Genome-wide association study identifies loci on 12q24 and 13q32 associated with Tetralogy of Fallot

Heather J. Cordell\textsuperscript{1}, Ana Töpf\textsuperscript{1}, Chrysovalanto Mamasoula\textsuperscript{1}, Alex V. Postma\textsuperscript{2}, Jamie Bentham\textsuperscript{3}, Diana Zelenika\textsuperscript{4,5}, Simon Heath\textsuperscript{4,6}, Gillian Blue\textsuperscript{7}, Catherine Cosgrove\textsuperscript{3}, Javier Granados Riveron\textsuperscript{8}, Rebecca Darlay\textsuperscript{1}, Rachel Soemedi\textsuperscript{1}, Ian J. Wilson\textsuperscript{1}, Kristin L. Ayers\textsuperscript{1}, Thahira J. Rahman\textsuperscript{1}, Darroch Hall\textsuperscript{1}, Barbara J.M. Mulder\textsuperscript{2}, Aelko H. Zwinderman\textsuperscript{2}, Klaartje van Engelen\textsuperscript{2}, J. David Brook\textsuperscript{8}, Kerry Setchfield\textsuperscript{8}, Frances A. Bu’Lock\textsuperscript{9}, Chris Thornborough\textsuperscript{9}, John O’Sullivan\textsuperscript{10}, A. Graham Stuart\textsuperscript{11}, Jonathan Parsons\textsuperscript{12}, Shoumo Bhattacharya\textsuperscript{3}, David Winlaw\textsuperscript{7}, Seema Mital\textsuperscript{13}, Marc Gewillig\textsuperscript{14}, Jeroen Breckpot\textsuperscript{15}, Koen Devriendt\textsuperscript{15}, Antoon F.M. Moorman\textsuperscript{2}, Anita Rauch\textsuperscript{16}, G. Mark Lathrop\textsuperscript{4,5}, Bernard D. Keavney\textsuperscript{1,*,†}, and Judith A. Goodship\textsuperscript{1,‡}

\textsuperscript{1}Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, UK, \textsuperscript{2}Academic Medical Center, Amsterdam, The Netherlands, \textsuperscript{3}Department of Cardiovascular Medicine, Oxford University, Oxford OX1 2JD, UK, \textsuperscript{4}Institut Genomique, Centre National de Genotypage, Commisariat à l’énergie Atomique (CEA), Evry, France, \textsuperscript{5}Centre d’Etude du Polymorphisme Humain, Fondation Jean Dausset, Paris 75010, France, \textsuperscript{6}Centro Nacional de Análisis Genómico, Barcelona, Spain, \textsuperscript{7}The Children’s Hospital at Westmead, Westmead, New South Wales, Australia, \textsuperscript{8}Institute of Genetics, Nottingham University, Nottingham NG7 2RD, UK, \textsuperscript{9}University Hospitals of Leicester NHS Trust, Leicester, UK, \textsuperscript{10}Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK, \textsuperscript{11}Bristol Royal Hospital for Children, Bristol, UK, \textsuperscript{12}Leeds Teaching Hospitals NHS Trust, Leeds, UK, \textsuperscript{13}Hospital for Sick Children, Toronto, Canada, \textsuperscript{14}Department of Pediatric Cardiology, \textsuperscript{15}Center for Human Genetics, University of Leuven, Leuven 3000, Belgium and \textsuperscript{16}Institute of Medical Genetics, University of Zurich, Zurich 8006, Switzerland

Received September 12, 2012; Revised December 17, 2012; Accepted December 24, 2012

We conducted a genome-wide association study to search for risk alleles associated with Tetralogy of Fallot (TOF), using a northern European discovery set of 835 cases and 5159 controls. A region on chromosome 12q24 was associated (\(P = 1.4 \times 10^{-7}\)) and replicated convincingly (\(P = 3.9 \times 10^{-5}\)) in 798 cases and 2931 controls [per allele odds ratio (OR) = 1.27 in replication cohort, \(P = 7.7 \times 10^{-11}\) in combined populations]. Single nucleotide polymorphisms in the glypican 5 gene on chromosome 13q32 were also associated (\(P = 1.7 \times 10^{-7}\)) and replicated convincingly (\(P = 1.2 \times 10^{-5}\)) in 789 cases and 2927 controls (per allele OR = 1.31 in replication cohort, \(P = 3.03 \times 10^{-11}\) in combined populations). Four additional regions on chromosomes 10, 15 and 16 showed suggestive association accompanied by nominal replication. This study, the first genome-wide association study of a congenital heart malformation phenotype, provides evidence that common genetic variation influences the risk of TOF.
INTRODUCTION

Congenital heart disease (CHD) affects ~1% of live births and is a major source of morbidity and mortality in childhood. Approximately 20% of CHD occurs in the setting of chromosomal conditions or multisystem malformation syndromes. Family studies in the remaining 80% of ‘sporadic’ cases indicate a significant complex genetic component to the disease (1). Previous studies have indicated that rare and de novo copy number variation in the human genome contribute 5–10% of the population risk of sporadic CHD (2–5), but genome-wide association studies (GWAS) assessing the relationship between common single nucleotide polymorphism (SNP) and CHD risk are yet to be reported. Tetralogy of Fallot (TOF) is the commonest form of cyanotic CHD, affecting ~3 per 10,000 newborns (6). Although TOF is usually repaired in infancy with low mortality, there is substantial late morbidity, in particular from pulmonary valvular insufficiency and atrial and ventricular arrhythmias. Population studies suggest a substantial familial recurrence risk in sporadic, non-syndromic TOF (1,7). We, therefore, undertook a GWAS to identify common genetic risk factors predisposing to TOF, given the high success of this design over the past 5 years (8).

RESULTS

Figure 1 shows the genome-wide association results obtained with two complementary approaches, case/control analysis in PLINK and family-based analysis in estimation of maternal, imprinting and interaction effects software (EMIM) (see Materials and Methods). With PLINK, the strongest signal of association occurred at a group of SNPs on chromosome 12q24 (top SNP rs11065987, \( P = 4.6 \times 10^{-8} \)), with several other regions, including SNPs on chromosomes 3, 10, 13 and 16 reaching more modest, but suggestive levels of significance (\( P \leq 1 \times 10^{-5} \)). Quantile–quantile (QQ) plots (Supplementary Material, Fig. S1A) indicated a slight inflation in the genome-wide distribution of test statistics (genomic control inflation factor \( l = 1.076 \)) (9), possibly due to unmodelled population substructure. Correction using the top 10 principal components from EIGENSOFT reduced the genomic control factor \( l \) to 1.037, with no systematic departure from the expected line in the resulting QQ plot (Supplementary Figure 1. Manhattan plots of the genome-wide association results. The top panel (PLINK results) shows the –log10 \( P \)-values from the Cochran–Armitage trend test (for autosomal SNPs) or logistic regression allowing for gender as a covariate (for X-chromosomal SNPs). The bottom panel (EMIM results) shows the \( -\log_{10} P \)-values from the child trend test (autosomal SNPs only). Dashed lines are shown at significance thresholds for suggestive (\( P = 1 \times 10^{-5} \)) and significant (\( P = 5 \times 10^{-8} \)) association, respectively.
The association was seen on chromosome 13q32 (rs7982677), with a per allele odds ratio of $1.7 \times 10^{-8}$, reducing to $P = 1.7 \times 10^{-7}$ with correction for overall significance. The chromosome 13q32 region strongly replicated in Dutch and Canadian TOF cases and controls of northern European ancestry. The chromosome 12q24 region strongly replicated in the European, North American, and Asian TOF GWAS. The chromosome 12q24 region also strongly replicated on chromosomes 10 (rs2228638, $P = 1.03 \times 10^{-10}$ at rs12593223) and 16 (rs17696736, $P = 1.3 \times 10^{-10}$ at rs11065987), with a per allele odds ratio (OR) of $1.7 \times 10^{-8}$ at rs12593223 and $1.3 \times 10^{-10}$ at rs17696736.

More modest levels of replication were seen at two separate regions of chromosome 10 ($P = 0.033$ at rs4901150) and on chromosomes 15 ($P = 0.0062$ at rs233722, combined discovery and replication, $P = 1.2 \times 10^{-8}$ at rs7982677) and on chromosome 13q32 ($P = 0.033$ at rs4901150, combined discovery and replication, $P = 3.03 \times 10^{-9}$ at rs12593223 and 16 ($P = 0.0018$ at rs233722, combined discovery and replication, $P = 9.7 \times 10^{-8}$ at rs11065987), with a per allele odds ratio (OR) of $1.3 \times 10^{-9}$ at rs233722, $1.5 \times 10^{-9}$ at rs11065987, and $1.1 \times 10^{-9}$ at rs12593223, respectively. These findings suggest that the association signal at chromosome 12q24 could largely be accounted for by a single variant (rs11065987), whereas only one SNP (rs11065987) retained nominal significance ($P = 1.4 \times 10^{-5}$ at rs7982677, combined discovery and replication, $P = 3.9 \times 10^{-5}$ at rs233722), suggesting that LD with rs11065987 could account for most of the strong associations seen at SNPs in this region. In the replication cohort, only two SNPs (rs233722 and rs17696736) retained nominal significance ($P = 0.0043$ at rs233722, $P = 0.0062$ at rs17696736), whereas only one SNP (rs11065987) retained nominal significance ($P = 0.057$ at rs11065987) in the discovery cohort. The chromosome 13q32 region also strongly replicated in the replication cohort. The chromosome 13q32 region strongly replicated in the replication cohort. The chromosome 13q32 region also strongly replicated in the replication cohort.

To further refine the association signal at chromosome 12q24, we carried out imputation in the discovery cohort within the LD block. More modest levels of replication were seen at two separate regions of chromosome 10 ($P = 0.033$ at rs4901150) and on chromosomes 15 ($P = 0.0062$ at rs233722, combined discovery and replication, $P = 1.2 \times 10^{-8}$ at rs7982677) and on chromosome 13q32 ($P = 0.033$ at rs4901150, combined discovery and replication, $P = 3.03 \times 10^{-9}$ at rs12593223 and 16 ($P = 0.0018$ at rs233722, combined discovery and replication, $P = 9.7 \times 10^{-8}$ at rs11065987), with a per allele odds ratio (OR) of $1.3 \times 10^{-9}$ at rs233722, $1.5 \times 10^{-9}$ at rs11065987, and $1.1 \times 10^{-9}$ at rs12593223, respectively. These findings suggest that the association signal at chromosome 12q24 could largely be accounted for by a single variant (rs11065987), whereas only one SNP (rs11065987) retained nominal significance ($P = 1.4 \times 10^{-5}$ at rs7982677, combined discovery and replication, $P = 3.9 \times 10^{-5}$ at rs233722), suggesting that LD with rs11065987 could account for most of the strong associations seen at SNPs in this region. In the replication cohort, only two SNPs (rs233722 and rs17696736) retained nominal significance ($P = 0.0043$ at rs233722, $P = 0.0062$ at rs17696736), whereas only one SNP (rs11065987) retained nominal significance ($P = 0.057$ at rs11065987) in the discovery cohort. The chromosome 13q32 region also strongly replicated in the replication cohort. The chromosome 13q32 region strongly replicated in the replication cohort. The chromosome 13q32 region also strongly replicated in the replication cohort.
Figure 2. LocusZoom plots for the replicating loci on chromosomes 12 (PLINK results) and 13 (EMIM results). The colour coding indicates the degree of LD of each SNP with the named index SNP (shown as a purple diamond), as estimated from the HapMap CEU population.
DISCUSSION

Our study provides compelling evidence that common genetic variation influences the risk of CHD. The top SNPs on chromosome 12 implicated by our discovery and replication analyses ($P = 7.7 \times 10^{-11}$) are located within a large LD block on 12q24. This region has previously been associated with a variety of complex genetic conditions, including type 1 diabetes (11), celiac disease (12), coronary artery disease (13), rheumatoid arthritis (14), systemic lupus erythematosus (15) multiple sclerosis (16) and with blood platelet count (13). The risk alleles at the top SNPs in our analysis all lie on a common (40% frequency in individuals of White European ancestry) 1.6 Mb haplotype at chromosome 12q24, containing 15 genes, that spans our associated interval (Fig. 2 and Supplementary Material, Table S3). A previous analysis (13) provided strong evidence that the haplotype has undergone recent positive selection in individuals of European ancestry. For all conditions associated with this haplotype described thus far, including TOF, the alleles conferring increased risk at the associated SNPs are the derived alleles, whose frequencies have increased in Europeans by a selective sweep (starting around 3000 years ago). Our chromosome 12 results add to the remarkable disease pleiotropy associated with variation in the 12q24 region of the human genome.

Imputation analysis in our discovery cohort indicated that, although our association signal could be driven by any one of a number of imputed SNPs, none show significantly stronger association than is already seen at our top genotyped SNP (rs11065987). Further interrogation of this region, ideally in non-European cohorts, showing varying LD patterns, will be required to further refine the association signal. Within the 12q24 region, the strongest candidate gene is tyrosine-protein phosphatase non-receptor type 11 (PTPN11), a regulator of Ras/mitogen-associated protein kinase signalling. Activating mutations of PTPN11 are a cause of Noonan’s syndrome (17), a Mendelian multisystem disorder in which malformation of the cardiac outflow tract is a typical feature. Association between a SNP in PTPN11 (rs11066320) and TOF was previously demonstrated (18) in a candidate gene study using a family-based association analysis approach on a set of 754 TOF cases that partially overlap with the set of cases described here. Further research will be necessary to investigate the hypothesis that the SNP association we have observed in this region results from upregulation of the activity of PTPN11 and the magnitude of effect compared with that of the gain-of-function mutations that lead to Noonan’s syndrome.

The observation of association between TOF and an evolutionarily selected haplotype at 12q24 raises some intriguing questions. In contrast with conditions of adult onset, such as coronary artery disease and rheumatoid arthritis, previously shown to be associated with 12q24 SNPs, the condition on which we have focussed carries substantial early mortality that significantly impacts reproductive fitness. Prior to the modern cardiac surgical era, mortality from TOF before the age of 10 years was around 80%. Given this, it might be expected that any variant conferring even a modest additional risk of TOF would have been eliminated, or driven to very low allele frequency, by natural selection. The nature of the selection event responsible for the emergence of the pleiotropic 12q24 risk haplotype is unknown, although given the role of genes in the region in T-cell function, it is thought likely to relate to enhanced resistance to infectious disease at a time of increased human population density in Europe (13). Our data on a childhood phenotype that until recently was highly lethal illustrate the impact of immunity as a selective force in human evolution—as has been recently shown for the effects of Y chromosome haplogroups on immune function and atherosclerosis risk (19).

Our findings on chromosome 13 [rs7982677 at 13q32, OR = 1.31, 95% CI = (1.16–1.48), $P = 3.03 \times 10^{-11}$] and in two separate regions of chromosome 10 [rs2388986 at 10p14, OR = 0.83, 95% CI = (0.74, 0.93), $P = 8.55 \times 10^{-8}$ and rs2228638 at 10p11.2, OR = 1.28, 95% CI = (1.07, 1.53), $P = 2.05 \times 10^{-7}$] are also significant. The top SNPs in the 13q32 region lie in intron 7 of the GPC5 gene that encodes for glypican 5. Glypicans are heparan sulphate proteoglycans that are bound to the outer surface of the plasma membrane by a glycosyl-phosphatidylinositol anchor. There are six members of the glypican family in mammals, and they serve as regulators of key developmental signalling pathways, including the Wnt, Hedgehog, fibroblast growth factor and bone morphogenetic protein pathways (20). Glypicans are, therefore, strong candidate genes for involvement in cellular growth control and morphogenesis during heart development. GPC5 has eight exons and occupies ~1.42 Mb at chromosome 13p32. The seventh intron, which harbours our top SNPs, is 735 Kb long and contains three putative non-coding transcripts, none of which are overlapped by our top SNPs. SNPs in GPC5 have been associated with the risk of nephrotic syndrome (21), with protection from sudden cardiac arrest (22) and with higher risks of lung cancer and multiple sclerosis (23,24). GPC5 is among the genes deleted in a subset of patients with 13q deletion syndrome, a multisystem disorder (typically characterized by holoprosencephaly) in which CHD can occur. In a French cohort of 12 patients with 13q deletion, 2 had TOF, and GPC3 was deleted in both of these patients (25); single patients with 13q deletion and TOF have been reported by two other groups (26,27). Defects in genes encoding for other members of the glypican protein family have been linked to cardiac malformation syndromes. CHD is common in Simpson–Golabi–Behmel syndrome, the condition associated with GPC3 mutation (28), and has been reported in omopalsydiasia, the condition associated with GPC6 mutation (29). Given the large size of the genomic interval spanned by the GPC5 gene, it seems likely that the association we have observed here results from an effect on the function of GPC5 itself, rather than a neighbouring gene.

On chromosome 10, rs2388986 lies in a gene desert [the nearest gene, GATA-binding transcription factor protein family, lies some 0.8 Mb distant], whereas rs2228638 is a non-synonymous coding SNP that results from the substitution of leucine for Valine at position 733 of the neuropilin-1 protein, a regulator of Ras/mitogen-associated protein kinase signalling (28), and has been reported in omodysplasia, the condition associated with 13q32 region lie in intron 7 of the GPC5 gene that encodes for glypican 5. Glypicans are heparan sulphate proteoglycans that are bound to the outer surface of the plasma membrane by a glycosyl-phosphatidylinositol anchor. There are six members of the glypican family in mammals, and they serve as regulators of key developmental signalling pathways, including the Wnt, Hedgehog, fibroblast growth factor and bone morphogenetic protein pathways (20). Glypicans are, therefore, strong candidate genes for involvement in cellular growth control and morphogenesis during heart development. GPC5 has eight exons and occupies ~1.42 Mb at chromosome 13p32. The seventh intron, which harbours our top SNPs, is 735 Kb long and contains three putative non-coding transcripts, none of which are overlapped by our top SNPs. SNPs in GPC5 have been associated with the risk of nephrotic syndrome (21), with protection from sudden cardiac arrest (22) and with higher risks of lung cancer and multiple sclerosis (23,24). GPC5 is among the genes deleted in a subset of patients with 13q deletion syndrome, a multisystem disorder (typically characterized by holoprosencephaly) in which CHD can occur. In a French cohort of 12 patients with 13q deletion, 2 had TOF, and GPC3 was deleted in both of these patients (25); single patients with 13q deletion and TOF have been reported by two other groups (26,27). Defects in genes encoding for other members of the glypican protein family have been linked to cardiac malformation syndromes. CHD is common in Simpson–Golabi–Behmel syndrome, the condition associated with GPC3 mutation (28), and has been reported in omopalsydiasia, the condition associated with GPC6 mutation (29). Given the large size of the genomic interval spanned by the GPC5 gene, it seems likely that the association we have observed here results from an effect on the function of GPC5 itself, rather than a neighbouring gene.

On chromosome 10, rs2388986 lies in a gene desert [the nearest gene, GATA-binding transcription factor protein family, lies some 0.8 Mb distant], whereas rs2228638 is a non-synonymous coding SNP that results from the substitution of leucine for Valine at position 733 of the neuropilin-1 protein, which is encoded by the gene NRP1. Neurphilin-1 is a receptor for ligands from the vascular endothelial growth factor and bone morphogenetic protein pathways (20). Glypicans are, therefore, strong candidate genes for involvement in cellular growth control and morphogenesis during heart development. GPC5 has eight exons and occupies ~1.42 Mb at chromosome 13p32. The seventh intron, which harbours our top SNPs, is 735 Kb long and contains three putative non-coding transcripts, none of which are overlapped by our top SNPs. SNPs in GPC5 have been associated with the risk of nephrotic syndrome (21), with protection from sudden cardiac arrest (22) and with higher risks of lung cancer and multiple sclerosis (23,24). GPC5 is among the genes deleted in a subset of patients with 13q deletion syndrome, a multisystem disorder (typically characterized by holoprosencephaly) in which CHD can occur. In a French cohort of 12 patients with 13q deletion, 2 had TOF, and GPC3 was deleted in both of these patients (25); single patients with 13q deletion and TOF have been reported by two other groups (26,27). Defects in genes encoding for other members of the glypican protein family have been linked to cardiac malformation syndromes. CHD is common in Simpson–Golabi–Behmel syndrome, the condition associated with GPC3 mutation (28), and has been reported in omopalsydiasia, the condition associated with GPC6 mutation (29). Given the large size of the genomic interval spanned by the GPC5 gene, it seems likely that the association we have observed here results from an effect on the function of GPC5 itself, rather than a neighbouring gene.
will be required to identify whether the association is due to a functional effect of rs2228638 or of some other variant in LD.

In conclusion, our study has identified two regions (12q24 and 13q32) that are strongly and replicably associated with TOF. As far as we are aware, this is the first genome-wide study focussing on SNP associations to have been reported in CHD. Pooling patient resources from several international groups was required to achieve sufficient numbers of cases for adequate power to detect and replicate a number of genetic effects; however, studies of commoner diseases have emphasized the importance of very large cohorts of cases and controls. The present relatively small GWAS cohort has in all likelihood detected only the strongest genome-wide influences on TOF risk. Our results suggest that larger international collaborative studies have the potential to discover additional significantly associated loci.

Complete summary association results from the GWAS component of our study (the EIGENSOFT corrected results) are available to interested researchers by contacting the corresponding author.

MATERIALS AND METHODS

Study subjects and genotyping

For the initial (discovery) phase, individuals with TOF of northern European ancestry, together with their parents and affected siblings (when available), were recruited from multiple centres in Newcastle, Leeds, Bristol, Liverpool, Oxford, Nottingham and Leicester (all UK), Leuven (Belgium), Erlangen (Germany) and Sydney (Australia). Ethical approval was obtained from the local institutional review boards at each of the participating centres prior to blood or saliva sample collections, and informed consent was obtained from all subjects, or from their parents/legal guardians, if the patient was a child too young to provide consent him/herself. Patients who exhibited clinical features of recognized malformation syndromes, developmental abnormalities or learning difficulties were excluded from the study. All samples were screened for the known 22q11.2 deletion associated with TOF (33,34) using multiplex ligation-dependent probe amplification (MRC-Holland) and excluded, if the deletion was present. SNP genotyping was carried out at the Centre National de Genotypage (Evry Cedex, France) using the Illumina 660wQUAD array, and the genotypes were compared with genotype data for UK population-based controls (5667 individuals genotyped on the Illumina 1.2M chip) obtained from the Wellcome Trust Case Control Consortium 2 (WTCCC2) (https://ccc.wtcc.org.uk/cc2).

For the replication phase, individuals with TOF of self-reported Caucasian ancestry were recruited from Oxford, Nottingham, Newcastle, The Netherlands and Canada. Five hundred and fifteen Dutch TOF cases were identified from the Netherlands national registry and DNA bank of patients with CHD (CONCOR), the design of which has been previously described (35). One hundred and forty-four Canadian TOF cases, together with Canadian controls of self-reported Caucasian ancestry, were obtained from the SickKids Heart Centre Biobank, an Ontario province-wide biorepository for CHD. These replication samples were genotyped at the Centre National de Genotypage using Sequenom matrix-assisted laser desorption/ionization—time of flight. Additional UK population-based control data for the replication were obtained from the TwinsUK resource (http://www.twinsuk.ac.uk), an adult twin registry comprising 12 000 (predominantly female) twins. Genotype data for 3512 twin individuals (genotyped using the Illumina 610K array) were obtained from the Department of Twin Research and Genetic Epidemiology at King’s College, London. Only the first twin from each pair of genotyped twins (2603 unrelated individuals) was used in the current study.

Statistical analysis

Following stringent quality control (QC) procedures (see below), the final discovery dataset for analysis consisted of 835 unrelated cases and 5159 controls (plus 717 additional family members, including both parents for 293 of the cases), genotyped at 516 131 autosomal and X-chromosomal SNPs. The primary analysis performed was a case/control analysis of the 835 unrelated cases and 5159 controls using either the Cochran–Armitage trend test (for autosomal SNPs) or logistic regression, allowing for gender as a covariate (for X-chromosomal SNPs), using the software PLINK (36). We also used two alternative approaches designed to correct for possible population stratification: Firstly, we used the ‘smartpca’ routine of the EIGENSOFT package (37) to calculate the top 10 principal components that were entered as covariates into a logistic regression analysis of each SNP across the genome. Secondly, we analysed each SNP using a score test from a linear mixed model (38), allowing for possible relatedness between individuals via consideration of their genome-wide estimated kinship coefficients, using the ‘mmscore’ function in the R package GenABEL (39). This approach has been previously proposed as a means of correcting for unknown population structure between apparently unrelated individuals in a genome-wide association study (40).

In addition to the case/control analyses, we also performed a family-based association analysis (of autosomal SNPs only) of all cases, their relatives and controls using the software package EMIM (41). EMIM is primarily designed for the investigation of complex genetic effects, including maternal genotype effects, maternal–fetal interactions and imprinting (41). In this instance, we used the ‘child trend’ model in EMIM to model only an effect of the child’s (case’s) own genotype. The resulting analysis was, thus, conceptually similar to the case/control analysis we had performed in PLINK, but with the advantage of allowing additional information from parental genotypes to be incorporated, where available.

The replication cohort (after QC) consisted of 798 cases and 2931 controls. Genotypes were analysed using logistic regression in PLINK, allowing for two levels of nationality (British/Dutch or Canadian) as a covariate. P-values for association in the combined (discovery and replication) datasets were calculated using Fisher’s trend approach for combining P-values as implemented in the online software MetaP (http://www.svaproject.org/metap.php).

To further refine the association signal at chromosome 12q24, we carried out imputation in the discovery cohort within the 5 Mb region centred around rs11065987. We used
the program IMPUTE version 2 (42) with data from the 1000 Genomes Project (Phase I interim data, released June 2011) as a reference panel. Data at 11 208 SNPs passing post-imputation QC (from an original 56 637 imputed SNPs) were analysed using the program SNPTEST to test for association with disease status.

QC procedures

Discovery cohort
We used stringent QC checks to ensure that only high-quality data were included in the final analysis. QC procedures were carried out in PLINK version 1.07 (36) with visualization performed in R (http://www.r-project.org/). For the current study, genotype data were generated at 557 124 SNPs across the genome for 1733 individuals (comprising 913 TOF cases plus a number of unaffected relatives). We excluded individuals with genotype call rates <98.78% and average heterozygosities outside the range (0.310, 0.336) (based on consideration of 540 241 autosomal SNPs passing loose QC, namely successfully genotyped in >95% of individuals and with a Hardy–Weinberg equilibrium test \( P \)-value >10\(^{-8} \)). These exclusion thresholds were chosen based on visual inspection of the call rates and heterozygosities (Supplementary Material, Fig. S5) to retain the majority of individuals, whereas excluding outlying individuals.

We generated a smaller set of 41 692 autosomal SNPs (successfully genotyped in >95% of individuals, with a Hardy–Weinberg equilibrium test \( P \)-value >10\(^{-8} \), with minor allele frequencies >0.4 and pruned to show low levels of LD using the PLINK command ‘--indep 50 5 2’) that were used to check relationships/sample duplications and ethnicities. Genome-wide identity-by-descent (IBD) sharing was calculated using the ‘--Z-genome’ command in PLINK, and one of each pair of related individuals (mean proportion of alleles shared IBD >0.1) was excluded. Multidimensional scaling of our samples together with 210 unrelated Phase II HapMap (43) individuals from 4 populations [Utah residents with ancestry from northern and western Europe (CEU), Japanese in Tokyo, Japan, Han Chinese in Beijing, China and Yoruba in Ibadan, Nigeria] (genotyped at same set of 41 692 autosomal SNPs) was performed and identified 33 individuals in our study who did not cluster with the CEU samples, suggesting non-European ancestry (Supplementary Material, Fig. S6). These individuals were excluded.

We used the ‘--check-sex’ option in PLINK to check (based on the average X chromosomal heterozygosity) that the gender of our samples matched its expected value and excluded samples for which we were unable to resolve inconsistencies.

Following QC, we were left with 835 unrelated Phase II HapMap (43) individuals from 4 populations [Utah residents with ancestry from northern and western Europe (CEU), Japanese in Tokyo, Japan, Han Chinese in Beijing, China and Yoruba in Ibadan, Nigeria] (genotyped at same set of 41 692 autosomal SNPs) was performed and identified 33 individuals in our study who did not cluster with the CEU samples, suggesting non-European ancestry (Supplementary Material, Fig. S6). These individuals were excluded.

We used the ‘--check-sex’ option in PLINK to check (based on the average X chromosomal heterozygosity) that the gender of our samples matched its expected value and excluded samples for which we were unable to resolve inconsistencies.

Replication cohort

QC was also performed on the genotype data from the TwinsUK replication sample. From the 2603 first twins considered, we excluded 43 showing genotype call rates <99% and average heterozygosities outside the range (0.312, 0.331) (based on the consideration of 576 610 autosomal SNPs passing loose QC, namely: successfully genotyped in >95% of individuals and with a Hardy–Weinberg equilibrium test \( P \)-value >10\(^{-8} \)). These exclusion thresholds were chosen based on visual inspection of the call rates and heterozygosities. We carried out testing of relationships/sample duplications and ethnicity using the same approach as described above for the TOF cohort and excluded first twins who did not cluster with the CEU HapMap samples and one of each pair of first twins who showed high IBD sharing (mean proportion of alleles IBD >0.05). This resulted in a final set of 2547 TwinsUK controls to be used in the replication study.

Post-imputation QC

We used the program IMPUTE version 2 (42) to carry out imputation in the discovery cohort within the 5 Mb region centred around rs11065987. Data from the 1000 Genomes Project (47) (Phase I interim data, released June 2011) were used as a reference panel. Post-imputation QC involved excluding any SNPs likely to be poorly imputed (specifically those with an ‘info’ score <0.5 or with minor allele frequency in controls <0.01). Data at 11 208 SNPs passing post-imputation QC (from an original 56 637 imputed SNPs) were analysed in the program SNPTEST version 2.1.1 (46), using the ‘-method ml’ option (a Newton–Raphson algorithm that maximizes the missing data likelihood) to account for genotype uncertainty.
SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

This study made use of data generated by the Wellcome Trust Case-Control Consortium 2 (WTCCC2). A full list of the investigators who contributed to the generation of the data is available from http://www.wtccc.org.uk. Access to genotyping data from the TwinsUK cohort was kindly provided by the Department of Twin Research (DTR) and Genetic Epidemiology at King’s College, London. We thank the staff of the Twin Research Unit for their help and support in undertaking this project. We would also like to thank all patients, their families and control subjects for their voluntary contribution to this research project. B.D.K. and J.A.G. acknowledge the support of the National Institute for Health Research, through the Northumberland, Tyne and Wear CLRN.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the British Heart Foundation, the European Community’s 7th Framework Programme contract ‘CHeartED’ (grant number HEALTH-F2-2008-223040) the Wellcome Trust (grant number 087436), the Heart and Stroke Foundation of Ontario, Ontario Ministry of Research and Innovation, McLaughlin Centre, Labatt Family Heart Centre and Canadian Foundation for Innovation (to S.M.). B.D.K. and S.B. hold BHF Personal Chairs. Funding for the Centre and Canadian Foundation for Innovation (to S.M.).

REFERENCES


