathogenicity, multiple *in silico* software such as REVEL (https://sites.google.com/site/revelgenomics/) (Ghosh, Oak, & Plon, 2017; Ioannidis et al., 2016), GERP++ (http://mendel.stanford.edu/SidowLab/downloads/gerp/) (Davydov et al., 2010), or dbscSNV (https://sites.google.com/site/jpopgen/dbNSFP) (Ghosh et al., 2017; Jian, Boerwinkle, & Liu, 2014) were used. In addition, the ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) for locus specific databases and the Pfam (https://pfam.xfam.org/) and the InterPro (https://www.ebi.ac.uk/interpro/) for functional domain databases were used.

Carrier frequency (CF) and predicted genetic prevalence analysis (pGP)

For CF and pGP analysis, only heterozygous PLPV (not homozygous PLPV) was considered (Hanany et al., 2018; Hanany, Rivolta, & Sharon, 2020). Therefore, the allele frequency of heterozygous PLPV (AF_V) and CF_V for a variant V were calculated as follows:

$$AF_V = \frac{allele\ count - 2*\text{homozygous count}}{\text{allele number}}$$

$$CF_V = \frac{AF_V * allele\ number}{\text{Number of individuals}} = 2AF_V$$

Where the allele count (number of variant alleles), allele number (number of genotyped alleles = 2 * number of individuals) and homozygous count (number of homozygous individuals) for a variant was provided by gnomAD.

For the CF and pGP in a gene level (CF_G and pGP_G, respectively), two methods were applied. The first method (method 1) followed CF_G and pGP_Gcalculations as previously described (Hanany et al., 2020) as follows:

$$CF_G = 1 - \prod_{k=1}^{n} (1 - CF_V)$$

$$\text{pGP}_G = \frac{\sum_{k=1}^{n} (CF_V)_{\text{ik}} (CF_V)_{\text{ik}}}{4}$$

The second method (method 2) was based on the Hardy–Weinberg equation.

The CF_G was calculated as follows:

$$CF_G = \sum_{k=1}^{n} \text{CF}_V(k)$$

Using the Hardy–Weinberg equation ($p^2 + 2pq + q^2 = 1$), the predicted genetic prevalence of congenital hypothyroidism (q^2) as pGP_G was calculated as follows:

$$CF_G = 2pq = 2\,(1-q)\,q$$

$$q = \frac{-(-2) - \sqrt{(-2)^2 - 4(2 * CF_G)}}{2 * 2}$$

$$q^2 = \ pGP_G = \left(\frac{2 - \sqrt{4 - (8 * CF_G)}}{4}\right)^2$$

Comparison analysis between method 1 and method 2 was performed using linear regression (IBM SPSS statistics version 25, Chicago, IL, USA).

The total pGP, in terms of the proportion of individuals in each ethnic group predicted to be affected by the PLPV in all candidate 10 genes, was calculated as the sum of each pGP_G.

Logistic regression analysis

To analyze which factors influence submission of the PLPV to ClinVar, multivariate logistic regression analysis was performed (IBM SPSS statistics v.25, USA). Among the stepwise methods, the forward likelihood ratio (LR) was selected. The independent variables (covariates) were number of allele count, the ethnic population and the genes. $P < 0.05$ indicated statistical significance.

Results

Presumed PLPVs in 10 candidate genes

The presumed PLPVs in the 10 candidate genes are described in Tables S1–S9. A total 557 variants were classified into PLPVs: 22 variants in *SLC5A5* , 76 variants in *TPO* , 178 variants in *TG* , 20 variants in *IYD* , 36 variants in *DUOXA2* , 138 variants in *DUOX2* , 55 variants in *TSHR* , 8 variants in *TSHB* , and 24 variants in *TRHR* (Table 1). Of the 557 variants, only 71 variants (12.6%, 95% CI 9.9%–15.9%) were registered in the ClinVar database: 18.2% (95% CI 4.9%–46.6%) in *SLC5A5* , 15.8% (95% CI 8.2%–27.6%) in *TPO* , 5.6% (95% CI 2.7%–10.3%) in *TG* , 20.0% (95% CI 5.5%–51.2%) in *IYD* , 8.3% (95% CI 1.7%–24.4%) in *DUOXA2* , 15.9% (95% CI 10%–24.1%) in *DUOX2* , 16.4% (95% CI 7.5%–31.1%) in *TSHR* , 50% in *TSHB* , 8.3% (95% CI 1.0%–30.1%) in *TRHR* . These 557 variants included 191 frameshift variants (34.3%, 95% CI 29.6%–39.5%), 188 nonsense variants (33.8%, 95% CI 29.1%–38.9%), 141 splice variants (25.3%, 95% CI 21.3%–29.9%), 29 missense variants (5.2%, 95% CI 3.5%–7.5%), 6 initiation codon variations (1.1%, 95% CI 0.4%–2.3%), and 2 in-frame deletion variants (0.4%, 95% CI 0.04%–1.3%) (Tables S1–S9).

Contributing factors for submission of presumed PLPVs to ClinVar

To assess the factors that influence submission of the presumed PLPV to ClinVar, multivariate logistic regression analysis was performed. The following variables correlated with submission of presumed PLPV to ClinVar: higher allele counts (X1), whether presumed PLPVs were detected in the Non-Finnish European population (X2), whether these PLPVs in *TSHB* (X3) were present, and whether these PLPVs in *TG* (X4) were absent (Table 2). The equation for this model is:

$\mathrm{Ln}(\frac{P}{1-P})$ = -3.461 + 0.129 * X1 + 1.331 * X2 + 1.934 * X3 – 0.882 * X4

Comparison of the two methods for analysis of carrier frequency and predicted genetic prevalence

The carrier frequency for a variant (CF_V) was calculated based on the allele frequency of each heterozygous PLPV except homozygous carriers. The CF_G and pGP_Gwere analyzed using two different methods as

described in the methods section for every gene in each ethnic group, and the $CF_G$and pGP_G values calculated by the two methods were compared. Notably, linear regression analysis showed a perfect fit: R^2=1 (y = 0.00001814 + 0.990x) for CF_G(Figure 1A) and R^2=1 (y = 0.0000001347 + 0.961x) for pGP_G (Figure 1B).

Distribution of carrier frequency and predicted genetic prevalence in each ethnic group

The total CF for all 10 candidate genes in the overall population was 2.9% (Figure 2A). This means that unaffected carriers among 10 candidate genes are predicted to be 2.9%. Among the 8 ethnic groups, the East Asian population showed the highest total CF (6.2%), followed by other (3.2%), Finnish (3.2%), Non-Finnish European (2.8%), African (2.4%), Latino (2.1%), South Asian (1.8%), and Ashkenazi Jewish (1.4%). In the Ashkenazi Jewish group, any presumed PLPV in *IYD*and all 4 genes (*TSHR* , *TSHB* , *TRHR, FOXE1*) associated with thyroid dysgenesis were not found. In the overall population, of the 10 candidate genes, *DUOX2* showed the highest CF (1.46%), followed by *TG* (0.47%), *TPO* (0.42%),*TSHR* (0.16%), *DUOXA2* (0.15%), *IYD* (0.08%),*TRHR* (0.07%), *SLC5A5* (0.07%), *TSHB* (0.04%), and*FOXE1* (0%). Interestingly, the distribution of the proportion of genes in the East Asian group was unique: *DUOX2* (3.96%) > *DUOXA2* (1.06%) > *TPO* (0.56%) > *TG* (0.50%) > *SLC5A5* (0.16%) > *TSHR* (0.04%) > *TRHR* (0.03%) > *TSHB* (0.01%) > *IYD* (0.01%).

Overall, the pGP caused by 10 candidate genes in the total population was 6.57 individuals per 100.000 births (1:15,216) (Figure 2B). The pGP of the East Asian group was 44.65 per 100,000 births (1:2,240), followed by Finnish (14.21 per 100,000 births, 1:7039), other (5.99 per 100,000 births, 1:16,683), Non-Finnish European (5.96 per 100,000 births, 1:16,762), African (3.80 per 100,000 births, 1:26,283), South Asian (3.07 per 100,000 births, 1:32,575), Latino (2.99 per 100,000 births, 1:33,453), and Ashkenazi Jewish (1.52 per 100,000 births, 1:65,972).

Discussion

Congenital hypothyroidism is the most common neonatal disorder (Feuchtbaum, Carter, Dowray, Currier, & Lorey, 2012). Prompt diagnosis and treatment help prevent patient intellectual disability (Cherella & Wassner, 2017). The newborn screening program for congenital hypothyroidism with detection of blood spot thyroid stimulating hormone (TSH) or thyroxine (T4) was implemented between 1970 and 1980 worldwide, especially in developed countries. This public health program has nearly eradicated the profound physical and cognitive impairments due to severe congenital hypothyroidism. Recent studies raised an issue that current screening criteria miss borderline or subclinical congenital hypothyroidism (Lain et al., 2016; Leonardi et al., 2008).

Primary congenital hypothyroidism is broadly caused by thyroid dysgenesis (including agenesis, hypoplasia, or abnormal location) or dyshormogenesis (when a normal thyroid gland produces abnormal amounts of thyroid hormone). Historically, the most common cause (approximately 85%) of primary hypothyroidism is thyroid dysgenesis (Cherella & Wassner, 2017; Deladoey, Ruel, Giguere, & Van Vliet, 2011; Olivieri, Fazzini, Medda, & Italian Study Group for Congenital, 2015; Wassner & Brown, 2015), with an incidence of about 1:4,000 births. However, thyroid dysgenesis occurs sporadically, and fewer than 5% of thyroid dysgenesis cases are attributable to genetic variations in the known genes. Dyshormogenesis accounts for approximately 15% of primary hypothyroidism and is mainly caused by a genetic defect. The proportion of dyshormogenesis cases within congenital hypothyroidism has been increasing up to over 30% (Cherella & Wassner, 2017; Wassner & Brown, 2015).

In this study, the carrier frequency and predicted genetic prevalence of primary congenital hypothyroidism were analyzed based on the general population database. Most of the general population is regarded to include individuals without severe hypothyroid conditions; therefore, only genes associated with primary congenital hypothyroidism (excluding central congenital hypothyroidism) inherited in an autosomal recessive pattern were included. To date, there are 6 genes (*SLC5A5* ,*TPO* , *TG* , *IYD* , *DUOXA2* , *DUOX2*) associated with thyroid dyshormonogenesis and 8 genes (*TSHR, PAX8, TSHB, NKX2-5, THRA, TRHR, TBL1X, IRS4*) associated with nongoitrous congenital hypothyroidism in the Online Mendelian Inheritance

in Man database. All thyroid dyshormonogenesis genes are inherited in an autosomal recessive pattern, while three nongoitrous congenital hypothyroidism genes (*TSHR, TSHB, TRHR*) are inherited in an autosomal recessive pattern, three genes (*PAX8, NKX2-5, THRA*) are autosomal dominant, and two (*TBL1X, IRS4*) are X-linked. Other genes associated with thyroid dysgenesis include *NKX2-1, CDCA8, JAG1,* and *NTN1* and their related conditions are autosomal dominant (Peters, van Trotsenburg, & Schoenmakers, 2018). Other genes,*FOXE1* or *GLIS3* , are inherited in an autosomal recessive manner. Even though a higher proportion of genes associated with dyshormonogenesis compared with thyroid dysgenesis genes were included in this study, the CF of all genes associated with dyshormonogenesis was approximately 90%.

Differences in the prevalence of congenital hypothyroidism by ethnicity have been reported (Feuchtbaum et al., 2012; Stoppa-Vaucher, Van Vliet, & Deladoey, 2011). The Asian and Latino (Hispanic) groups showed higher rates while the African population had a lower rate compared with the prevalence of congenital hypothyroidism in the European group. In this study, the pGP of congenital hypothyroidism in East Asians (1:2,240) was notably higher than other populations and consistent with the prevalence of congenital hypothyroidism calculated based on number of patients in Asians (1:918–1:4,464). However, in contrast to the previous studies, the Latino population in this study showed the lowest rate of pGP for congenital hypothyroidism among all populations except Ashkenazi Jewish. In addition, there was a difference between the pGP and real prevalence of congenital hypothyroidism in other populations except the East Asian group.

The difference between the pGP based on the population database and the real prevalence might be determined by how many genes following autosomal recessive inheritance patterns were associated with their diseases by ethnic group, because the pGP in this study was calculated not considering autosomal dominant inheritance: a larger proportion of genes that follow an autosomal recessive inheritance pattern within the entire genetic portion, the gap between the pGP and real prevalence is narrowing. There are differences between the proportions of thyroid dysgenesis and dyshormonogenesis between ethnic groups (Stoppa-Vaucher et al., 2011; Sun et al., 2018). Since all pathogenic variations associated with dyshormonogenesis are inherited in an autosomal recessive manner, if the proportion of dyshormonogenesis is higher in the specific population, the pGP would be more consistent with the real prevalence. Recent studies using NGS have reported that more than 50% of congenital hypothyroidism in East Asians was caused by thyroid dyshormonogenesis (Park et al., 2016; Sun et al., 2018; Yu et al., 2018). These results may indicate why the pGP of the East Asian group in this study was consistent with the real prevalence. Interestingly, the pGP of the sum of *DUOX* and *DUOXA2* was 43.2 per 100,000 births (1:2317) and 96.7% of total pGP in the East Asian group. In contrast, if the proportion of dyshormonogenesis in a population is lower, the difference between pGP and real prevalence would be bigger because the genetic cause from thyroid dysgenesis would be underestimated; many of the thyroid dysgenesis genes are inherited in an autosomal dominant manner.

Most of the studies on the genetic epidemiology of congenital hypothyroidism were based on European populations. Generally, if the specific variant is submitted to ClinVar as PLPV, it means that clinical patients with those PLPVs are present as well as those PLPVs are the main cause of their disease development. In this study, the results showed that whether the presumed PLPV was detected in a European group (specifically Non-Finnish European) significantly affected the submission to ClinVar. Therefore, the CF and pGP might be underestimated in the population that showed the biggest difference between the pGP and real prevalence, because many variants would be classified as variants of uncertain significance and not as PLPVs due to insufficiency of genetic and clinical information.

Additionally, unknown genetic factors (including causative variants in unknown genes or unrecognized variants in known genes) and epidemiologic or environmental factors (Hinton et al., 2010; Medda et al., 2005) also are attributable to the difference between pGP and prevalence.

In conclusion, this is the first study that assessed congenital hypothyroidism based on general population data and estimated CF and pGP by ethnicity. The feasibility of genetic screening for congenital hypothyroidism may be determined by ethnicity. In particular, in comparing the pGP with the real prevalence of congenital hypothyroidism, genetic screening in East Asian populations may be feasible in the future; this may be caused by a higher proportion of thyroid dyshormonogenesis due the contribution of*DUOX2* and *DUOXA2*

compared with other populations. The approach to obtain genomic information of a general population would allow an additional and helpful direction for preventive medicine. However, when using genomic information from the general population, the pathogenesis of particular diseases should be considered by ethnic group.

Acknowledgments

The author thanks Professor Jong-Won Kim at the Department of Laboratory Medicine and Genetics, Samsung Medical Center, Sungkyunkwan University School of Medicine for his valuable comments on this research.

Data Availability Statement

The data that supports the findings of this study are available in the supplementary material of this article.

Web resources

gnomAD, https://gnomad.broadinstitute.org/

HGVS nomenclature, https://mutalyzer.nl/

LOFTEE, https://github.com/konradjk/loftee

MANE, https://www.ncbi.nlm.nih.gov/refseq/MANE/

REVEL, https://sites.google.com/site/revelgenomics/)

GERP++, http://mendel.stanford.edu/SidowLab/downloads/gerp/

dbscSNV, https://sites.google.com/site/jpopgen/dbNSFP

ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/

Pfam, https://pfam.xfam.org/

InterPro, https://www.ebi.ac.uk/interpro/

References

Andermann, A., & Blancquaert, I. (2010). Genetic screening: A primer for primary care. *Can Fam Physician, 56* (4), 333-339.

Buyukgebiz, A. (2013). Newborn screening for congenital hypothyroidism.*J Clin Res Pediatr Endocrinol, 5 Suppl 1* , 8-12. doi:10.4274/jcrpe.845

Cherella, C. E., & Wassner, A. J. (2017). Congenital hypothyroidism: insights into pathogenesis and treatment. *Int J Pediatr Endocrinol, 2017* , 11. doi:10.1186/s13633-017-0051-0

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol, 6* (12), e1001025. doi:10.1371/journal.pcbi.1001025

Deladoey, J., Ruel, J., Giguere, Y., & Van Vliet, G. (2011). Is the incidence of congenital hypothyroidism really increasing? A 20-year retrospective population-based study in Quebec. *J Clin Endocrinol Metab, 96* (8), 2422-2429. doi:10.1210/jc.2011-1073

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., . . . Taschner, P. E. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat, 37* (6), 564-569. doi:10.1002/humu.22981

Feuchtbaum, L., Carter, J., Dowray, S., Currier, R. J., & Lorey, F. (2012). Birth prevalence of disorders detectable through newborn screening by race/ethnicity. *Genet Med, 14* (11), 937-945. doi:10.1038/gim.2012.76

Ghosh, R., Oak, N., & Plon, S. E. (2017). Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biol, 18* (1), 225. doi:10.1186/s13059-017-1353-5

Hanany, M., Allon, G., Kimchi, A., Blumenfeld, A., Newman, H., Pras, E., . . . Sharon, D. (2018). Carrier frequency analysis of mutations causing autosomal-recessive-inherited retinal diseases in the Israeli population. *Eur J Hum Genet, 26* (8), 1159-1166. doi:10.1038/s41431-018-0152-0

Hanany, M., Rivolta, C., & Sharon, D. (2020). Worldwide carrier frequency and genetic prevalence of autosomal recessive inherited retinal diseases. *Proc Natl Acad Sci U S A, 117* (5), 2710-2716. doi:10.1073/pnas.1913179117

Hinton, C. F., Harris, K. B., Borgfeld, L., Drummond-Borg, M., Eaton, R., Lorey, F., . . . Pass, K. A. (2010). Trends in incidence rates of congenital hypothyroidism related to select demographic factors: data from the United States, California, Massachusetts, New York, and Texas.*Pediatrics, 125 Suppl 2* , S37-47. doi:10.1542/peds.2009-1975D

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., . . . Sieh, W. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet, 99* (4), 877-885. doi:10.1016/j.ajhg.2016.08.016

Jian, X., Boerwinkle, E., & Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome.*Nucleic Acids Res, 42* (22), 13534-13544. doi:10.1093/nar/gku1206

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., . . . MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature, 581* (7809), 434-443. doi:10.1038/s41586-020-2308-7

Lain, S. J., Bentley, J. P., Wiley, V., Roberts, C. L., Jack, M., Wilcken, B., & Nassar, N. (2016). Association between borderline neonatal thyroid-stimulating hormone concentrations and educational and developmental outcomes: a population-based record-linkage study.*Lancet Diabetes Endocrinol, 4* (9), 756-765. doi:10.1016/S2213-8587(16)30122-X

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536* (7616), 285-291. doi:10.1038/nature19057

Leonardi, D., Polizzotti, N., Carta, A., Gelsomino, R., Sava, L., Vigneri, R., & Calaciura, F. (2008). Longitudinal study of thyroid function in children with mild hyperthyrotropinemia at neonatal screening for congenital hypothyroidism. *J Clin Endocrinol Metab, 93* (7), 2679-2685. doi:10.1210/jc.2007-2612

Medda, E., Olivieri, A., Stazi, M. A., Grandolfo, M. E., Fazzini, C., Baserga, M., . . . Sorcini, M. (2005). Risk factors for congenital hypothyroidism: results of a population case-control study (1997-2003).*Eur J Endocrinol, 153* (6), 765-773. doi:10.1530/eje.1.02048

Olivieri, A., Fazzini, C., Medda, E., & Italian Study Group for Congenital, H. (2015). Multiple factors influencing the incidence of congenital hypothyroidism detected by neonatal screening. *Horm Res Paediatr, 83* (2), 86-93. doi:10.1159/000369394

Park, K. J., Park, H. K., Kim, Y. J., Lee, K. R., Park, J. H., Park, J. H., . . . Kim, J. W. (2016). DUOX2 Mutations Are Frequently Associated With Congenital Hypothyroidism in the Korean Population. *Ann Lab Med, 36* (2), 145-153. doi:10.3343/alm.2016.36.2.145

Peters, C., van Trotsenburg, A. S. P., & Schoenmakers, N. (2018). DIAGNOSIS OF ENDOCRINE DISEASE: Congenital hypothyroidism: update and perspectives. *Eur J Endocrinol, 179* (6), R297-R317. doi:10.1530/EJE-18-0383

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of

the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med, 17* (5), 405-424. doi:10.1038/gim.2015.30

Stoppa-Vaucher, S., Van Vliet, G., & Deladoey, J. (2011). Variation by ethnicity in the prevalence of congenital hypothyroidism due to thyroid dysgenesis. *Thyroid, 21* (1), 13-18. doi:10.1089/thy.2010.0205

Sun, F., Zhang, J. X., Yang, C. Y., Gao, G. Q., Zhu, W. B., Han, B., . . . Song, H. D. (2018). The genetic characteristics of congenital hypothyroidism in China by comprehensive screening of 21 candidate genes. *Eur J Endocrinol, 178* (6), 623-633. doi:10.1530/EJE-17-1017

Wassner, A. J., & Brown, R. S. (2015). Congenital hypothyroidism: recent advances. *Curr Opin Endocrinol Diabetes Obes, 22* (5), 407-412. doi:10.1097/MED.0000000000000181

Yu, B., Long, W., Yang, Y., Wang, Y., Jiang, L., Cai, Z., & Wang, H. (2018). Newborn Screening and Molecular Profile of Congenital Hypothyroidism in a Chinese Population. *Front Genet, 9* , 509. doi:10.3389/fgene.2018.00509

Hosted file

`Congenital hypothyroidism_Tables.docx` available at https://authorea.com/users/351676/articles/476211-analysis-of-worldwide-carrier-frequency-and-predicted-genetic-prevalence-of-congenital-hypothyroidism-based-on-a-general-population-database

`figures/Figure-1/Figure-1-eps-converted-to.pdf`