

ARTICLE

Association of the TSHR gene with Graves' disease: the first disease specific locus

Bryan M Dechairo¹, Delilah Zabaneh¹, Joanne Collins², Oliver Brand², Gary J Dawson¹, Angie P Green¹, Ian Mackay¹, Jayne A Franklyn², John M Connell³, John AH Wass⁴, Wilmar M Wiersinga⁵, Laszlo Hegedus⁶, Thomas Brix⁶, Bruce G Robinson⁷, Penny J Hunt⁸, Anthony P Weetman⁹, Alisoun H Carey¹ and Stephen C Gough^{*,2}

¹Target Discovery, Oxagen Ltd., Abingdon OX14 4RY, UK; ²Division of Medical Sciences, Institute of Biomedical Research, University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham B15 2TH, UK; ³University Department of Medicine & Therapeutics, Western Infirmary, Glasgow, G11 6NT, UK; ⁴Department of Metabolic Medicine, Nuffield Orthopaedic Centre, Oxford, OX3 7LD, UK; ⁵Department of Endocrinology, University Hospital Academisch Medisch Centrum, University of Amsterdam, PO Box 22700, 1100 DE Amsterdam, The Netherlands; ⁶Department of Endocrinology and Metabolism, Odense University Hospital, DK-5000 Odense C, Denmark; ⁷Kolling Institute and Department of Endocrinology, Royal North Shore Hospital, University of Sydney, Australia; ⁸Department of Endocrinology, Christchurch Hospital, Private Bag 4710, Christchurch, NZ; ⁹University of Sheffield Clinical Sciences Centre, Northern General Hospital, Sheffield S5 7AU, UK

The development of autoimmune thyroid disease (AITD) is associated with autoantibodies directed against the thyroid stimulating hormone receptor (TSHR). Previous studies have failed to demonstrate a consistent association between the *TSHR* and AITD, or any of its sub-phenotypes. In the present study, we analysed the linkage disequilibrium (LD) structure encompassing the *TSHR*, to identify LD 'blocks' and SNPs, which capture the majority of intra-block haplotype diversity. The haplotype tagging SNPs, plus all common SNPs reported in previous studies were genotyped in 1059 AITD Caucasian cases and 971 Caucasian controls. A haplotype, across two LD blocks, showed association ($P < 1 \times 10^{-6}$, OR 1.7) with Graves' disease (GD) but not autoimmune hypothyroidism (AIH). We replicated these findings by genotyping the most associated GD SNP, rs2268458, in a separate UK Caucasian cohort of 1366 AITD cases and 1061 controls (GD, $P = 2 \times 10^{-6}$, OR 1.3; AIH, $P = \text{NS}$). These results in two independent Caucasian data sets suggest that the *TSHR* is the first replicated GD-specific locus meriting further fine mapping and functional analysis to identify the aetiological variants.

European Journal of Human Genetics (2005) 13, 1223–1230. doi:10.1038/sj.ejhg.5201485;
published online 17 August 2005

Keywords: TSHR; GD; AITD; Graves; genetic; htSNP

Introduction

Autoimmune thyroid disease (AITD) is a common disorder, encompassing both Graves' disease (GD) and autoimmune

hypothyroidism (AIH) each with a prevalence of around 1%.¹ GD manifests through the production of autoantibodies that bind and stimulate the thyroid-stimulating hormone receptor (TSHR) resulting in hyperthyroidism.² Similarly, in AIH autoimmune targeting of the thyroid leads to tissue destruction and hypothyroidism, although less than 15% of cases are believed to be due to the effect of blocking autoantibodies against the TSHR² with other thyroid autoantibodies likely to be exerting a primary disease-causing effect.³

*Correspondence: SC Gough, Division of Medical Sciences, Institute of Biomedical Research, University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham B15 2TH, UK.

Tel: +44 121 4243593; Fax: +44 121 4240593;

E-mail: s.c.gough@bham.ac.uk

Received 8 March 2005; revised 26 May 2005; accepted 15 July 2005;

published online 17 August 2005

Evidence for a genetic predisposition to AITD is well established from familial clustering and twin studies, with sibling-risk estimates of between 5 and 10 for individual diseases.^{4,5} Genome scans across extended families and affected sibling pair cohorts have detected linkage to several chromosome regions, but have yet to convincingly replicate a single locus.^{6,7} Candidate genes studies have met with greater success, with replicated associations between AITD and HLA on chromosome 6p21,⁸ *CTLA4* on chromosome 2q33⁹ and *LYP* on chromosome 1p13.^{10,11}

The well-known biological role of the TSHR and the link between TSHR autoantibodies and disease have made the *TSHR* an obvious candidate for AITD. While family studies have yet to confirm the linkage between the *TSHR* region on chromosome 14 and an AITD phenotype,^{6,7,12} several small association studies have examined known amino-acid residue changing polymorphisms within the *TSHR*^{13–21} and produced conflicting results.

In the present study, we sought to determine whether the *TSHR* is a susceptibility locus for an AITD phenotype. First, by using 40 single nucleotide polymorphisms (SNPs), we defined linkage disequilibrium (LD) 'blocks' across the *TSHR* genomic interval and subsequently selected haplotype tagging SNPs (htSNPs).^{22,23} We then conducted a two-stage case-control association study (1059 AITD cases vs 971 controls) using the htSNPs as well as polymorphisms across the *TSHR* previously analysed by other groups.^{13–21} Finally, we replicated the most associated SNP in an independent cohort (1366 AITD cases and 1061 controls).

Materials and methods

Study design

For assay validation and LD analysis, 86 control samples from the UK were genotyped. In order to limit the amount of genotyping, save DNA resources and decrease reagent expenditure, a two-staged approach was used in the initial association study. For Stage 1, 366 AITD cases (183 AIH and 183 GD) and 350 controls were genotyped. Only SNPs showing association, $P < 0.05$, in the Stage 1 analysis were taken forward to Stage 2. For Stage 2, 730 AITD cases (367 AIH and 363 GD) and 621 controls were genotyped. Both index cases and controls were selected randomly from their respective cohorts. The most associated SNP from the combined data set was then analysed in the UK replication cohort.

Subjects (initial study 1)

Unrelated index cases with GD and AIH were taken from 1119 relative pairs with AITD. The relative pairs were collected as part of an international collaboration from eight participating centres in the UK, Continental Europe (Denmark) and Australasia (Australia and New Zealand) all with similar prevalence data for AITD. The study received

ethical approval from each institution's local review ethics committee and additionally from the regional multi-centre ethics committees where appropriate. Written informed consent was obtained from every participant. To minimise genetic heterogeneity, only subjects with at least a three-generation history of white Caucasian origin were recruited. All subjects satisfied identical inclusion and diagnostic criteria for the respective disease states and were recruited by the local endocrinologist. Families were recruited if they had either a sibling pair or at least two affected members (including avuncular and grandparent/grandchild pairs, but excluding parent/offspring relationships) affected with AITD.

GD was defined by the presence of documented biochemical hyperthyroidism in combination with either (1) a diffuse goiter on a scan or (2) positive autoantibodies (to TSHR, TG or TPO) or (3) Graves' ophthalmopathy or (4) confirmation of a lymphocytic infiltrate in thyroid histology. The diagnostic criteria for AIH were defined as documented biochemical hypothyroidism and either (1) positive autoantibodies to TG or TPO or (2) histological confirmation of a lymphocytic infiltrate in a fine needle aspirate or (3) presence of a goiter on clinical examination. Clinical criteria reflected those seen in the general population^{1–3} with a mean age at onset within the fourth decade and, for example, 50% of patients with GD having thyroid eye disease (TED) and an AITD F:M ratio of 6:1.

Gender matched controls with no known family history of AITD, of white Caucasian descent, were collected from seven large centres within the UK via general practice.

Subjects (replication study 2)

Unrelated Caucasian patients of UK origin with GD and AIH were recruited from thyroid clinics in the UK, as previously described.^{24,25} Briefly, patients were defined as having GD by the presence of biochemical hyperthyroidism together with either the presence of TED or two of the following criteria: diffuse goiter, a significant titre of microsomal (TPO), thyroglobulin (Tg) or TSHR autoantibodies. AIH was diagnosed by the presence of biochemical hypothyroidism together with a significant titre of TPO, Tg or TSHR autoantibodies, with or without a diffuse goiter. Ethnically and gender-matched control subjects with no family history of autoimmune disease were bled at geographically matched sites. All control subjects had normal thyroid function and were negative for thyroid autoantibodies. All patients and subjects gave informed written consent and the local ethics committee approved the project. In total, 1366 AITD cases (307 AIH and 1059 GD) and 1061 controls were recruited.

Genotyping

Prior to genotyping, venous blood samples were taken from all subjects and DNA was prepared centrally (Oxagen

Ltd) from whole blood using the Puregene kit (Gentra Systems) according to the manufacturer's instructions. DNA concentration was checked by performing three independent Picogreen assays. Each sample was checked for amplification success using three different sets of primers and products on an agarose gel with 99.9% of DNA samples extracted passing the amplification QC test.

A total of 40 SNPs across a 600 kb genomic DNA interval (NCBI, build 34, chromosome 14) encompassing the *TSHR* transcript (GenBank NM_000369) were selected for validation from public databases (dbSNP) and previous publications^{13–21} (Table S1).

Genotyping for the LD analysis and two-stage experiment was performed using TaqMan[®] chemistry. All 40 SNPs were ordered as either pre-validated TaqMan[®] Assays-on-Demand[™] or newly designed TaqMan[®] Assays-by-Design[®] (Applied Biosystems). Using the RAPIDGene[™] automated genotyping system (Oxygen), 2.5 μ l of 2 \times TaqMan[®] assay and 2.5 μ l of Absolute[™] QPCR ROX Mix (ABgene) were added to 10 ng of dried genomic DNA. The TaqMan[®] reaction was carried out following Applied Biosystems standard protocol. Fluorescent emissions were captured using the ABI Prism[®] 7900HT Sequence Detection System and converted into genotypes using the ABI Prism[®] SDS 2.0 software package. Genotype data were electronically linked with phenotype data for each individual.

For the replication study, DNA was prepared in Birmingham UK using Nucleon Bacc II kit from Trepnel Life Sciences PLC, Manchester, UK. rs2268458 was designed and ordered as a PCR-RFLP assay (Sigma Genosys; F-CCAGCAGAGG GAGCACA, R- TAGAGAATAGAGCAGCAAGACT). The 415 bp product was amplified from 20 ng of genomic DNA using 35 cycles of PCR (Taq DNA polymerase, Biorline; dNTPs, Amersham). Postamplification, 5 μ l of the 25 μ l PCR reaction were digested for 2 h at 37°C using the *AluI* restriction endonuclease (New England Biolabs). The digested DNA fragments were separated on a 3% agarose gel, visualised using ethidium bromide and converted into genotypes.

Statistical analysis

Pairwise coefficients of LD were estimated using the expectation-maximization (EM) algorithm²⁶ implemented in LDMAX, and the resulting patterns of LD across the *TSHR* genomic interval were visualised using GOLD 1.1.²⁷ LD 'blocks' were defined with a pairwise r^2 cutoff of >0.25 . For each block, haplotype frequencies across all SNPs were estimated using the EM algorithm in SNP-HAP v.0.2.1. The minimal set of SNPs which tag all nontag SNPs with a haplotypic $r^2 \geq 0.9$ was then selected using the program htsearch, implemented within the statistical package STATA v.1.8²³ (Table S1).

Significance of single SNP allelic and haplotype association analyses were assessed by a log-likelihood ratio (LLR) test, as described by Clayton.²⁸

Estimation of haplotype frequencies and likelihoods, for both randomised and actual data, was with the software SNP-HAP. A PERL script (RASH – randomised association using SNP-HAP) was written to automate this procedure and is available on request. In addition to accounting for problems associated with rare haplotypes, the use of an empirical significance level derived by randomising affection status of individuals (rather than of chromosomes) does not require any assumption of Hardy–Weinberg equilibrium in cases or/and controls.²⁸ Finally, for haplotypes, an approximate Z-score was computed, to compare the frequency of any individual haplotype in cases and controls. These analysis procedures follow those implemented in the program GENECOUNTING (2002), with the exception that we used SNP-HAP to implement haplotype and likelihood estimation. In all, 1 000 000 permutations were used for combined and replication single SNP *P*-value calculations, and 10 000 permutations were used for haplotype and Stage 1 single SNP *P*-value calculations.

Quality of genotypes was assessed with tests of Hardy–Weinberg equilibrium and tests of allelic and haplotypic heterogeneity between the Stages 1 and 2 cohorts. The tests of heterogeneity between the two stages were also conducted by LLR tests with empirical significance levels, here using 10 000 permutations. Odds ratios and corresponding 95% confidence intervals were calculated using GenStat v.6.2 from VSN International Ltd.

Results

SNP validation, LD block determination, and htSNP selection

Three blocks of LD were defined from the pairwise r^2 data, from 17 SNPs across the *TSHR* genomic interval (~ 250 kb) in 86 individuals. To investigate additional haplotype diversity within the middle block, three further SNPs were validated. To define the 'boundaries' of blocks 1 and 3, 17 additional SNPs were validated extending 5' from block 1 and 3' from block 3. Finally, to compare current results with those from previous publications, all known amino-acid residue altering polymorphisms were genotyped. In total, 38 SNPs with minor allele frequencies (MAF) greater than 0.01 were used for LD block structure analysis across the *TSHR* region (Figures 1 and 2).

Using the National Center of Biotechnology Information (NCBI) build 34 draft of the human genome sequence, the three LD blocks spanning the *TSHR* were calculated to be roughly 355, 87, and 40 kb, respectively. Block 1 covered the region approximately 280 kb upstream from exon 1 and extended into intron 1. Block 2 started in intron one and terminated at an undefined point between intron 6 and 8. Block 3 covered the remainder of the gene and extended about 10 kb 3' to exon 10 (Figure 1).

Altogether 18 htSNPs were chosen, nine, four and five SNPs across blocks 1, 2 and 3, respectively. In addition, two

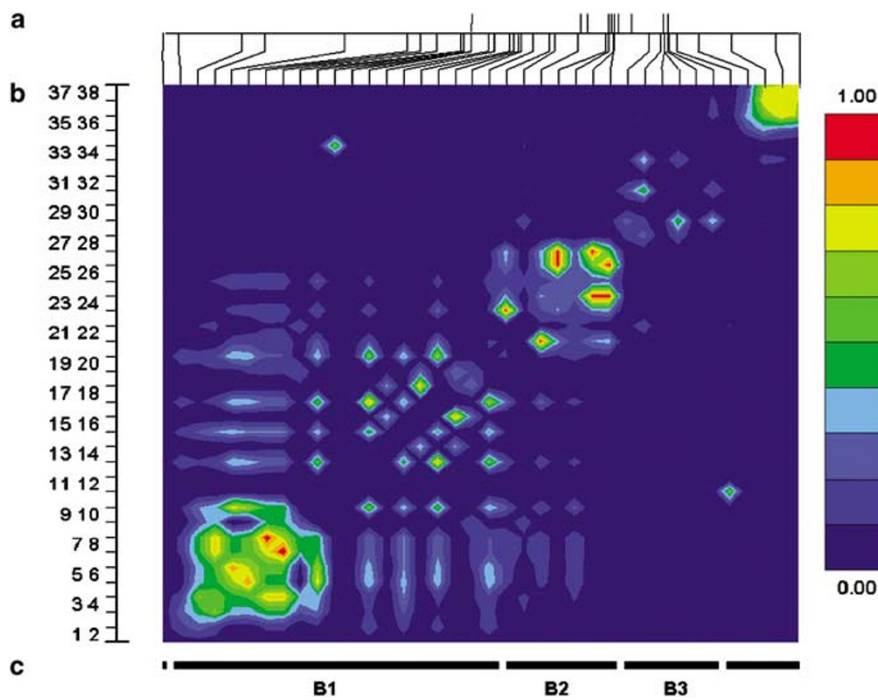


Figure 1 LD analysis across the *TSHR* genomic region on chromosome 14. (a) Location of the 10 *TSHR* exons (top) and 38 SNPs (bottom) across the ~600 kb genomic interval. (b) Graphical interpretation of pairwise r^2 statistics. (c) LD blocks determined using an r^2 cutoff of >0.25 .

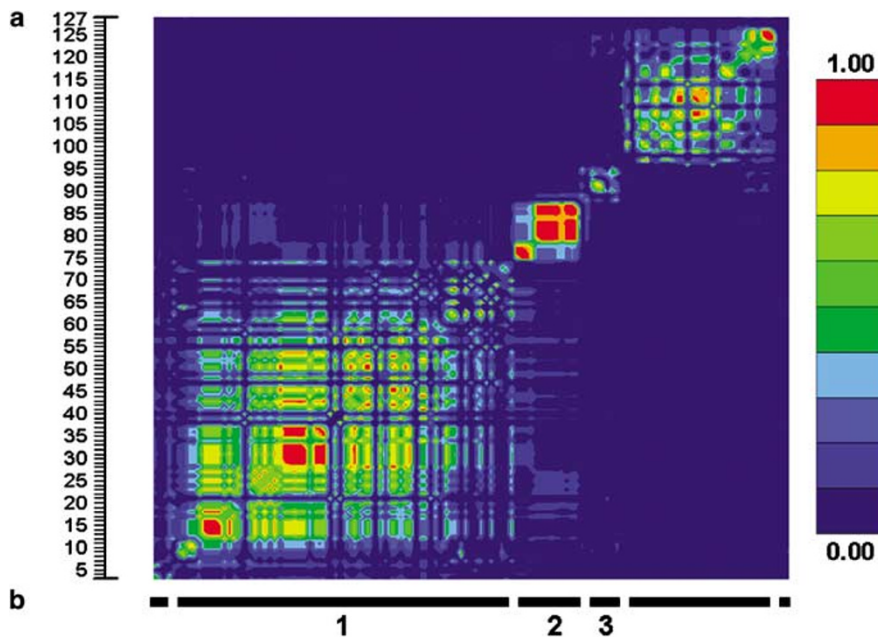


Figure 2 LD analysis across the *TSHR* genomic region using the Caucasian HAPMAP data. (a) Graphical interpretation of pairwise r^2 statistics. (b) LD blocks determined using an r^2 cutoff of >0.25 .

extra nontagged SNPs, selected for genotyping prior to defining the 5' boundary of block 1, were genotyped (Table S1).

Two-stage association study

Stage 1 A total of 20 SNPs were genotyped across 365 AITD cases and 350 controls with an average genotyping

call rate of 99.23% (SD 0.78) per marker. Single SNP association ($P < 0.05$) was detected for both GD and AITD, but not for AIH (Table 1). In GD, nine SNPs across blocks 1 and 2 showed single SNP association ($P < 0.05$), including three SNPs with censored P -values of 0, due to the use of only 10 000 permutations during analysis ($P < 1 \times 10^{-4}$).

The haplotype analysis in Stage 1 gave similar results to the single SNP data, in that GD and AITD showed association with no contribution from AIH (Table 2). While interpretation of the global haplotype results suggested that only block 2 was associated, the Z -scores of the individual haplotype results indicated that a single haplotype across blocks 1 and 2 was driving the association (Table 3). When comparing the haplotype data with the single SNP results, this remained the most parsimonious explanation. In fact, all SNPs in blocks 1 and 2 with their minor allele on the associated haplotype showed single SNP association to GD. Only the weak single SNP association with the rare amino-acid-altering polymorphism P52T appeared to be independent to the haplotype.

As none of the SNPs in block 3 showed association with disease, including rs2288495, which had its minor allele on the associated haplotype, these were not taken forward for further analysis.

Stage 2 The power of the analysis was increased by doubling the number of individuals genotyped and combining the data with Stage 1. An additional 694 AITD cases and 621 controls were genotyped using the 15 SNPs from blocks 1 and 2, with an average genotype call rate of 99.62% (SD 0.69) per marker.

In total, 10 of the 15 SNPs showed association ($P < 0.05$) with GD, three of which were more significant than could be calculated using 1 000 000 permutations ($P < 1 \times 10^{-6}$) (Table 4). Investigation into the consistency of the SNP data revealed one deviation from Hardy–Weinberg equilibrium in the combined GD cases ($P = 0.005$, rs2284722), one deviation from homogeneity between the two stages of genotyping for the GD cases ($P = 0.012$, rs2300523) and one deviation from homogeneity between the

Table 1 Stage 1 single SNP association results

SNP	Blocks	CON		GD			AIH				AITD			
		MAF	MAF	Allelic P	OR	95% CI	MAF	Allelic P	OR	95% CI	MAF	Allelic P	OR	95% CI
rs2167061	1	0.338	<i>0.407</i>	<i>0.0126</i>	1.39	1.07–1.81	0.316	0.5979	0.93	0.71–1.23	0.362	0.2150	1.15	0.92–1.44
rs2590480	1	0.399	0.344	0.0712	0.79	0.60–1.02	0.448	0.1466	1.21	0.94–1.57	0.396	0.8763	0.98	0.79–1.21
rs2371461	1	0.370	0.363	0.7530	0.96	0.74–1.25	0.353	0.5393	0.92	0.70–1.19	0.358	0.5783	0.94	0.75–1.16
rs2217177	1	0.427	<i>0.475</i>	<i>0.0013</i>	1.50	1.16–1.94	0.398	0.4285	0.90	0.69–1.17	0.462	0.1506	1.17	0.95–1.45
D36H	1	0.004	0.011	0.2369	2.56	0.57–11.5	0.008	0.6717	1.92	0.38–9.53	0.010	0.3406	2.24	0.58–8.71
P52T	1	0.056	<i>0.028</i>	<i>0.0274</i>	0.47	0.23–0.94	0.058	1.0000	1.01	0.59–1.74	0.043	0.2153	0.74	0.46–1.19
rs2110695	1	0.165	0.156	0.6146	0.91	0.64–1.29	0.168	0.9557	0.99	0.71–1.39	0.162	0.7249	0.94	0.71–1.25
rs2284722	1	0.325	<i>0.437</i>	<i>0.0002</i>	1.63	1.25–2.12	0.318	0.8934	0.98	0.75–1.29	0.377	<i>0.0302</i>	1.28	1.03–1.59
rs2268458	1	0.251	0.385	<i>0.0000^a</i>	1.90	1.44–2.49	0.257	0.8146	1.05	0.78–1.40	0.321	<i>0.0014</i>	1.44	1.14–1.81
rs917986	1	0.262	0.271	0.7869	1.04	0.78–1.38	0.261	0.9433	0.99	0.74–1.32	0.266	0.9781	1.01	0.79–1.28
rs2300521	1	0.237	<i>0.354</i>	<i>0.0000^a</i>	1.77	1.34–2.34	0.239	0.9445	1.02	0.76–1.37	0.296	<i>0.0114</i>	1.37	1.08–1.74
rs2300523	2	0.224	<i>0.350</i>	<i>0.0000^a</i>	1.84	1.39–2.43	0.258	0.2833	1.19	0.88–1.60	0.304	<i>0.0013</i>	1.50	1.18–1.90
rs1005292	2	0.205	0.191	0.5609	0.90	0.66–1.24	0.187	0.4169	0.88	0.64–1.21	0.189	0.3766	0.88	0.68–1.15
rs2284734	2	0.312	<i>0.399</i>	<i>0.0056</i>	1.45	1.11–1.88	0.290	0.4454	0.89	0.67–1.17	0.344	0.2247	1.15	0.92–1.43
rs2241119	2	0.116	<i>0.200</i>	<i>0.0007</i>	1.87	1.33–2.64	0.118	1.0000	1.01	0.68–1.49	0.159	<i>0.0291</i>	1.42	1.05–1.93
rs2268476	3	0.261	0.238	0.3633	0.87	0.65–1.16	0.254	0.7246	0.95	0.71–1.27	0.246	0.4711	0.91	0.72–1.16
rs1957547	3	0.407	0.361	0.0784	0.80	0.61–1.03	0.399	0.6415	0.94	0.72–1.21	0.380	0.1867	0.87	0.70–1.07
rs2300540	3	0.284	0.283	0.9204	0.98	0.74–1.30	0.295	0.8036	1.04	0.78–1.38	0.289	0.9487	1.01	0.81–1.28
D727E	3	0.079	0.082	0.8896	1.05	0.65–1.68	0.099	0.2811	1.30	0.83–2.05	0.090	0.4506	1.18	0.80–1.72
rs2288495	3	0.451	0.500	0.1266	1.22	0.94–1.57	0.481	0.3520	1.13	0.87–1.45	0.490	0.1652	1.17	0.95–1.44

^aThe use of 10 000 permutations limits P -value calculations to four decimal places, $P < 1 \times 10^{-4}$.
 Nine associated SNPs in italics.

Table 2 Global haplotype association results (P -values)

Block	GD	Stage 1		AITD	GD	Combined	
		AIH	AITD			AIH	AITD
1	0.1933	0.8674	0.6356	0.0023	0.9539	0.2545	
2	<i>0.0003</i>	0.2111	<i>0.0030</i>	<i>0.0002</i>	0.2019	<i>0.0006</i>	
3	0.4471	0.8776	0.5894	NA	NA	NA	
1 and 2	0.1155	0.6483	0.7313	0.0565	0.7070	0.2731	
1, 2 and 3	<i>0.0141</i>	0.6909	<i>0.0271</i>	NA	NA	NA	

Table 3 Best individual GD haplotype results ($P < 1 \times 10^{-4}$)

Block	HAP ^a	Z-score	Stage 1		Z-score	Combined	
			GD MAF	CON MAF		GD MAF	CON MAF
1	21121112212	4.53	0.197	0.098	4.99	0.183	0.118
2	2122	4.42	0.190	0.095	4.83	0.155	0.096
3	NS	NA	NA	NA	NA	NA	NA
1 and 2	211211122122122	4.81	0.117	0.040	4.46	0.088	0.047
1, 2 and 3	21121112212212211112	3.67	0.088	0.035	NA	NA	NA

^aThe common allele is coded 1 and the minor allele is coded 2.

Table 4 Combined single SNP association results

SNP	Blocks	CON		GD			AIH				AITD			
		MAF	MAF	Allelic P	OR	95% CI	MAF	Allelic P	OR	95% CI	MAF	Allelic P	OR	95% CI
rs2167061	1	0.320	0.401	0.000002	1.45	1.24–1.70	0.342	0.127796	1.13	0.96–1.32	0.371	0.000317	1.27	1.11–1.45
rs2590480	1	0.418	0.361	0.001674	0.79	0.67–0.92	0.417	0.938138	0.99	0.85–1.15	0.389	0.074069	0.89	0.79–1.01
rs2371461	1	0.368	0.363	0.641216	0.96	0.82–1.12	0.363	0.643662	0.96	0.83–1.12	0.363	0.528635	0.96	0.84–1.09
rs2217177	1	0.437	0.490	0.000034	1.36	1.17–1.58	0.435	0.928828	1.01	0.87–1.17	0.473	0.010718	1.18	1.04–1.33
D36H	1	0.007	0.010	0.530208	1.40	0.63–3.09	0.009	0.669258	1.26	0.56–2.85	0.010	0.395705	1.38	0.70–2.71
P52T	1	0.054	0.037	0.024187	0.66	0.45–0.96	0.056	0.933141	1.02	0.73–1.40	0.046	0.309851	0.86	0.65–1.14
rs2110695	1	0.152	0.147	0.631294	0.95	0.77–1.17	0.162	0.568674	1.06	0.87–1.30	0.154	1.000000	1.00	0.84–1.19
rs2284722	1	0.335	0.435	0.000000 ^a	1.55	1.33–1.81	0.352	0.247731	1.10	0.94–1.28	0.394	0.000027	1.31	1.15–1.49
rs2268458	1	0.265	0.370	0.000000 ^a	1.66	1.41–1.95	0.285	0.155354	1.13	0.96–1.33	0.327	0.000002	1.37	1.20–1.57
rs917986	1	0.247	0.244	0.661025	0.96	0.81–1.14	0.254	0.851135	1.02	0.86–1.21	0.249	0.810276	0.98	0.85–1.13
rs2300521	1	0.249	0.346	0.000000 ^a	1.60	1.36–1.88	0.259	0.526493	1.06	0.89–1.25	0.302	0.000119	1.31	1.14–1.51
rs2300523	2	0.232	0.300	0.000040	1.44	1.22–1.70	0.253	0.148826	1.14	0.96–1.35	0.277	0.000430	1.29	1.12–1.49
rs1005292	2	0.196	0.199	0.913369	1.01	0.84–1.22	0.204	0.664647	1.05	0.87–1.26	0.202	0.839089	1.02	0.87–1.19
rs2284734	2	0.306	0.378	0.000036	1.39	1.19–1.62	0.336	0.080559	1.15	0.99–1.35	0.357	0.000391	1.27	1.12–1.45
rs2241119	2	0.113	0.171	0.000021	1.62	1.31–2.00	0.136	0.077009	1.23	0.98–1.54	0.153	0.000123	1.44	1.20–1.73

^aThe use of 1 000 000 permutations limits P -value calculations to six decimal places, $P < 1 \times 10^{-6}$. Ten associated SNPs in italics.

various international collection centres for the GD cases ($P = 0.003$, rs2300523). No deviations from either Hardy–Weinberg equilibrium or homogeneity were detected in the controls. As in Stage 1, there was no association with AIH.

Stage 2 haplotype association results were also comparable to Stage 1, albeit more significant and with both blocks showing global association with the GD phenotype (Table 2). Analysis of the individual haplotype data indicated that the same haplotype seen in Stage 1 was driving both the global haplotype associations and the majority (9/10) of single SNP associations (Table 3). However, instead of the combined haplotype across blocks 1 and 2 showing the strongest association, analysis of the Z -scores suggested that individual haplotypes from each block were driving the combined block association and that the haplotype in block 1 was more associated with GD than the haplotype in block 2. Comparison of the MAFs and single SNP results between blocks 1 and 2 supported this conclusion. SNPs in block 1 with minor alleles on the associated haplotype had MAFs of greater discordance compared to the frequency of the associated haplotype, yet

showed stronger single SNP association. Finally, as in Stage 1, the P52T polymorphism was again the only marker to show single SNP association to GD independent of the associated haplotypes.

Replication study

To verify the association between the *TSHR* and GD detected in the two-stage study, the SNP showing the strongest combined single SNP association, rs2268458 (Table 4), was selected for further investigation in an independent UK AITD cohort. Using alternative genotyping chemistry, rs2268458 was genotyped across 1366 AITD cases and 1061 controls with a genotype call rate of 83.2%. Replicating the previous results, rs2268458 showed association to GD ($P = 2 \times 10^{-6}$) and no association to AIH (Table 5). Investigation into the consistency of the SNP data between the two studies revealed no heterogeneity ($P < 0.05$) between either the GD cases or the controls. The accuracy of this genotyping method was subsequently validated using TaqMan[®] chemistry as described above with 100% genotyping replication.

Table 5 Replication results for rs2268458

	GD	AIH	AITD
Allelic p	0.000002	0.935581	0.000070
MAF	0.350	0.294	0.340
OR	1.31	1.010	1.250
95% CI	1.12–1.53	0.79–1.29	1.08–1.45

NB – control MAF = 0.292.

Discussion

While several studies have previously examined the *TSHR* for association with AITD, they have met with limited success and consistency, due to a focus on rare amino-acid-altering polymorphisms in relatively small data sets.^{13–21} We focused on common SNPs that tagged the limited haplotype diversity in the region, while also including previously studied amino-acid-altering SNPs. Using this approach, we detected highly significant associations to GD but not AIH. P52T was the only previously associated amino-acid-altering SNP to show a weak association to GD in our two-stage experiment, although in the opposite direction to the original study.¹⁶ With this discrepancy and the other negative results,^{13,15,17–19,21} it would appear that both the initial and our own P52T associations are most likely Type I errors. However, a protective effect of P52T cannot be excluded.

Only one prior AITD genetic study analysed an intronic polymorphism of the *TSHR*.¹⁷ In that study, rs2239610 showed association with an Asian GD cohort. Although this SNP was not analysed in the present study, we did type rs2268458, a SNP in perfect LD ($r^2 = 1$), with rs2239610 and this showed the strongest single SNP association with GD in our combined data set. In our replication cohort, the association between rs2268458 and GD was confirmed, although the odds ratio was smaller than that in our initial study. Bias in the measurement of effect size, normally leading to inflated risk estimates, is common in studies detecting novel association.²⁹ Thus, the odds ratio from our replication study may be a better measure of the effect of this locus on Caucasian GD susceptibility.

We adopted a staged strategy for the study of a candidate gene region for AITD. The initial association study was conducted in two stages to save valuable resources. Analysis of Stage 1, Stage 2 and the combined data sets showed minor differences in allele and genotype distributions confirming the validity of this approach. A common failing of many genetic studies is the inadequate size of data sets.³⁰ This was overcome in the present study with the establishment of an international collaboration in which all recruiting centres worked to identical inclusion criteria and a central laboratory was used for DNA extraction, genotyping and data analysis. Furthermore, the use of an independent replication cohort addressed a

number of potential limitations of the initial study. Not only did it confirm the finding of association of the *TSHR* region with GD, and lack of association with AIH, but it validated our overall strategy and addressed possible sampling problems that could have arisen over the use of unrelated index cases recruited as part of a family-based study and the potential bias of using samples from different geographical locations with control DNA from only the UK.

Although we have provided strong evidence for association of the *TSHR* region with GD, further studies are required to identify the primary disease-causing DNA variants within this region. Within the associated region spanning LD blocks 1 and 2, two other novel genes of unknown function reside, c14orf61 and c14orf145. Although these novel genes cannot be excluded as harbouring disease-causing polymorphisms, the *TSHR* is the most attractive candidate for GD susceptibility in the region. The difference observed between GD and AIH in both studies is, however, an important finding. GD and AIH have similar prevalence data¹ showing that we had the same power to detect association with both disease states. The fact that the *TSHR* has been shown to act as the primary autoantigen for GD² and not AIH^{3,31} supports the notion that polymorphism of the *TSHR* could be a disease-specific locus for GD.

While further studies are required to confirm the exact location of disease-causing polymorphism in this region, intronic *TSHR* polymorphisms in LD block 1, altering the splicing of exons coding the extracellular domain of the protein, may play a major role in *TSHR* autoantibody production. The extracellular domain of the *TSHR* is cleaved, forming α and β subunits. Studies have shown that the α subunit is occasionally shed into the plasma.³¹ Potentially, the derivation of an alternative extracellular domain through alternative splicing could either instigate shedding or convert the wild-type α subunit into a potent autoantigen, both of which may elicit the autoimmune response.

Acknowledgements

We are grateful to research nurses for assistance with sample collection, Jenny Taylor for initial project management, Jeremy Martin for software development and support, Anna Barrows for DNA extraction and Sarah Cranston for data management. L Hegedus is supported by the Agnes and Knut Moerk Foundation. J Collins is a Diabetes Research and Wellness Foundation PhD student. This work was also supported by Oxagen Ltd and the Wellcome Trust.

References

- 1 Tunbridge WM, Evered DC, Hall R *et al*: The spectrum of thyroid disease in a community: the Wickham survey. *Clin Endocrinol (Oxford)* 1977; 7: 481–493.
- 2 Weetman AP: Graves' disease. *N Engl J Med* 2000; 343: 1236–1248.

- 3 Tunbridge WM, Brewis M, French JM *et al*: Natural history of autoimmune thyroiditis. *BMJ* 1981; **282**: 258–262.
- 4 Brix TH, Kyvik KO, Hegedus L: A population-based study of chronic autoimmune hypothyroidism in Danish twins. *J Clin Endocrinol Metab* 2000; **85**: 536–539.
- 5 Brix TH, Kyvik KO, Christensen K, Hegedus L: Evidence for a major role of heredity in Graves' disease: a population-based study of two Danish twin cohorts. *J Clin Endocrinol Metab* 2001; **86**: 930–934.
- 6 Sakai K, Shirasawa S, Ishikawa N *et al*: Identification of susceptibility loci for autoimmune thyroid disease to 5q31–q33 and Hashimoto's thyroiditis to 8q23–q24 by multipoint affected sib-pair linkage analysis in Japanese. *Hum Mol Genet* 2001; **10**: 1379–1386.
- 7 Tomer Y, Ban Y, Concepcion E *et al*: Common and unique susceptibility loci in Graves and Hashimoto diseases: results of whole-genome screening in a data set of 102 multiplex families. *Am J Hum Genet* 2003; **73**: 736–747.
- 8 Simmonds MJ, Howson JM, Heward JM *et al*: Regression mapping of association between the human leukocyte antigen region and Graves disease. *Am J Hum Genet* 2005; **76**: 157–163.
- 9 Ueda H, Howson JM, Esposito L *et al*: Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 2003; **423**: 506–511.
- 10 Smyth D, Cooper JD, Collins JE *et al*: 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* 2004; **53**: 3020–3023.
- 11 Velaga MR, Wilson V, Jennings CE *et al*: The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. *J Clin Endocrinol Metab* 2004; **89**: 5862–5865.
- 12 de Roux N, Shields DC, Misrahi M *et al*: Analysis of the thyrotropin receptor as a candidate gene in familial Graves' disease. *J Clin Endocrinol Metab* 1996; **81**: 3483–3486.
- 13 Allahabadia A, Heward JM, Mijovic C *et al*: Lack of association between polymorphism of the thyrotropin receptor gene and Graves' disease in United Kingdom and Hong Kong Chinese patients: case control and family-based studies. *Thyroid* 1998; **8**: 777–780.
- 14 Ban Y, Greenberg DA, Concepcion ES, Tomer Y: 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* 2002; **12**: 1079–1083.
- 15 Chistyakov DA, Savost'yanov KV, Turakulov RI *et al*: Complex association analysis of graves disease using a set of polymorphic markers. *Mol Genet Metab* 2000; **70**: 214–218.
- 16 Cuddihy RM, Dutton CM, Bahn RS: A polymorphism in the extracellular domain of the thyrotropin receptor is highly associated with autoimmune thyroid disease in females. *Thyroid* 1995; **5**: 89–95.
- 17 Ho SC, Goh SS, Khoo DH: Association of Graves' disease with intragenic polymorphism of the thyrotropin receptor gene in a cohort of Singapore patients of multi-ethnic origins. *Thyroid* 2003; **13**: 523–528.
- 18 Kaczur V, Takacs M, Szalai C *et al*: 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* 2000; **27**: 17–23.
- 19 Kotsa KD, Watson PF, Weetman AP: No association between a thyrotropin receptor gene polymorphism and Graves' disease in the female population. *Thyroid* 1997; **7**: 31–33.
- 20 Sale MM, Akamizu T, Howard TD *et al*: Association of autoimmune thyroid disease with a microsatellite marker for the thyrotropin receptor gene and CTLA-4 in a Japanese population. *Proc Assoc Am Physicians* 1997; **109**: 453–461.
- 21 Simanainen J, Kinch A, Westermark K *et al*: Analysis of mutations in exon 1 of the human thyrotropin receptor gene: high frequency of the D36H and P52T polymorphic variants. *Thyroid* 1999; **9**: 7–11.
- 22 Daly MJ, Rioux JD, Schaffner SF *et al*: High-resolution haplotype structure in the human genome. *Nat Genet* 2001; **29**: 229–232.
- 23 Johnson GC, Esposito L, Barratt BJ *et al*: Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001; **29**: 233–237.
- 24 Collins JE, Heward JM, Howson JM *et al*: Common allelic variants of exons 10, 12, and 33 of the thyroglobulin gene are not associated with autoimmune thyroid disease in the United Kingdom. *J Clin Endocrinol Metab* 2004; **89**: 6336–6339.
- 25 Heward JM, Allahabadia A, Daykin J *et al*: 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* 1998; **83**: 3394–3397.
- 26 Excoffier L, Slatkin M: Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995; **12**: 921–927.
- 27 Abecasis GR, Cookson WO: GOLD – graphical overview of linkage disequilibrium. *Bioinformatics* 2000; **16**: 182–183.
- 28 Clayton DG: Population association; in Balding DJ, Bishop M, Cannings C (eds): *Handbook of Statistical Genetics*. Chichester: Wiley, 2003, 2nd edn, pp 939–960.
- 29 Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG: Replication validity of genetic association studies. *Nat Genet* 2001; **29**: 306–309.
- 30 Lohmueller KE, Pearce CL, Pike M *et al*: Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003; **33**: 177–182.
- 31 Kopp P: The TSH receptor and its role in thyroid disease. *Cell Mol Life Sci* 2001; **58**: 1301–1322.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)