*Human Molecular Genetics, 2009, Vol. 18, No. 9* ***1704–1713***
*doi:10.1093/hmg/ddp087*
*Advance Access published on February 25, 2009*

# Association of the thyroid stimulating hormone receptor gene (*TSHR*) with Graves' disease

**Oliver J. Brand[1], Jeffrey C. Barrett[2], Matthew J. Simmonds[1], Paul R. Newby[1], Christopher J. McCabe[1], Christopher K. Bruce[1], Boris Kysela[1], Jackie D. Carr-Smith[1], Thomas Brix[4], Penny J. Hunt[1], Wilmar M. Wiersinga[5], Laszlo Hegedüs[4], John Connell[6], John A.H. Wass[3], Jayne A. Franklyn[1], Anthony P. Weetman[7], Joanne M. Heward[1] and Stephen C.L. Gough[1,*]**

[1]School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, Institute of Biomedical Research, University of Birmingham, Edgbaston, Birmingham, West Midlands B15 2TT, UK, [2]Wellcome Trust Centre for Human Genetics and [3]Department of Endocrinology, Churchill Hospital, University of Oxford, Oxford, UK, [4]Department of Endocrinology and Metabolism, Odense University Hospital, Denmark, [5]Department of Endocrinology and Metabolism, Academic Medical Centre, University of Amsterdam, The Netherlands, [6]University of Glasgow, Glasgow, UK and [7]Medical School, University of Sheffield, Sheffield, UK

Received December 23, 2008; Revised and Accepted February 20, 2009

**Graves' disease (GD) is a common autoimmune disease (AID) that shares many of its susceptibility loci with other AIDs. The thyroid stimulating hormone receptor (TSHR) represents the primary autoantigen in GD, in which autoantibodies bind to the receptor and mimic its ligand, thyroid stimulating hormone, causing the characteristic clinical phenotype. Although early studies investigating the *TSHR* and GD proved inconclusive, more recently we provided convincing evidence for association of the *TSHR* region with disease. In the current study, we investigated a combined panel of 98 SNPs, including 70 tag SNPs, across an extended 800 kb region of the *TSHR* to refine association in a cohort of 768 GD subjects and 768 matched controls. In total, 28 SNPs revealed association with GD ($P < 0.05$), with strongest SNP associations at rs179247 ($\chi^2 = 32.45$, $P = 8.90 \times 10^{-8}$, OR = 1.53, 95% CI = 1.32–1.78) and rs12101255 ($\chi^2 = 30.91$, $P = 1.95 \times 10^{-7}$, OR = 1.55, 95% CI = 1.33–1.81), both located in intron 1 of the *TSHR*. Association of the most associated SNP, rs179247, was replicated in 303 GD families ($P = 7.8 \times 10^{-4}$). In addition, we provide preliminary evidence that the disease-associated genotypes of rs179247 (AA) and rs12101255 (TT) show reduced mRNA expression ratios of flTSHR relative to two alternate TSHR mRNA splice variants.**

## INTRODUCTION

Graves' disease (GD) is a common autoimmune disease (AID) affecting 0.5–1.0% of the general population, with autoantibodies targeting thyroid antigens (1,2). Genetic loci conferring risk to GD are shared among other AIDs, particularly type 1 diabetes and encode molecules vital for T cell activation and function. These include the HLA class I and class II genes on chromosome 6p21 (3,4), the *CTLA-4* region on chromosome 2q33 (5), the *PTPN22* region on chromosome 1p13 (6) and the recently identified *IL2Rα*/CD25 region on chromosome 10p15 (7). Despite a shared genetic contribution to autoimmunity, disease specific susceptibility loci are likely to contribute to the phenotypic differences between AIDs. A subject of debate in AID is the role of the target autoantigen in influencing or triggering the autoimmune process (8).

The thyroid stimulating hormone receptor (TSHR) is expressed on the surface of thyroid follicular cells and is the primary autoantigen in GD. Although early genetic studies provided inconclusive evidence of *TSHR* association with GD (9–20), recent studies by ourselves and a Japanese group have provided more convincing evidence of association in the *TSHR* region (21,22). Previously, we analysed 20 SNPs covering a 600 kb region encompassing *C14ORF145*/*TSHR*/

*To whom correspondence should be addressed. Tel: +44 1214158819; Fax: +44 1214158712; Email: s.c.gough@bham.ac.uk

*GTF2A1* in a GD case–control cohort consisting of 730 multinational autoimmune thyroid disease (AITD) patients [367 Hashimoto's thyroiditis (HT) and 363 GD] and 621 controls (21). We provided evidence for 10 single SNP associations with AITD spanning a 300 kb region around *TSHR* intron 1 and the neighbouring, *C14ORF145*, 5′ of the *TSHR* itself (21). Associations were also shown to be driven by GD and not HT patients (21). In addition, the Wellcome Trust Case Control Consortium (WTCCC) recently provided further evidence of association of the *TSHR* nsSNP, rs3783941 with GD, although since this study only investigated coding SNPs, the association signal within intron 1 was not examined (23).

Results from the previous investigations by our own group and others, strongly suggest the existence of a susceptibility locus within the *TSHR* region. The aim of the current study was, therefore, to refine *TSHR* association with GD, by analysing a panel of 98 SNPs across an extended 800 kb region of the *TSHR*, giving an average marker density of one SNP every 8 kb.

## RESULTS

### Identification of individual SNP and haplotype associations

The SNP panel consisted of 70 tag SNPs capturing most of the common genetic variation in the HapMap phase I CEU population across an extended 800 kb region of chromosome 14q31. Previously, we analysed a 600 kb region, however the analysis window was extended 100 kb 5′ and 3′ to ensure the association boundaries could be confidently defined. During the study, further HapMap genotype information became available (HapMap phase II) and so an additional 28 SNPs were selected to improve the amount of genetic variation captured. All SNPs were genotyped in 768 GD patients and 768 controls of UK European ancestry as described previously and all were within Hardy–Weinberg equilibrium bounds ($P > 0.05$) (Supplementary Material, Table S1) (24). The 768 GD patients and controls investigated in the current study were independent of the AITD (consisting of both GD and HT) and control cohorts investigated previously (21). In total, 28 SNPs revealed evidence of association with GD ($P < 0.05$), all of which are located within a 340 kb window spanning *TSHR* intron 1 and neighbouring *C14ORF145* (Fig. 1). The strongest SNP associations are all located within intron 1 of the *TSHR* and are separated by just 18.5 kb; these include rs179247 ($\chi^2 = 32.45$, $P = 8.90 \times 10^{-8}$, OR = 1.53, 95% CI = 1.32–1.78), rs3783948 ($\chi^2 = 28.50$, $P = 6.5 \times 10^{-5}$, OR = 1.52, 95% CI = 1.30–1.78) and rs12101255 ($\chi^2 = 30.91$, $P = 1.95 \times 10^{-7}$, OR = 1.55, 95% CI = 1.33–1.81) (Table 1). To further validate our data, the most associated SNP, rs179247, was investigated in our dataset of 303 GD families of which 160 were informative, revealing preferential transmission of the associated allele A in affected individuals, whereas unaffected family members exhibited an exactly equal transmission of alleles A and G ($P = 7.80 \times 10^{-4}$) (Table 2). SNPs located within *C14ORF145* also show strong evidence of association however with a slightly reduced magnitude, including rs759916 ($\chi^2 = 24.27$, $P = 5.3 \times 10^{-6}$, OR = 1.43, 95% CI = 1.23–1.67) and rs1458993 ($\chi^2 = 21.12$, $P = 2.6 \times 10^{-5}$, OR = 1.4, 95% CI = 1.20–1.63).

### Linkage disequilibrium and haplotype analysis

All linkage disequilibrium (LD) measures were calculated using the raw SNP genotype data generated in the present study. LD blocks were defined by the algorithm proposed by Gabriel *et al*. (25) and from each LD block haplotypes were constructed (Fig. 2). This analysis identified 10 haplotypes all of which were located within LD blocks 5–12 that revealed evidence of association with GD (Table 3). Haplotype number 4 (Table 3) in LD block 7, consisting of SNPs rs179247 with the protective G allele and rs179249 with the T allele present, revealed an increased frequency in controls (0.48) compared with GD (0.38) ($\chi^2 = 31.39$, $P = 2.1 \times 10^{-8}$). Haplotype number 5, also in LD block 7, consisting of the same two SNPs with both alleles switched, rs179247 (A) and rs179249 (C), revealed an increase in frequency in GD (0.46) compared with controls (0.38) ($\chi^2 = 19.82$, $P = 8.5 \times 10^{-6}$) (Table 3). Notably, with the exception of haplotype number 1 in LD block 5, all other major haplotype associations are located in LD blocks 7–12, within *TSHR* intron 1.

The majority of individual SNP and all haplotype associations are located between LD blocks 5 and 12 which covers a 250 kb region encompassing *C14ORF145* and *TSHR* intron 1 (Fig. 2). Importantly, between LD blocks 5 and 7 and similarly 3′ of LD block 12, there is a breakdown in LD with reduced $D'$ and $r^2$-values ($D' < 0.80$, $r^2 < 0.70$) suggesting that the strongest single SNP and haplotype association signals identified between LD blocks 7 and 12 may harbour the primary disease predisposing variant/s (Fig. 2). Interestingly, the strongest individual SNP associations, rs179247 and rs12101255, are within LD blocks 8 and 10, respectively, which seem to be located within a 40 kb region of strong but incomplete LD ($D' > 0.80$, $r^2 = 0.02–0.98$).

In order to better understand the pattern of association in this region, we performed a stepwise forward logistic regression analysis to assess whether there was evidence for independent effects, as opposed to one effect driving all signals via LD. Logistic regression was performed using the PLINK software package (26). We first tested all SNPs in the region, then re-analysed the data conditioning on the two most associated SNPs. We found that no additional SNPs remained associated after conditioning on either of the two most strongly associated SNPs (rs179247 and rs12101255). These two SNPs were each nominally significant after conditioning on the other, indicating that the causal variant is likely to be in strong, but not necessarily complete LD ($r^2 = 1.00$) with both of these SNPs. It is important to note that these two SNPs (rs179247 and rs12101255) share an $r^2 = 0.46$. On the basis of these results, it seems likely that the causal variant resides within the 40 kb region of *TSHR* intron 1. However, the incomplete LD and similar allele frequencies of the two most associated SNPs suggest that the causal variant is unlikely to be either of these SNPs.

### Correlated expression TSHR mRNA variants with associated risk alleles

Having identified two SNPs highly associated with GD and localized within a 40 kb region of distinct LD within *TSHR* intron 1, we sought to determine how these associated *TSHR*

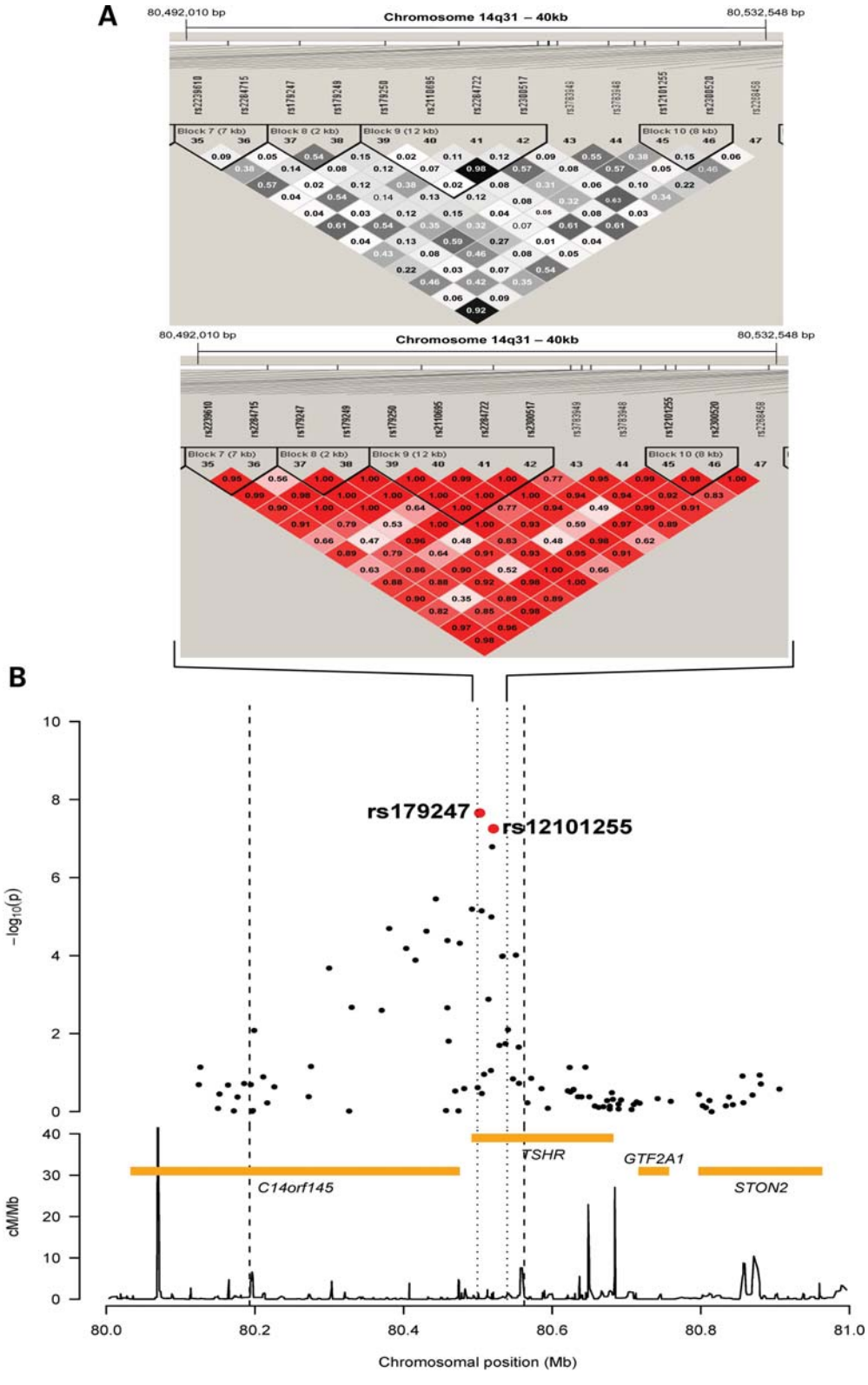


**Figure 1.** Case–control association results for all 98 SNPs. (**A**) Shows the $r^2$ (top LD chart) and $D'$ (lower LD chart) measures of pairwise linkage disequilibrium within the highly associated 40 kb region. The darker shades of black in the top LD chart represent larger $r^2$-values and values between each SNP pair are displayed in each block. The darker shades of red in the lower LD chart represent larger $D'$-values and actual $D'$-values between each SNP pair are displayed in each block. The LD displayed in both charts is calculated from the raw genotype data generated in the present study. (**B**) Displays log of the $\chi^2$ *P*-value in all 98 SNPs. The lower chart displays the recombination rate as estimated with HapMap data. Orange bars show genes in the region. Dashed vertical lines represent the greatest possible extent of LD around our association. The dotted vertical lines show the 40 kb region where the disease causing SNP in the region is likely to be located.

**Table 1.** Case–control association results for the three most highly associated SNPs

| | Genotype | Controls (%) | Graves' disease (%) | Genotype $\chi^2$ | *P*-value | Odds ratio | 95% Confidence intervals |
|---|---|---|---|---|---|---|---|
| rs179247 | A/A | 182 (27.4) | 279 (37.8) | 32.45 | $8.9 \times 10^{-8}$ | 1.53 | 1.32–1.78 |
| | A/G | 322 (48.5) | 359 (48.6) | | | | |
| | G/G | 160 (24.1) | 100 (13.6) | | | | |
| rs12101255 | C/C | 268 (41.1) | 197 (28.5) | 30.91 | $1.9 \times 10^{-7}$ | 1.55 | 1.33–1.81 |
| | C/T | 295 (45.2) | 345 (49.8) | | | | |
| | T/T | 89 (13.7) | 150 (21.7) | | | | |
| rs3783948 | A/A | 108 (16.6) | 61 (8.5) | 28.50 | $6.5 \times 10^{-5}$ | 1.52 | 1.30–1.78 |
| | A/G | 305 (46.8) | 314 (43.9) | | | | |
| | G/G | 238 (36.6) | 341 (47.6) | | | | |

Displays genotype counts (%) with $\chi^2$ and OR (95% confidence intervals) for three of the most highly associated SNPs, all located within the highly important 40 kb region of *TSHR* intron 1.

**Table 2.** Transmission of alleles A and G for SNP, rs179247 in Graves' disease affected and unaffected family members

| | Transmissions of A (%) | Transmissions of G (%) | Total transmissions |
|---|---|---|---|
| Graves' disease affected | 126 (61.8%) | 78 (38.2%) | 204 |
| Unaffected family members | 85 (50%) | 85 (50%) | 170 |

$2 \times 2$ Chi-squared: $\chi^2 = 5.22$, $P = 0.022$
TDT affected family members: $\chi^2 = 11.29$, $P = 7.8 \times 10^{-4}$
TDT unaffected family members: $\chi^2 = 0$, $P = 1$

intron 1 SNPs may affect biological mechanisms that could lead to GD onset. The human *TSHR* transcript was first cloned in 1990 (27). Northern blot analysis provided preliminary evidence that up to eight truncated *TSHR* mRNA transcripts may exist (28,29). By scrutinizing the literature, we identified two truncated mRNA transcripts, previously termed ST4 and ST5 that contain unique amino acid coding regions allowing them to be specifically detected by PCR (28,29). Primers designed specifically to ST4 and ST5 and subsequent sequencing of the PCR product confirmed their existence in a single sample of thyroid tissue (Supplementary Material, Fig. S1). Both ST4 and ST5 contain the first eight exons of the fl*TSHR* and neither contains exons 9 or 10, but has unique amino acid coding regions, utilizing different parts of intron 8. Both ST4 and ST5 share the same ATG start codon as the fl*TSHR* but have unique in frame stop codons that are spliced from the fl*TSHR*, suggesting mRNAs might be translated, although it is possible they are degraded. We measured the expression of fl*TSHR*, ST4 and ST5 mRNAs in 12 samples of thyroid tissue by real-time quantitative PCR, with the use of standard curves created with plasmids containing the relevant inserts for fl*TSHR*, ST4 and ST5. In each tissue sample, we genotyped the two most highly associated SNPs (rs179247 and rs12101255) as well as one showing no association with GD (rs7145447). Using the Kruscall Wallace test, we compared the expression ratios of flTSHR:ST4 and flTSHR:ST5 with genotype of the three different SNPs.

We found that genotypes of the GD-associated rs179247 were associated with disruption of the relative expression ratios of fl*TSHR* compared with ST4 (Fig. 3). The presence of the disease-associated AA genotype was associated with a reduced ratio of flTSHR:ST4, thus reducing the absolute amount of fl*TSHR* mRNA compared with ST4 (Fig. 3). The heterozygote AG genotype increased the expression of fl*TSHR* relative to ST4 and the GG genotype increased it still further ($P = 0.02$) (Fig. 3). A similar trend of disrupted expression ratio was seen between fl*TSHR* and ST5 for rs179247 genotypes, although not significant ($P = 0.071$). The other SNP highly associated with GD, rs12101255, also revealed evidence of disrupted fl*TSHR*/ST5 mRNA expression ratios with genotype. Specifically, the GD-associated TT genotype revealed a reduced flTSHR:ST5 ratio (reduced flTSHR relative to ST5). Similar to the results of rs179247, the AT heterozygote of rs12101255 increased the expression of flTSHR relative to ST5 and CC genotype increased it even further, with $P = 0.04$ (Supplementary Material, Fig. S2). No significant difference in expression ratios was seen between fl*TSHR* and ST4 or ST5 in the non-associated SNP (rs7145447). As a result of the relatively small number of 12 thyroid tissue samples available to us, we also attempted to measure flTSHR, ST4 and ST5 expression in peripheral blood lymphocytes from 12 GD patients, using the same methods as those employed for thyroid mRNA expression analysis. Expression of *TSHR* transcripts in these cells was however undetectable (data not shown but available upon request). Validation of the expression data of fl*TSHR* ST4 and ST5 will, therefore, need to come from further thyroid tissue, currently in the process of being collected in our laboratory.

## DISCUSSION

This study, including a panel of 98 SNPs capturing the majority of genetic variation in the HapMap CEU population, across a large 800 kb region of chromosome 14q31, encompassing *C14ORF145-TSHR-GTF2A1-STN2* provides further evidence for association of the *TSHR* region with GD. Although numerous associations were identified across a wide 340 kb region, the strongest single SNP associations were detected within *TSHR* intron 1. Similarly, haplotype associations support these results, with the most associated haplotypes also present within this region. Logistic regression analysis suggests that two SNPs, rs179247 and rs12101255, which are part of a 40 kb region of distinct LD, account for

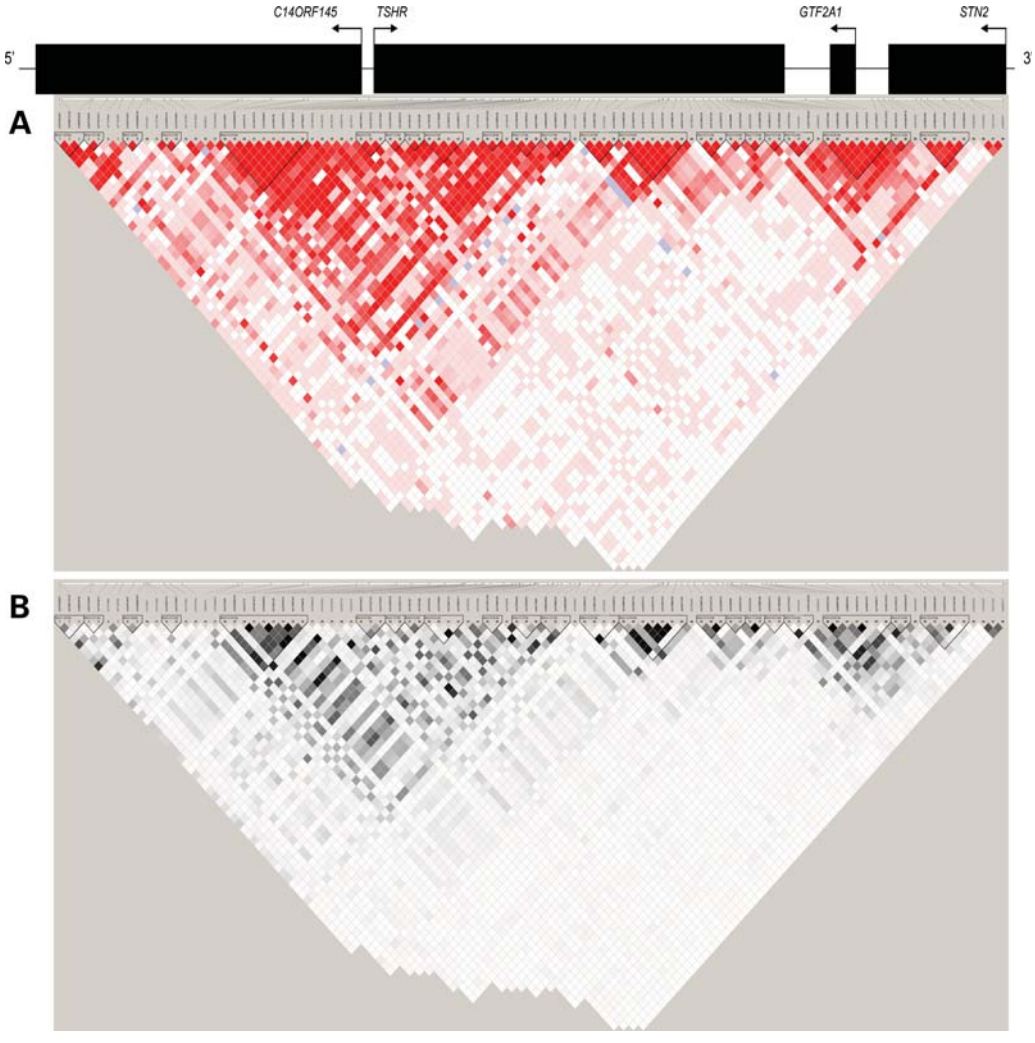


**Figure 2.** Linkage disequilibrium from the raw SNP genotype data generated in the present study between all SNPs investigated across the 800 kb region. The relative gene locations are displayed at the top of the figure. The black outlines on both LD plots represent LD blocks as defined by Gabriel *et al*. (25) algorithm. (**A**) Displays the LD parameter $D'$ where darker shades of red represent greater $D'$-values. (**B**) Shows the correlation coefficient of LD, $r^2$, darker shades of black represent larger $r^2$-values.

all association signals identified in the region. Logistic regression was unable to split association between rs179247 and rs12101255, despite incomplete LD ($D' = 0.92$, $r^2 = 0.46$). Investigation of HapMap phase II data (30) reveals a number of SNPs, tagged but not investigated in the current study, in strong LD ($r^2 > 0.80$) with rs17927, located within 5 kb of rs179247 itself within *TSHR* intron 1. In the case of rs12101255, HapMap phase II data show at least four additional SNPs in strong LD ($r^2 > 0.80$), all located within the proposed 40 kb region, separated by just 18.5 kb from rs12101255. Although analysis of these SNPs is not only attractive but also mandatory in our attempts at identifying the GD causal variant in this region, the aim of the current study was to refine a large region of disease association based on a tag SNP strategy.

Although the current data demonstrate the strongest signals of association to be within a 40 kb region of distinct LD, we are not only limited in our ability to determine the causal variant within this region but also unable to confidently conclude whether one or more disease causing variants are located outside this region. The all important next step in the localization of disease causing variants will require comprehensive re-sequencing and extensive high throughput genotyping of all SNPs with appropriate statistical analysis in large GD sample datasets, such as those now available in UK (7,23). The recent advances in accurate, high throughput whole genome sequencing technologies could be implemented within this region (31,32), or more likely, the first data release from the 1000 genomes project (http://www.1000genomes.org/page.php) may provide the required detailed information on the location and LD patterns of all structural variants. This could reduce the need for further re-sequencing of the region and help direct the location and extent of future genotyping. This will address some of the limitations created by potential associations outside blocks of strong association and the effects of rare causal variants with high penetrance. Although exact distances may be difficult to predict, it seems likely that further mapping will need to be extended

**Table 3.** Haplotypes showing association with $P < 1.0 \times 10^{-4}$

| Haplotype number | LD block | SNP | Allele | GD frequency | Control frequency | $\chi^2$ | *P*-value |
|---|---|---|---|---|---|---|---|
| 1 | 5 | rs327465 | T | 0.31 | 0.38 | 18.18 | $2.0 \times 10^{5}$ |
| | | rs10498551 | A | | | | |
| | | rs162171 | G | | | | |
| | | rs1197474 | G | | | | |
| | | rs1458993 | C | | | | |
| | | rs228120 | C | | | | |
| | | rs2556611 | G | | | | |
| | | rs2288347 | C | | | | |
| | | rs759916 | C | | | | |
| 2 | 7 | rs2239610 | G | 0.46 | 0.56 | 25.47 | $4.5 \times 10^{7}$ |
| | | rs2284715 | G | | | | |
| 3 | 7 | rs2239610 | C | 0.37 | 0.29 | 19.53 | $9.9 \times 10^{6}$ |
| | | rs2284715 | G | | | | |
| 4 | 8 | rs179247 | G | 0.38 | 0.48 | 31.39 | $2.1 \times 10^{8}$ |
| | | rs179249 | T | | | | |
| 5 | 8 | rs179247 | A | 0.46 | 0.38 | 19.82 | $8.5 \times 10^{6}$ |
| | | rs179249 | C | | | | |
| 6 | 9 | rs179250 | G | 0.32 | 0.40 | 22.74 | $1.9 \times 10^{6}$ |
| | | rs2110695 | C | | | | |
| | | rs2284722 | G | | | | |
| | | rs2300517 | T | | | | |
| 7 | 10 | rs12101255 | C | 0.54 | 0.63 | 26.63 | $2.5 \times 10^{7}$ |
| | | rs2300520 | G | | | | |
| 8 | 10 | rs12101255 | T | 0.35 | 0.28 | 15.48 | $8.4 \times 10^{5}$ |
| | | rs2300520 | G | | | | |
| 9 | 11 | rs2268460 | G | 0.34 | 0.42 | 29.97 | $4.4 \times 10^{8}$ |
| | | rs2268462 | A | | | | |
| | | rs917986 | A | | | | |
| 10 | 12 | rs4903965 | G | 0.34 | 0.42 | 21.13 | $4.3 \times 10^{6}$ |
| | | rs2300521 | A | | | | |
| | | rs4903967 | A | | | | |

Displays haplotypes that show a difference in frequency between GD patients compared with controls with $P < 1.0 \times 10^{-4}$.

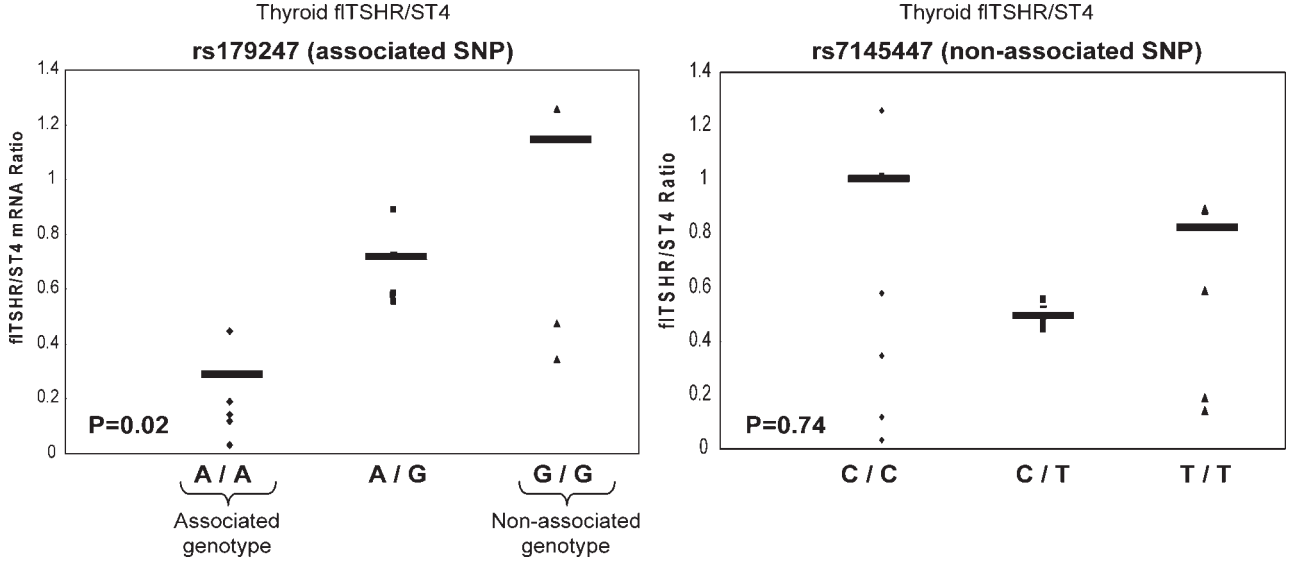


**Figure 3.** TSHR mRNA isoform expression ratios for SNPs rs179247 and rs7145447. Absolute expression of full length *TSHR* mRNA relative to *TSHR* ST4, expressed as a ratio (flTSHR/ST4). Expression ratios are split by genotype of two *TSHR* SNPs rs179247 (most associated SNP) and rs7145447 (non-associated SNP). Dark horizontal lines display the ratios of the average fl*TSHR* expression level relative to the average absolute expression level of ST4. Kruscall Wallace test *P*-values comparing the difference in expression ratios between genotype are displayed on the bottom left corner of each chart.

both 5′ and 3′ of our 40 kb block to include the 250 kb region harbouring the strongest haplotype association signals, encompassing both *C14ORF145* and *TSHR*.

Although previous small-sized association studies proved inconclusive (9–20), a more recent and detailed study performed in a Japanese GD cohort investigated 22 intronic SNPs across the *TSHR* region and identified single SNP associations with GD primarily located within intron 7 of the *TSHR* (22). In the current study, we have captured the majority of common genetic variation within intron 7 and identified no evidence of association with GD in our cohort of UK European ancestry. Moreover, we directly analysed the *TSHR* intron 7 SNP, rs2300530, and found no evidence for association in our cohort ($P = 0.70$). This once again serves to highlight the presence of geographical genetic variation and the need for detailed analysis in different populations.

The recently published WTCCC non-synonymous SNP study has also replicated association between GD and the *TSHR* region (rs3783941) in an independent UK cohort (23). We did not type rs3783941, although it is tagged by rs3783942 ($D' = 1.00$, $r^2 = 1.00$), and this SNP was not associated with GD in our study ($\chi^2 = 3.97$, $P = 0.14$). Interestingly the initial WTCCC analysis comparing 1000 GD cases and 1500 healthy controls revealed no evidence of GD association with the *TSHR* nsSNP rs3783941 (23). The larger control reference panel however used by the WTCCC, which included genotypes from three other diseases (1000 cases from each of ankylosing spondylitis, multiple sclerosis and breast cancer) in addition to the 1500 controls (4500 individuals), provided a sufficiently sized cohort for comparison with the GD cohort to detect association at rs3783941 with GD ($P = 2.1 \times 10^{-5}$) (23). Although stringent control measures involving comparisons with the different HapMap populations were implemented to ensure all individuals were of Western European ancestry, all GD cases and controls could not be matched by exact UK geographic location, which may explain the differences observed between the WTCCC and the current study (23).

Our preliminary *TSHR* mRNA expression analysis in human thyroid tissue of fl*TSHR* and two truncated *TSHR* mRNA isoforms (ST4 and ST5), suggests that the highly associated SNPs (rs179247 and rs12101255) are associated with changes in expression levels of ST4 and ST5 relative to flTSHR, and may support a previously proposed hypothesis for disease pathogenesis. The TSHR protein can undergo a post-translation intramolecular cleavage event resulting in shedding of the TSHR A-subunit. It has been suggested that this may initiate or exacerbate autoimmunity in GD. This is based on observations that TSHR autoantibodies preferentially target the extracellular A-subunit (33) and that intramuscular injection of TSHR A-subunit into a mouse model is an absolute requirement to induce the production of TSHR autoantibodies and hyperthyroidism (34). The truncated *TSHR* mRNA transcripts, ST4 and ST5 investigated in the current study, possess exons 1–8 and a ninth exon unique to each of ST4 and ST5 (retained introns) and, therefore, if translated could produce a soluble A-subunit directly. Excess mRNA expression of *TSHR* ST4 or ST5 may result in an increase in soluble A-subunit protein, independently of post-translational cleavage and shedding. It has been proposed that excessive A-subunit could influence autoantibody production. The expression of ST4 and ST5 would result in even higher levels of soluble A-subunit in the periphery, thus increasing the chances of autoantibody production against the TSHR. As a result, however, of limited access to thyroid tissue additional *TSHR* expression studies were performed on peripheral blood lymphocytes. Unfortunately, there was no evidence of expression of the different *TSHR* mRNA transcripts in these cells. At the present time, therefore, caution must be exercised in the interpretation of these data until further samples are able to be analysed. Moreover, although the association between the associated intron 1 SNPs and transcript expression ratios are interesting, concluding a causal link to disease is not possible until the exact disease causing DNA variants have been identified by further fine-mapping studies.

Our findings in GD highlight some of the future challenges facing association studies in all common diseases when moving from genetic association to the identification of aetiological variants. Extensive LD across the genome is likely to lead to the identification of a handful of variants with similar magnitudes of association that cannot easily be split by regression analysis. The similar effect size and magnitude of associations identified in the current study between SNPs rs179247 and rs12101255 illustrates this problem. This maybe resolved by performing future fine-mapping studies in multiple populations including, for example, those of African or recent African origin, due to the commonly observed reduced LD in these populations (35–37). Employing such a strategy may reveal a smaller subset of SNPs with stronger association (38). Furthermore, detailed-SNP screens such as that performed in the present study, based on common variants, are generally limited in their ability to capture genetic variation of rare SNPs, with minor allele frequencies $<5\%$ (38). The potential role and penetrance of such rare variants are presently unclear although, as previously mentioned, information from the 1000 genomes project and further case–control genotyping in large disease cohorts may help resolve this issue.

In summary, we provide evidence that SNPs within a 40 kb region of *TSHR* intron 1 are highly associated with GD in UK. Although we have not excluded an effect outside intron 1 and the mechanisms underlying the association are unclear, we have preliminary evidence to suggest that the most disease-associated SNPs could be associated with expression of the variant *TSHR* mRNA isoforms and hence the expression of TSHR proteins. Further genetic and functional studies are now required to identify the most important associations in this region to determine the consequences of changes in *TSHR* expression and the development of the autoimmune process in GD.

## MATERIALS AND METHODS

### Subjects

For the case–control cohort, we used 768 UK GD patients of European UK ancestry recruited from thyroid clinics in Birmingham, Bournemouth, Exeter and Walsall. GD patients were defined by the presence of biochemical hyperthyroidism

together with the presence of either thyroid eye disease (NOSPECS classification 2–6) or with a diffuse goitre and a significant titre of microsomal, thyroglobulin or TSHR autoantibodies. Seven hundred and sixty-eight control subjects with no personal or family history of AID were recruited from various sites including the blood transfusion service, Birmingham Heartlands Hospital and Queen Elizabeth Hospital, Birmingham. All control subjects had grandparents of European UK ancestry and were age and sex matched with the GD patients. For the family analysis, 303 European UK ancestry GD patients were recruited from the same clinics, along with both parents and unaffected siblings, as described previously (39).

For *TSHR* mRNA isoform expression analysis, 12 samples of thyroid tissue were collected. These samples were collected at the time of thyroidectomy and were surplus to pathological requirements; all studies were performed on histologically normal European ancestry samples.

We obtained informed consent from all participant patients and the study was approved by the local Ethics Committee.

### SNP selection

To define the boundaries and refine the region of association by capturing the majority genetic variation, SNP genotype information was downloaded in December 2005 (HapMap phase I) from the HapMap CEU population from an extended 800 kb region of chromosome 14. HapMap genotypes were analysed within Haploview (40) and Tag SNPs were selected using the Tagger (41) function. Seventy tag SNPs were selected with $r^2 \geq 0.80$ to capture 80% of genotype information in the region. The average tag SNP $r^2 = 0.934$. During the course of the study, additional genetic information became available (HapMap phase II) and a further 28 SNPs were selected to improve the amount of genetic variation captured.

The most associated SNP, rs179247, was investigated in our family GD cohort ($n = 303$) to confirm association.

The two most highly associated SNPs (rs179247 and rs12101255) and one non-associated SNP (rs7145447) were selected for genotyping in our 12 thyroid samples for correlation with our gene expression data.

### Genotyping

All SNP genotyping in case–control, family and expression datasets were genotyped using TaqMan® SNP genotyping chemistries (Applied Biosystems, UK). All SNPs were genotyped with a success rate >95% for all SNPs.

### Total RNA extraction from thyroid tissue

Total thyroid RNA was extracted after homogenization using the Tri-reagent Kit (Sigma, UK), following the manufacturers protocol, which is a single step acid guanidinium phenol–chloroform extraction procedure. RNA quality was assessed spectrophotometically and cDNA quality was assessed by checking the quality of 18s amplification curves during the subsequent real-time PCR. Any samples that did not have a smooth looking amplification curve with clear early amplification, exponential and plateau phases were removed. RNA was reverse transcribed in a 20 μl reaction containing 1 μg of total RNA using the high-capacity cDNA reverse transcription kit (Applied Biosystems).

### Confirming TSHR ST4 and ST5 existence

Early experiments using northern blotting procedures have suggested the existence of up to eight TSHR mRNA variants (28,42). Most are truncated isoforms of wild-type flTSHR with unique 3′-UTRs. According to Hunt *et al*. (42), TSHR ST4 and ST5 reveal unique amino acid coding regions, which could be detected feasibly by PCR and sequencing. Forward primers were designed to span the *TSHR* exon 1–2 boundary and reverse primers were designed in a unique amino acid coding regions of ST4 and ST5, respectively, which after sequencing of the PCR product confirmed their existence (Supplementary Material, Fig. S1). We found TSHR ST4 contains the first eight exons of the flTSHR, then splices 159 bp of intron 8, then utilizes the next 66 bp before a stop codon. Suggesting, the novel 66 bp on the 3′ end is a novel exon. TSHR ST5 also utilizes the first eight exons and then splices 28 418 bp of intron 8 before utilizing the next 61 bp on its 3′ end.

### Absolute quantification by TaqMan real-time PCR

To detect flTSHR, we obtained a gene expression assay designed to span *TSHR* exons 9 and 10 (Applied Biosystems) (Supplementary Material, Fig. S1). TaqMan primers and probes specific to TSHR ST4 and ST5 were designed using primer express software® (Applied Biosystems) (Supplementary Material, Fig. S1).

To create a standard curve for real-time absolute quantification (aqPCR) of the flTSHR, we obtained a flTSHR clone in pcDNA3.1 (Missouri S&T cDNA resource centre). For absolute quantification of TSHR ST4 and ST5, we cloned the TaqMan PCR products into the pGEM® T-Easy plasmid, as per the manufacturers protocol (Promega, UK). All clones were grown up and maxipreped using the Genopure plasmid maxi kit using the manufacturers protocol (Roche Applied Science, UK). Each plasmid was sequenced to check the insert contained the correct sequence with no introduced mutations. Plasmid concentrations were determined using picogreen dye (invitrogen, UK) and read at $A_{260}$ on a Wallac Victor III multilabel counter (Perkin Elmer, UK). Each plasmid was diluted to a working concentration of 10 ng/μl. Eight further serial dilutions were preformed to create the standard curve (1:10, $1{:}10^1$, $1{:}10^2$, $1{:}10^3$, $1{:}10^4$, $1{:}10^5$, $1{:}10^6$, $1{:}10^7$, $1{:}10^8$), therefore plasmid concentrations ranged from 10 ng/μl to 0.0001 fg/μl. The molecular weight of each plasmid was determined using the following equation:

$$M = [n]\left[1.096 \times 10^{-21}\,\mathrm{g/bp}\right]$$

where *M*, mass; *n,* plasmid size (bp). The size of pcDNA3.1 including the flTSHR insert was 7723 bp, the size of PGEM clones containing ST4 or ST5 inserts were 3095 and 3102 bp, respectively (equation details from Applied Biosystems product literature entitled ‘creating standard curves

with genomic DNA or plasmid DNA templates for use in quantitative PCR'). After determining the molecular weight of one plasmid, it was then possible to determine the number of plasmid copies in each dilution series. The real-time aqPCR was performed on the Applied Biosystems 7900 HT sequence detection system (Applied Biosystems). The standard curve was created in the aqPCR function of the Applied Biosystems SDS 2.2 software where each Ct-value correlates with a known quantity or copy number of plasmid. Each dilution series of the standard curve was run in triplicate, and all thyroid target samples were analysed in duplicate. For each sample, the Ct-value was normalized to the standard curve to give an absolute copy number.

### Statistical analysis

All genotyping results were tested for Hardy–Weinberg equilibrium to ensure the accuracy of genotyping. Single SNP associations were tested using the $\chi^2$-test with the statistical package MINITAB (MINITAB Release 14, ©1994, Minitab Inc., PA, USA). A result was deemed significant with a $P < 0.05$ and all $P$-values were uncorrected. The transmission disequilibrium test was used to confirm association of results in our family dataset. An excess of transmissions of the associated allele from heterozygous parents to affected offspring, where $P < 0.05$ was taken as confirmation of association.

Pairwise $D'$ and $r^2$ measures of LD were calculated between all SNP results within Haploview version 3.4 (40). LD blocks were defined by the default algorithm within Haploview proposed by Gabriel *et al*. (25). Logistic regression was performed using the PLINK software package (26). We first tested all SNPs in the region, then re-analysed the data conditioning on the most associated SNP. We repeated this process once more for the only remaining associated SNP. No additional SNPs were even nominally associated after conditioning on these two.

Absolute expression values for TSHR ST4 and ST5 were compared with a ratio of the flTSHR expression values. Expression ratios of flTSHR:TSHR ST4 or flTSHR:TSHR ST5 were split by the three genotypes in five of the SNPs (rs179247, rs12101255, rs3783948, rs3783949 and rs7145447). The Kruscall Wallace test was implemented within MINITAB (MINITAB Release 14, ©1994, Minitab Inc.) and used to test for significant differences between expression ratios with genotype. Differences in ratios with genotype were deemed significant with a $P < 0.05$.

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

## ACKNOWLEDGEMENTS

We thank all doctors and nurses for recruiting patients.

*Conflict of Interest statement*. None of the authors have any conflicts of interest.

## FUNDING

We would like to thank the Wellcome Trust for funding this study.

## REFERENCES

1. Gough, S.C. (2000) The genetics of Graves' disease. *Endocrinol. Metab. Clin. North Am.*, **29**, 255–266.
2. Tunbridge, W.M., Evered, D.C., Hall, R., Appleton, D., Brewis, M., Clark, F., Evans, J.G., Young, E., Bird, T. and Smith, P.A. (1977) The spectrum of thyroid disease in a community: the Whickham survey. *Clin. Endocrinol. (Oxf)*, **7**, 481–493.
3. Simmonds, M.J., Howson, J.M., Heward, J.M., Cordell, H.J., Foxall, H., Carr-Smith, J., Gibson, S.M., Walker, N., Tomer, Y., Franklyn, J.A. *et al.* (2005) Regression mapping of association between the human leukocyte antigen region and Graves disease. *Am. J. Hum. Genet.*, **76**, 157–163.
4. Simmonds, M.J., Howson, J.M., Heward, J.M., Carr-Smith, J., Franklyn, J.A., Todd, J.A. and Gough, S.C. (2007) A novel and major association of HLA-C in Graves' disease that eclipses the classical HLA-DRB1 effect. *Hum. Mol. Genet.*, **16**, 2149–2153.
5. Ueda, H., Howson, J.M., Esposito, L., Heward, J., Snook, H., Chamberlain, G., Rainbow, D.B., Hunter, K.M., Smith, A.N., Di Genova, G. *et al.* (2003) Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature*, **423**, 506–511.
6. Smyth, D., Cooper, J.D., Collins, J.E., Heward, J.M., Franklyn, J.A., Howson, J.M., Vella, A., Nutland, S., Rance, H.E., Maier, L. *et al.* (2004) Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. *Diabetes*, **53**, 3020–3023.
7. Brand, O.J., Lowe, C.E., Heward, J.M., Franklyn, J.A., Cooper, J.D., Todd, J.A. and Gough, S.C. (2007) Association of the interleukin-2 receptor alpha (IL-2Ralpha)/CD25 gene region with Graves' disease using a multilocus test and tag SNPs. *Clin. Endocrinol. (Oxf)*, **66**, 508–512.
8. Eisenberg, R. (2005) Do autoantigens define autoimmunity or vice versa? *Eur. J. Immunol.*, **35**, 367–370.
9. Cuddihy, R.M., Dutton, C.M. and Bahn, R.S. (1995) A polymorphism in the extracellular domain of the thyrotropin receptor is highly associated with autoimmune thyroid disease in females. *Thyroid*, **5**, 89–95.
10. Kotsa, K.D., Watson, P.F. and Weetman, A.P. (1997) No association between a thyrotropin receptor gene polymorphism and Graves' disease in the female population. *Thyroid*, **7**, 31–33.
11. Allahabadia, A., Heward, J.M., Mijovic, C., Carr-Smith, J., Daykin, J., Cockram, C., Barnett, A.H., Sheppard, M.C., Franklyn, J.A. and Gough, S.C. (1998) Lack of association between polymorphism of the thyrotropin receptor gene and Graves' disease in UK and Hong Kong Chinese patients: case control and family-based studies. *Thyroid*, **8**, 777–780.
12. Gabriel, E.M., Bergert, E.R., Grant, C.S., van Heerden, J.A., Thompson, G.B. and Morris, J.C. (1999) Germline polymorphism of codon 727 of human thyroid-stimulating hormone receptor is associated with toxic multinodular goiter. *J. Clin. Endocrinol. Metab.*, **84**, 3328–3335.
13. Simanainen, J., Kinch, A., Westermark, K., Winsa, B., Bengtsson, M., Schuppert, F., Westermark, B. and Heldin, N.E. (1999) Analysis of mutations in exon 1 of the human thyrotropin receptor gene: high frequency of the D36H and P52T polymorphic variants. *Thyroid*, **9**, 7–11.
14. Kaczur, V., Takacs, M., Szalai, C., Falus, A., Nagy, Z., Berencsi, G. and Balazs, C. (2000) Analysis of the genetic variability of the 1st (CCC/ACC, P52T) and the 10th exons (bp 1012-1704) of the TSH receptor gene in Graves' disease. *Eur. J. Immunogenet.*, **27**, 17–23.
15. Chistyakov, D.A., Savost'anov, K.V., Turakulov, R.I., Petunina, N.A., Trukhina, L.V., Kudinova, A.V., Balabolkin, M.I. and Nosikov, V.V. (2000) Complex association analysis of Graves' disease using a set of polymorphic markers. *Mol. Genet. Metab.*, **70**, 214–218.
16. Watson, P.F., French, A., Pickerill, A.P., McIntosh, R.S. and Weetman, A.P. (1995) Lack of association between a polymorphism in the coding region of the thyrotropin receptor gene and Graves' disease. *J. Clin. Endocrinol. Metab.*, **80**, 1032–1035.
17. Ho, S.C., Goh, S.S. and Khoo, D.H. (2003) Association of Graves' disease with intragenic polymorphism of the thyrotropin receptor gene in a cohort of Singapore patients of multi-ethnic origins. *Thyroid*, **13**, 523–528.
18. Ban, Y., Greenberg, D.A., Concepcion, E.S. and Tomer, Y. (2002) A germline single nucleotide polymorphism at the intracellular domain of

the human thyrotropin receptor does not have a major effect on the development of Graves' disease. *Thyroid*, **12**, 1079–1083.
19. Chistiakov, D.A., Savost'anov, K.V., Turakulov, R.I., Petunina, N., Balabolkin, M.I. and Nosikov, V.V. (2002) Further studies of genetic susceptibility to Graves' disease in a Russian population. *Med. Sci. Monit.*, **8**, CR180–184.
20. Muhlberg, T., Herrmann, K., Joba, W., Kirchberger, M., Heberling, H.J. and Heufelder, A.E. (2000) Lack of association of nonautoimmune hyperfunctioning thyroid disorders and a germline polymorphism of codon 727 of the human thyrotropin receptor in a European Caucasian population. *J. Clin. Endocrinol. Metab.*, **85**, 2640–2643.
21. Dechairo, B.M., Zabaneh, D., Collins, J., Brand, O., Dawson, G.J., Green, A.P., Mackay, I., Franklyn, J.A., Connell, J.M., Wass, J.A. *et al.* (2005) Association of the *TSHR* gene with Graves' disease: the first disease specific locus. *Eur. J. Hum. Genet.*, **13**, 1223–1230.
22. Hiratani, H., Bowden, D.W., Ikegami, S., Shirasawa, S., Shimizu, A., Iwatani, Y. and Akamizu, T. (2005) Multiple SNPs in intron 7 of thyrotropin receptor are associated with Graves' disease. *J. Clin. Endocrinol. Metab.*, **90**, 2898–2903.
23. Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H. *et al.* WTCCC (2007) Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.*, **39**, 1329–1337.
24. Simmonds, M.J., Heward, J.M., Carr-Smith, J., Foxall, H., Franklyn, J.A. and Gough, S.C. (2006) Contribution of single nucleotide polymorphisms within FCRL3 and MAP3K7IP2 to the pathogenesis of Graves' disease. *J. Clin. Endocrinol. Metab.*, **91**, 1056–1061.
25. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2229.
26. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
27. Frazier, A.L., Robbins, L.S., Stork, P.J., Sprengel, R., Segaloff, D.L. and Cone, R.D. (1990) Isolation of TSH and LH/CG receptor cDNAs from human thyroid: regulation by tissue specific splicing. *Mol. Endocrinol.*, **4**, 1264–1276.
28. Kakinuma, A. and Nagayama, Y. (2002) Multiple messenger ribonucleic acid transcripts and revised gene organization of the human TSH receptor. *Endocr. J.*, **49**, 175–180.
29. Hunt, N., Wiley, K.P., Abend, N., Balvers, M., Jahner, D., Northemann, W. and Ivell, R. (1995) Novel splicing variants of the human thyrotropin receptor encode truncated polypeptides without a membrane-spanning domain. *Endocrine*, **3**, 233–240.
30. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
31. Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R. *et al.* (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**, 53–59.
32. Bentley, D.R. (2006) Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.*, **16**, 545–552.
33. Costagliola, S., Khoo, D. and Vassart, G. (1998) Production of bioactive amino-terminal domain of the thyrotropin receptor via insertion in the plasma membrane by a glycosylphosphatidylinositol anchor. *FEBS Lett.*, **436**, 427–433.
34. Chen, C.R., Pichurin, P., Nagayama, Y., Latrofa, F., Rapoport, B. and McLachlan, S.M. (2003) The thyrotropin receptor autoantigen in Graves disease is the culprit as well as the victim. *J. Clin. Invest.*, **111**, 1897–1904.
35. Nalls, M.A., Wilson, J.G., Patterson, N.J., Tandon, A., Zmuda, J.M., Huntsman, S., Garcia, M., Hu, D., Li, R., Beamer, B.A. *et al.* (2008) Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. *Am. J. Hum. Genet.*, **82**, 81–87.
36. Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R. *et al.* (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, **451**, 998–1003.
37. Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., Ng, D.P., Holmkvist, J., Borch-Johnsen, K., Jorgensen, T. *et al.* (2008) SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.*, **40**, 1039–1040.
38. McCarthy, M.I. and Hirschhorn, J.N. (2008) Genome-wide association studies: potential next steps on a genetic journey. *Hum. Mol. Genet.*, **17**, R156–R165.
39. Heward, J.M., Allahabadia, A., Daykin, J., Carr-Smith, J., Daly, A., Armitage, M., Dodson, P.M., Sheppard, M.C., Barnett, A.H., Franklyn, J.A. *et al.* (1998) Linkage disequilibrium between the human leukocyte antigen class II region of the major histocompatibility complex and Graves' disease: replication using a population case control and family-based study. *J. Clin. Endocrinol. Metab.*, **83**, 3394–3397.
40. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
41. de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J. and Altshuler, D. (2005) Efficiency and power in genetic association studies. *Nat. Genet.*, **37**, 1217–1223.
42. Nicholas Hunt, K.P.W., Abend, N., Balvers, M., Jahner, D., Northemann, W. and Ivell, R. (1995) Novel splicing variants of the human thyrotropin receptor encode truncated polypeptides without a membrane-spanning domain. *Endocrine*, **3**, 233–240.