Published in final edited form as:
Nat Genet. 2013 July ; 45(7): 822–824. doi:10.1038/ng.2637.

Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16

Heather J. Cordell[1,*], **Jamie Bentham**[2,*], **Ana Topf**[1], **Diana Zelenika**[3,4], **Simon Heath**[3,5], **Chrysovalanto Mamasoula**[1], **Catherine Cosgrove**[2], **Gillian Blue**[6], **Javier Granados-Riveron**[7,§], **Kerry Setchfield**[7], **Chris Thornborough**[8], **Jeroen Breckpot**[9], **Rachel Soemedi**[1], **Ruairidh Martin**[1], **Thahira J. Rahman**[1], **Darroch Hall**[1], **Klaartje van Engelen**[10], **Antoon F.M. Moorman**[11], **Aelko H Zwinderman**[12], **Phil Barnett**[11], **Tamara T. Koopmann**[13], **Michiel E. Adriaens**[13], **Andras Varro**[14], **Alfred L. George Jr.**[15], **Christobal dos Remedios**[16], **Nanette H. Bishopric**[17], **Connie R. Bezzina**[13], **John O'Sullivan**[18], **Marc Gewillig**[19], **Frances A. Bu'Lock**[8], **David Winlaw**[6], **Shoumo Bhattacharya**[2], **Koen Devriendt**[9], **J. David Brook**[7], **Barbara J.M. Mulder**[20], **Seema Mital**[21], **Alex V. Postma**[11], **G. Mark Lathrop**[3,4], **Martin Farrall**[2], **Judith A. Goodship**[1], and **Bernard D. Keavney**[1,22]

[1]Institute of Genetic Medicine, Newcastle University, Newcastle-upon-Tyne, UK [2]Department of Cardiovascular Medicine, University of Oxford, Oxford, UK [3]Commisariat a l'energie Atomique (CEA), Institut Genomique, Centre National de Genotypage, Evry, France [4]Fondation Jean Dausset, Centre d'Etude du Polymorphisme Humain, Paris, France [5]Centro Nacional de Analisis Genomico, Barcelona, Spain [6]The Children's Hospital at Westmead, Westmead, Australia [7]Institute of Genetics, Nottingham University, Nottingham, UK [8]East Midlands Congenital Heart Centre, University Hospitals of Leicester NHS Trust, Leicester, UK [9]Center for Human Genetics, University of Leuven, Leuven, Belgium [10]Department of Clinical Genetics, Academic Medical Centre, Amsterdam, The Netherlands [11]Heart Failure Research Center, Academic Medical Center, Amsterdam, The Netherlands [12]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre, Amsterdam, The Netherlands [13]Heart Failure

Correspondence: Bernard Keavney Institute of Cardiovascular Sciences The University of Manchester Core Technology Facility 46 Grafton Street Manchester M13 9NT bernard.keavney@manchester.ac.uk Or: Judith Goodship Institute of Genetic Medicine Newcastle University Central Parkway Newcastle NE1 3BZ UK judith.goodship@ncl.ac.uk.

*Contributed equally

§Current address: Department of Genetics, Genomics and Bioinformatics, Hospital Infantil de Mexico "Federico Gomez", Mexico City, Mexico.

URLs:
Wellcome Trust Case-Control Consortium 2: https://ccc.wtccc.org.uk/ccc2
Twins UK: http://www.twinsuk.ac.uk
Online Mendelian Inheritance in Man (OMIM): http://www.ncbi.nlm.nih.gov/omim
R Project for Statistical Computing : www.r-project.org

AUTHOR CONTRIBUTIONS:
Conceived and designed study: HJC, JB, SB, JDB, GML, MF, JAG, BK.
Wrote grant: JB, MF, BK.
Supervised/coordinated discovery sample collection: JB, AT, CC, GB, JG-R, KS, CT, JB, JOS, MG, FAB, DW, KD, JAG, BK.
Supervised/coordinated replication sample collection: AT, GB, JG-R, KS, KvE, AFMM, AHZ, FAB, DW, BJMM, SM, AVP, JAG, BK
Supervised/coordinated human tissue sample collection: AV, ALG, CDR, NHB, CRB
Performed or supervised laboratory work: JB, AT, DZ, SH, JG-R, KS, RS, TTK, PB, CRB, GML, JAG, BK
Performed statistical analyses: HJC, CM, MEA, MF
Wrote first draft: HJC, JB, JAG, BK
Commented critically and revised draft: All.

The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

Research Center, Department of Clinical and Experimental Cardiology, Academic Medical Center, Amsterdam, The Netherlands [14]Department of Pharmacology and Pharmacotherapy, Faculty of Medicine, University of Szeged, Szeged, Hungary [15]Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN, USA [16]Muscle Research Unit, Department of Anatomy, Bosch Institute, The University of Sydney, Sydney, Australia [17]Department of Medicine, Molecular and Cellular Pharmacology, University of Miami School of Medicine, Miami, USA [18]Newcastle Hospitals NHS Foundation Trust, Newcastle-upon-Tyne, UK [19]Pediatric Cardiology, University of Leuven, Leuven, Belgium [20]Heart Centre, Academic Medical Centre, Amsterdam, The Netherlands [21]Hospital for Sick Children, Toronto, Canada [22]Institute of Cardiovascular Sciences, The University of Manchester, Manchester, UK

Abstract

We carried out a genome-wide association study (GWAS) of congenital heart disease (CHD). Our discovery cohort comprised 1,995 CHD cases and 5,159 controls, and included patients from each of the three major clinical CHD categories (septal, obstructive and cyanotic defects). When all CHD phenotypes were considered together, no regions achieved genome-wide significant association. However, a region on chromosome 4p16, adjacent to the *MSX1* and *STX18* genes, was associated ($P=9.5\times10^{-7}$) with the risk of ostium secundum atrial septal defect (ASD) in the discovery cohort (N=340 cases), and this was replicated in a further 417 ASD cases and 2520 controls (replication $P=5.0\times10^{-5}$; OR in replication cohort 1.40 [95% CI 1.19-1.65]; combined $P=2.6\times10^{-10}$). Genotype accounted for ~9% of the population attributable risk of ASD.

Congenital heart disease (CHD) is the most frequent congenital disorder in newborns affecting 7/1000 live births; it is a significant cause of childhood death and long-term morbidity [1]. Chromosomal abnormalities, rare genomic copy number variants, Mendelian disorders and *in utero* exposures together account for approximately a quarter of CHD cases; among the remaining "sporadic" cases there is significant heritability which is currently unexplained [2]. We conducted a genome-wide association study (GWAS) to determine whether we could detect common genetic variants that influence CHD risk.

A discovery cohort comprising CHD cases of self-reported European Caucasian ancestry was recruited from multiple centres in the United Kingdom, and from centres in Leuven (Belgium) and Sydney (Australia). All diagnoses were established by congenital heart disease specialists in the contributing centres and patients were classified using European Paediatric Cardiac Codes. Patients exhibiting clinical features of recognized malformation syndromes, multiple developmental abnormalities or learning difficulties were excluded from the study. 1,995 CHD cases were genotyped in the discovery cohort, with a distribution of phenotypes as shown in Supplementary Table 1. SNP genotyping in the cases was carried out using the Illumina 660wQUAD array, and the genotypes were compared with 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). After stringent QC, 1,819 unrelated CHD cases and 5,159 WTCCC controls with genotypes at 514,952 autosomal and X-chromsomal SNPs were included in the primary analyses. When all CHD phenotypes were considered together in the case/control analyses, no SNPs reliably achieved conventionally accepted genome-wide significance (the only two SNPs achieving $P<10^{-6}$ were not supported by any signal in the surrounding regions and were deemed most likely false positives, see Supplementary Figure 1).

We carried out pre-specified subsidiary analyses in the five largest diagnostic groups: ventricular septal defect (VSD), ostium secundum atrial septal defect (ASD), transposition

of the great arteries (TGA), conotruncal malformations, and left-sided malformations. Although no SNPs achieved conventional GWAS significance ($p=5\times10^{-8}$) in these analyses, signals of $P<1\times10^{-6}$ supported by evidence at three or more neighbouring SNPs were present for the common malformations VSD and ASD (Supplementary Figure 2). For replication purposes, we followed up these signals and any others achieving $P<2\times10^{-5}$ within the VSD or ASD subgroups. We genotyped 21 SNPs in 11 different genomic regions in 209 VSD replication samples and 10 SNPs in 6 different regions in 417 secundum ASD replication samples, all of Caucasian ancestry and originating from The Netherlands and Canada; and compared the genotypes with data for 2520 individuals of Caucasian ancestry from the TwinsUK resource genotyped using the Illumina 610K array. No significant replicated association was observed for VSD. In contrast, association with ASD was replicated at three SNPs on chromosome 4p16, where the top SNP rs870142 (which had minor allele frequency in controls = 0.23) conferred an OR 1.52 ($p=9.5\times10^{-7}$) per copy of the minor allele in the discovery cohort, and OR 1.40 ($p=5.0\times10^{-5}$) in the replication cohort; when these results were combined the overall OR was 1.46 ($p=2.6\times10^{-10}$). Results for the top 3 ASD SNPs are shown in Table 1, and results for all SNPs typed in the replication cohort are shown in Supplementary Tables 2 and 3. Using Levin's formula, genotype at rs870142 accounted for 9% of the population attributable risk of ASD. A LocusZoom [3] plot of the 4p16 region, illustrating the strong linkage disequilibrium between the three replicating SNPs, is shown in Figure 1. Imputation analysis in the discovery cohort, using the 1000 Genomes as a reference panel, also provided strong support for the signal in this region (Supplementary Figure 3) with the top imputed SNP (rs4689904; $P=5.0\times10^{-7}$) achieving slightly higher significance than had been seen at the top genotyped SNPs. Finally, in an exploratory analysis pooling our limited number of atrio-ventricular septal defect (AVSD) cases (N=73) with ASD cases in the discovery cohort, the statistical significance of the top ASD SNPs was reduced (data not shown), suggesting that within the spectrum of atrial septal and AV canal defects the association is specific to secundum ASD.

ASD accounts for 7-10% of CHD in children, but for 25-30% of CHD in adult populations (due to childhood mortality from other CHD conditions, and diagnosis of ASD in adult life) [4]. People with ASD have a higher morbidity and mortality than those without, although this tends to be evident only at older ages [5]. Our top SNP rs870142 lies in the 300kb interval between *STX18* and *MSX1*. Large (typically >1.9Mb) deletions encompassing this region of chromosome 4 are responsible for Wolf-Hirschhorn syndrome (WHS, OMIM #194190), a rare developmental disorder which includes CHD (typically ASD) in around 50% of cases. *STX18* is involved in transport between the ER and Golgi [6] and is not an obvious candidate for ASD. By contrast, *MSX1* encodes a homeobox transcription factor which we show to be expressed in the atrial septum during development both in mouse and chick (Supplementary Figure 4). *MSX1* functionally interacts with *TBX5*, a transcription factor known to be critical in atrial septal development [7,8]. In the mouse, due to functional redundancy between the *Msx1* and *Msx2* genes, only double knockout animals display CHD, which involves abnormalities both of the outflow tract and the atrioventricular junction [9,10]. *MSX1* loss-of-function mutations in man cause tooth agenesis, cleft lip and palate, and Witkop (tooth-and-nail) syndrome, but CHD is not typically seen [11-13], making it somewhat unlikely that our top SNPs act solely by regulating the expression of *MSX1*.

One of the three associated SNPs (rs6824295) is located within a non-coding RNA transcript (GenBank Accession: *BI192733.1*) that is of unknown function; *BI192733.1* lies within an intron of the non-coding RNA gene *LOC100507266* and is transcribed in the same direction. We showed (Supplementary Figure 5) that the *BI192733.1* transcript is expressed in the developing human heart between the 9 and 20 week stage. Gene-expression studies of *LOC100507266,* which we performed in adult human cardiac tissue from transplant donor hearts, showed that the risk alleles at our ASD SNPs were associated (p=0.02 at top SNP),

with lower expression of *LOC100507266* (Supplementary Figure 6). These observations suggest that *cis* and/or *trans* acting influences of these non-coding RNAs on the transcription of other genes may be involved in the relationship between genotype at our associated SNPs and ASD risk. However, further work conducted in the appropriate developmental context will be required to definitively identify the mechanism responsible for the association we have observed in the 4p16 region.

We did not observe genome-wide significant association with CHD risk in all 1995 cases considered together, despite having had sufficient power to detect moderate-sized effects had they been present. Our rationale for this study design was that if such loci had been detectable, their population impact would have been much larger than any locus influencing only one phenotypic subgroup. The region of chromosome 12 that we have previously shown to be associated with the risk of the CHD condition Tetralogy of Fallot (TOF) [14] was not significantly associated with CHD risk, either overall or in any subgroup, in the present study (which did not include patients with TOF). Similarly, the association between SNPs at 4p16 and ASD was not apparent for CHD conditions other than ASD. Our work therefore adds to recent data from studies of copy number variants (CNVs) suggesting that genetic associations with CHD exhibit a considerable degree of phenotypic specificity [15,16]. Our analyses of even the commoner CHD conditions (in the case of ASD, 340 discovery and 417 replication cases) was of low power in comparison to the large-scale GWAS of commoner diseases now reported in the literature; replication of our findings in independent cohorts will be of value in future studies. Our findings emphasize the ongoing need for the establishment of large collections of homogeneous clinical CHD cases to detect additional associations.

In conclusion, we present evidence for association between common SNPs at 4p16 and the risk of ASD, a common CHD condition; the association accounts for around 9% of the population attributable risk. To the best of our knowledge, this is the first reported GWAS significant association with ASD.

ONLINE METHODS

Study subjects and genotyping

For the case cohort, ethical approval was obtained from the local institutional review boards at each of the participating centres. Informed consent was obtained from all participants, or from their parents/legal guardians in the case of children. In the discovery phase, control genotype data from healthy individuals of UK ancestry were obtained from the Wellcome Trust Case Control Consortium 2 (WTCCC2). In the replication phase, control genotype data from healthy individuals of Northern European ancestry were obtained from the TwinsUK resource. Only the first twin from each pair of genotyped twins (2603 unrelated individuals) contributed to the present study. Although no specific measures were taken to exclude CHD in the control cohorts, the prevalence of CHD in adult populations (~0.31%) indicates that any loss of power due to misspecification of controls would be negligible.

Genotyping of the discovery cohort used the Illumina660WQUAD array, and genotyping of replication SNPs used SEQUENOM MALDI-TOF. Genotyping was carried out at the Centre National de Genotypage (Evry Cedex), France.

Quality control Procedures and Statistical Analysis

Discovery Cohort—QC procedures were carried out in PLINK version 1.07 [17] with visualisation performed in R. Genotype data was initially generated at 557124 SNPs across the genome for 1995 individuals with CHD. We excluded individuals with genotype call

rates <98.5% and average heterozygosities outside the range [0.310, 0.331] (based on consideration of 538029 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 in order to retain the majority of individuals while excluding outlying individuals (Supplementary Figure 7).

We generated a smaller set of 40521 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 (defined as having probability > 8% of sharing 0 alleles IBD) was excluded. Multidimensional scaling of our samples together with 210 unrelated Phase II HapMap [18] individuals from four populations (CEU, JPT, CHB, YRI) (genotyped at the same set of 40521 autosomal SNPs) was performed and identified 22 individuals in our study that did not cluster with the CEU samples, suggesting non-European ancestry (Supplementary Figure 7). These individuals were excluded. We used the '--check-sex' option in PLINK to check (on the basis of 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 1819 unrelated CHD cases, whose genotypes were compared with genotype data from 5159 UK population-based controls obtained from the WTCCC2. These controls comprised 2673 samples from the 1958 British Birth Cohort (58C) and 2486 National Blood Service (NBS) samples (selected from an initially genotyped set of 2930 58C samples and 2737 NBS samples). We excluded the same controls as had been excluded in the WTCCC2 [19] and WTCCC3 [20] studies, plus an additional 4 controls that we found to be outliers following a principal components analysis using the 'smartpca' routine of the EIGENSOFT package [21].

Within each of the case and control cohorts, we excluded any SNPs with minor allele frequencies <0.01, that were successfully genotyped in <95% of individuals or that had a Hardy-Weinberg equilibrium test *P* value $<10^{-8}$. Within the two control cohorts, we also implemented SNP exclusions recommended by WTCCC2 relating to a measure of the statistical information in the genotype data about allele frequency (exclude if <0.975), missingness (exclude if >2% missing genotypes) and plate effects (exclude if *P* value from an n-degree of freedom test of plate association $<1\times10^{-5}$). This resulted in a final set of 514,952 autosomal and X-chromsomal SNPs typed in both case and control cohorts that were tested for association.

Following an initial association analysis (performed using the Cochran-Armitage trend test implemented in PLINK) using all CHD cases combined, we performed a separate analysis in each of five sub-phenotypes (comparing cases for each sub-phenotype to the same 5159 WTCCC2 controls). Given that the strongest signals ($P<1\times10^{-6}$) occurred in VSD and ASD, those SNPs passing significance level $<2\times10^{-5}$ in VSD and/or ASD (plus a few additional neighbouring SNPs that just failed to meet this threshold) were chosen to take forward for replication. No inflation of genome-wide test-statistics due to unmodelled population substructure was observed (genomic control factors λ=0.99 for VSD and λ=1.01 for ASD, Supplementary Figure 8) and so no correction on this account was made. Visual inspection of intensity cluster plots was performed for all SNPs to be taken forward for replication (Supplementary Figures 9-11), and only those SNPs for which the genotype calls appeared

reliable (well-clustered into three distinct groups) and which showed no evidence of departure from Hardy-Weinberg equilibrium were taken forward.

Replication Cohort—The replication cohort comprised an independently ascertained 417 secundum ASD cases and 209 VSD cases. Genotype data at those SNPs chosen for replication was compared to genotype data obtained from the TwinsUK resource, an adult twin registry comprising 12,000 (predominantly female) British twins. Genotype data for 3512 twin individuals (genotyped using the Illumina 610K array) were obtained from the Department of Twin Research (DTR) 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. 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 consideration of 576610 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 CHD cohort, and excluded first twins that did not cluster with the CEU HapMap samples and one of each pair of first twins that showed high IBD sharing (mean proportion of alleles IBD >0.05). We also used PLINK to perform multidimensional scaling of the TwinsUK samples together with the discovery cases and controls, and excluded those twins that did not cluster sufficiently closely with the discovery cases and controls. This resulted in a final set of 2520 TwinsUK controls to be used in the replication study. Multidimensional scaling plots for all discovery samples (cases and controls) and replication controls that were included in the final analyses (calculated after the exclusion of any outlying individuals) are shown in Supplementary Figure 12.

Association in the replication cohort was assessed initially using the Cochran-Armitage trend test implemented in PLINK, and subsequently (for all SNPs taken forward for replication) via logistic regression analysis in PLINK. To combine the discovery and replication results we performed a standard fixed-effects meta analysis on the basis of the estimated log odds ratios and their standard errors, implemented via the '--meta-analysis' command in PLINK.

We used the program IMPUTE version 2 [22] to carry out imputation in the discovery cohort across the 4p16 region, using the "-pgs" option to replace genotyped SNPs with their imputed values. Data from the 1000 Genomes Project [23] (Phase I version 3 integrated data, released March 2012) was used as a reference panel, with 392 SNPs that had been genotyped in both cases and controls in the 2MB region around rs870142 used to inform the imputation. Post-imputation QC involved excluding any SNPs likely to be poorly imputed (specifically those with an 'info' score <0.5). Data at 8405 SNPs passing post-imputation QC (from an original 36461 imputed SNPs) were analysed via a Frequentist allelic association test in the program SNPTEST version 2.1.1 [24] using the '-method threshold' option.

Expression studies of transcripts in associated region

In situ hybridisation to demonstrate expression of *Msx1* in developing mouse heart was performed as previously described [25]. A 500bp EcoNI-SphI fragment of the 3′UTR of mouse *Msx1* (NM_010835; bp1161-1697) was used as a template for the *Msx1* antisense probe. Sections were photographed on a Zeiss Axiophot microscope.

Segments of genomic sequence spanning each of the three SNPs showing association with ASD in 4p16 were used as queries for a BLAST search of human expressed sequence tags

(ESTs). An unspliced EST (GenBank accession number *BI192733.1*) derived from an epithelioid carcinoma cell line was found to span rs682495. The corresponding full-length EST clone was obtained (Source Bioscience, Nottingham, UK) and the insert sequenced completely, showing no significant open reading frame. A poly-A tail was present in the transcript despite the absence of a consensus polyadenylation signal. The poly-A tail of this transcript was shown to be present in the genomic sequence. An RT-PCR assay was performed to assess expression of the transcript in the developing human heart. 1 μg of Human Fetal Heart Total RNA, which was extracted from pooled heart tissue derived from fetuses between 9 and 20 weeks (Clontech, Mountain View, CA, catalogue no. 636583), and 1μg of Human Testis Total RNA (Clontech, catalogue no. 636533) were reverse transcribed using 400ng random hexamers (Thermo Scientific, Thermo Fisher Scientific Inc, USA) and 200U M-MuLV reverse transcriptase (New England Biolabs, UK), in a final volume of 25 μL. DEPC water was used for reverse transcriptase negative controls. Synthesised cDNA was used as the template for rtPCR, which was carried out in a reaction volume of 25μl containing 1μl cDNA, 21μl Megamix (Microzone Ltd, UK) and 10μM of each primer (Supplementary Table 4). Thermocycling consisted of an initial step at 96°C for 5 minutes then 34 cycles consisting of denaturing at 96°C for 45 seconds, annealing at 58.3°C for 45 seconds and extension at 72°C for 1 minute with a final extension at 72°C for 5 minutes. The T-box transcription factor gene *TBX5* was used as an internal control.

Human left ventricular samples were obtained from 181 non-diseased hearts of unrelated organ donors of European descent whose hearts were either explanted to obtain pulmonary and aortic valves for transplant surgery, or intended for transplantation but not used for logistical reasons. The tissues were collected at the University of Szeged, Hungary (N=65), Vanderbilt University, USA (N=54), the University of Miami, USA (N=37) and the University of Sydney, Australia (N=25). Procurement and handling of the material was approved by the ethical review boards of each centre. RNA and DNA were isolated using standard protocols. Preparation of cRNA (TotalPrep-96 RNA Amplification Kit) and chip hybridisation (Illumina HumanHT-12 v4) for genome-wide expression analyses were performed at ServiceXS (Leiden, The Netherlands), according to the manufacturer's instructions. Probes containing common SNPs (HapMap Phase III release 2) and probes whose sequence did not align unambiguously to the human reference genome (HG19) were excluded. The raw data was imported in R version 2.15.1 using the *beadarray* package [26]. QC was performed using the *ArrayQualityMetrics* package [27]. The data were normalised using the *neqc* algorithm [28]. SNP genotyping was carried out using Illumina HumanOmniExpress BeadChips at the Genome Analysis Center, Helmholz Center, Munich, Germany. QC of genotypic data was performed in the GenABEL package [29]. PCA identified several samples showing population stratification which were removed. Imputation was performed using MACH [30] and the HapMap Phase III release 2 data. Only SNPs imputed with high confidence were retained. After pre-processing and QC, a total of 129 samples remained for eQTL analysis. *LOC100507266* transcript levels were tested for association with genotypes at rs870142, rs6824295 and rs16835979 using linear models, with age, gender and recruitment centre as covariates. P-values were calculated according to an additive genetic model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The principal acknowledgement is to the congenital heart disease patients, their families, and healthy individuals who participated in the research. We acknowledge the contributions of C. de Gier-de Vries, L.M. Sneddon, M.

Westerveld, C.Q. Simmons, T. Zhang, R. Hussein, and V. Otero. This study was funded by the Wellcome Trust (Grants BH100708 and 087436), the British Heart Foundation (BHF), the European Union FP7 Programme "CHeartED" (HEALTH-F2-2008-223040), the SickKids Labatt Family Heart Centre Biobank (to SM), the Netherlands Heart Foundation (NHS2010B175 to TTK and CRB), Heart Research UK, National Institutes of Health (HL068880 to ALG), and the ANR Labex project "Medical Genomics" (to GML). MF acknowledges support from the Wellcome Trust core award (090532/Z/09/Z) and BHF Centre of Research Excellence in Oxford. SB and BK hold BHF Personal Chairs. This study makes use of data generated by the Wellcome Trust Case-Control Consortium 2, funded by the Wellcome Trust under award 085475. Access to genotype data from the TwinsUK cohort was kindly provided by the Department of Twin Research (DTR) and Genetic Epidemiology at King's College London. TwinsUK is funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013), and also receives support from the UK Department of Health via the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. TwinsUK SNP genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR.

REFERENCES

1. Tennant PW, Pearce MS, Bythell M, Rankin J. 20-year survival of children born with congenital anomalies: a population-based study. Lancet. 2010; 375:649–56. [PubMed: 20092884]
2. Soemedi R, et al. Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. American journal of human genetics. 2012; 91:489–501. [PubMed: 22939634]
3. Pruim RJ, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010; 26:2336–7. [PubMed: 20634204]
4. Webb G, Gatzoulis MA. Atrial septal defects in the adult: recent progress and overview. Circulation. 2006; 114:1645–53. [PubMed: 17030704]
5. Campbell M. Natural history of atrial septal defect. British heart journal. 1970; 32:820–6. [PubMed: 5212356]
6. Iinuma T, et al. Role of syntaxin 18 in the organization of endoplasmic reticulum subdomains. Journal of cell science. 2009; 122:1680–90. [PubMed: 19401338]
7. Boogerd KJ, et al. Msx1 and Msx2 are functional interacting partners of T-box factors in the regulation of Connexin43. Cardiovascular research. 2008; 78:485–93. [PubMed: 18285513]
8. Li QY, et al. Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. Nature genetics. 1997; 15:21–9. [PubMed: 8988164]
9. Chen YH, Ishii M, Sucov HM, Maxson RE Jr. Msx1 and Msx2 are required for endothelial-mesenchymal transformation of the atrioventricular cushions and patterning of the atrioventricular myocardium. BMC developmental biology. 2008; 8:75. [PubMed: 18667074]
10. Ishii M, et al. Combined deficiencies of Msx1 and Msx2 cause impaired patterning and survival of the cranial neural crest. Development. 2005; 132:4937–50. [PubMed: 16221730]
11. Jumlongras D, et al. A nonsense mutation in MSX1 causes Witkop syndrome. American journal of human genetics. 2001; 69:67–74. [PubMed: 11369996]
12. van den Boogaard MJ, Dorland M, Beemer FA, van Amstel HK. MSX1 mutation is associated with orofacial clefting and tooth agenesis in humans. Nature genetics. 2000; 24:342–3. [PubMed: 10742093]
13. Vastardis H, Karimbux N, Guthua SW, Seidman JG, Seidman CE. A human MSX1 homeodomain missense mutation causes selective tooth agenesis. Nature genetics. 1996; 13:417–21. [PubMed: 8696335]
14. Goodship JA, et al. A common variant in the PTPN11 gene contributes to the risk of tetralogy of Fallot. Circulation. Cardiovascular genetics. 2012; 5:287–92. [PubMed: 22503907]
15. Greenway SC, et al. De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. Nature genetics. 2009; 41:931–5. [PubMed: 19597493]
16. Soemedi R, et al. Phenotype-specific effect of chromosome 1q21.1 rearrangements and GJA5 duplications in 2436 congenital heart disease patients and 6760 controls. Human molecular genetics. 2012; 21:1513–20. [PubMed: 22199024]
17. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81:559–75. [PubMed: 17701901]

18. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]
19. Barrett JC, et al. Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. Nature genetics. 2009; 41:1330–4. [PubMed: 19915572]
20. Mells GF, et al. Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. Nature genetics. 2011; 43:329–32. [PubMed: 21399635]
21. Price AL, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics. 2006; 38:904–9. [PubMed: 16862161]
22. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nature genetics. 2007; 39:906–13. [PubMed: 17572673]
23. Abecasis GR, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–73. [PubMed: 20981092]
24. Wellcome Trust Case-Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–78. [PubMed: 17554300]
25. Moorman AF, Houweling AC, de Boer PA, Christoffels VM. Sensitive nonradioactive detection of mRNA in tissue sections: novel application of the whole-mount in situ hybridization protocol. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society. 2001; 49:1–8. [PubMed: 11118473]
26. Dunning MJ, Smith ML, Ritchie ME, Tavare S. beadarray: R classes and methods for Illumina bead-based data. Bioinformatics. 2007; 23:2183–4. [PubMed: 17586828]
27. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. Bioinformatics. 2009; 25:415–6. [PubMed: 19106121]
28. Shi W, Oshlack A, Smyth GK. Optimizing the noise versus bias trade-off for Illumina whole genome expression BeadChips. Nucleic acids research. 2010; 38:e204. [PubMed: 20929874]
29. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23:1294–6. [PubMed: 17384015]
30. Scott LJ, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science. 2007; 316:1341–5. [PubMed: 17463248]

Europe PMC Funders Author Manuscripts

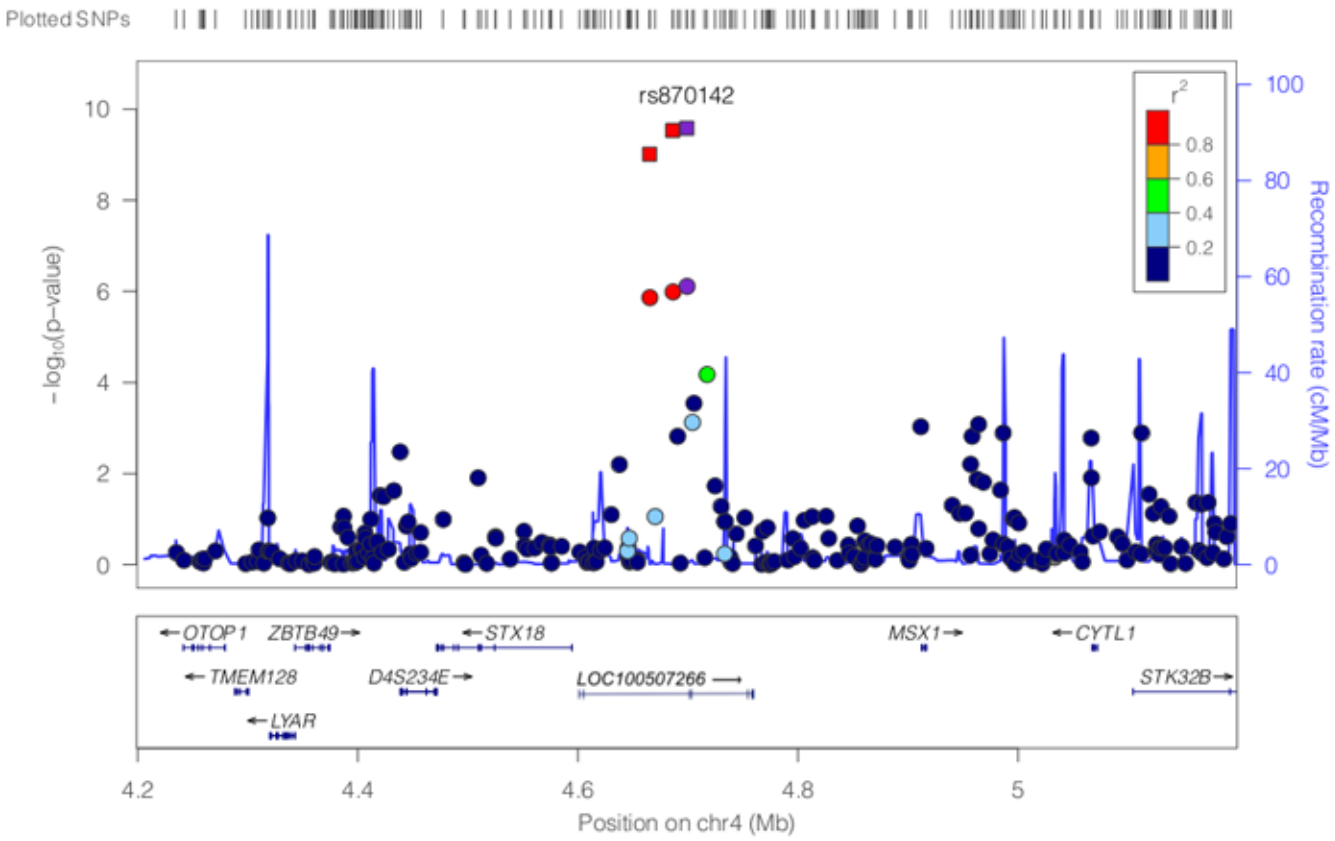

Figure 1.

LocusZoom plot of region of association with ASD at chromosome 4p16. Genes and ESTs within the region are shown in the lower panel, and the unbroken blue line indicates recombination rate within the region. Each filled circle represents the P-value for one SNP in the discovery cohort, with the top SNP rs870142 shown in purple and SNPs in the region colour-coded depending upon their degree of correlation (r^2) with rs870142 (as estimated internally by LocusZoom on the basis of CEU HapMap haplotypes). The P-values for the three SNPs in this region when analysed in the combined discovery and replication cohorts are shown as filled squares.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Top replicating SNPs for ASD in GWAS and replication cohorts

Locus					Discovery (GWAS) results (340 ASD cases, 5159 controls)				Replication results (417 ASD cases, 2520 controls)				Combined results (combined via fixed effects meta-analysis)		
CHR	SNP	BP	Minor allele	Major allele	MAF in cases	MAF in controls	OR	P	MAF in cases	MAF in controls	OR	P	P	OR	Heterogeneity P (Cochran's Q)
4	rs6824295	4665181	A	G	0.312	0.230	1.505	1.66E-06	0.314	0.250	1.376	0.00011	9.73E-10	1.437	0.4501
4	rs16835979	4686177	A	C	0.312	0.229	1.511	1.24E-06	0.315	0.248	1.399	4.47E-05	2.94E-10	1.452	0.5155
4	rs870142	4698948	A	G	0.312	0.228	1.519	9.52E-07	0.312	0.246	1.399	4.99E-05	2.61E-10	1.456	0.4890