Population-specific risk of type 2 diabetes (T2D) conferred by HNF4A P2 promoter variants: a lesson for replication studies

Inês Barroso1*, Jian’an Luan2, Eleanor Wheeler1, Pamela Whittaker1, Jon Wasson3, Eleftheria Zeggini4,5, Michael N. Weedon6, Sarah Hunt1, Ranganath Venkatesh1, Timothy M Frayling6, Marcos Delgado1, Rosalind J. Neuman3, Jinghua Zhao2, Richard Sherva3, Benjamin Glaser7, Mark Walker8, Graham Hitman9, Mark I McCarthy4,5, Andrew T Hattersley6, M. Alan Permutt4, Nicholas J. Wareham2, Panagiotis Deloukas1

1Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
2MRC Epidemiology Unit, Institute of Metabolic Science, Cambridge, UK
3Washington University School of Medicine, St. Louis, MO, USA.
4Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, UK
5Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
6Institute of Biomedical and Clinical Science, Peninsula Medical School, Exeter, UK
7Endocrine and Metabolism Service, The Hadassah Hebrew University Medical Center, Israel
8Diabetes Research Group, School of Clinical Medical Sciences, Newcastle University, UK
9Centre for Diabetes and Metabolic Medicine, Barts and the London, Queen Mary School of Medicine and Dentistry, University of London, London, UK

*Correspondence should be addressed to:
Inês Barroso, Metabolic Disease Group, Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail: ib1@sanger.ac.uk.

Received 12 April 2007 and accepted 15 August 2008.

Additional information for this article can be found in an online appendix at http://diabetes.diabetesjournals.org

This is an uncopied electronic version of an article accepted for publication in Diabetes. The American Diabetes Association, publisher of Diabetes, is not responsible for any errors or omissions in this version of the manuscript or any version derived from it by third parties. The definitive publisher-authenticated version will be available in a future issue of Diabetes in print and online at http://diabetes.diabetesjournals.org.
ABSTRACT

Objective: SNPs in the P2 promoter region of HNF4A associated with T2D predisposition originally in Finnish and Ashkenazim and more recently in Scandinavian populations, but generated conflicting results in additional populations. We aimed to investigate whether data from a large-scale mapping approach would replicate this association in novel Ashkenazi samples and in UK populations, and whether these data would allow us to refine the association signal.

Research Design and Methods: Using a dense linkage disequilibrium (LD) map of 20q we selected SNPs from a 10Mb interval centred on HNF4A. In a staged approach we first typed 4608 SNPs in case-control populations from four UK and an Ashkenazi population (N=2516). In phase 2, a subset of 763 SNPs were genotyped in 2513 additional samples from the same populations.

Results: Combined analysis of both phases demonstrated association between HNF4A P2 SNPs (rs1884613 and rs2144908) and T2D in the Ashkenazim (N=991, p<1.6x10^-6). Importantly these associations are significant in a subset of Ashkenazi samples (N=531) not previously tested for association with P2 SNPs (OR~1.7, p<0.002), thus providing replication within the Ashkenazim. In the UK populations this association was not significant (N=4022, p>0.5) and the estimate for the odds ratio was much smaller OR=1.04 (95%CI 0.91 – 1.19).

Conclusions: These data indicate that the risk conferred by HNF4A P2 is significantly different between UK and Ashkenazi populations (p<0.00007) suggesting that the underlying causal variant remains unidentified. Interactions with other genetic or environmental factors may also contribute to this difference in risk between populations.
The presence of type 2 diabetes (T2D) susceptibility genes on chromosome 20 has been suggested by linkage scans in several populations. The 20q12-q13 region (OMIM 603694) is the best replicated and harbours the gene HNF4A, mutations in which lead to MODY 1 (OMIM 125850). Evidence for association between SNPs in the β-cell P2 promoter region of HNF4A has been recognized in Finnish (1) and Ashkenazi (2) populations, with data suggesting that the HNF4A P2 SNPs (or variants in strong LD with them) contribute to the linkage signal on chromosome 20q (1; 2). Association with HNF4A promoter SNPs has been replicated in some (3-7) but not all (8-12) populations tested. In other populations there was evidence for association with SNPs or haplotypes in the HNF4A region other than the P2 SNPs (10; 13-15). More recently, the association between HNF4A promoter SNPs and T2D has been confirmed in Scandinavians but not in a broader meta-analysis with additional populations (16), suggesting that P2 SNPs confer varying risk effects in different populations possibly due to the underlying causal variant not having been identified. We investigated a 10Mb interval (38.1Mb-48.2Mb NCBI build 35), centred around HNF4A, including genotypes from 4608 non-redundant (r²<1) SNPs (1 SNP per 2Kb on average) in five T2D case-control populations to evaluate whether we could confirm and refine the association signal in Ashkenazim, and evaluate whether this association was also present in UK populations. We were also interested in assessing whether there was evidence for additional association signals within this broader interval. We tested an Ashkenazi T2D case-control study (N=998) including novel samples (N=531) not previously tested for linkage or association with HNF4A P2 SNPs (2), two UK population-based case-control studies where linkage and association studies with HNF4A P2 had not been carried out (N=2189), and two additional UK case-control collections (N=1842), one enriched for earlier onset T2D where linkage studies had not been done but with suggestive association with HNF4A P2 SNPs (4) and one study which included samples where despite no evidence of linkage to chromosome 20q, association of HNF4A P2 SNPs with T2D risk had previously been suggested (4; 17).

RESEARCH DESIGN AND METHODS

Additional details are provided in Supplementary material.

Populations

Cambridgeshire Case Control. Population based Cambridgeshire case-control cohort of randomly selected unrelated patients with T2D (n=555) and matched unrelated controls (n=541) (18).

EPIC- Norfolk. Population based cohort study of Norfolk, UK from which 354 T2D cases and 739 unrelated control were used in this study (19; 20).

Young onset type 2 diabetes patients from Exeter. Consecutive-case series of unrelated patients with T2D diagnosed before 45 years from North and East Devon (21; 22) and gender matched controls from parents in the Exeter Family (22; 23). We included 414 cases and 425 controls in this study.

Warren 2 study diabetes UK. Warren 2 index cases (all Europid) passing all the stringent genome scan criteria as described (17), totalling 528 probands from 573 families. The 475 controls are from the ECACC Human Control Resource.

Ashkenazi Case-Control Study Sample. Study comprised of 143 cases from multiplex-affected sibships ascertained for published genome scan (24) and 393 newly ascertained unrelated cases. Control samples are 149 unrelated subjects with no personal or first degree family history of T2D and 313 additional unrelated samples from The National Laboratory for the Genetics of Israeli Populations at Tel Aviv University, Israel. Of the controls 138 were novel.

For this study each of the UK populations was randomly subdivided into
two halves and each used in phase 1 or phase 2 of the study. At the onset of the study many of the novel Ashkenazi samples were not available for testing, consequently 155 cases included in phase 1 were novel, while all 238 cases included in phase 2 were novel. Of the 138 novel controls, 137 were from phase 2.

Ethical Approvals. Ethical permission for all studies was granted by their respective local research ethics committees, and study participants provided informed consent.

Study Design. We applied a two stage design; in phase 1 we genotyped 934 T2D patients and 1120 controls from the four UK populations and 298 T2D patients and 164 Ashkenazi controls. Single SNP associations within the UK populations (collectively referred to as UK4) were performed in a joint analysis including a term for study. The Ashkenazi data was analysed independently. Phase 2 included 917 T2D patients and 1060 controls from the UK and 238 T2D patients and 298 Ashkenazi controls and tested a subset of SNPs tested in phase 1. Results from phase 1 and 2 were combined in joint analysis. Phase 2 in the Ashkenazi set includes controls previously tested for association with \( HNF4A \) P2 SNPs (2) so cannot be considered replication. To obtain independent evidence of replication we analysed results from the 531 novel Ashkenazi samples separately. In this set we conducted permutation testing (10,000 permutations) to obtain empirical p-values.

Genotyping. Based on a dense linkage disequilibrium (LD) map of chr20 (25; 26) which included a 10Mb region (38.1Mb-48.2Mb NCBI build 35) centred on \( HNF4A \) we performed pairwise LD analysis on all 5324 SNP markers with a minor allele frequency (MAF) >4% (CEPH panel) and excluded SNPs with an \( r^2 = 1 \). This identified 4608 SNPs (1 SNP per 2Kb on average) for genotyping in phase 1, of which 763 were retested in phase 2 on additional samples. By design, the 4608 SNPs tagged all variation with MAF>4% in this region in HapMap phase 1 data. SNPs were assayed with Illumina’s Golden Gate assay as described elsewhere (25; 27) and standard sample and genotyping QC were applied (Supplementary Information).

Statistical Analysis. All analyses were conducted using Stata/SE 9.2 for windows or Unix (Stata Corporation, Texas, U.S.A.). Genotype frequencies were tested for Hardy-Weinberg equilibrium (HWE) using a chi-square goodness-of-fit test in controls and cases, separately. For each SNP, association with T2D was tested using an additive model on one degree of freedom (df). The equality of allele frequencies between phases was tested using large-sample statistics in controls and cases, separately.

In all joint analyses when no significant heterogeneity (p>0.05) was detected (between populations and between phases) an odds ratio (OR) using Mantel-Haenszel (MH) method was estimated. However, when significant heterogeneity was detected, a random-effect meta-analysis was applied.

RESULTS

Phase 1 results in UK4 for all SNPs that survived QC are shown in Figure 1a. SNPs with nominal significant results (p<0.05, unadjusted for multiple testing) in the Ashkenazi population are also shown. There was very modest overlap between the nominal significant results in UK4 and Ashkenazim. To avoid both false-positive association claims (type I error) and false-negative claims (type II error), we selected all SNPs with p<0.15 in the UK4 to test in a further set of independent samples in phase 2 of the study. Additional SNPs with p<0.01 in Ashkenazim not included in the previous selection were also added. In total 763 SNPs were tested in phase 2 on a further 916 UK T2D cases and 1059 controls and 238 Ashkenazi cases and 298 controls.

In the Ashkenazim there was significant association between SNPs in the \( HNF4A \) P2 region and T2D (Figure 1b), rs1884613 (p=6.8x10^{-7}) and rs2144908 (p=1.3x10^{-6}), and
in 10 SNPs in the GDAP1L1 gene (p<5.3x10^{-3} – Supplementary Table 2). Importantly the association with SNP rs1884613 (OR 1.70, p=0.0014, empirical p=0.0012) is independently replicated (Table 1; similar results were obtained for rs2144908) in the subset of novel Ashkenazi samples (393 cases and 138 controls). In contrast, of the 16 PREX1 SNPs with significant results in phase 1 (p<0.01) in the joint analysis none achieved equivalent significance (Figure 1 a, b). There was also no evidence for association with other SNPs (rs2425637, rs2425639, rs2425640, rs6130609 and rs745975, p>0.2) mapping to the coding sequence or P1 promoter of HNF4A (data not shown). In contrast with the Ashkenazi data, in UK4 there was no evidence of stronger association based on the joint analysis as compared to the results from phase 1 alone (Figure 1 a and b and Supplementary Table 3), furthermore none of the results would survive any kind of multiple-testing correction.

We used log-likelihood ratio tests to assess whether SNPs rs1884613 and rs2144908 explained all of the observed associations with T2D in this region. The correlation between SNPs rs1884613 and rs2144908 is very high (r^2=0.99) therefore we cannot separate their effects on T2D risk. We consecutively added the other SNPs in this region in a log additive form to a model containing these SNPs (1df assuming no dominance at the test locus). This analysis demonstrated that none of the other SNPs in this region of high LD improved the model containing just one of rs1884613 or rs2144908 SNPs, and indicates that no additional genotyped SNP is independently contributing to T2D risk (data not shown). The reciprocal analysis, showed that addition of either rs1884613 or rs2144908 improved all models containing any of the other SNPs. Repeating this analysis with phase 1 data (higher SNP density although on a smaller sample size – data not shown) confirms these results and suggests that no other genotyped SNP in this region is independently associated with T2D risk.

**DISCUSSION**

This study provides clear evidence for association of HNF4A P2 promoter SNPs with T2D in the Ashkenazim. Importantly, these results are replicated in the subset of novel samples (393 cases and 138 controls) that are independent to those used in the original report (2), with a point estimate of OR=1.70 (p=0.0014) for rs1884613 (Table 1). However, despite the relatively high density of SNPs used to span the 10Mb candidate interval (1SNP/ 2Kb on average) we were unable to refine the association signal beyond the two originally associated SNPs (rs1884613 and rs2144908) in the Ashkenazim. Thus, further fine-mapping in this region will be required to identify the underlying causal variant. In contrast, in UK4 this association result could not be replicated (OR~ 1.04, p>0.5). Furthermore, the confidence intervals for the odds ratio in UK4 and the Ashkenazim did not overlap (Supplementary Table 2), showing that the risk of P2 SNPs on T2D in these two groups is significantly different (p<0.00007). Independent analysis of each UK population also did not provide evidence for association of HNF4A P2 SNPs and diabetes risk in any of the individual populations (Supplementary Table 4).

Within the UK, the previously reported effect size for nominally associating SNPs (rs4810424 and rs2144908) was small (OR=1.09–1.13) (4), with MAF=16%, alpha=0.05 and a sample size of 4022, our UK populations were underpowered to detect such effects (power = 54%) which could underlie our lack of association. Importantly, both the current and previously reported confidence intervals for the effect size in the UK (4) do not overlap those in the Ashkenazim, suggesting population-specific effects. We observed significant heterogeneity of effects at these SNPs (p<0.0001) between UK and Ashkenazim populations, by comparing logistic regression models, with and without a
genotype times study interaction term, using a likelihood ratio test.

Allele frequency and Fst values across the region suggested there was no evidence of significant genetic distance at these loci (Supplementary Information). Furthermore, LD plots and LDU maps across this region (Supplementary Information) also demonstrated that overall the magnitude of LD was not materially different between Ashkenazi and UK populations and was unlikely to significantly contribute to divergent results. Thus, our data suggest that if HNF4A P2 SNPs are a risk factor for T2D in UK European populations the risk they confer is considerably smaller than that which they impart on the Ashkenazi population, this is consistent with linkage having been observed in Ashkenazim(24) but not in UK samples (17) and with other more recent results (16).

This result provides further support for the involvement of HNF4A P2 SNPs (or others in high LD) in conferring risk of T2D in Ashkenazim, and has important wider implications as it demonstrates the benefit of performing replication testing in samples that are recruited from the same population sampling-frame as those used in the exploratory early stages of an association study. Population-specific effects will have a significant impact in the interpretation of results from replication studies.

ACKNOWLEDGEMENTS

We thank T Dibling and C Hind for technical assistance in genotyping. We thank all study participants. The Ashkenazi population was collected by the Israeli Diabetes Research Group, funded by the Russell Berrie Foundation, D-Cure, Diabetes Care in Israel, grant #5040 from the Chief Scientist Office of the Israel Ministry of Health, and an unrestricted research grant from Novo-Nordisk, Denmark. JW, RN, RS, BG, MAP are funded by NIH R01 DK049583. The Wellcome Trust and Diabetes UK funded the Warren 2 study. The work on the Cambridgeshire Case-control study and EPIC-Norfolk studies was funded by support from the Wellcome Trust and MRC. Diabetes UK and NHS Research and Development funded the Exeter case control population. IB,PW, SH, RV and PD are funded by the Wellcome Trust. IB and EW also acknowledge support from EU FP6 funding (contract no LSHM-CT-2003-503041). ATH is a Wellcome Trust Research Leave Fellow. MNW is a Vandervell Foundation Research Fellow. EZ is a Wellcome Trust Research Career Development Fellow.
REFERENCES


Table 1– Genotype counts for *HNF4A* P2 SNP rs1884613 in each UK and Ashkenazim populations.
ASH_new, novel samples not previously tested for association (138 controls/ 393 cases); ASH-old samples overlapping those in the original report love-Gregory et al., 2004. (324 controls/ 143 cases), CCC, Cambridgeshire study; W2, Warren 2 study. Genotype counts (and frequencies) are shown for homozygous for common allele (11), heterozygous (12) and homozygous for the minor allele (22). Odds Ratios (OR), 95% confidence interval (95%CI) and p-values are based on an additive model. For SNP rs1884613 there is a lower call rate for Exeter samples. * p-values <0.01.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Population</th>
<th>Controls</th>
<th>Cases</th>
<th>OR</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>12</td>
<td>22</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>rs1884613</td>
<td>ASH_new</td>
<td>90</td>
<td>41</td>
<td>6</td>
<td>200</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>ASH_old</td>
<td>217</td>
<td>88</td>
<td>16</td>
<td>76</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>CCC</td>
<td>381</td>
<td>138</td>
<td>9</td>
<td>375</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>Exeter</td>
<td>147</td>
<td>55</td>
<td>3</td>
<td>152</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>EPIC-Norfolk</td>
<td>506</td>
<td>209</td>
<td>19</td>
<td>239</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>W2</td>
<td>318</td>
<td>110</td>
<td>18</td>
<td>338</td>
<td>144</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>All data</td>
<td>1659</td>
<td>641</td>
<td>71</td>
<td>1380</td>
<td>651</td>
</tr>
</tbody>
</table>
**Figure 1 – Single SNP Association Results**

A. Phase 1 results in UK4 (all results) and nominal significant association results in Ashkenazi (p<0.05).

B. Joint analysis of Phase 1 and 2 results in UK4 (all results) and nominal significant association results in Ashkenazi (p<0.05).

UK4 data (green circles), Ashkenazi data (red crosses), results for two *HNF4A* P2 SNPs (rs1884613 and rs2144908) in the Ashkenazi (blue triangle); -log10 of the unadjusted p-value is shown on the Y-axis while chromosome coordinates are shown on the X-axis. Dashed line represents nominal significance p=0.05. Genes with most significant results are shown as black bars, while location of *HNF4A* on the chromosome is shown as grey bar for reference.
HNF4A P2 associates with T2D in Ashkenazim