Whole-genome sequence-based analysis of thyroid function. Nature Communications 6, 5681. 10.1038/ncomms6681 file

Publishers page: http://dx.doi.org/10.1038/ncomms6681
<http://dx.doi.org/10.1038/ncomms6681>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.
Whole-genome sequence-based analysis of thyroid function

Peter N. Taylor, Eleonora Porcu, Shelby Chew, Purdey J. Campbell, Michela Traglia, Suzanne J. Brown, Benjamin H. Mullin, Hashem A. Shihab, Josine Min, Klaudia Walter, Yasin Memari, Jie Huang, Michael R. Barnes, John P. Beilby, Pimphen Charoen, Petr Danecek, Frank Dudbridge, Vincenzo Foretta, Celia Greenwood, Elin Grundberg, Andrew D. Johnson, Jennie Hui, Ee M. Lim, Shane McCarthy, Dawn Muddymann, Vijay Panicker, John R. Perry, Jordana T. Bell, Wei Yuan, Caroline Relton, Tom Gaunt, David Schlessinger, Goncalo Abecasis, Francesco Cucca, Gabriela L. Surdulescu, Wolfram Woltersdorf, Eleftheria Zeggini, Hou-Feng Zheng, Daniela Toniolo, Colin M. Dayan, Silvia Naitza, John P. Walsh, Tim Spector, George Davey Smith, Richard Durbin, J. Brent Richards, Serena Sanna, Nicole Soranzo, Nicholas J. Timpson, Scott G. Wilson, & the UK10K Consortium

Normal thyroid function is essential for health, but its genetic architecture remains poorly understood. Here, for the heritable thyroid traits thyrotropin (TSH) and free thyroxine (FT4), we analyse whole-genome sequence data from the UK10K project (N = 2,287). Using additional whole-genome sequence and deeply imputed data sets, we report meta-analysis results for common variants (MAF ≥ 1%) associated with TSH and FT4 (N = 16,335). For TSH, we identify a novel variant in SYN2 (MAF = 23.5%, P = 6.15 × 10⁻⁵) and a new independent variant in PDE8B (MAF = 10.4%, P = 5.94 × 10⁻¹⁴). For FT4, we report a low-frequency variant near B4GALT6/SLC25AS2 (MAF = 3.2%, P = 1.27 × 10⁻⁹) tagging a rare TTR variant (MAF = 0.4%, P = 2.14 × 10⁻¹¹). All common variants explain ≥ 20% of the variance in TSH and FT4. Analysis of rare variants (MAF < 1%) using sequence kernel association testing reveals a novel association with FT4 in NRG1. Our results demonstrate that increased coverage in whole-genome sequence association studies identifies novel variants associated with thyroid function.
thyroid hormones have fundamental but diverse physiological roles in vertebrate physiology, ranging from induction of metamorphosis in amphibians to photoperiodic regulation of seasonal breeding in birds. In humans, they are essential for adult health and childhood development and levothyroxine is one of the commonest drugs prescribed worldwide. Clinically, thyroid function is assessed by measuring circulating concentrations of free thyroxine (FT4) and the pituitary hormone thyrotropin (TSH); the complex inverse relationship between them renders TSH the more sensitive marker of thyroid status. Even small differences in TSH and FT4, within the normal population reference range, are associated with a wide range of clinical parameters, including blood pressure, lipids and cardiovascular mortality, as well as obesity, bone mineral density and lifetime cancer risk.

Twin and family studies estimate the heritability of TSH and FT4 as up to 65%. Genome-wide association studies (GWAS) identified common variants associated with TSH and FT4 in a recent HapMap-based meta-analysis, in which we identified 19 loci associated with TSH and 4 with FT4. However, these accounted for only 6% of the variance in TSH and 2.3% in FT4. Therefore, most of the heritability of these important traits remains unexplained.

The unidentified genetic component of variance might be explained by common variants poorly tagged by markers assessed in previous studies, or those with small effects. However, rarer variants within the minor allele frequency (MAF) spectrum might also account for a substantial proportion of the missing heritability as has been proposed for many polygenic traits. These variants, although individually rare (MAF <1%), are collectively frequent, and while their effects may be insufficient to produce clear familial aggregation, effect sizes for individual variants are potentially much greater than those observed for common variants. In addition, a greater understanding of the relative proportion of thyroid function explained by common variants is now possible with the availability of whole-genome sequencing (WGS) and this is essential to refine future research and analysis strategies when appraising the genetic architecture of thyroid function.

In this study, the first to utilize WGS to examine the genetic architecture of TSH and FT4, we perform single-point association analysis in two discovery cohorts in the UK10K project with WGS data available and a meta-analysis using genome wide association data (GWAS) with deep imputation from five additional data sets. We report three new loci associated with thyroid function in healthy individuals, undertake quantitative trait loci and DNA methylation analyses to further study these relationships and undertake genome-wide complex trait analyses (GCTA) to assess the contributions of common variants (MAF >1%) to variance in thyroid function. We also explore whether there is a shared polygenic basis between TSH and FT4. In individuals with WGS data, we perform sequence kernel-based association testing (SKAT) analysis to identify regions of the genome where rare variants have the strongest association with thyroid function and identify a novel locus associated with FT4. The results demonstrate that WGS-based analyses can identify rare functional variants and associations derived from rare aggregates. Larger meta-analyses of studies with WGS data are now required to identify additional common and rare variants, which may explain the missing heritability of thyroid function.

Results

Single-point association analysis. In the discovery study, using a meta-analysis of WGS data from the Avon Longitudinal Study of Parents and Children (ALSPAC) and TwinsUK cohorts (N = 2,287) analysing up to 8,816,734 markers (Supplementary Tables 1 and 2; Supplementary Methods), we find associations at two previously described loci for TSH. These are NR3C2 (rs11728154; MAF = 21.0%, B = 0.21, s.e. = 0.037, P = 8.21 × 10−9; r² = 0.99 with the previously reported rs10028213) and FOXE1 (rs1877431; MAF = 39.5%, B = −0.19, s.e. = 0.30, P = 2.29 × 10−11; r² = 0.99 with the previously reported rs965513). We find one borderline signal (between P = 5.0 × 10−08 and P = 1.17 × 10−06) at a novel locus FAM222A (rs11067829; MAF = 18.3%, B = 0.210, s.e. = 0.038, P = 3.73 × 10−5; Supplementary Figs 1a and 2; Supplementary Table 3). No variants show genome-wide significant association for FT4 (Supplementary Figs 1a and 3).

In a meta-analysis of the discovery cohorts and five additional cohorts, we find associations for 13 SNPs at 11 loci for TSH (N = 16,335) of which 11 loci have been identified previously and 4 SNPs at 4 loci for FT4 (N = 13,651) of which 3 have been identified previously (Table 1; Figs 1a–c,2a,b and 3; Supplementary Figs 1b and 3–6).

To determine whether our identified associations at established loci represented previous associated signals, we analysed the linkage disequilibrium (LD) between the strongest associated variants from this study and those from our previous study (Supplementary Table 4). The top variants from loci in both studies were in strong LD (r² >0.6), apart from MBIP and FOXE1, although these were in strong LD with variants previously associated with TSH by others. Two SNPs associated with TSH in our study are novel, one at SYN2 (rs310763; MAF = 23.5%, B = 0.082, s.e. = 0.014, P = 6.15 × 10−9; Fig. 1a–c). SYN2 is a member of a family of neuron-specific phosphoproteins involved in the regulation of neurotransmitter release with expression in the pituitary and hypothalamus (http://biogps.org/#goto=genereport&id=6854).

We also identify one novel variant at PDE8B (MAF = 10.4%, B = −0.145, s.e. = 0.019, P = 5.94 × 10−14) in linkage equilibrium (r² = 0.002, D’ = 0.17) with the previously described variant rs6885099 (ref. 10) and independent from our top SNP rs2046045 (P = 1.93 × 10−11) after conditional analysis. In the overall meta-analysis, we are unable to replicate the association between FAM222A and TSH in the discovery analysis (B = 0.014, s.e. = 0.015, P = 0.378); however, we observe evidence of heterogeneity between cohorts (test for heterogeneity P = 4.70 × 10−6; Supplementary Table 5), so potentially this locus may find support in future WGS studies.

In our meta-analysis, we also identify four SNPs associated with FT4, three at previously established loci (DIO1, LHX3 and AADAT; Table 1; Fig. 3; Supplementary Figs 1b, 4e and 6; Supplementary Table 4). We find a novel uncommon variant at B4GALT6/SLC25A52 associated with FT4 (rs113107469; MAF = 3.20%, B = 0.225, s.e. = 0.037, P = 1.27 × 10−9; Fig. 2a). B4GALT6 is in the ceramide metabolic pathway, which inhibits cyclic AMP production in TSH-stimulated cells. However, the B4GALT6 signal (rs113107469) is in weak LD (r² <0.1, D’ = 0.66) with the Thr139Met substitution (rs28933981; MAF = 0.4%) and it may therefore be a marker for this functional change in TTR. The Thr139Met substitution was associated with FT4 levels in our single-point meta-analysis (P = 2.14 × 10−11), however, was not originally observed as the MAF was lower than our 1% threshold. Conditional analysis of the TTR region using rs28933981 as the conditioning marker in the ALSPAC WGS cohort reveals no evidence of association between rs113107469 in B4GALT6 and FT4 (P = 0.124; Fig. 2b). Analysis using direct genotyping in the ALSPAC WGS and replication cohorts confirms the effect of the Thr139Met substitution on FT4 levels. Here, 0.79% of children were heterozygous for the Thr139Met substitution, which is positively associated with FT4 (B = 1.70, s.e. = 0.17, 95% CI 1.37,
Table 1 | Independent SNPs with MAF ≥1% associated with serum TSH and FT4 levels in the overall meta-analysis.

| Gene   | SNP               | Chromosome | Position | A1/A2 | Freq A1 | Effect | Std Err | N      | P      | Het P   |
|--------|-------------------|------------|----------|-------|---------|--------|---------|--------|--------|---------|
| TSH    | CAP2B rs12410532  | 1          | 19845279 | T/C   | 0.164   | –0.090 | 0.016   | 16,332 | 9.41×10⁻⁹ | 0.003   |
|        | IGBPBP rs7568039  | 2          | 21761231 | A/C   | 0.250   | –0.122 | 0.014   | 16,335 | 2.11×10⁻¹⁹ | 0.370   |
|        | SYN2 rs310763     | 3          | 12230704 | T/C   | 0.235   | 0.083  | 0.014   | 16,334 | 6.15×10⁻⁹ | 0.252   |
|        | NR3C2 rs28435578  | 4          | 14964653 | C/T   | 0.227   | –0.166 | 0.014   | 16,333 | 4.59×10⁻¹² | 0.109   |
|        | PDEBB rs2046045   | 5          | 7653581  | G/T   | 0.414   | 0.142  | 0.012   | 16,334 | 4.05×10⁻⁹ | 0.653   |
|        | PDE8B rs2928167   | 5          | 76477820 | G/A   | 0.104   | –0.145 | 0.019   | 16,334 | 5.94×10⁻⁴ | 0.994   |
| VEGFA  | rs6923866         | 6          | 43901184 | C/T   | 0.280   | –0.102 | 0.013   | 16,333 | 7.55×10⁻¹⁵ | 0.646   |
| VEGFA  | rs2996084         | 6          | 43804625 | A/G   | 0.287   | –0.096 | 0.013   | 16,333 | 4.33×10⁻¹³ | 0.422   |
| PDE10A | rs3008034         | 6          | 16604382 | C/T   | 0.312   | –0.131 | 0.012   | 16,335 | 4.68×10⁻²⁶ | 0.084   |
| FOXE1  | rs112817873       | 9          | 100548934| T/A   | 0.323   | –0.140 | 0.015   | 11,544 | 6.15×10⁻²⁰ | 2.02×10⁻⁶ |
| ABO    | rs116552240       | 9          | 136149098| A/T   | 0.239   | 0.120  | 0.016   | 14,047 | 1.92×10⁻¹⁴ | 4.11×10⁻⁴ |
| MBIP   | rs16909374        | 14         | 36738361 | T/C   | 0.043   | –0.208 | 0.032   | 15,037 | 4.69×10⁻¹¹ | 0.179   |
| MAF    | rs17767742        | 16         | 79740541 | G/C   | 0.354   | –0.113 | 0.012   | 16,335 | 5.64×10⁻²⁰ | 0.447   |

Table 1 continued...

Expression quantitative trait locus analysis. Expression quantitative trait locus (eQTL) analysis reveals that our SYN2 variant modulates SYN2 transcription in adipose, skin and whole-blood cells, but not lymphoblastoid cell lines (Supplementary Table 7). Furthermore, bioinformatics analysis suggests that the C-allele at rs310763 attenuates an EGR1 regulatory motif, EGR1 is expressed in thyrocytes, regulates pituitary development and may influence thyroid status via LH3 promoter activity. Several other variants in the SYN2 gene region are in strong LD (r² > 0.8) with rs310763, including the non-synonymous coding variant rs794999. Although predicted to be benign (PolyPhen-2 score = 0.002 (ref. 22)), rs794999 is located in a DNaSe hypersensitivity cluster, influences four predicted regulatory motifs, and appears to be under evolutionary constraint. SNPs identified in our study, or those in LD, also showed strong eQTL associations with PDE8B (P = 8.69×10⁻²⁵), FOXE1 (P = 9.10×10⁻³⁴) and AADAT (P = 7.86×10⁻⁹) gene expressions (Supplementary Table 7).

DNA methylation analysis. To further explore cis-regulatory effects of variants identified in our study, we carried out analysis of DNA methylation profiles in whole-blood samples in 279 individuals from the TwinsUK cohort. We find evidence for a methylation quantitative trait locus (meQTL) at the novel TSH-associated variant rs2928167 in PDE8B (P = 4.38×10⁻⁷, Supplementary Table 8), which are also eQTLs in multiple tissues (Supplementary Table 7). Recently, meQTL effects using the same probe (cg16418800) in adipose tissue also identified a peak signal at rs2359775 (P = 6×10⁻¹⁵), which is in LD with rs2928167 (r² = 0.5). We find that variants in ABO (P = 2.02×10⁻²³) and AADAT (P = 1.80×10⁻⁸) also show strong evidence for cis-meQTL effects (Supplementary Table 8). In additional analyses in 745 ALSPAC children, we find strong meQTL associations for rs2359775 in PDE8B (P = 3.03×10⁻²⁸) and variants in ABO (P = 1.01×10⁻¹⁰¹) and AADAT (P = 4.18×10⁻³⁴) (Supplementary Table 8).

SKAT analysis. Tests of the association between aggregates of rare variants (MAF < 1%) in the WGS cohorts were restricted to genes relevant to thyroid function. We find no evidence of association from SKAT analyses with TSH, however, for FT4 we identify one SKAT bin with multiple-testing-corrected evidence for association (P = 1.55×10⁻⁵) in NRG1 (P = 2.53×10⁻⁶, Fig. 4; Supplementary Table 9). NRG1 is a glycoprotein that interacts with the NEU/ERBB2 receptor tyrosine kinase, and is critical in organ development.

GCTA and polygenic score analysis. SNPs were thinned to a set of 2,203,581 approximately independent SNPs with an LD threshold of r² > 0.2, a window size of 5,000 SNPs and step of 1,000 SNPs. A genomic relationship matrix was then generated for unrelated individuals. We fitted linear mixed-effect models and estimate that all assessed common SNPs (MAF > 1%) explain 24% (95% CI 19, 29) and 20% (95% CI 14, 26) of TSH and FT4 variance, respectively (P < 0.0001; Supplementary Table 10). Polygenic score analyses based on SNPs with P values under a fixed threshold do not detect evidence of a polygenic signal for TSH or FT4, nor of a shared polygenic basis between thyroid function and key metabolic outcomes. However, a genetic score based on 67 SNPs previously associated with thyroid function in

---

2.03, P = 3.89×10⁻²⁴. In the ALSPAC replication data set, rs113107469 in B4GALT6 was also positively associated with FT4 (P = 0.0002); however, when conditioned on the Thr139Met substitution there was no longer any evidence of association (P = 0.20). The Thr139Met substitution also appears to be functional: this mutation has increased protein stability compared (PolyPhen-2 score 23) and may influence thyroid status via thyroxine affinity, which is in LD with rs2928167 (r² = 0.5). We find that variants in ABO (P = 2.02×10⁻²³) and AADAT (P = 1.80×10⁻⁸) also show strong evidence for cis-meQTL effects (Supplementary Table 8). In additional analyses in 745 ALSPAC children, we find strong meQTL associations for rs2359775 in PDE8B (P = 3.03×10⁻²⁸) and variants in ABO (P = 1.01×10⁻¹⁰¹) and AADAT (P = 4.18×10⁻³⁴) (Supplementary Table 8).
Figure 1 | Regional and genome-wide association plots for TSH. (a) Regional association plot showing genome-wide significant locus for serum TSH at the SYN2, TIMP4 gene region. Inset is in vitro expression QTL data for the lead SNP rs310763 in adipose cells (A), lymphoblastoid cell lines (L), skin cells (S) and whole blood (W). Dotted line denotes genome-wide significance threshold. (b) Regional association plot after conditional analysis on rs2046045 in PDE8B showing our novel association with TSH at rs2928167 in PDE8B remained genome-wide significant. (c) Annotated Manhattan plot from the overall analysis for serum TSH levels. SNPs (MAF > 1%) are plotted on the x axis according to their position on each chromosome against association with TSH on the y axis (shown as \(-\log_{10}(P\text{-value})\)). The loci are regarded as genome-wide significant at \(P < 5 \times 10^{-8}\). Variants with 1% < MAF < 5% are shown as open diamond symbols. Common SNPs (MAF > 5%) are shown as solid circles with those present in Hapmap II reference panels in grey and those derived from WGS or deeply imputed using WGS and 1000 genomes reference panels in blue. SNPs shown as a red asterisk represent novel genome-wide significant findings.
GWAS shows strong evidence of association with TSH \( (P = 7.9 \times 10^{-10}) \) and FT4 \( (P = 2.7 \times 10^{-4}) \) and we observe evidence of shared genetic pathways with TSH associated with the FT4 gene score \( (P = 7.0 \times 10^{-4}) \). These 67 SNPs explain 7.1% (95% CI 5.2, 9.0) of the variance in TSH and 1.9% (95% CI 1.1, 3.0) of the variance in FT4. Taken together, this suggests that many loci underlying thyroid function remain unknown.

Chemogenomic analysis. We undertook a database analysis of differential gene expression in cultured cells in response to hormone stimulation. We find SYN2 (rank 64 of 22283 (HL60 cells)) rates highest among the genes studied in the experiment, providing strong support for the role of this newly discovered locus in thyroid metabolism. Two other genes, NRG1 and CAPZB, also show evidence of levothyroxine responsiveness in at least one cell line on the basis of a genome-wide differential expression and rank in the top 5th percentile (Supplementary Table 11). Publicly available data on altered SYN2 expression in brain, limb and tail from control and levothyroxine-treated Xenopus laevis during metamorphosis also provide evidence for the relevance of SYN2 in thyroid function.

Discussion
In this study, we demonstrate the utility of WGS data (and SNP array data when deeply imputed to WGS reference panels) in appraising the genetic architecture of thyroid function. Using WGS data, we identify a rare functional variant in TTR that appears to drive the observed association between an uncommon novel variant near B4GALT6 and FT4, and we demonstrate a novel association with FT4 arising from rare aggregates in NRG1. We also show that common variants collectively account for over 20% of the variance in TSH and FT4, a substantial advance on using only the 'top SNPs' from earlier GWA studies. Taken together, this work indicates that both common variants with
modest effects and rare variants with larger effects might explain a substantial proportion of the missing heritability of thyroid function, but larger studies are required to identify these variants. Studies including individuals with subclinical thyroid disease, particularly those who are negative for thyroid autoantibodies, may be particularly rewarding, as rare genetic variants with large effect sizes may be associated with serum TSH and FT4 concentrations outside the inclusion ranges we used and therefore would not be detected in our analyses.

Such endeavours are clinically relevant, as there has been a dramatic increase in levothyroxine prescribing for borderline TSH levels\(^29\). At least three loci identified in this study show evidence of responsiveness to levothyroxine in cell line models, underscoring that borderline TSH levels often reflect the influence of genetic variation rather than overt autoimmune thyroid disease, in which case thyroid hormone replacement may not be appropriate. Our results indicate that further investigation of TSH heterogeneity at the population level is necessary.

**Methods**

**Cohorts.** Seven populations were used in this study. They are known as the TwinsUK WGS cohort, the TwinsUK GWAS cohort, the ALSPAC WGS cohort, the ALSPAC GWAS cohort, the SardiNIA cohort, the ValBorbera cohort and the Busselton Health Study cohort. Summary statistics of each cohort and full descriptions are given in Supplementary Methods, Supplementary Tables 1 and 2. All human research was approved by the relevant institutional ethics committees.

**WGS data generation.** Low-read depth WGS was performed in the TwinsUK and ALSPAC as part of the UK10K project. The SardiNIA cohort also had WGS data available (see Supplementary Methods).

**Statistical analysis.** An inverse normal transformation was applied to each trait within each cohort. Age, gender and any other cohort-specific variables (Supplementary Table 1) were applied as covariates. Genotype imputation was performed for relevant cohorts using the IMPUTE\(^{30}\), MaCH\(^{31}\) or Minimac\(^{32}\) software packages, with poorly imputed variants excluded. See Supplementary Table 1 for cohort-specific details.

**Single-point association analysis.** Association analysis within each cohort was performed using the SNPTEST v2 (ref. 33), GEMMA (genome-wide efficient mixed model association)\(^{34}\), EPACTS (efficient and parallelizable association container tool box) or ProbABEL\(^{35}\) software packages. Cohort-specific quality control filters relating to call rate and Hardy–Weinberg equilibrium were applied (Supplementary Table 1). In our analysis, we assessed the change in standardized thyroid measure by allele using a MAF threshold \(\geq 1\%\) and a genome-wide significance threshold of \(P = 1.17 \times 10^{-8}\) (ref. 36). Meta-analyses were performed using the GWAMA (genome-wide association meta analysis) software\(^2\), which was used to perform fixed-effect meta-analyses using estimates of the allelic effect size and s.e. Two meta-analyses were performed for each phenotype: a meta-analysis of the two UK10K WGS cohorts and a meta-analysis of all seven cohorts. The ValBorbera cohort does not have FT4 phenotype data, so this cohort was not included in the meta-analyses for this phenotype. In the meta-analyses, any variants that were missing from \(> 2\) cohorts or with a combined MAF \(\leq 1\%\) were excluded. However, in the discovery analyses, a MAF of 0.5% in either cohort was accepted to prevent marginal MAF drops; the MAF \(< 1\%\) exclusion was then applied during the meta-analysis.

**Conditional analysis.** A conditional analysis was performed to identify independent association signals. Each study re-analysed significant loci using the lead SNP identified in the primary analysis (Table 1) as the conditioning marker. In cohorts where the lead SNP was not available, the best proxy was included (\(r^2 > 0.8\)). A meta-analysis was then performed on these conditional results, using the same methods and filters as described above. The standard genome-wide significant cut-off (\(P < 5 \times 10^{-8}\)) was used to identify secondary associations.

**Estimation of phenotypic variance explained by genetic variants.** We undertook GCTA using WGS data in the ALSPAC and TwinsUK discovery cohorts and data from the SardiNIA and Busselton cohorts to estimate the variance explained by all common SNPs (MAF \(> 1\%\)) in the genome for TSH and FT4, using the GCTA method of Yang et al.\(^12\) We fitted linear mixed-effect models to
obtain evidence of a common genetic basis for related disorders. We ranked SNPs by their marginal association with TSH and FT4 using the meta-analysis data set, with TwinsUK samples excluded (excluding *N* = 13,874 for TSH and *N* = 12,561 for FT4). SNPs were thinned to a set of 2,203,881 approximately independent SNPs using the *--indep-pairwise* option in PLINK with an LD threshold of *r*² > 0.2, window size of 5,000 SNPs and step of 1,000 SNPs. A genomic relationship matrix was constructed for subjects in the TwinsUK data sets by forming the weighted sum of trait-increasing alleles, with the weights taken as the effect size in the meta-analysis data set. To construct polygenic scores, we used 67 SNPs (*rs10028213, rs10030589, rs10032216, rs10402008, rs10499559, rs10519227, rs10799824, rs10917469, rs10914777, rs1103377, rs113107496, rs11642776, rs116552240, rs11699374, rs1194732, rs11726248, rs11755845, rs12140532, rs13015993, rs1537424, rs1571583, rs171720214, rs17723470, rs17775063, rs2046045, rs22355344, rs2239302, rs22435578, rs22982167, rs3008034, rs3008040, rs310763, rs334699, rs334721, rs34248920, rs3813382, rs4074397, rs4084416, rs45738967, rs5608276, rs61988381, rs6499766, rs68859099, rs68923666, rs69777660, rs7128207, rs7190187, rs7240777, rs7279671, rs73392602, rs73755800, rs76580393, rs7694879, rs7825175, rs7860634, rs7864322, rs79131135, rs8922817, rs9442829, rs9472138, rs9497965, rs965513, rs9662423 and rs9913657) that have been shown to be associated with thyroid hormone levels. The polygenic score was then tested for association with relevant thyroid and other phenotypes in the TwinsUK sample.

**Chemogenomic analysis.** To identify putative thyroxine-responsive genes among the candidate loci (AADAAT, ABO, B4GALT6, CAP2B, DIO1, FOXE1, IGF2BP2, LHFA, MAF, MBIP, MFAP3L, NR3C2, NR1G1, PDE10A, PDE8B, QSOX2, SLCA25A2, SLC25A5, SLC25A52, SLCO1B1, TSHRB, TR, TR2, TR4, TR6, VEGFA), gene expression data measured in response to levothyroxine treatment in a range of cell lines were retrieved from the Connectivity Map resource [30].

We conducted a genome-wide differential expression analysis in the top 5% percentile among 22,232 probes as evidence of differential expression.

### References

1. Dumont, J. et al. Ontogeny, anatomy, metabolism and physiology of the thyroid. *Dis. Mark. Available at http://www.thyroidmanager.org/chemogenomics/chemistry-ontogeny-anatomy-physiology-of-the-thyroid* (2011).
2. Haddow, J. E. et al. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *New Engl. J. Med.* 341, 549–555 (1999).
3. Vanderpump, M. P. The epidemiology of thyroid disease. *Br. Med. Bull.* 99, 39–51 (2011).
4. Haddow, N. C. et al. The relationship between TSH and free T4 in a large population is complex and nonlinear and differs by age and sex. *J. Clin. Endocrinol. Metab.* 98, 2936–2943 (2013).
5. Taylor, P. N., Razvi, S., Pearce, S. H. & Dayan, C. M. A review of the clinical consequences of variation in thyroid function within the reference range. *J. Clin. Endocrinol. Metab.* 98, 3562–3571 (2013).
6. Panicker, V. et al. Hereditry expression data measured in response to levothyroxine treatment in a range of cell lines were retrieved from the Connectivity Map resource [30].
7. Arnaud-Lopez, L. et al. Phosphodiesterase 8B gene variants are associated with serum TSH levels and thyroid function. *Am. J. Hum. Genet.* 82, 1270–1280 (2008).
8. Gudmundsson, J. et al. Discovery of common variants associated with low TSH levels and thyroid cancer risk. *Nat. Genet.* 44, 319–322 (2012).
9. Panicker, V. et al. A locus on chromosome 1p36 is associated with thyrotropin and thyroid function as identified by genome-wide association study. *Am. J. Hum. Genet.* 87, 430–435 (2010).
10. Porcu, E. et al. A meta-analysis of thyroid-related traits reveals novel loci and gender-specific differences in the regulation of thyroid function. *PLoS Genet.* 9, e1003626 (2013).
11. Rodman, W. & Bonilla, C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* 40, 695–701 (2008).
12. Yang, J. & E. Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 86, 76–82 (2011).
13. Alves L. et al. Thyroxine binding in a TTR Met 119 kindred. *J. Mol. Biol.* 277, 484–488 (1993).
14. Sebastiao, M. P., Lamzin, V., Saravia, M. J. & Damas, A. M. Transhyretin stability as a key factor in amyloidogenesis: X-ray analysis at atomic resolution. *J. Mol. Biol.* 306, 733–744 (2001).
15. Curtis, A. et al. Thyroxine binding by human transthyretin variants: mutations at position 119, but not position 54, increase thyroxine binding affinity. *Hum. Mol. Genet.* 12, 459–465 (2003).
16. Hamilton, J. A. & Benson, M. D. Transhyretin: a review from a structural perspective. *Cell. Mol. Life Sci.* 58, 1491–1521 (2001).
17. Yang, T. P. et al. Eanver: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics 26, 2474–2476 (2010).
18. Grundberg, E. et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 44, 1084–1089 (2012).
19. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40, D930–D934 (2012).
20. Valen, B. C., Garcia, 3rd M., Smith, T. P. & Rhodes, S. J. Two promoters mediate transcription from the human LHX3 gene: involvement of nuclear factor I and specificity protein 1. Endocrinology 147, 324–337 (2006).
21. Savage, J. J., Yaden, B. C., Kiratipranon, P. & Rhodes, S. J. Transcriptional control during mammalian anterior pituitary development. Gene 319, 1–19 (2003).
22. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249 (2010).
23. Bernstein, B. E. et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
24. Davydov, E. V. et al. Identifying a high fraction of the human genome to be under selective constraint using GERF + + . PLoS Comput. Biol. 6, e1001025 (2010).
25. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478, 476–482 (2011).
26. Medici, M. et al. A large-scale association analysis of 68 thyroid hormone pathway genes with serum TSH and FT4 levels. Eur. J. Endocrinol. 164, 781–788 (2011).
27. Lamb, J. et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929–1935 (2006).
28. Das, B. et al. Gene expression changes at metamorphosis induced by thyroid hormone in Xenopus laevis tadpoles. Dev. Biol. 291, 342–355 (2006).
29. Taylor, P. N. et al. Falling threshold for treatment of borderline elevated thyrotropin levels—balancing benefits and risks: evidence from a large community-based study. JAMA Intern. Med. 174, 32–39 (2015).
30. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
31. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved alleles in a heterogeneous panel. Genet. Epidemiol. 34, 816–834 (2010).
32. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. 44, 955–959 (2012).
33. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906–913 (2007).
34. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821–824 (2012).
35. Aukunchoung, Y. S., Struchalin, M. V. & van Duijn, C. M. ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics 11, 134 (2010).
36. Xu, C. et al. Estimating genome-wide significance for whole-genome sequencing studies. Genet. Epidemiol. 38, 281–290 (2014).
37. Magi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 11, 288 (2010).
38. Emerson, V. et al. Genetics of gene expression and its effect on disease. Nature 452, 423–428 (2008).
39. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).
40. Wu, M. C. et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. 89, 82–93 (2011).
41. Voorman, A., Brody, I. & Lumley, T. SkatMeta: an R package for meta analyzing region-based tests of rare DNA variants. Available at (http://cran.r-project.org/web/packages/skatMeta) (2013).
42. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. PLoS Genet. 9, e1003348 (2013).

Acknowledgements
We are grateful to all the participants in the cohort studies and the staff involved including interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. This study makes use of the data generated by the UK10K Consortium. Funding for UK10K was provided by the Wellcome Trust under award WT093100. A full list of the investigators who contributed to the generation of the data is available at www.UK10K.org. Further acknowledgments from all the cohorts and details on cohort and investigator funding can be found in the Supplementary Methods.

Author contributions
Cohort collection was done by P.N.T., E.P., G.A., C.M.D., S.N., J.P.B., J.H., E.M.L., Y.P., W.W., D.T., J.P.W., C.M.D., T.D.S., G.D.S., R.J.B., S.S., N.S., N.J.T. and S.G.W. Phenotype cleaning was done by P.N.T., E.P., G.A., C.M.D., T.D.S., J.P.B., J.H., E.M.L., V.P., I.R.B., S.S., N.S., N.J.T. and S.G.W. Phenotype data processing and cleaning was done by S.I.B., J.M., K.W., Y.M., P.J.W., H., S.M., D.S. and E.Z. Genotype-phenotype association testing was done by P.N.T., E.P., S.C., P.J.C., M.T., S.I.B., B.H.M., H.A.S., M.R.B., P.C., P.D., F.D., V.C., E.G., A.D.J., J.H., V.P., I.R.B., J.T.B., W.Y., C.T., G.L.S. and H.-F.Z. Bioinformatics were done by S.C., B.H.M., S.I.B., J.M., K.W., Y.M., S.G.W., J.R.B., M.R.B., P.G. and M.G. Manuscript drafting and revision was done by P.N.T., E.P., S.C., P.J.C., M.T., S.I.B., B.H.M., J.P.W., C.M.D., J.P.T., M.R.B., F.D., S.S., N.J.T. and S.G.W. All authors critically revised the manuscript.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Taylor, P. N. et al. Whole-genome sequence-based analysis of thyroid function. Nat. Commun. 6:5681 doi: 10.1038/ncomms6681 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

The UK10K Consortium
Saeed Al Turki9,27, Carl Anderson9, Richard Anney26, Dinu Antony29, Maria Soler Artigas28, Muhammad Ayub30, Senduran Balasubramaniam9, Jeffrey C. Barrett9, Inês Barroso9,31, Phil Beales29, Jamie Bentham32, Shoumo Bhattacharya32, Ewan Birney31, Douglas Blackwood34, Martin Bobrow35, Elena Bochukova31, Patrick Bolton36, Rebecca Bounds31, Chris Boustrid8, Gerome Breen36,37, Mattia Calissano38, Keren Carss9, Krishna Chatterjee31, Lu Chen39,39, Antonio Ciampi16, Sebhattin Cirak38,40, Peter Clapham9, Gail Clement22, Guy Coates9, David Collie41,42, Catherine Cosgrove32, Tony Cox9, Nick Craddock43, Lucy Crooks9,44, Sarah Curran36,45,46, David Curtis47, Allan Daly9, Aaron Day-Williams9,48, Ian N.M. Day9, Thomas Down9,49, Yuanping Du50, Ian Dunham32, Sarah Edkins9, Peter Ellis9, David Evans8,51, Sadaf Farooqi31, Ghazaleh Fatemifar9, David R. Fitzpatrick52, Paul Flice9,33, James Flyod9,53, A. Reghan Foley16, Christopher S. Franklin9, Marta Futema54, Louise Gallagher28, Matthias Geis9, Daniel Geschwind55, Heather Griffin56, Detelina Grozeva35, Xueqin Guo50,
Xiaosen Guo, Hugh Gurling, Deborah Hart, Audrey Hendricks, Peter Holmans, Bryan Howie, Liren Huang, Tim Hubbard, Steve E. Humphries, Matthew E. Hurles, Pirro Hysi, David K. Jackson, Yalda Jamshidi, Tian Jing, Chris Joyce, Jane Kaye, Thomas Keane, Julia Keogh, John Kemp, Karen Kennedy, Anja Kolb-Kokocinski, Genevieve Lachance, Cordelia Langford, Daniel Lawson, Irene Lee, Monkol Lek, Jieqin Liang, Hong Lin, Rui Li, Yingrui Li, Ryan Liu, Jouko Lönnqvist, Margarida Lopes, Valentina Lotchkova, Daniel MacArthur, Jonathan Marchini, John Maslen, Mangino Massimo, Iain Mathieson, Gaëlle Marenne, Peter McGuffin, Andrew McIntosh, Andrew G. McKechanie, Sarah Mettrus, Hannah Mitchison, Alireza Moayeri, James Morris, Francesco Munton, Kate Northstone, Michael O’Donovan, Alexandros Onoufriadis, Stephen O’Rahilly, Karim Oualkachaa, Michael J. Owen, Aarno Palotie, Kalliopi Panoutsopoulou, Victoria Parker, Jeremy R. Parr, Lavinia Paternoster, Tiina Paunio, Felicity Payne, Olli Pietilainen, Vincent Plagnol, Stephen Plom, Lydia Quaye, Michael A. Quail, Lucy Raymond, Karola Rehnström, Brent Richards, Susan Ring, Graham R.S. Ritchie, Nicola Roberts, David B. Savage, Peter Scambler, Stephen Schiffers, Miriam Schmids, Nadia Schoenmakers, Robert K. Semple, Eva Serra, Sally I. Sharp, So-Youn Shin, David Skuse, Kerrin Small, Lorraine Southam, Olivera Spasic-Boskovic, David St Clair, Jim Stalker, Elizabeth Stevens, Beate St Pourcian, Jianping Sun, Jaana Suvisaari, Ionna Tachmazidou, Martin D. Tobin, Ana Valdes, Margriet Van Kogelenberg, Parthiban Vijayarangakannan, Peter M. Visscher, Louise V. Wain, James T.R. Walters, Guangbiao Wang, Jun Wang, Yu Wang, Kirsten Ward, Elanor Wheeler, Tamieka Whyte, Hywel Williams, Kathleen A. Williamson, Crispian Wilson, Kim Wong, Changjiang Xu, Jian Yang, Feng Zhang, Pingbo Zhang.

21Department of Pathology, King Abdulaziz Medical City, Riyadh 11426, Saudi Arabia. 22Department of Psychiatry, Trinity Centre for Health Sciences, St James Hospital, James’s Street, Dublin 8, Ireland. 23Genetics and Genomic Medicine and Birth Defects Research Centre, UCL Institute of Child Health, London WC1N 1EH, UK. 24Division of Developmental Disabilities, Department of Psychiatry, Queen’s University, Kingston, Ontario, Canada K7L7X3. 25University of Cambridge Metabolic Research Laboratories, NIHR Cambridge Biomedical Research Centre, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge CB2 0QQ, UK. 26Department of Cardiovascular Medicine and Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. 27European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. 28Division of Psychiatry, The University of Edinburgh, Royal Edinburgh Hospital, Edinburgh EH10 5HF, UK. 29Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK. 30Institute of Psychiatry, King’s College London, 16 De Crespigny Park, London SE5 8AF, UK. 31NIHR BRC for Mental Health, Institute of Psychiatry and SLaM NHS Trust, King’s College London, 16 De Crespigny Park, London SE5 8AF, UK. 32Computational Biology and Genomics, Biogen Idec, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA. 33Division of Genetics and Molecular Genetics, King’s College London School of Medicine, Guy’s Hospital, London SE1 9RT, UK. 34BGI-Shenzhen, Shenzhen 518083, China. 35University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland 4102, Australia. 36MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, the University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. 37The Genome Centre, John Vane Science Centre, Queen Mary, University of London, Charterhouse Square, London EC1M 6BQ, UK. 38Cardiovascular Genetics, BHF Laboratories, Rayne Building, Institute Cardiovascular Sciences, University College London, London WC1E 6JY, UK. 39SUCLA David Geffen School of Medicine, Los Angeles, California 90095, USA. 40HelLEX—Centre for Health, Law and Emerging Technologies, Department of Public Health, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK. 41Department of Mathematical and Statistical Sciences, University of Colorado, Denver, Colorado 80220, USA. 42Adaptive Biotechnologies Corporation, Seattle, Washington 98102, USA. 43Human Genetics Research Centre, St George’s University of London SW17 0RE, UK. 44Behavioural and Brain Sciences Unit, UCL Institute of Child Health, London WC1N 1EH, UK. 45Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. 46BGI-Europe, London, UK. 47National Institute for Health and Welfare (THL), Helsinki, Finland. 48Welcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. 49Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02132, USA. 50Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK. 51Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. 52The Patrick Wild Centre, The University of Edinburgh, Edinburgh EH10 5HF, UK. 53The Department of Epidemiology and Biostatistics, Imperial College London, St Mary’s campus, Norfolk Place, Paddington, London W2 1PQ, UK. 54Department of Mathematics, Université de Québec À Montréal, Montréal, Québec, Canada PK-5151. 55Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki FI-00014, Finland. 56Institute of Neuroscience, Henry Wellcome Building for Neurobiology,
Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK. 73Department of Psychiatry, University of Helsinki, Helsinki, Finland. 74University College London (UCL) Genetics Institute (UGI), Gower Street, London WC1E 6BT, UK. 75ALSPAC School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK. 76Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK. 77School of Oral and Dental Sciences, University of Bristol, Lower Maudlin Street, Bristol BS1 2LY, UK. 78School of Experimental Psychology, University of Bristol, 12a Priory Road, Bristol BS8 1TU, UK. 79Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia. 80Departments of Health Sciences and Genetics, University of Leicester, Leicester LE1 6TP, UK. 81Department of Medicine, State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, 21 Sassoon Road, Hong Kong, China. 82Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark. 83Princess Al Jawhara Albrahim Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia. 84Macau University of Science and Technology, Avenida Wai Long, Taipa, Macau 999078, China.