Title: Comprehensive screening of eight known causative genes in congenital hypothyroidism with gland-in-situ.

Authors: AK Nicholas*[1], EG Serra*[2], H Cangul[3], S Alyaarubi[4], I Ullah[4], E Schoenmakers[1], A Deeb[5], AM Habeb[6], M AlMaghamsi[7], C Peters[8], N Nathwani[9], Z Aycan[10], H Saglam[11], E Bober[12], M Dattani[13], S Shenoy[14], PG Murray[15], A Babiker[16], R Willemsen[17], A Thankamony[17], G Lyons[1], R Irwin[18], R Padidela[19], K Tharian[20], JH Davies[21], V Puthi[22], S-M Park[23], AF Massoud[24], JW Gregory[25], A Albanese[26], E Pease-Gevers[27], H Martin[28], K Brugger[28], ER Maher[28], K Chatterjee[1], CA Anderson[2], N Schoenmakers[1]

*These authors contributed equally

1. University of Cambridge Metabolic Research Laboratories, Wellcome Trust-Medical Research Council Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK

2. Department of Human Genetics, The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

3. Research Centre for Regenerative and Restorative Medicine Department of Medical Genetics Istanbul Medipol University, Kavacık, Istanbul, Turkey

4. Pediatric Endocrine Unit, Department of Child Health, Sultan Qaboos University Hospital, Muscat, Oman

5. Paediatric Endocrinology Department, Mafraq Hospital, AbuDhabi, United Arab Emirates.

6. Pediatric Department Prince Mohamed Bin Abdulaziz Hospital, Madinah, KSA

7. Department of Paediatrics, Madina Maternity & Children's Hospital Madina Munawara, Saudi Arabia

8. Department of Endocrinology, Great Ormond Street Hospital for Children, London UK

9. Department of Paediatrics, Luton and Dunstable University Hospital, Luton, UK

10. Division of Paediatric Endocrinology Dr Sami Ulus Woman Health and Children Research Hospital Ankara, Turkey

11. Department of Paediatric Endocrinology Uludağ University, School of Medicine Bursa, Turkey

12. Department of Paediatric Endocrinology Dokuz Eylül University, Faculty of Medicine Izmir, Turkey

13. Developmental Endocrinology Research Group, Section of Genetics and Epigenetics in Health and Disease, Genetics and Genomic Medicine Programme, University College London Institute of Child Health, London UK

14. Department of Paediatrics, Leicester Royal infirmary, Leicester UK

15. Centre for Paediatrics and Child Health, Institute of Human Development University of Manchester, and Royal Manchester Children's Hospital, Manchester, UK

16. Paediatric Endocrinology Fellowship King Saud University and King Saud University Medical City, Riyadh, Saudi Arabia

17. Department of Paediatrics, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK

18. West Midlands Regional Genetics Laboratory, Birmingham Women's Hospital NHS Foundation Trust, Birmingham UK

19. Department of Paediatric Endocrinology, Central Manchester University Hospitals NHS Foundation Trust, Manchester

20. Department of Paediatrics, Diana Princess Of Wales Hospital, Grimsby, UK

21. Department of Paediatric Endocrinology, University Hospital Southampton, Southampton, UK

22. Department of Paediatrics, Peterborough and Stamford Hospitals NHS Foundation Trust, Peterborough, UK

23. Department of Clinical Genetics, Cambridge University Hospitals NHS Foundation Trust, Cambridge UK

24. London North West Healthcare NHS Trust, Harrow, Middlesex, UK

25. Division of Population Medicine, School of Medicine, Cardiff University, Heath Park Cardiff, UK

26. Department of Paediatric Endocrinology, St George's University Hospitals NHS Foundation Trust, London, UK

27. Centre for Endocrinology, William Harvey Research Institute, Queen Mary University London and Children's Hospital, Barts Health NHS Trust, London, UK

28. Department of Medical Genetics, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, United Kingdom

Abbreviated title: Genetics of Congenital Hypothyroidism with GIS

Keywords: Congenital hypothyroidism, Dyshormonogenesis, genetics, gland-in-situ

Word count Abstract 250; Text 4309

Number of Figures and Tables: 5 main figures, 2 supplemental figures, 3 supplemental tables

Corresponding author and person to whom reprint requests should be addressed: Dr N Schoenmakers, University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Level 4, Box 289, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ, E-mail: **naaa2@cam.ac.uk**

Disclosures: the authors have nothing to disclose

Funding Information and Acknowledgements: This study made use of data generated by the UK10K Project and we acknowledge the contribution of the UK10K Consortium. This work was supported by Wellcome Trust Grants 100585/Z/12/Z (to N.S.), and 095564/Z/11/Z (to V.K.C.) and the National Institute for Health Research Cambridge Biomedical Research Center (to V.K.C., N.S.). E.G.S and C.A.A. are supported by the Wellcome Trust (098051). Funding for the UK10K Project was provided by the Wellcome Trust under award WT091310

Abstract

Context: Lower thyroid-stimulating hormone (TSH) screening cut-offs have doubled the ascertainment of congenital hypothyroidism (CH), particularly cases with a eutopically-located gland-in-situ (GIS). Although mutations in known dyshormonogenesis genes, or the thyroid-stimulating hormone receptor (TSHR) underlie some cases of CH with GIS, systematic screening of these eight genes has not previously been undertaken.

Objective: To evaluate the contribution and molecular spectrum of mutations in eight known causative genes (*TG*, *TPO*, *DUOX2*, *DUOXA2*, *SLC5A5*, *SLC26A4*, *IYD* and *TSHR*) in CH cases with GIS.

Patients, Design and Setting: We screened forty-nine CH cases with GIS from thirty-four ethnically diverse families, using next-generation sequencing. Pathogenicity of novel mutations was assessed in silico.

Results: Twenty-nine cases harbored likely disease-causing mutations. Monogenic defects (nineteen cases) most commonly involved *TG* (twelve), *TPO* (four), *DUOX2* (two) and *TSHR* (one case). Ten cases harboured triallelic (digenic) mutations: *TG* and *TPO* (one); *SLC26A4* and *TPO* (three) and *DUOX2* and *TG* (six cases). Novel variants overall included fifteen *TG*, six *TPO*, and three *DUOX2* mutations. Genetic basis was not ascertained in twenty patients, including fourteen familial cases.

Conclusions: The aetiology of CH with GIS remains elusive, with only 59% attributable to mutations in TSHR or known dyshormonogenesis-associated genes in a cohort enriched for familial cases. Biallelic *TG* or *TPO* mutations most commonly underlie severe CH. Triallelic defects are frequent, mandating future segregation studies in larger kindreds to assess their contribution to variable phenotype. A high proportion (~41%) of unsolved or ambiguous cases suggests novel genetic aetiologies that remain to be elucidated.

Introduction

Congenital hypothyroidism (CH) is the most common neonatal endocrine disorder, and historically thyroid dysgenesis was thought to account for approximately 80% of cases (1). However, recent studies have reported a change in the epidemiology of CH, with a doubling in incidence to around 1 in 1500 live newborns, predominantly driven by an increase in CH with eutopic GIS, which accounted for almost two-thirds of recently diagnosed cases in Lombardy, Italy (2). Lower TSH screening cut-offs may be the major driver for this increase in diagnosis although altered ethnicities of the screened population, increased multiple and premature births, iodine status and hitherto uncharacterized factors may also contribute (3, 4).

The molecular basis of CH with GIS remains poorly understood (5, 6). Genetic variation in seven genes involved in thyroid hormone biosynthesis (*TG, TPO, DUOX2, DUOXA2, IYD, SLC5A5* and *SLC26A4*) and *TSHR* mediates some cases. Disease-causing mutations are usually biallelic, with the exception of monoallelic *DUOX2, IYD* and *TSHR* mutations, which may also confer a phenotype (1). Phenotypic heterogeneity in cases harbouring similar causative mutations suggests that mono and polygenic factors and environmental modulators may also play a role in determining disease severity (7, 8).

Genetic characterization of CH with GIS has been limited by the cost and labour implications of Sanger sequencing multiple exons. Previous studies have generally focused on either a small number of genes (e.g. *TG, TPO, TSHR* and *DUOX2* in 43 Korean cases) (6), specific phenotypic subsets of cases (5, 8) or multiple genes in a small subset of patients (9). There are occasional reports of digenic mutations involving *TSHR* and either *DUOX2* (6, 10, 11) or *TPO* (12), or combined *DUOX2* and *DUOXA2* mutations (13). However, the role of oligogenicity in disease development and penetrance remains unclear, with no evidence for an additive effect of digenic mutations in one large published kindred (12).

Next-generation sequencing (NGS) technologies increase sequencing capacity and speed, enabling efficient screening of multiple genes simultaneously. A recent publication describes large-scale multiplexed genetic screening of *TPO, TSHR, DUOX2, DUOXA2, PAX8* and *SLC5A5* in 170 Korean CH cases. However, cases were from a single ethnic background and not selected on the basis of thyroid morphology; moreover *TG, IYD* and *SLC26A4* were not sequenced (11). We undertook comprehensive screening of *TG, TPO, DUOX2, DUOXA2, IYD, SLC5A5, SLC26A4* and *TSHR* in an ethnically and biochemically heterogeneous CH cohort with GIS. As well as reporting known and novel mutations in these genes, we document the frequent occurrence of potential oligogenicity, with triallelic variation in two candidate genes, in a population enriched for familial and consanguineous cases.

Patients and Methods

Patients

All investigations were part of an ethically approved protocol and/or clinically indicated, being undertaken with written informed consent from patients and/or next of kin including specific consent for WES (MREC 98/5/024). Forty-nine cases were included in the study, from thirty-four families referred from centers in the UK, Oman, Saudi Arabia, UAE and Turkey. Inclusion required clinical evidence of goitre or radiological evidence of a normally-sited thyroid gland in the proband. In five cases without goiter who had not undergone thyroid imaging at diagnosis, we accepted goiter or radiological evidence of GIS in at least one affected family member with CH, assuming a common underlying genetic aetiology. A diagnosis of overt or subclinical primary congenital hypothyroidism was made on the basis of referral through newborn screening and/or a raised venous TSH. Newborn screening blood spot cut offs were as follows: 6-10mU/L (UK), 10mU/L (UAE) or cord blood TSH 40mU/L (Oman). Childhood TSH normal range was 0.35-5.5mU/L. Thyroid biochemistry was measured using local analyzers in the referring hospitals.

DNA sequencing

Three different NGS-based strategies (whole-exome sequencing, WES, and two different targeted sequencing protocols) were used to screen Thyroglobulin (*TG),* Thyroid peroxidase (*TPO*), Thyroid stimulating hormone receptor (*TSHR*), Dual oxidase 2, (*DUOX2*), Dual oxidase maturation factor 2 (*DUOXA2*), Iodotyrosine deiodinase (*IYD*), solute carrier family 5, member 5, NIS (*SLC5A5*) and Solute carrier family 26 member 4, Pendrin (*SLC26A4*). Detailed methods, coverage and quality control data is available in the Supplementary methods and results. We sought to identify rare variants (MAF <0.02 in all control databases) with likely pathogenic consequences predicted by in silico algorithms. Given the ethnic heterogeneity of our cohort we selected the maximum number of control exomes (n~80, 000) matched as closely for ethnicity as we could achieve (Supplementary methods). All positive results were validated by Sanger sequencing.

Nomenclature

Variants were described using nomenclature approved by the Human Genome Variation Society (HGVS; www.hgvs.org/mutnomen). Further details are available in the Supplementary methods.

Structural model for TPO & DUOX2

The models for TPO and DUOX2 were generated using the phyre2 (Protein Homology/analogy Recognition Engine 2) web portal which predicts and analyses protein structures based on homology/analogy recognition to solved protein crystal structures (14). The figures were generated with MacPyMOL Molecular Graphics System, Schrödinger, LLC.

Results

Sequencing data quality

Detailed information regarding individual gene coverage is summarized in the Supplementary section. In the samples sequenced by WES or HiSeq targeted sequencing panel, optimal median coverage (>30 fold) was achieved for all genes except *DUOXA2* and *SLC5A5* in the eleven samples screened by targeted sequencing (median coverage 5-fold and 24-fold respectively) (Supplementary Figure 1A, B). Exons screened using the MiSeq targeted sequencing panel either achieved >20-fold coverage (in

house validation had demonstrated 100% sensitivity for detecting variants at this sequencing depth), or were repeated by Sanger sequencing, such that this approach was expected to be highly sensitive. In the WES and HiSeq protocols, in common with previous studies employing similar techniques, although median coverage was generally acceptable, coverage was non-uniform across individual genes (Supplementary Figure 2). This was most marked with the HiSeq targeted sequencing panel in which specific exons exhibited <10-fold coverage including *DUOXA2* (exons 1, 2, 4, 5 and 6), *SLC5A*5 (exons 1-3, 5, 6, 11, 12 and 15), *DUOX2* (exons 2, 5, 6, 8, 15 and 34), *TG* (exons 13, 15, and 16), *TPO* (exons 3, 7, 8, and 16), *SLC26A4* (exon 21) and *IYD* (exon 6). A detailed comparison of the sequencing techniques is provided in Supplementary Figure 2.

Mutation frequencies (Figure 1)

Forty-nine cases from thirty-four families of European, Asian, Middle Eastern and Afro-Caribbean origin were investigated and twenty-nine cases (twenty families, 59%) were considered 'solved' following identification of a decisive link between genotype and phenotype. In eleven 'ambiguous' cases (22%) it was felt that the ascertained genotype could plausibly be contributing to the phenotype, but the evidence to support a causal link was weaker than in the 'solved' group. Finally, nine cases were considered 'unsolved' as they carried no mutations in any of the listed genes. Detailed genetic and phenotype data is supplied in Supplementary Tables 1, 2 and 3.

CH was more severe biochemically in solved cases than in unsolved or ambiguous cases (mean TSH 100mU/L vs 36mU/L at diagnosis, p=0.02, Welch's t-test) and solved cases were more frequently from consanguineous backgrounds (69% cases vs. 40% cases). This likely reflects the increased incidence of recessive disease in the presence of consanguinity, since CH- associated mutations in five of the eight targeted genes (*TG, TPO, DUOXA2, SLC5A5* and *SLC26A4*) are usually biallelic. Cases with affected siblings were common in both solved and unsolved or ambiguous categories (79% vs. 70% cases) (Figure 1, Supplementary Tables 2, 3).

'Solved' kindreds harbouring mutations in one gene (monogenic kindreds)

Nineteen cases had a monogenic basis of disease, most commonly involving biallelic mutations in *TG* (twelve cases), followed by *TPO* (four cases), *DUOX2* (one monoallelic and one biallelic mutation) and *TSHR* (one case). There were no cases with CH attributable to mutations in *IYD*, *SLC5A5* or *SLC26A4* (Figure 1).

TG mutations (Figure 2)

TG is the secretory protein upon which thyroid hormone is synthesized, and the 12 cases with monogenic *TG* mutations predominantly exhibited moderate-severe CH (Figure 2). One known and three novel homozygous nonsense or frameshift mutations were identified which truncate TG before the carboxy-terminal acetyl cholinesterase (ACHE)-like domain, which has a crucial role in normal conformational maturation and intracellular trafficking of TG (F1, 2, 3, 4) (15). Two siblings (F5 a, b) were compound heterozygous for a known nonsense mutation (p.R296*) and a rare, novel missense variant, (p.C160S) which affects a highly conserved cysteine residue in TG (GERP score 5.84). Cysteine residues within repetitive domains in TG form intramolecular disulphide bonds needed for protein folding, thus p.C160S may be deleterious to TG affecting the tertiary structure as predicted by PolyPhen (16, 17, 18). Two siblings (F7a, b) harbored the same homozygous *TG* splice region variant (c.638+5 G>A) inherited from heterozygous parents; although the pathogenicity of this cannot be ascertained *in silico*, it is unique to the affected siblings, and adjacent to a known pathogenic mutation (c.638+1G>A) (19), supporting causality, albeit in association with a mild CH phenotype.

TPO mutations (Figure 3)

TPO is the heme peroxidase catalyzing the final steps of thyroid hormone synthesis and biallelic mutations (Figure 3) were identified in four monogenic kindreds. These included two known pathogenic missense mutations (F16; p.R491H, F17; p.R665Q), two novel frame shift (F20; p.C808Afs*24, F16; p.A397Pfs*76) and two novel missense variants (F18; p.R291H, p.G331V) (Table 2). The p.R291H variant is predicted to disrupt a hydrogen bond network close to the TPO heme group thereby destabilizing the catalytic domain. G331 is located close to the substrate binding domain, and mutation to the larger valine amino acid will likely cause steric hindrance impeding

substrate binding (Figure 3). Two cases were compound heterozygous: F16 p.[A397Pfs*76];[R491H], associated with dyshormonogenic goitre requiring thyroidectomy and F18 p.[R291H];[G331V], who also exhibited goitre.

DUOX2 mutations (Figure 4)

DUOX2 is the NADPH oxidase, which generates H_2O_2 required for thyroid hormone biosynthesis. Two solved cases with monogenic *DUOX2* mutations were identified (Figure 4), including one known heterozygous mutation (F23; p.F966Sfs*29) and one novel homozygous mutation (F24; p.L1028Afs*3), both of which would truncate DUOX2 before the NADPH oxidase domain, thereby abrogating protein function. Affected cases generally had a milder or transient (F23) CH phenotype compared with cases harbouring monogenic *TG* and *TPO* mutations.

TSHR mutation

A single individual from the UAE with mild CH harbored a known pathogenic heterozygous *TSHR* mutation (F26; p.P68S) (Supplementary table 2), previously identified in an Arab population. Parental DNA was not available, however, the mild CH phenotype was consistent with previously reported biochemistry associated with this mutation (20).

'Solved' kindreds harboring mutations in two genes (oligogenic kindreds, Figure 5)

Ten solved cases from seven families harbored digenic pathogenic variants. These were predominantly triallelic, and most commonly comprised biallelic *TG* mutations in association with a monoallelic *DUOX2* mutation. Such digenic mutations were detected in consanguineous Turkish kindreds F6, 8 and 9 (Figure 5). In these kindreds, although defined as variants of uncertain significance by ACMG criteria, the biallelic *TG* mutations were rare (p. W1051L; MAF <0.001 in 1KG Europeans, and absent in all other population datasets, including ExAC East Asians) or unique, affected conserved amino acids and were predicted to be pathogenic by PolyPhen and SIFT. In F6, two siblings (a, b) with CH were both homozygous for *TG* p.W1051L and p.C726Y but one sibling (F6b) harbored an additional, maternally-inherited heterozygous *DUOX2* mutation (p.Q686*),

previously described in association with transient CH (21). Biochemistry at diagnosis could not be retrieved from F6b for comparison with F6a, however both presented with neonatal goitre and had similar levothyroxine requirements. Their mother exhibited adult-onset hypothyroidism of unknown etiology. Two unrelated sibling pairs also harbored homozygous *TG* mutations in association with a heterozygous *DUOX2* mutation: *TG* p.1493Y and *DUOX2* p.Q686* in F8 a, b and *TG* p.W2685L and *DUOX2* R354W (predicted to perturb the DUOX2 peroxidase-like domain) in F9a, b (Figure 4). There was also a strong history of goitre (mother and maternal aunt) in F8 but maternal DNA was not available to confirm *DUOX2* genotype. In all three kindreds, the most severe phenotype was observed in individuals harbouring biallelic *TG* or triallelic (biallelic *TG* and monoallelic *DUOX2*) mutations, however it was impossible to distinguish the effects of the mutations in the two genes reliably in these small pedigrees with limited subphenotype data.

Since monogenic, heterozygous *DUOX2* mutations (including p.Q686*) are frequently associated with CH, we hypothesized that an additive phenotypic contribution of all three mutations was very plausible. Calculation of the number of East Asian individuals in the ExAC database (n=8,654) harburing similarly rare, predicted damaging variants in *DUOX2* yielded a population mutation frequency of 0.06%. The observed proportion of *TG* mutation carriers with a monoallelic *DUOX2* variant in our cohort (8.8% families) was therefore significantly higher (p=0.0233, Fisher's exact one-tail test), supporting a potential phenotypic contribution of the *DUOX2* mutation in these individuals. Much larger cohorts of sequenced CH individuals will be required to assess the phenotypic consequences of digenicity in CH thoroughly.

Biallelic mutations in *TPO* were identified in two kindreds in addition to heterozygous known *SLC26A4* mutations, previously associated with recessive disease: F19a: *TPO* p.R584Q (homozygous) and *SLC26A4* p.N324Y (heterozygous); F19b: *TPO* p.R584Q (homozygous) and *SLC26A4* p.I713M (heterozygous); F21: *TPO* p.[E17Dfs*77];[Y453D] (compound heterozygous) and *SLC26A4* p.E384G (heterozygous) (Figure 5). The novel *TPO* p.R584Q missense variant is predicted to perturb polar contacts possibly affecting the catalytic domain (Figure 4).

The occurrence of Pendred syndrome usually mandates biallelic *SLC26A4* mutations, and manifests universally with congenital or postnatal progressive sensorineural hearing loss, whereas thyroid dysfunction is usually mild or absent. In both these kindreds, only the biallelic *TPO* mutations segregated with CH; this was severe whereas hearing was normal. In F11, a known homozygous pathogenic *TPO* mutation (p.R491H) was inherited together with a heterozygous *TG* variant (p.Q1644E). Since biallelic inheritance is also usually required for CH due to *TG* mutations, these observations suggest the *TPO* mutations are predominant drivers of the CH phenotype in these three kindreds, although we cannot definitively exclude a contribution of the heterozygous *SLC26A4* or *TG* mutation. Comparison with population mutation frequencies in *TG* and *SLC26A4* in the ExAC cohort (non-Finnish Europeans, N=66,740), suggested that congruence of *TPO* mutations with *TG* or *SLC26A4* mutations was not increased in our cohort (p=0.2280, p=0.0951 respectively).

Detailed investigation of the contribution of oligogenicity to genotype-phenotype variability mandates the study of large kindreds with a spectrum of genotypes, e.g. F10 (Figure 5). In this large, consanguineous Pakistani kindred the proband harbours a known pathogenic *DUOX2* mutation (p.Q570L, previously published in 8). Homozygosity for this mutation segregates with permanent CH (F10a), whereas *DUOX2* p.Q570L heterozygotes exhibit either euthyroidism or transient CH. Two novel, rare *TG* variants (p.L2547Q, predicted to be pathogenic by PolyPhen and SIFT, and p.R1691C, of less certain significance) were also identified in this kindred, yet neither of these variants segregated with transient CH in the DUOX2 p.Q570L heterozygotes, suggesting digenic mutations in the genes screened did not explain the phenotypic variability associated with this genotype.

Unsolved or ambiguous kindreds (Figure 1, Supplementary Table 3)

This group included two cases harboring heterozygous pathogenic *TG* variants; a novel nonsense mutation in F13 (p.Q771*) and a previously described missense mutation in F12 (p.Q870H). An additional case was heterozygous for a frameshift mutation in *TPO* (p.E510Afs*14, F22). Previous reports of CH due to *TG* and *TPO* mutations most commonly involve biallelic mutations, therefore it is unclear whether the mild or subclinical hypothyroidism was attributable to the monoallelic

mutation or whether they harbored a second 'hit' not detected by our sequencing methods. Other cases in this category harbored novel heterozygous *TG* missense (p.Y759C, F14) or splice region (c.3433+3_3433+6delGAGT, F15) variants, a novel heterozygous *DUOX2* variant (p.R764W, F25) inherited from a healthy parent and a homozygous *DUOXA2* splice site (c.555-5G>A) variant for which *in silico* predictions were inconclusive (F27). Nine cases (seven families) remained completely unsolved with no likely disease-causing variants identified. Copy number variant (CNV) analysis was undertaken in individuals who had undergone whole exome sequencing: F13, 15, 33 (ambiguous or unsolved cases) and F3, 6-10 (solved cases), however no rare CNVs were identified that segregated with disease phenotype in each pedigree.

Discussion

In this study, next-generation sequencing technologies enabled efficient screening of eight genes associated with CH and GIS in forty-nine cases from the UK, Turkey, Middle East and Asia, and with a spectrum of biochemical phenotypes. In addition to single-gene mutations, the contribution of oligogenic variants was assessed. Previous genetic evaluations of cohorts of CH with GIS have been less comprehensive, screening fewer genes, or fewer cases with restricted ethnicities (6, 9, 22, 23). The only large-scale multiplex study in CH did not select cases on the basis of thyroid morphology and excluded *TG*, *SLC26A4* and *IYD* from its sequencing panel (11). Direct sequencing of *DUOX2*, *TG*, *TPO* and *TSHR* has been undertaken in 43 Korean CH cases with GIS (6); in common with our study, only around 50% of cases harbored causative, pathogenic variants in one or more genes.

The relative frequency of mutations in known CH causative genes depends on selection criteria and ethnic origin of the cohort (24, 6). Our cohort included individuals of diverse ethnicities, in whom the biochemical diagnosis of CH was achieved using different, country-specific, screening protocols, or following neonatal or early childhood presentation with clinical hypothyroidism. These multiple variables preclude detailed comparison of relative mutation frequencies with other studies of populations with more uniform ethnicity or biochemical diagnostic approach. The heterogeneous

population screened in this study also mandated the use of ethnically-matched controls in order to prevent 'false-positive results' due to incorrect classification of ethnically-specific SNP's as pathogenic mutations. The paucity of West Asian exomes in publically-accessible databases precluded this for 17 non-Turkish West Asian cases. However, the large number of controls used (~80,000) and the fact that 8 of the 10 solved West Asian cases harboured truncating or previously reported CH-associated mutations, made false positive results unlikely.

In our study, mutations were most frequently found in *TG*, followed by *TPO*, whereas *DUOX2* mutations were relatively infrequent compared with findings by *Jin et al* (mutations in 35% all cases), probably reflecting the higher prevalence of *DUOX2* mutations in individuals of East Asian ethnicity, who were poorly represented in our study (6, 11, 25). No definitively pathogenic mutations were found in *DUOXA2*, *IYD* or *SLC5A5*, which is in keeping with previous reports suggesting that these are rare genetic causes of dyshormonogenesis, with the exception of a recurrent *DUOXA2* mutation in Korean cases (26, 11). The paucity of *TSHR* mutations in a CH cohort with GIS is surprising; however, the high incidence of consanguinity in our cohort predicts occurrence of biallelic mutations that, in the case of *TSHR*, may cause thyroid hypoplasia, with such cases possibly being excluded from recruitment to our GIS CH cohort (6, 27). Despite unselected recruitment of either sporadic or familial cases, our cohort was greatly enriched for familial CH (76% cases), and consanguinity, which may have increased the percentage of cases harboring an underlying genetic etiology. In a standard UK clinic population with a greater proportion of sporadic, non-consanguineous cases, the proportion of mutation-negative cases could be higher.

Interpretation of novel genetic variants requires functional studies *in vitro* (or *in vivo* evidence of impaired TSH-stimulated mutant thyroglobulin production for *TG* mutations) in order to confirm pathogenicity (18). Although such analyses were not undertaken, the novel variants identified are rare, segregate with phenotype, and have strong bioinformatic or structural (TPO) predictions of pathogenicity, supporting a causal role. Moreover, the location of novel variants in *TPO* (heme-binding region or substrate-binding region) and *DUOX2* (R354W; peroxidase-like domain) mirrors

that of previously described pathogenic mutations. Analysis of novel variants in *TG* is hindered by an incomplete knowledge of its functional domains or crystal structure, but those identified affect similar regions to previously documented mutations (N-terminal cysteine-rich repetitive elements, C-terminal ACHE-like domain) also supporting causality (8, 16, 18, 28).

The associated clinical phenotypes in our mutation-positive patients were similar to published cases. *TG* mutations may result in euthyroid goitre and mild or severe hypothyroidism (18), and monoallelic and biallelic *DUOX2* mutations may both cause permanent or transient CH (8, 21, 23, 25). Even *TPO* mutations, although classically associated with total iodide organification defects, can cause milder phenotypes (28). Solved cases usually had a more severe phenotype than unsolved or ambiguous cases, however the latter group included four cases of subclinical or mild CH harbouring heterozygous mutations in *TPO* or *TG*. Such monoallelic mutations have previously been described in association with CH, but are usually assumed to coexist with an additional undetected CNV, intronic or regulatory mutation on the other chromosome (16, 24, 29). This may be the case in our patients as well; our sequencing techniques would not have detected mutations in non-coding regions of the genome and although CNVs were not detected in F15, 13 and 33, they could not be excluded in the remaining families. Our observations highlight the fact that mutations in *TPO* or *TG* may underlie subclinical hypothyroidism as well as cases with overt CH. Despite elevated TSH levels, several of our non-TSHR mutation positive cases (mainly detected in the neonatal period) did not exhibit goiter. Quantitation of thyroid volume radiologically at this age is technically challenging, such that mild thyroid enlargement may not have been detected. However, TSH-driven goitrogenesis in these cases will have been dependent on fetal TSH levels – whose role in thyroid follicular cell growth remains unclear. In common with our findings, others have demonstrated that dyshormonogenetic CH, even associated with total iodide organification defect, is not always associated with thyroid enlargement (30).

Oligogenicity has often been proposed to underlie the intrafamilial variability seen in known genetic causes of CH, especially in association with *DUOX2* mutations (8). The *Pax8/Titf1* murine model

exemplifies the role of polygenicity in thyroid dysgenesis, since only mice doubly heterozygous for the two null alleles and bred on a C57BL/6 background exhibit a phenotype (31). Despite reports of digenic GIS cases in the literature, pedigree studies have either not been performed (11, 6), or have not confirmed a genotype-phenotype correlation (12). Our study detected likely pathogenic variants in more than one CH-associated gene, especially in consanguineous kindreds, most commonly involving *TG* and *DUOX2*. It is possible that this is a conservative estimate of the frequency of oligogenicity in CH with GIS; the high percentage of consanguinity in our study facilitates identification of potentially pathogenic variants in a disease model with recessive inheritance, but also increases the likelihood of detecting variants which are contributory to the CH phenotype but not causative, due to the occurrence of genomic regions with loss of heterozygosity involving CH-associated genes. Accordingly, we cannot discount the possibility that some of our monogenic, consanguineous, 'solved' cases harbour additional mutations in genes which were not screened in our study, that could contribute to the CH phenotype. Small pedigree sizes, poor information about mutation frequencies in populations matched to our CH cases, and a paucity of subphenotype data preclude definitive statements regarding the relative aetiological contribution of digenicity in CH. Further studies with large pedigrees and clear phenotypic variability are required to ascertain the role of polygenic modulators in CH with GIS. Alternative candidate genes involved in the same biological pathways as known causative genes may be implicated, either exacerbating or playing a compensatory role in the context of loss-of-function mutations. Examples include *DUOX1, DUOXA1,* and *NOX*, which are also involved in H_2O_2 production and whose expression may be upregulated in the context of *DUOX2* deficiency (12, 32).

It is conceivable that despite adequate median coverage, non-uniform coverage of genes could have resulted in failure to detect variants. This is most likely to be significant for the eleven cases (eight families) in which coverage of specific exons was <10 fold (predominantly affecting *DUOXA2* and *SLC5A5*). Suboptimal coverage of these regions raises the possibility of a Type II error. However, undetected variants in these cases are unlikely to affect the conclusions of this study since five cases harbored mutations which explained their CH (F26, F2a, b, F11, F17), and two ambiguous cases

harbored heterozygous TG variants (F12 a, b). Additionally, although the study was not designed to allow direct comparison of different sequencing methods, the rate of causative mutations in cases screened using either the most sensitive technique (MiSeq targeted sequencing, in which exons with <20 fold coverage were individually re-sequenced using Sanger sequencing) or WES, was similar and supported our conclusion that ~ 40% cases are unsolved. Previous studies have also reported considerable variability in uniformity and depth of coverage across the exome, and this data, together with our sequencing depth analysis, highlights a limitation of targeted sequencing, which may impact and limit variant identification (33). High-depth, whole-genome sequencing can improve exon coverage and the advent of recent sequencing technologies (such as the Illumina X10 system) make this possible at large scale.

The aetiology of CH with GIS remains elusive and factors other than known dyshormonogenesis-associated genes or the *TSHR* must be implicated. CH with GIS may be transient, and most of our cases did not undergo a formal trial off levothyroxine withdrawal. However, requirement for ongoing levothyroxine replacement in significant dosage, or continuing TSH elevation, suggested persistent CH in at least twelve unsolved cases. Biochemical CH did tend to be more severe in genetically-ascertained cases, which argues against the routine screening of *TG* and *TPO* in milder GIS CH cases. Iodine status was not assessed; however the high familial component in the unsolved case category favors an etiological contribution of genetic factors rather than environmental modulators, including regulatory region or intronic mutations, or CNVs in the genes screened. Genes associated with syndromic CH (eg *GLIS3, GNAS*), were not analysed. Not quantitating thyroid gland size formally might have failed to ascertain cases with mild thyroid hypoplasia, harbouring mutations in some thyroid-dysgenesis associated genes (eg *PAX8, Nkx2-1*). Our aim in using the HiSeq targeted sequencing and MiSeq protocols was to exclude mutations in known CH-associated genes in order to identify a smaller, mutation-negative cohort, which could then be analysed by WES. Thus, future studies with WES/WGS (whole genome sequencing) in familial cases may identify novel genetic aetiologies for CH with GIS, elucidating novel pathways in thyroid development and physiology.

Acknowledgements

This work was supported by Wellcome Trust Grants 100585/Z/12/Z (to N.S.), and 095564/Z/11/Z (to V.K.C.); the National Institute for Health Research Cambridge Biomedical Research Center (to V.K.C., N.S.). E.G.S and C.A.A. are supported by the Wellcome Trust (098051).

This study made use of data generated by the UK10K Project (www.uk10k.org). Funding for the UK10K Project was provided by the Wellcome Trust under award WT091310. A full list of UK10K Consortium members can be found at the UK10K Project website. Specific contributions are as follows: as part of the UK10K project, Shane McCarthy performed the read alignment, improvement, variant calling and quality control as detailed in Supplementary methods for both the exome sequencing and the HiSeq targeted sequencing. James Floyd designed the custom array pull-down for the HiSeq targeted sequencing experiment. CNV calls were generated and quality controlled by Shane McCarthy and Parthiban Vijayarangakannan. These individuals are all affiliated to the Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

References

1. Szinnai G 2014 Clinical Genetics of Congenital Hypothyroidism. *Paediatric Thyroidology. Endocr Dev. Basel, Karger*, 2014, vol 26, pp 60–78

2. Corbetta C, Weber G, Cortinovis F, Calebiro D, Passoni A, Vigone MC, Beck-Peccoz P, Chiumello G, Persani L. A 7-year experience with low blood TSH cutoff levels for neonatal screening reveals an unsuspected frequency of congenital hypothyroidism (CH). *Clin Endocrinol (Oxf).* 2009; 71:739-45.

3. Harris KB, Pass KA. Increase in congenital hypothyroidism in New York State and in the United

States. *Mol Genet Metab*. 2007; 91:268-77.

4. Persani L. Congenital Hypothyroidism with Gland in situ is more Frequent than Previously Thought *Front Endocrinol (Lausanne)*. 2012; 3:18.

5. Rabbiosi S, Vigone MC, Cortinovis F, Zamproni I, Fugazzola L, Persani L, Corbetta C, Chiumello G, Weber G. Congenital hypothyroidism with eutopic thyroid gland: analysis of clinical and biochemical features at diagnosis and after re-evaluation. *J Clin Endocrinol Metab*. 2013; 98:1395-402.

6. Jin HY, Heo SH, Kim YM, Kim GH, Choi JH, Lee BH, Yoo HW. High frequency of DUOX2 mutation in transient or permanent congenital hypothyroidism with eutopic thyroid glands. *Horm Res Paediatr*. 2014; 82:252-60.

7. Grasberger H, Refetoff S. Genetic causes of congenital hypothyroidism due to dyshormonogenesis. *Curr Opin Pediatr*.2011; 23:421-8.

8. Muzza M, Rabbiosi S, Vigone MC, Zamproni I, Cirello V, Maffini MA, Maruca K, Schoenmakers N, Beccaria L, Gallo F, Park SM, Beck-Peccoz P, Persani L, Weber G, Fugazzola L. The clinical and molecular characterization of patients with dyshormonogenic congenital hypothyroidism reveals specific diagnostic clues for DUOX2 defects. *J Clin Endocrinol Metab*. 2014; 99:E544-53

9. Narumi S, Muroya K, Asakura Y, Aachi M, Hasegawa T Molecular basis of thyroid dyshormonogenesis: genetic screening in population- based Japanese patients. *J Clin Endocrinol Metab.* 2011; 96:E1838–E1842.

10. Satoh M, Aso K, Ogikubo S, Yoshizawa-Ogasawara A, Saji T Hypothyroidism caused by the combination of two heterozygous mutations:one on the TSH receptor gene the other in the DUOX2

gene. *J Pediatr Endocrinol Metab* 2015; 28:657-61

11. Park KJ, Park HK, Kim YL, 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; 36:145-53

12. Sriphrapradang C, Tenenbaum-Rakover Y,Weiss M, Barkoff MS, Admoni O, Kawthar D, Caltabiano G, Pardo L, Dumitrescu AM, Refetoff S The coexistence of a novel inactivating mutant thyrotropin receptor allele with two thyroid peroxidase mutations: a genotype- phenotype correlation. *J Clin Endocrinol Metab* 2011; 96:E1001–E1006.

13. Zheng X, Ma SG, Qiu YL, Guo ML, Shao XJ. Novel c.554+5C>T Mutation in the DUOXA2 gene Combinated with p.R885Q Mutation in the DUOX2 Gene Causes Congenital Hypothyroidism. *J Clin Res Pediatr Endocrinol*. 2015 doi: 10.4274/jcrpe.2380.

14. Kelley LA, Mezulis S, Yates CM Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 2015; 10: 845-58.

15. Lee, J., Di Jeso, B., Arvan, P.,The cholinesterase-like domain of thyroglobulin functions as an intramolecular chaperone. *J. Clin. Invest*. 2008; 118: 2950–2958.

16. Citterio CE, Machiavelli GA, Miras MB, Gruñeiro-Papendieck L, Lachlan K, Sobrero G, Chiesa A, Walker J, Muñoz L, Testa G, Belforte FS, González-Sarmiento R, Rivolta CM, Targovnik HM. *Mol Cell Endocrinol*. 2013; 365:277-91.

17. Molina F, Bouanani M, Pau B, Granier C. Characterization of the type-1 repeat from thyroglobulin, a cysteine-rich module found in proteins from different families. *Eur J Biochem*. 1996; 240:125-33.

18. Targovnik HM, Citterio CE, Rivolta CM Thyroglobulin gene mutations in congenital hypothyroidism. *Horm Res Paedi*atr. 2011; 75:311-21.

19. Alzahrani AS, Baitei EY, Zou M, Shi Y. Clinical case seminar: metastatic follicular thyroid carcinoma arising from congenital goiter as a result of a novel splice donor site mutation in the thyroglobulin gene. *J Clin Endocrinol Metab.* 2006; 91:740-6.

20. Tenenbaum-Rakover Y, Grasberger H, Mamanasiri S, Ringkananont U, Montanelli L, Barkoff MS, Dahood AM, Refetoff S. Loss-of-function mutations in the thyrotropin receptor gene as a major determinant of hyperthyrotropinemia in a consanguineous community. *J Clin Endocrinol Metab* 2009; 94:1706-12.

21. Moreno JC, Bikker H, Kempers MJ, van Trotsenburg AS, Baas F, de Vijlder JJ, Vulsma T, Ris-Stalpers C. Inactivating mutations in the gene for thyroid oxidase 2 (THOX2) and congenital hypothyroidism. *N Engl J Med*. 2002; 347:95-102.

22. Wang F, Lu K, Yang Z, Zhang S, Lu W, Zhang L, Liu S, Yan S 2014 Genotypes and phenotypes of congenital goitre and hypothyroidism caused by mutations in dual oxidase 2 genes *Clin Endocrinol (Oxf).* 2014; 81:452-7.

23. De Marco G, Agretti P, Montanelli L, Di Cosmo C, Bagattini B, De Servi M, Ferrarini E, Dimida A, Freitas Ferreira AC, Molinaro A, Ceccarelli C, Brozzi F, Pinchera A, Vitti P, Tonacchera M. Identification and functional analysis of novel dual oxidase 2 (DUOX2) mutations in children with congenital or subclinical hypothyroidism. *J Clin Endocrinol Metab*. 2011; 96:E1335-9.

24. Bakker B, Bikker H, Vulsma T, de Randamie JS, Wiedijk BM, De Vijlder JJ. 2000 Two decades

of screening for congenital hypothyroidism in The Netherlands: TPO gene mutations in total iodide organification defects (an update). *J Clin Endocrinol Metab*. 2000; 85:3708-12.

25. Maruo Y1, Takahashi H, Soeda I, Nishikura N, Matsui K, Ota Y, Mimura Y, Mori A, Sato H, Takeuchi Y.Transient congenital hypothyroidism caused by biallelic mutations of the dual oxidase 2 gene in Japanese patients detected by a neonatal screening program. *J Clin Endocrinol Metab*. 2008; 93:4261-7.

26. Fu C, Chen S, Chen R, Fan X, Luo J, Li C, Qian J. 2014 Mutation screening of the sodium iodide symporter gene in a cohort of 105 China patients with congenital hypothyroidism. *Arq Bras Endocrinol Metabol*. 2014; 58:828-32.

27. Persani L, Calebiro D, Cordella D, Weber G, Gelmini G, Libri D, de Filippis T, Bonomi M. Genetics and phenomics of hypothyroidism due to TSH resistance. *Mol Cell Endocrinol*. 2010; 322:72-82

28. Ris-Stalpers C, Bikker H. Genetics and phenomics of hypothyroidism and goiter due to TPO mutations. *Mol Cell Endocrinol.* 2010; 322:38-43.

29. Fugazzola L, Cerutti N, Mannavola D, Vannucchi G, Fallini C, Persani L, Beck-Peccoz P. Monoallelic expression of mutant thyroid peroxidase allele causing total iodide organification defect. *J Clin Endocrinol Metab* 2003; 88:3264-71

30. Cavarzere P, Castanet M, Polak M, Raux-Demay MC, Cabrol S, Carel JC, Leger J, Czernichow P 2008 Clinical description of infants with congenital hypothyroidism and iodide or ganification defects *Horm Res* **70** 240-8

31. Amendola E, De Luca P, Macchia PE, Terracciano D, Rosica A, Chiappetta G, Kimura S,

Mansouri A, Affuso A, Arra C, Macchia V, Di Lauro R, De Felice M. A mouse model demonstrates a multigenic origin of congenital hypothyroidism. *Endocrinology*. 2005; 146:5038-47.

32. Hulur I, Hermanns P, Nestoris C, Heger S, Refetoff S, Pohlenz J, Grasberger H. A single copy of the recently identified dual oxidase maturation factor (DUOXA) 1 gene produces only mild transient hypothyroidism in a patient with a novel biallelic DUOXA2 mutation and monoallelic DUOXA1 deletion. *J Clin Endocrinol Metab.* 2011; 96:E841-5.

33. Manase D, D'Alessandro LC, Manickarai AK, Al Turki S, Hurles ME, Mital S High throughput exome coverage of clinically relevant cardiac genes. *BMC Med Genomics* 2014; 7:67 doi: 10.1186/s12920-014-0067-8.

34. Léger J, Olivieri A, Donaldson M, Torresani T, Krude H, van Vliet G, Polak M, Butler G; ESPE-PES-SLEP-JSPE-APEG-APPES-ISPAE; Congenital Hypothyroidism Consensus Conference Group. European Society for Paediatric Endocrinology consensus guidelines on screening, diagnosis and management of congenital hypothyroidism. *J Clin Endocrinol Metab*. 2014 ;99:363-84.

35. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 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* 2015; **17**:405-24

641

642 **Figure Legends**

643 **Figure 1**

644 Schematic illustrating case selection, variant filtering and distribution of mutations in the cohort of
645 patients studied with CH and GIS. 'Solved' cases refers to cases in whom a definitive link was
646 established between genotype and CH phenotype. In 'ambiguous' cases, the ascertained genotype
647 could plausibly be contributing to the phenotype, but the evidence to support a causal link was weaker
648 than in the 'solved' group, and 'unsolved' cases carried no mutations in any of the listed genes. The
649 number of cases harbouring monoallelic or biallelic mutations in each gene are listed beneath the
650 corresponding gene name for the 'solved' cases. Numbers in the intersect between circles denote
651 triallelic cases harbouring mutations in both genes. In the 'ambiguous' category, the number of
652 mutations in each gene are classified by mutation type beneath the relevant gene name; all except
653 DUOXA2 were monoallelic. 'Solved' and 'ambiguous' or 'unsolved' cases were equally likely to be
654 familial, but CH was generally more severe in the 'solved' cases. Splice; splice region variant, fs*;
655 frameshift mutation resulting in a premature stop codon, VUS; variant of uncertain significance.

656

657 **Figure 2**

658 Summary of *TG* mutations identified in the study and the associated biochemical phenotype. CH
659 severity is classified according to ESPE criteria on the basis of serum fT4 levels; severe, <5, moderate
660 5 to <10, and mild > 10 pmol/l, respectively (33) and pathogenicity is predicted according to ACMG
661 guidelines (34). A schematic of the TG protein illustrates the position of the mutations relative to the
662 key structural domains of TG including the repetitive type 1, 2 and 3 cysteine-rich regions,
663 acetylcholinesterase homology (ACHE-like) domain and hormonogenic domains. Known mutations
664 are shown in grey, novel mutations in black. *; cases for which complete biochemical data at
665 diagnosis is not available. CH severity refers to sibling. bs; blood spot.

666

667

668

Figure 3

Summary of *TPO* mutations identified in the study and the associated biochemical phenotype. CH severity is classified according to ESPE criteria (33) and pathogenicity is predicted according to ACMG guidelines (34). The effect of the novel missense mutations was modelled using the phyre2-server. Figures in the top row show the wild-type (WT) model, with amino acids of interest in green, figures on bottom row show the model with the mutant amino acid (orange); local polar contacts are shown with black broken lines.

The R291H and R584Q mutations affect amino acids contributing to an intensive network of H-bond contacts close to the catalytic domain involving the heme-group. R291 makes polar contacts with R585 and R582, interacting directly with the heme-group and R584 makes direct polar contacts with the heme-group itself as well as P203 and D633. The mutations R291H (increased hydrophobicity) and R584Q (resulting in a smaller polar group) are likely to disrupt polar contacts affecting local structure and are predicted to affect catalytic activity.

The G331V mutation affects local space filling with the larger valine predicted to impair substrate binding by displacement of the nearby helix and/or disruption of polar contacts (orange amino acids, H_2O molecules in blue), affecting the local structure of TPO.

Figure 4

Summary of *DUOX2* mutations identified in the study and the associated biochemical phenotype. CH severity is classified according to ESPE criteria (33) and pathogenicity is predicted according to ACMG guidelines (34). Mutation position is illustrated using a schematic representation of the domain structure of the DUOX2 protein. Known mutations are shown in grey and novel mutations in black.

The structural model of the peroxidase domain suggests that R354 is part of an intensive hydrogen network. The novel missense mutation R354W replaces the hydrophilic Arginine by the hydrophobic tryptophan disrupting this network and also results in a possible repositioning of the loop containing

R354 and C351, which mediates interactions between the peroxidase domain and extracellular loops obligatory for DUOX2 function.

Figure 5

Genotype-phenotype segregation in six kindreds with oligogenic variants. Horizontal bars denote individuals who have been genotyped. Black shading denotes homozygous individuals and half-black shading denotes heterozygotes for *TG* mutations (F9, F6, F8), TPO mutations (F19, F21) and *DUOX*2 mutations (F10). Potential oligogenic modulators are included by aligning genotype and phenotype data with the individual to whom they refer in the pedigree. *; cases for whom complete biochemical data at diagnosis is not available (F6b, F8a) and CH severity refers to sibling.

In F10, black, half-black and white shading denote the *DUOX2* genotype (Q570L homozygous, heterozygous or wild-type respectively). The pedigree is annotated with *TG* genotype in those cases harbouring variants (L2547Q, R1691C), and phenotype (euthyroid, transient or permanent CH) with venous screening TSH results for CH cases. Cases annotated (euthyroid) were born in Pakistan and although euthyroid in adulthood, the fact that they were not screened neonatally for CH may have precluded detection of transient CH.

Supplementary Figure 1

Proportion of gene length covered at various depth intervals. (A) WES (B) HiSeq targeted sequencing panel. SAMtools mpileup was used to determine the depth at each base within every bait region of every gene for all samples. The median coverage across samples per gene (at exonic sequences only) is represented on top of each bar.

Supplementary Figure 2

Table describing translated exons with inadequate median sequencing depth (<10 fold) in the samples sequenced by WES, HiSeq and MiSeq targeted sequencing and comparing the number of cases solved by each method

Supplementary Table 1

Known and novel mutations detected in the eight genes screened with the predicted consequences by SIFT ('Sorting Tolerant from Intolerant' Algorithm, Ng 2001) and PolyPhen (Polymorphism Phenotyping v2, Adzhubei 2010) and the Genomic Evolutionary Rate Profiling (GERP) score (Davydov 2010). c.DNA; Variant represented in the HGVS notation at the coding DNA level for a transcript. MAF (minor allele frequency) refers to the maximum MAF in all the databases interrogated, with the database from which the MAF is derived in brackets. Blank entries denote unique variants. NA; Not available

1000 Genomes Phase I (1KG_AF), NHLBI GO Exome Sequencing Project 6,500I (ESP_AF), UK10K low-coverage study (UK10K), other UK10K whole-exome sequencing studies (UK10K cohorts) and Exome Aggregation Consortium r0.3 (ExAC).

Supplementary Table 2

Biochemical, radiological and demographic details at diagnosis in cases in whom mutations are detected which are thought to explain the CH phenotype. The column entitled 'Background' indicates the ethnic background. The column entitled 'Diagnosis' includes age at diagnosis as well as indicating the country in which newborn screening was performed for cases detected by this method. TSH values are venous unless otherwise stated; F20 was on L-T4 at the time of sampling. C; known to be consanguineous, bs; blood spot, U; ultrasound, I; isotope scan. KSA; Saudi Arabia, A-C; African – Caribbean, NSP; newborn screening programme, het; heterozygous mutation, hom; homozygous mutation, wt; wild-type, sib; sibling, T.CH; transient CH, NA; Not available, L-T4; levothyroxine Additional clinical features: F17: Hypospadias, hand contractures, hypertrichosis, hyperpigmentation, gynaecomastia, F9a, b:? Cohen syndrome (congenital glucosidation defect and neurometabolic disease), F9a: Bicuspid aortic valve, F21: Polydactyly, macroglossia. F24: the mother was screened from this highly consanguineous family, and the *DUOX2* mutation identified in heterozygosity; this was homozygous in her affected daughter. Newborn screening blood spot cut offs were as follows: 6-10mU/L (UK), 10mU/L (UAE) or cord blood TSH 40mU/L (Oman). Childhood TSH normal range was 0.35-5.5mU/L.

Supplementary Table 3

Biochemical, radiological and demographic details at diagnosis in cases in whom mutations are detected which are thought significant without fully explaining the CH phenotype (ambiguous cases), and cases in whom no significant mutations were found. The column entitled 'Background' indicates the ethnic background. The column entitled 'Diagnosis' includes age at diagnosis as well as indicating the country in which newborn screening was performed for cases detected by this method. TSH values are venous unless otherwise stated. C; known to be consanguineous, bs; blood spot, U; ultrasound, I; isotope scan, KSA; Saudi Arabia, NSP; newborn screening programme, het; heterozygous mutation, hom; homozygous mutation, NA; Not available. F25*; No investigations but had goitre, and on 50mcg levothyroxine aged 5. Two siblings had dyshormonogenesis with TSH>100, fT4 6.9, and avid uptake in a normally-sited gland (i).

Additional clinical features: F12a,b: IUGR, F14a,b: Papillon-Lefevre syndrome, F21, postaxial polydactyly, F27: Monogenic skeletal dysplasia unrelated to CH, F32: Metabolic acidosis, developmental delay, F34: Hypospadias, undescended testes, laryngomalacia, hypertelorism.

Newborn screening blood spot cut offs were as follows: 6-10mU/L (UK), 10mU/L (UAE) or cord blood TSH 40mU/L (Oman). Childhood TSH normal range was 0.35-5.5mU/L.