BMC Genetics

RESEARCH ARTICLE

Open Access

Genome-wide linkage analysis of congenital heart defects using MOD score analysis identifies two novel loci

Antònia Flaquer[1,2*], Clemens Baumbach[1,2], Estefania Piñero[3], Fernando García Algas[4], María Angeles de la Fuente Sanchez[4], Jordi Rosell[5], Jorge Toquero[6], Luis Alonso-Pulpon[6], Pablo Garcia-Pavia[6], Konstantin Strauch[1,2] and Damian Heine-Suñer[5]

Abstract

Background: Congenital heart defects (CHD) is the most common cause of death from a congenital structure abnormality in newborns and is often associated with fetal loss. There are many types of CHD. Human genetic studies have identified genes that are responsible for the inheritance of a particular type of CHD and for some types of CHD previously thought to be sporadic. However, occasionally different members of the same family might have anatomically distinct defects — for instance, one member with atrial septal defect, one with tetralogy of Fallot, and one with ventricular septal defect. Our objective is to identify susceptibility loci for CHD in families affected by distinct defects. The occurrence of these apparently discordant clinical phenotypes within one family might hint at a genetic framework common to most types of CHD.

Results: We performed a genome-wide linkage analysis using MOD score analysis in families with diverse CHD. Significant linkage was obtained in two regions, at chromosome 15 (15q26.3, $P_{empirical} = 0.0004$) and at chromosome 18 (18q21.2, $P_{empirical} = 0.0005$).

Conclusions: In these two novel regions four candidate genes are located: *SELS, SNRPA1,* and *PCSK6* on 15q26.3, and *TCF4* on 18q21.2. The new loci reported here have not previously been described in connection with CHD. Although further studies in other cohorts are needed to confirm these findings, the results presented here together with recent insight into how the heart normally develops will improve the understanding of CHD.

Keywords: Congenital heart defects, Genetics, Linkage analysis, Genome-wide study

Background

Congenital heart defects (CHD) refer to abnormalities in the heart structure or function that arise at the fetal stages and affect approximately 1% of newborns [1]. Multiple surgeries are almost always required to correct many of the anatomical defects, and quality of life is often greatly compromised. There are many types of CHD. Examples include transposition of the great arteries/vessels (TGA/TGV), tetralogy of Fallot (TOF), double outlet right ventricle (DORV), atrial septal defects (ASD), ventricular septal defects (VSD), bicuspid aortic valve (BAV), and Ebstein's anomaly, among many others.

Normal heart development involves many regulatory pathways including receptor-ligand interactions (JAGGED/NOTCH, TGFB-BMP/TGFBR, VEGF/FLT1-FLK1, NODAL/ACVRA-ACVRB and RTK/RAS), signal transduction (kinases or phosphatases such as MAPK, ERK1/2, calcineurin, or GSK), and transcription factors that determine the expression of cardio-specific genes (protein families with a T-box domain *TBX1, TBX5, and TBX20*, the GATA family *GATA4* and *FOG2* or homeobox domain *NKX2.5* and *NKX2.6*). Mutations that show a greater penetrance and that therefore approximate a

---

* Correspondence: antonia.flaquer@lmu.de
[1]Institute of Medical Informatics, Biometry, and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
[2]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
Full list of author information is available at the end of the article

monogenic inheritance are those affecting transcription factors or genes transcribed by them [2].

Although the major underlying defects that cause CHD are thought to be mutations in regulators of heart development during embryogenesis, epidemiological data also indicate an environmental influence [2-4]. However, these epidemiological studies mostly suggest risk factors rather than underlying disease mechanisms. Genetic factors for some of the CHD include Mendelian mutations, copy number variants, translocations, and single nucleotide polymorphisms (SNPs) [5-7].

A genetic component of CHD diseases was initially implicated by recurrence in families, and by studies showing a co-segregation of CHD with the deletion 22q11.2 [8]. It has also been shown that individuals with an affected parent are at twofold greater risk, which is even greater if siblings are affected [9]. While many studies have investigated the role of genes in the etiology of inherited and sporadic CHD [10-14] these studies have been focusing on specific cardiac lesions separately. Recently, rare variants have also been reported to be associated with CHD [15]. These rare variants are usually present in less than 1% of the normal population but are overrepresented in selected patients. They are typically inherited from an asymptomatic parent and are therefore not believed to be a sufficient cause of CHD in these patients; other mutations need to be present to lead to CHD.

However, occasionally, different members of one family are observed to suffer from anatomically distinct defects — for example, one member with ASD, one with TOF, and one with VSD. These apparently discordant clinical phenotypes arising within one family are difficult to rationalize. Benson *et al.* [16] observed a coinheritance of different *NKX2.5* mutations, i.e., compound heterozygotes, in members of families with ASD, VSD, and cardiac conduction abnormalities, suggesting that mutations in the same gene can affect different parts of the heart, and therefore cause different types of CHD. McGregor *et al.* [17] reported a statistically significant linkage evidence for a locus on chromosome 14 (within the *HOMEZ* gene) in South Indian cases born to consanguineous parents. However, the linkage finding was not robust in a genetic association follow-up study in a general United States population. Despite the successes over the past few years, much remains unknown about the genetics of CHD. It is believed that much of the missing heritability is likely to be due to rare variants with larger effect sizes as well as epigenetic effects. Their systematic characterization is beyond the means of studies that currently rely on linkage disequilibrium (LD) patterns such as genome-wide association analysis. In such situations taking family information into consideration using a direct mapping approach complemented by segregation techniques will be a useful s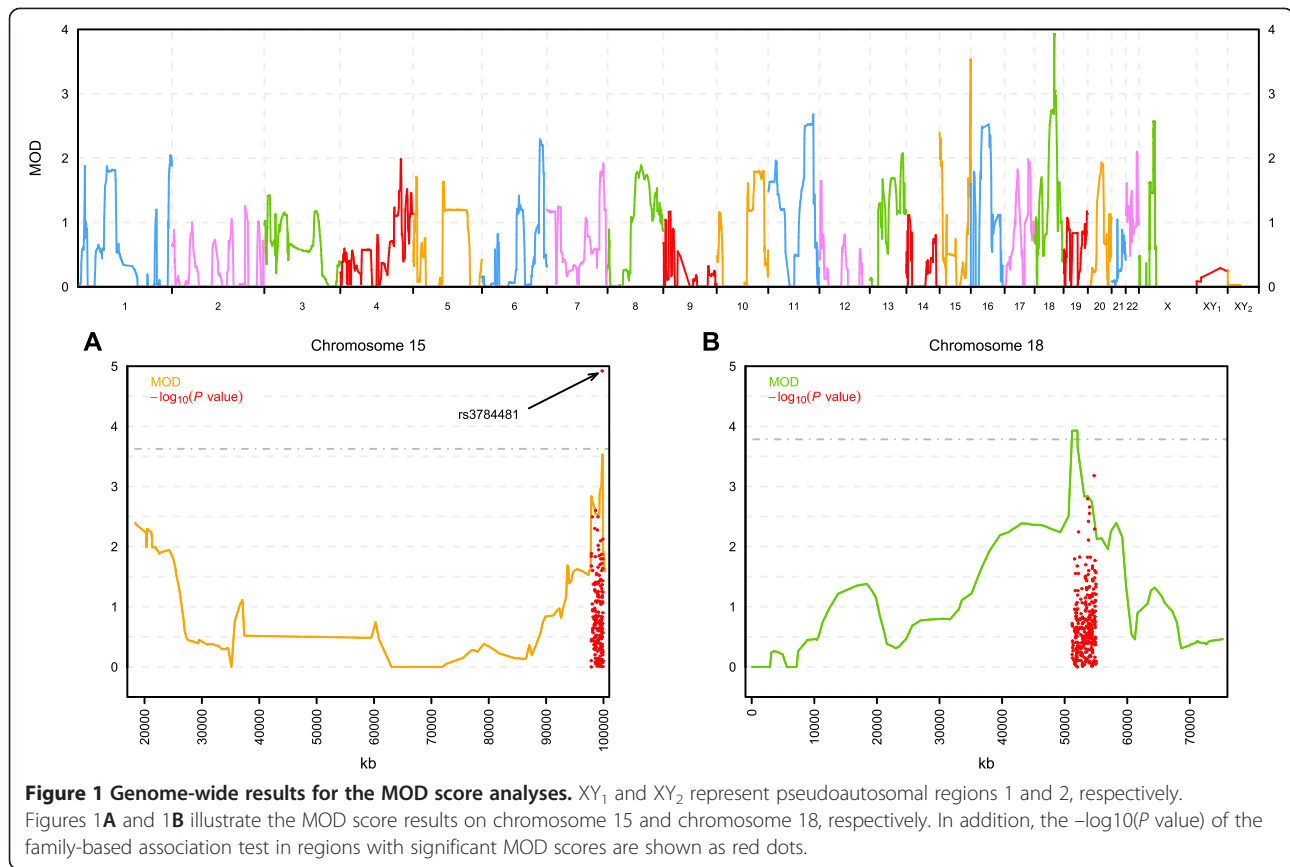trategy. For complex diseases linkage analysis is a powerful method for detecting gene effects [18,19]. Especially when looking at rare variants, linkage has the advantage over association of not being prone to allelic heterogeneity. That is, the combination of many weak association signals obtained in a certain region for single variants by means of collapsing methods is automatically performed in the context of linkage analysis.

In this study we performed a detailed genome-wide linkage analysis using MOD scores and 6 families affected by different CHD to test for a common genetic background among different types of heart defects.

Results and discussion

The genome-wide results for MOD scores are plotted in Figure 1. No significant results were obtained when modeling parent-specific heterozygote penetrances, resulting in no evidence of parental genomic imprinting in our set of pedigrees. The best results were obtained on chromosome 15 and 18. The results for these two chromosomes are plotted in Figures 1A and 1B, respectively. The highest scores (MOD = 3.9) were detected on chromosome 18 (18q21.2-18q21.3; 51197–52023 kb) comprising 57 SNPs with $P_{empirical} = (0.0005\text{-}0.0006)$, where $P_{empirical}$ refers to the P value obtained after simulations were performed. The associated parametric model is a recessive mode of inheritance with incomplete penetrance with a probability of 0.27 to be affected for individuals carrying two copies of the disease allele. Incomplete penetrance refers to the failure of individuals who carry the high risk genotype to exhibit the trait. This phenomenon is often observed in complex diseases. Besides different ways of gene action or degrees of expression, incomplete penetrance can arise from further genetic or environmental factors. The second significant region was at chromosome 15 (15q26.3; 99749–99883 kb) comprising 27 SNPs, all of them with a MOD score of 3.5 and $P_{empirical} = (0.0004\text{-}0.0005)$. The best-fitting model points to a recessive mode of inheritance, with a probability of 0.98 to be affected for individuals carrying two copies of the disease allele. It should be noted that the estimate of the disease allele frequency obtained by a MOD-score analysis has the largest variance of all trait-model parameters. Furthermore, in some cases, the estimated disease allele frequency will be markedly higher than the true value. This is due to the fact that specifying a higher disease allele frequency can compensate for a general model misspecification and hence lead to robustness in a multipoint analysis [20].

To corroborate these findings, family-based association analysis was performed in these two significant regions with SNPs displaying a MOD >2.5 (P value ≤ 0.01). Using this criterion a total number of 304 SNPs (P values between 5×10^{-04} and 1×10^{-02}) and 211 SNPs (P values between 4×10^{-04} and 4×10^{-03}) were included in the association analysis for 18q21.2-18q21.3 and 15q26.3,

Figure 1 Genome-wide results for the MOD score analyses. XY_1 and XY_2 represent pseudoautosomal regions 1 and 2, respectively. Figures 1**A** and 1**B** illustrate the MOD score results on chromosome 15 and chromosome 18, respectively. In addition, the $-\log10(P$ value) of the family-based association test in regions with significant MOD scores are shown as red dots.

respectively. A total number of 34 SNPs resulted to be significant at a type I error rate of 0.05 on chromosome 18. Nevertheless, none of them was significant after using the Bonferroni correction to adjust for multiple testing. On the other hand, on chromosome 15 a total number of 39 SNPs were found to be nominally significant and only one (rs3784481 at 99749 kb) was significant after Bonferroni correction (P value = 0.0025).

To see if one of the two linkage regions may be the causal genetic factor for a particular family we performed parametric linkage (LOD score) analysis of individual pedigrees under the best fitted model for the two linked regions. As shown in Table 1, all the pedigrees contribute positively to the LOD score in both regions (except pedigree 5, chr15). These positive contributions indicate that all types of CHD present in our pedigrees are linked to both regions, supporting the initial hypothesis of this study of a common genetic background for diverse CHD. The negative contribution of pedigree 5 to the LOD score on chromosome 15 has to be treated carefully. Due to the small number of meioses in this pedigree there is a lack of power to detect linkage, and no conclusion should be drawn when analyzing this pedigree individually.

When looking in detail at the region around the significant SNPs on chromosome 15, three genes are found

very closely to each other (*SELS*, *SNRPA1*, and *PCSK6*). In addition, one gene is located in the linkage region identified on chromosome 18 (*TCF4*).

One gene at 18q21.2: TCF4

The potentially relevant *TCF4* gene is located at the linkage peak of chromosome 18. This gene encodes transcription factor 4, a basic helix-loop-helix transcription factor. To date, no connection between this gene and CHD has been established. However, it is interesting to note that fairly recently it has been found that in mice the *tcf4* gene is strongly expressed in connective tissue fibroblast. It was demonstrated that connective tissue

Table 1 Parametric LOD scores for each family separately

Family	Heart defect	Chr15	Chr18
		15q26.3	18q21.2-18q21.3
1	VSD	0.615	0.632
2	TOF,TGV	0.324	0.299
3	TOF	0.545	0.602
4	ASD	1.670	1.297
5	VSD,DORV	−0.307	0.602
6	ASD	0.681	0.496
All	**VSD,TOF,TGV,ASD,DORV**	**3.529**	**3.929**

fibroblasts and *tcf4* critically regulate two aspects of myogenesis: muscle fiber type development and maturation [21]. Taken together this finding in mice and our results could suggest that malfunction of the *TCF4* gene may produce a malformation of the heart muscle tissue which in turn predisposes to a higher risk to develop some type of CHD.

Three genes at 15q26.3: SELS, SNRPA1, and PCSK6

The *SELS* gene encodes a selenoprotein, which contains a selenocysteine residue at its active site. Studies suggest that this protein may regulate cytokine production, and thus play a key role in the control of the inflammatory response decreasing pro-inflammatory cytokines [22]. Production of cytokines is one of the major mechanisms employed by T-cells to coordinate immune responses. It has been demonstrated that the developing immune system of the neonate not only differs significantly from that of an adult, but also varies based on gestational age [23]. A malfunction of the *SELS* gene might provoke deficient cytokine production by neonatal lymphocytes, and somehow predispose the newborn to a higher risk to develop some CHD. It was also shown that *SELS* is secreted from hepatoma cells and that it associates with low-density lipoprotein particles [24] suggesting also a role in lipid metabolism. Lundell *et al.* [25] had already shown in their study that lipid metabolism was disturbed in infants with CHD. Furthermore, genetic studies in relation to cardiovascular disease have revealed that variation in the *SELS* locus is associated with coronary heart disease and ischemic stroke in women [26].

The *SNRPA1* gene encodes the U2 small ribonucleoprotein A' (U2 snRNP A'). The action of snRNPs is essential to the removal of introns from pre-mRNA, a critical aspect of post-transcriptional modification of RNA. To date no association has been reported between *SNRPA1* and heart defects. However, post-transcriptional modification is one of the multiple mechanisms that regulate the level and activity of transcription factors which are responsible for regulating formation and function of the heart. Perturbations of transcription factor expression and regulation disrupt normal heart structure and function [27].

The protein encoded by the *PCSK6* gene belongs to the subtilisin-like proproteinconvertase family. The members of this family are proproteinconvertases that process latent precursor proteins into their biologically active products. This encoded protein is a calcium-dependent serine endoprotease that can cleave precursor protein at their paired basic amino acid processing sites. Some of its substrates are transforming growth factor beta related proteins, proalbumin, and von Willebrand factor. In 1998, a study reported abnormalities of von Willebrand factor in adults with CHD [28]. Interestingly, a protein from the same family, proproteinconvertasesubtilisin-like/kexin type 9, which is encoded by the *PCSK9* gene, is a key regulator of plasma low density lipoprotein cholesterol and has emerged as a promising target for prevention and treatment of coronary heart diseases. A recent genome-wide association study further bolstered the importance of *PCSK9* by establishing a link between a nearby SNP and early onset myocardial infarction [29].

Conclusions

We have conducted a whole genome linkage analysis for CHD using MOD score analysis. We used a total number of 177,568 SNPs providing a good coverage of the genome in a sample of 6 families of Spanish origin with members affected with anatomically distinct CHD: ASD, VSD, TOF, TGV, and DORV. The aim of this study was the identification of chromosomal regions that are implicated in a possible common genetic framework shared by distinct types of CHD. Two interesting linkage regions were found, on chromosome 15 and 18. The highest linkage peak was obtained at position 18q21.2 (LOD = 3.9, $P_{\text{empirical}}$ = 0.0005) and the second highest peak at 15q26.3 (LOD = 3.5, $P_{\text{empirical}}$ = 0.0004) using MOD-score analysis. In addition, at 15q26.3 a significant association was obtained for a SNP (rs3784481, P value = 0.0025) using family-based association analysis. Imprinting analysis of the data in the present study did not confirm the hypothesis of a possibly imprinted locus. When looking at these two linkage regions in detail, four genes are located there, three very close to each other at 15q26.3 and one at 18q21.2.

As in all research work, our results have some unavoidable limitations. They have been obtained using linkage techniques with a small sample size. It is well known that large, multi-generation pedigrees are the most informative for linkage analysis. However, such pedigrees are difficult to find when investigating complex diseases with a high and early mortality such as CHD. That might explain why to date there have been so few genome-wide linkage scans involving affected families with CHD. Therefore, despite the fact that our study comprising 6 families is small compared with linkage studies for other disorders, it is, thus far, the largest family study performed in CHD. Moreover, our findings on chromosome 15 are corroborated by family-based association analysis (with 7 additional trios). Clearly, recruiting more families with CHD will help to verify this linkage evidence as well as to identify new linkage regions.

We report in this study two genetic regions linked to CHD predisposition. There are three candidate genes located in the 15q26.3 region very close to each other (*SELS*, *SNRPA1* and *PCSK6*) and one gene at 18q21.2 (*TCF4)*. Only one of these genes, *SELS*, was previously

found to be implicated in cardiovascular disorders. The other three genes reported in this study may be considered as novel candidate loci for CHD.

While several genome scans have reported susceptibility genes for specific types of CHD, very few (only one [17], and the one presented here) have identified susceptibility genes that predispose to developing anatomically distinct types of CHD and should form the basis of new analyses of both linkage and association studies on CHD. Future linkage studies may benefit in power from larger sample size, and also case–control studies will benefit from direction or corroboration given by linkage scans. For example, the extensive multiple testing problem in genome-wide association studies can be partially alleviated by a weighted Bonferroni correction or Bayesian analysis that employs linkage data [30]. The results presented here should facilitate the discovery of both common and rare variants that confer genetic susceptibility for CHD and help the understanding of heart mechanisms. Understanding the etiology of CHD will also help to identify possible complications and risk factors for surgery or treatment, as patients (babies and children) with genetic syndromes or extracardiac anomalies are generally at higher risk of operative mortality and morbidity.

Methods

Subjects

A total number of 6 families comprising 48 individuals (19 females, 29 males) and 7 trios (an affected proband and both parents) of Spanish origin were recruited at the University Hospital Son Espases (HUSE) of the Balearic Islands and at the Puerta del Hierro University Hospital in Madrid. Pedigree structures are illustrated in Figure 2. The inclusion criterion for all patients was a CHD present at birth. Patients and their families were asked to participate in the study within the clinical practice and blood samples were taken. All members included in the study were checked for the presence of deletions or duplications using the same SNP-array analysis. Individuals that were found to carry a suggestive or clearly causative deletion or duplication (22q11.2 or other) were excluded from the analysis. All participants provided their written informed consent to participate in the study. A statement of informed consent form was signed by the parents of all participating children under 18 years old, following protocols approved by the ethical committee for clinical research of the Balearic Islands (IB 963/08; Comité ético de investigación clínica de las Islas Baleares; CEIC). The study conformed to all ethical guidelines of the institutions
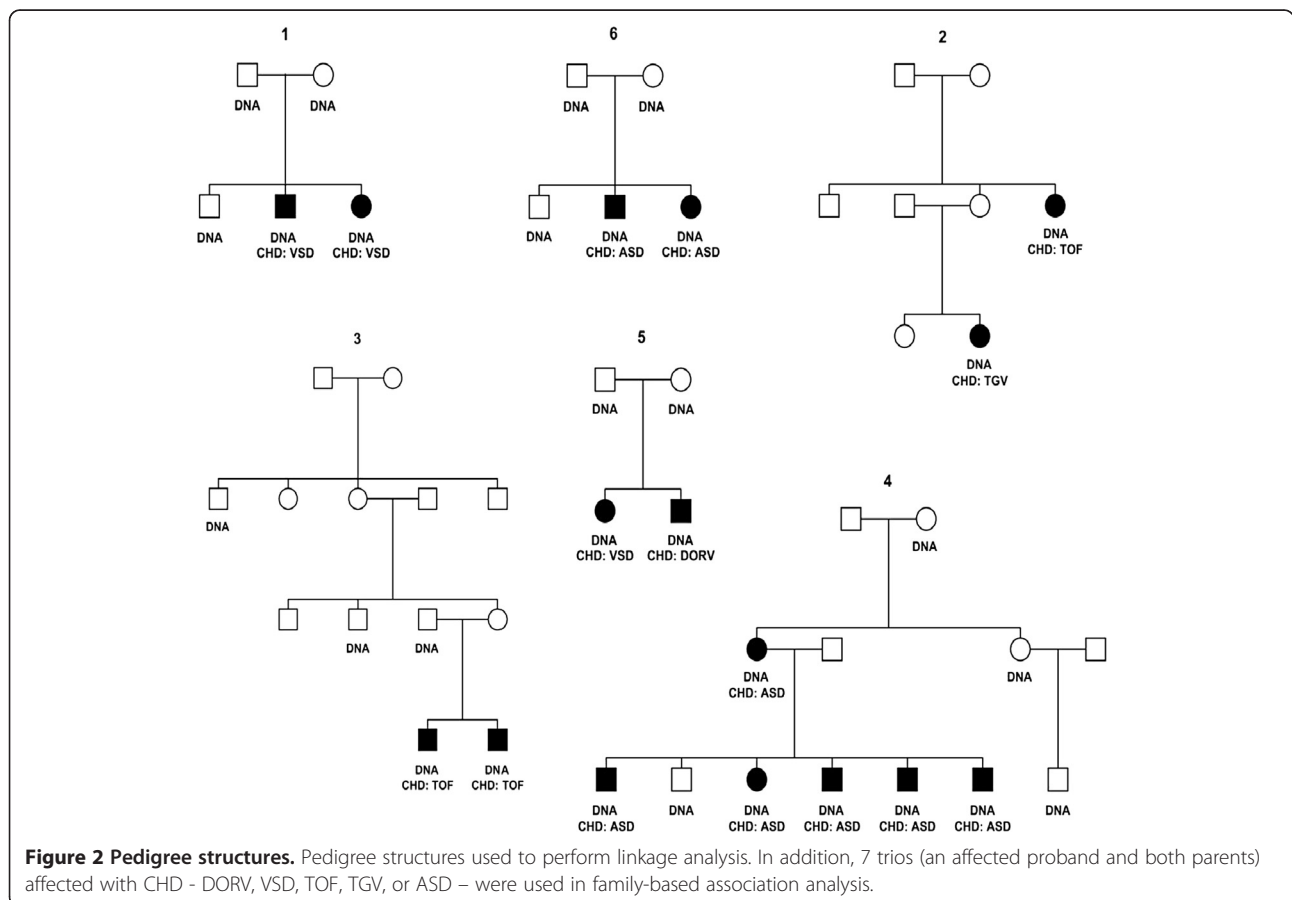

Figure 2 Pedigree structures. Pedigree structures used to perform linkage analysis. In addition, 7 trios (an affected proband and both parents) affected with CHD - DORV, VSD, TOF, TGV, or ASD – were used in family-based association analysis.

involved. Research at the HUSE and Puerta del Hierro follow the guidelines included in the June 1964 Declaration of Helsinki (entitled "Ethical Principles for Medical Research Involving Human Subjects"). All sample identifiers were re-encoded in a way that only the clinicians are able to access personal information.

DNA isolation and genotyping

All patients had a single tube of blood drawn at the time of enrollment, which was collected centrally for subsequent DNA extraction. Genotyping reactions of 52 individuals were performed at the Spanish genotyping center (CeGen) facilities in Barcelona using the Illumina (California, USA) Human660W-Quad chip, following the manufacturer's instructions.

Data quality control

Because genotyping errors can lead to incorrect inferences about gene flow in pedigrees and may reduce the effectiveness of linkage analysis, strict quality controls (QC) were performed in our data. First of all, the software Graphical Representation of Relationship Errors (GRR) was used to identify errors in the structure of the pedigrees and to eliminate duplicate samples, monozygotic twins, and unrelated subjects [31]. In one family GRR detected an inconsistency involving a mother and her two children. Eventually it turned out that the children's genotypes were mutually consistent and only the mother's genotypes were removed from the analysis. All individuals were also tested for sex consistency. Afterwards, the PEDCHECK software was used to detect marker typing incompatibilities [32]. A total number of 3271 (0.6%) incompatibilities were detected and those genotypes were removed from the data. Additional QC was undertaken to check for unlikely double recombinants. Two or more close recombination events on the same chromosome are uncommon due to the chance distribution of chiasmata, and because of interference. To detect such problematic genotypes the "error detection" option from the MERLIN v1.1.2 software was used and the unlikely genotypes were subsequently removed from the data using the "pedwipe" option [33]. Once Mendelian inconsistencies and unlikely genotypes were resolved, SNPs with call rates below 95%, monomorphic SNPs, and SNPs with a missing genetic position were also removed from the analysis. At the end, 93.95% of the original SNPs survived the QC. Allele frequencies were then estimated with MERLIN v1.1.2 using a maximum-likelihood procedure.

Data preparation for linkage analysis

In linkage, when analyzing SNPs in pedigrees where some parental genotypes are missing it is necessary to model for marker-marker LD prior to the analysis. Ignoring such LD

can result in severe biases in linkage results [34]. That is why pairwise LD between each pair of SNP loci was analyzed, using the procedure ldmax in the GOLD software [35]. ldmax calculates r^2 values from haplotype frequencies estimated via the expectation maximization (EM) algorithm of Excoffier and Slatkin [36]. In our data, whenever a pair of SNPs was in LD ($r^2 > 0.20$), one of the two SNPs was removed from the analysis. With the resulting data set a genome-wide linkage analysis was performed on a total number of 6 pedigrees with sizes varying from 4 to 13 members. After QC, a total number of 177,568 SNPs was analyzed, with 174,475 from the autosomes, 3078 from the X-chromosome, and 15 SNPs from the pseudoautosomal regions (PAR1 and PAR2).

Statistical methods

In parametric linkage analysis (LOD scores) the trait-model parameters, i.e., the penetrances and disease allele frequency, must be specified prior to the analysis. It is well known that the power to detect linkage decreases when the specified model is not sufficiently close to the true one. Consequently, we calculated multipoint MOD scores, i.e., LOD scores that were maximized over trait-model parameters using the program GENEHUNTER-MODSCORE (GH-MOD v3.0) [37]. This program calculates MOD scores automatically by varying the disease-allele frequency and three penetrances. We used the option "modcalc single" to perform a separate maximization for each genetic position assumed for the putative disease locus. This procedure yields a MOD score in conjunction with the penetrances and disease allele frequency of the best-fitting trait model for every genetic position. MOD score analysis represents a comprehensive way to analyze linkage data when the mode of inheritance is unknown; in general this procedure is particularly well suited for the genetic dissection of a complex trait [38].

It has also been observed that the CHD risk for children of mothers with a CHD is consistently higher than for those of fathers with a CHD [39,40]. One possible explanation for the increased transmission through females is a paternal imprinting mode of inheritance, i.e., offspring will be affected only if the inherited disease allele comes from the mother. Hence, we also checked for a parent-of-origin effect using two different penetrances for heterozygous individuals to take the parent who transmitted the disease allele into account (see "imprinting" option of GH-MOD v3.0).

A MOD-score analysis leads to an inflation of the type I error when compared to LOD scores calculated under a single parametric model. Therefore, in the context of MOD-score analysis, significance criteria for LOD scores cannot be applied without correction. To assess significance for a MOD score empirical *P* values need to be calculated. We used 5000 replicates, each one generated

with Monte-Carlo simulations of marker alleles under the assumption of no linkage and Hardy-Weinberg equilibrium, using the same pedigree structure, affection status, marker spacing, and allele frequencies as in the real data set. Replicates were analyzed like the original data. The empirical *P* value was obtained as the proportion of replicates showing a MOD score equal to or higher than the one observed in the real data set.

We computed MOD and MOD-imprinting for all autosomes. The MOD score has not been implemented so far for the X-chromosome. Therefore, we performed LOD-score analysis assuming different genetic models for the X-chromosome, ranging from completely dominant to completely recessive modes of inheritance. In addition, the PAR1 and PAR2 were also investigated. Genetic markers in the PARs follow the same inheritance pattern as autosomal markers, becoming progressively more sex-linked (i.e., recombination events are becoming rarer between X and Y) as they approach the pseudoautosomal boundary. When using multipoint linkage analysis in the PARs, it is crucial to use sex-specific maps since there exists a marked difference in recombination frequencies between males and females [41]. We used pseudoautosomal sex-specific genetic coordinates [41] implemented in the Rutgers map [42]. The sex-specific genetic positions were interpolated using a local regression with a quadratic fit as implemented by the "locfit" package [43] of the statistical software R. In order to obtain a mapping that is monotonous in the base pair positions we used R's "monoProc" package which monotonizes the local regression fit [44].

Finally, only in those regions where the most significant linkage signals were obtained, to corroborate and narrow the significant regions, family-based association analysis together with the trio families was performed using the LAMP software [45]. LAMP uses a maximum likelihood model to extract the exact information on genetic linkage and association from samples of unrelated individuals, sib pairs, trios, and larger pedigrees. *P* values obtained by the family-based association analysis were adjusted using Bonferroni correction (using the total number of SNPs in each specific region of linkage) to account for multiple testing. We used an overall type I error rate of 0.05.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
AF carried out all statistical analyses, interpretation of the results and wrote the manuscript. CB carried out all the simulation analyses, prepared the figures and edited the manuscript. EP carried out the SNP-array DNA preparation and partial analysis. FGA, MADLF, JT, LAP and PGP characterized clinically heart defects of patients. JR characterized clinical dysmorphology and evaluated inheritance. KS critically revised the manuscript and all statistical content. DHS provided samples and revised the manuscript critically for important molecular content. All authors gave final approval of the version to be published.

Author details
[1]Institute of Medical Informatics, Biometry, and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [2]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [3]Research Unit, Son Espases University Hospital, Palma de Mallorca, Spain. [4]Pediatric Cardiology, Son Espases University Hospital, Palma de Mallorca, Spain. [5]Genetic Section, Son Espases University Hospital, Palma de Mallorca, Spain. [6]Cardiomyopathy Unit, Department of Cardiology, Puerta de Hierro University Hospital, Madrid, Spain.

References
1. Hoffman JI, Kaplan S: **The incidence of congenital heart disease.** *J Am Coll Cardiol* 2002, **39**:1890–1900.
2. Wessels MW, Willems PJ: **Genetic factors in non-syndromic congenital heart malformations.** *Clin Genet* 2010, **78**:103–123.
3. Burn J, Brennan P, Little J, Holloway S, Coffey R, Somerville J, Dennis NR, Allan L, Arnold R, Deanfield JE, Godman M, Houston A, Keeton B, Oakley C, Scott O, Silove E, Wilkinson J, Pembrey M, Hunter AS: **Recurrence risks in offspring of adults with major heart defects: results from first cohort of British collaborative study.** *Lancet* 1998, **351**:311–316.
4. Oyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PK, Melbye M: **Recurrence of congenital heart defects in families.** *Circulation* 2009, **120**:295–301.
5. Greenway SC, Pereira AC, Lin JC, DePalma SR, Israel SJ, Mesquita SM, Ergul E, Conta JH, Korn JM, McCarroll SA, Gorham JM, Gabriel S, Altshuler DM, Quintanilla-Dieck Mde L, Artunduaga MA, Eavey RD, Plenge RM, Shadick NA, Weinblatt ME, De Jager PL, Hafler DA, Breitbart RE, Seidman JG, Seidman CE: **De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot.** *Nat Genet* 2009, **41**:931–935.
6. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D: **Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs.** *Nat Genet* 2008, **40**:1253–1260.
7. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliott AL, Parkin M, Hubbell E, Webster T, Mei R, Veitch J, Collins PJ, Handsaker R, Lincoln S, Nizzari M, Blume J, Jones KW, Rava R, Daly MJ, Gabriel SB, Altshuler D: **Integrated detection and population-genetic analysis of SNPs and copy number variation.** *Nat Genet* 2008, **40**:1166–1174.
8. Wilson DI, Cross IE, Wren C, Scambler PJ, Burn J, Goodship J: **Minimum prevalence of chromosome 22q11 deletions.** *Am J Hum Genet* 1994, **55**:A169.975.
9. Murabito JM, Pencina MJ, Nam BH, D'Agostino RB Sr, Wang TJ, Lloyd-Jones D, Wilson PW, O'Donnell CJ: **Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults.** *JAMA* 2005, **294**:3117–3123.
10. Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, Soults J, Grayzel D, Kroumpouzou E, Traill TA, Leblanc-Straceski J, Renault B, Kucherlapati R, Seidman JG, Seidman CE: **Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome.** *Nat Genet* 1997, **15**:30–35.
11. Li QY, Newbury-Ecob RA, Terrett JA, Wilson DI, Curtis AR, Yi CH, Gebuhr T, Bullen PJ, Robson SC, Strachan T, Bonnet D, Lyonnet S, Young ID, Raeburn JA, Buckler AJ, Law DJ, Brook JD: **Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family.** *Nat Genet* 1997, **15**:21–29.
12. Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen JC, Srivastava D: **GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5.** *Nature* 2003, **424**:443–447.

13. Schott JJ, Benson DW, Basson CT, Pease W, Silberbach GM, Moak JP, Maron BJ, Seidman CE, Seidman JG: Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* 1998, **281**:108–111.

14. Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, Grossfeld PD, Srivastava D: Mutations in NOTCH1 cause aortic valve disease. *Nature* 2005, **437**:270–274.

15. Roessler E, Ouspenskaia MV, Karkera JD, Velez JI, Kantipong A, Lacbawan F, Bowers P, Belmont JW, Towbin JA, Goldmuntz E, Feldman B, Muenke M: Reduced NODAL signaling strength via mutation of several pathway members including FOXH1 is linked to human heart defects and holoprosencephaly. *Am J Hum Genet* 2008, **83**:18–29.

16. Benson DW, Silberbach GM, Kavanaugh-McHugh A, Cottrill C, Zhang Y, Riggs S, Smalls O, Johnson MC, Watson MS, Seidman JG, Seidman CE, Plowden J, Kugler JD: Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. *J Clin Investig* 1999, **104**:1567–1573.

17. McGregor TL, Misri A, Bartlett J, Orabona G, Friedman RD, Sexton D, Maheshwari S, Morgan TM: Consanguinity mapping of congenital heart disease in a South Indian population. *PLoS One* 2010, **5**:e10286.

18. Ott J, Kamatani Y, Lathrop M: Family-based designs for genome-wide association studies. *Nat Rev Genet* 2011, **12**:465–474.

19. Zhu X, Feng T, Li Y, Lu Q, Elston RC: Detecting rare variants for complex traits using family and unrelated data. *Genetic epidemiology* 2010, **34**:171–187.

20. Risch N, Giuffra L: Model misspecification and multipoint linkage analysis. *Hum Hered* 1992, **42**:77–92.

21. Mathew SJ, Hansen JM, Merrell AJ, Murphy MM, Lawson JA, Hutcheson DA, Hansen MS, Angus-Hill M, Kardon G: Connective tissue fibroblasts and Tcf4 regulate myogenesis. *Development* 2011, **138**:371–384.

22. Fradejas N, Serrano-Perez Mdel C, Tranque P, Calvo S: Selenoprotein S expression in reactive astrocytes following brain injury. *Glia* 2011, **59**:959–972.

23. West LJ: Defining critical windows in the development of the human immune system. *Hum Exp Toxicol* 2002, **21**:499–505.

24. Gao Y, Pagnon J, Feng HC, Konstantopolous N, Jowett JB, Walder K, Collier GR: Secretion of the glucose-regulated selenoprotein SEPS1 from hepatoma cells. *Biochem Biophys Res Commun* 2007, **356**:636–641.

25. Lundell KH, Sabel KG, Eriksson BO: Plasma metabolites after a lipid load in infants with congenital heart disease. *Acta Paediatr* 1999, **88**:718–723.

26. Alanne M, Kristiansson K, Auro K, Silander K, Kuulasmaa K, Peltonen L, Salomaa V, Perola M: Variation in the selenoprotein S gene locus is associated with coronary heart disease and ischemic stroke in two independent Finnish cohorts. *Hum Genet* 2007, **122**:355–365.

27. Zhou P, He A, Pu WT: Regulation of GATA4 transcriptional activity in cardiovascular development and disease. *Curr Top Dev Biol* 2012, **100**:143–169.

28. Territo MC, Perloff JK, Rosove MH, Moake JL, Runge A: Acquired von Willebrand factor abnormalities in adults with congenital heart disease: Dependence upon cardiopulmonary pathophysiological subtype. *Clin Appl Thromb-Hem* 1998, **4**:257–261.

29. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, Anand S, Engert JC, Samani NJ, Schunkert H, Erdmann J, Reilly MP, Rader DJ, Morgan T, Spertus JA, Stoll M, Girelli D, McKeown PP, Patterson CC, Siscovick DS, O'Donnell CJ, Elosua R, Peltonen L, Salomaa V, Schwartz SM, Melander O, Altshuler D, Ardissino D, Merlini PA, Berzuini C, *et al*: Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 2009, **41**:334–341.

30. Roeder K, Bacanu SA, Wasserman L, Devlin B: Using linkage genome scans to improve power of association in genome scans. *Am J Hum Genet* 2006, **78**:243–252.

31. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: GRR: graphical representation of relationship errors. *Bioinformatics* 2001, **17**:742–743.

32. O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998, **63**:259–266.

33. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002, **30**:97–101.

34. Abecasis GR, Wigginton JE: Handling marker-marker linkage disequilibrium: pedigree analysis with clustered markers. *Am J Hum Genet* 2005, **77**:754–767.

35. Abecasis GR, Cookson WO: GOLD–graphical overview of linkage disequilibrium. *Bioinformatics* 2000, **16**:182–183.

36. Excoffier L, Slatkin M: Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995, **12**:921–927.

37. Mattheisen M, Dietter J, Knapp M, Baur MP, Strauch K: Inferential testing for linkage with GENEHUNTER-MODSCORE: the impact of the pedigree structure on the null distribution of multipoint MOD scores. *Genetic Epidemiol* 2008, **32**:73–83.

38. Strauch K, Furst R, Ruschendorf F, Windemuth C, Dietter J, Flaquer A, Baur MP, Wienker TF: Linkage analysis of alcohol dependence using MOD scores. *BMC Genet* 2005, **6**(Suppl 1):S162.

39. Gill HK, Splitt M, Sharland GK, Simpson JM: Patterns of recurrence of congenital heart disease: an analysis of 6,640 consecutive pregnancies evaluated by detailed fetal echocardiography. *J Am Coll Cardiol* 2003, **42**:923–929.

40. Romano-Zelekha O, Hirsh R, Blieden L, Green M, Shohat T: The risk for congenital heart defects in offspring of individuals with congenital heart defects. *Clin Genet* 2001, **59**:325–329.

41. Flaquer A, Fischer C, Wienker TF: A new sex-specific genetic map of the human pseudoautosomal regions (PAR1 and PAR2). *Hum Hered* 2009, **68**:192–200.

42. Matise TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, Hyland FC, Kennedy GC, Kong X, Murray SS, Ziegle JS, Stewart WC, Buyske S: A second-generation combined linkage physical map of the human genome. *Genome Res* 2007, **17**:1783–1786.

43. Loader C: *Local regression and likelihood.* New York: Springer; 1999.

44. Dette H, Neumeyer N, Pilz KF: A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* 2006, **12**:469–490.

45. Li M, Boehnke M, Abecasis GR: Efficient study designs for test of genetic association using sibship data and unrelated cases and controls. *Am J Hum Genet* 2006, **78**:778–792.