nature genetics

A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations

Zhibin Hu[1,2,14], Yongyong Shi[3,14], Xuming Mo[4,14], Jing Xu[5,14], Bijun Zhao[6,14], Yuan Lin[1,2,14], Shiwei Yang[7], Zhengfeng Xu[8], Juncheng Dai[2], Shandong Pan[2], Min Da[4], Xiaowei Wang[5], Bo Qian[4], Yang Wen[1,2], Juan Wen[1,2], Jinliang Xing[9], Xuejiang Guo[1,10], Yankai Xia[1,11,12], Hongxia Ma[2], Guangfu Jin[2], Shiqiang Yu[6], Jiayin Liu[13], Zuomin Zhou[1,10], Xinru Wang[1,11,12], Yijiang Chen[5], Jiahao Sha[1,10] & Hongbing Shen[1,2,12]

Congenital heart malformation (CHM) is the most common form of congenital human birth anomaly and is the leading cause of infant mortality. Although some causative genes have been identified, little progress has been made in identifying genes in which low-penetrance susceptibility variants occur in the majority of sporadic CHM cases. To identify common genetic variants associated with sporadic non-syndromic CHM in Han Chinese populations, we performed a multistage genome-wide association study (GWAS) in a total of 4,225 CHM cases and 5,112 non-CHM controls. The GWAS stage included 945 cases and 1,246 controls and was followed by 2-stage validation with 2,160 cases and 3,866 controls. The combined analyses identified significant associations ($P < 5.0 \times 10^{-8}$) at 1p12 (rs2474937 near *TBX15*; odds ratio (OR) = 1.40; $P = 8.44 \times 10^{-10}$) and 4q31.1 (rs1531070 in *MAML3*; OR = 1.40; $P = 4.99 \times 10^{-12}$). These results extend current knowledge of genetic contributions to CHM in Han Chinese populations.

Congenital heart disease (CHD) is characterized by structural and conduction abnormalities. Structural malformations of the heart are the leading cause of infant mortality[1]. CHM can be classified into three broad categories—cyanotic heart disease, left-sided obstruction defects and the more common septation defects[2]—but the proportion of each category varies greatly by geographic region.

During the past decade, studies of animal models have elucidated many fundamental pathways that genetically govern early cardiac patterns and differentiation, such as the bone morphogenetic protein (BMP), transforming growth factor (TGF)-β and Notch signaling pathways[3–5]. Studies have uncovered the genetic basis for some forms of CHM and have provided insights into how the heart develops and how the dysregulation of heart development leads to CHM[6,7]. The majority of CHM with monogenic inheritance is associated with non-cardiac malformations and thereby constitutes syndromic forms of CHM. These include well-known examples, such as Holt-Oram syndrome (*TBX5* mutations; CHM and limb anomalies), Alagille syndrome (*NOTCH2* and *JAG1* mutations; CHM and butterfly vertebrae), DiGeorge syndrome (*TBX1* deletion; CHM and cleft palate) and Noonan syndrome (*PTPN11* mutations; CHM, short stature and typical facial dysmorphology)[8,9].

Epidemiological studies indicate that syndromic CHM comprises approximately 25% of cases, and the remaining cases are represented by isolated non-syndromic CHM[10,11]. In recent years, several genes associated with monogenic forms of non-syndromic CHM have been reported[12], including those encoding T-box transcription factors (*TBX1*, *TBX5* and *TBX20*), homeobox transcription factors (*NKX2.5* and *NKX2.6*), basic helix-loop-helix transcription factors (*HAND1* and *HAND2*) and various SMAD transcription factors. In contrast to syndromic CHM, most non-syndromic CHM occurs sporadically and may result from a multifactorial inheritance model that involves a multitude of susceptibility genes with low-penetrance mutations (common variants) or intermediate-penetrance mutations (rare variants) superposed onto unfavorable environmental factors[13].

[1]State Key Laboratory of Reproductive Medicine, School of Public Health, Nanjing Medical University, Nanjing, China. [2]Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, China. [3]Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Genome Pilot Institutes for Genomics and Human Health, Shanghai Jiao Tong University, Shanghai, China. [4]Department of Cardiothoracic Surgery, Nanjing Children's Hospital, Nanjing Medical University, Nanjing, China. [5]Department of Thoracic and Cardiovascular Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China. [6]Department of Cardiovascular Surgery, Xijing Hospital, The Fourth Military Medical University, Xi'an, China. [7]Department of Cardiology, Nanjing Children's Hospital, Nanjing Medical University, Nanjing, China. [8]Center of Prenatal Diagnosis, The Affiliated Nanjing Maternity and Child Health Hospital of Nanjing Medical University, Nanjing, China. [9]State Key Laboratory of Cancer Biology, Department of Cell Biology, Cell Engineering Research Center, The Fourth Military Medical University, Xi'an, China. [10]Department of Histology and Embryology, Nanjing Medical University, Nanjing, China. [11]Department of Toxicology, School of Public Health, Nanjing Medical University, Nanjing, China. [12]Key Laboratory of Modern Toxicology of the Ministry of Education, School of Public Health, Nanjing Medical University, Nanjing, China. [13]Center of Clinical Reproductive Medicine, First Affiliated Hospital, Nanjing Medical University, Nanjing, China. [14]These authors contributed equally to this work. Correspondence should be addressed to H.S. (hbshen@njmu.edu.cn), Z.Z. (zhouzm@njmu.edu.cn), Y.C. (yjchen@njmu.edu.cn) or Z.H. (zhibin_hu@njmu.edu.cn).

Received 20 November 2012; accepted 12 April 2013; published online 26 May 2013; doi:10.1038/ng.2636

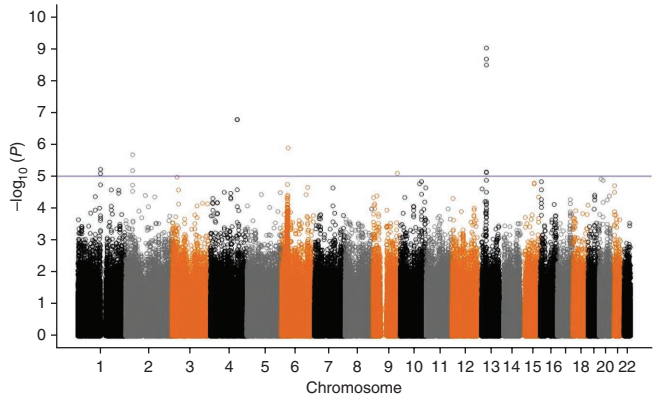

Figure 1 Genome-wide association results for CHM in Han Chinese populations. Scatter plot of *P* values on a $-\log_{10}$ scale from the logistic regression model in an additive model with adjustment for the top eigenvector. The blue horizontal line represents $P = 1.0 \times 10^{-5}$.

Although this hypothesis has been widely accepted and small-scale case-control studies have been reported to identify common variants[14–17], little progress has been made in identifying genes with low-penetrance susceptibility variants, and a GWAS has not yet been performed on sporadic CHM.

The broad phenotypic spectrum of CHM suggests a complex underlying genetic network with a large number of modifier genes. To reduce the heterogeneity between different phenotypes, we only included CHM cases with septation defects, including atrial septal defects (ASD), ventricular septal defects (VSD) and atrial septal defects combined with ventricular septal defects (ASD/VSD) in the GWAS stage of the current study. We performed the GWAS in 957 ASD, VSD and ASD/VSD cases and 1,308 non-CHM controls of Han Chinese ancestry using Illumina Omni Zhonghua chips with 900,015 SNPs. Principal-component analysis (PCA) showed little evidence of population stratification in our study populations (Online Methods and **Supplementary Fig. 1**). After quality control procedures, a total of 708,275 SNPs in 945 CHM cases and 1,246 controls were included in the subsequent genetic association analysis (Online Methods and **Supplementary Table 1**).

Associations were assessed in an additive model using logistic regression analyses with adjustment for the top eigenvector (**Fig. 1**). On the basis of having an additive *P* value of $\leq 1.0 \times 10^{-5}$ in the GWAS stage, 8 SNPs were selected to be genotyped in the first validation stage (validation 1) with an additional 1,578 ASD, VSD and ASD/VSD cases and 2,301 non-CHM controls (Online Methods and **Supplementary Tables 2** and **3**), whereas 5 SNPs in strong linkage disequilibrium (LD) with the selected SNPs were excluded from genotyping (**Supplementary Table 4**). Of the eight tested SNPs, two (rs2474937 at 1p12 and rs1531070 at 4q31.1) showed significant associations in the same direction as observed in the GWAS stage (**Table 1** and **Supplementary Table 3**). Furthermore, these 2 SNPs were also consistently replicated in another independent validation sample set (validation 2) including 582 ASD, VSD and ASD/VSD cases and 1,565 controls (**Table 1** and **Supplementary Table 3**). Combined analysis in which the two-stage validation and GWAS samples were pooled showed that the associations of rs2474937 and rs1531070 with risk of CHM achieved genome-wide significance (OR = 1.40, $P = 8.44 \times 10^{-10}$ and OR = 1.40, $P = 4.99 \times 10^{-12}$, respectively), without significant heterogeneity between stages (**Table 1**). Moreover, similar associations were also observed for the two SNPs in 1,120 CHM cases with phenotypes different from isolated ASD, isolated VSD or ASD/VSD and 3,866 controls (rs2474937 at 1p12: OR = 1.32, $P = 1.00 \times 10^{-3}$; rs1531070 at 4q31.1: OR = 1.20, $P = 1.20 \times 10^{-2}$) (validation 3; **Supplementary Table 3**). In stratified analysis, associations of these two loci were not significantly different in the major subtypes of ASD, VSD, ASD/VSD and patent ductus arteriosus (PDA) (**Supplementary Table 5**).

We then performed imputation analyses to fine map signals associated with CHM risk around each locus on the basis of our GWAS data (imputed $r^2 > 0.3$, quality threshold > 0.9, minor allele frequency (MAF) > 0.05, located within 800 kb of the two marker SNPs) (**Fig. 2**). rs2474937 was one of the top signals between *SPAG17* (encoding sperm-associated antigen 17) and *TBX15* (encoding T-box 15) at 1p12 (**Fig. 2a** and **Supplementary Table 6**). A series of signals in strong LD with rs1531070 were also found within the *MAML3* gene (encoding mastermind like 3) at 4q31.1 (**Fig. 2b** and **Supplementary Table 6**). Further fine-mapping studies based on data from the resequencing of these two regions may provide a clearer understanding of the associations at these two loci.

We found that *TBX15* and *MAML3* were expressed in human cardiac tissues at high levels, and similar expression was also observed in lung and liver tissues, whereas *SPAG17* was expressed at relatively low levels in cardiac tissues (**Supplementary Fig. 2**). We functionally annotated the LD region containing the SNP rs2474937 at 1p12 using the UCSC Genome Browser (see URLs; **Supplementary Fig. 3**) and did not observe any known coding or noncoding genes in this region. The SNP rs2474937 is located 175 kb upstream of *SPAG17* and 523 kb downstream of *TBX15*. This locus has been reported to be associated with height, pediatric stature and waist-hip ratio in humans[18–20]. T-box transcription factors have key roles in the development of

Table 1 Summary of associations with CHM for the 1p12 and 4q31.1 loci in GWAS and validation stages

				Genotype distribution[b]		MAF			
Locus	Study	Cases	Controls	Cases	Controls	Cases	Controls	OR_{add} (95% CI)[c]	P_{add}[c]
1p12 rs2474937	GWAS	941	1,245	14/191/736	2/182/1,061	0.12	0.07	1.63(1.32–2.01)	6.07×10^{-6}
T/C[a]	Validation 1	1,575	2,297	25/287/1,263	23/323/1,951	0.11	0.08	1.35(1.16–1.57)	9.97×10^{-5}
	Validation 2	580	1,556	8/104/468	10/239/1,307	0.10	0.08	1.27(1.01–1.59)	3.99×10^{-2}
	Validation 1 and 2	2,155	3,853	33/391/1,731	33/562/3,258	0.11	0.08	1.32(1.17–1.50)	1.04×10^{-5}
	All combined	3,096	5,098	47/582/2,467	35/744/4,319	0.11	0.08	1.40(1.26–1.56)	8.44×10^{-10}
4q31.1 rs1531070	GWAS	939	1,243	22/252/665	9/243/991	0.16	0.11	1.63(1.36–1.96)	1.68×10^{-7}
G/A[a]	Validation 1	1,576	2,301	35/357/1,184	30/436/1,835	0.14	0.11	1.29(1.13–1.48)	2.70×10^{-4}
	Validation 2	582	1,555	6/151/425	27/276/1,252	0.14	0.11	1.36(1.12–1.67)	2.32×10^{-3}
	Validation 1 and 2	2,158	3,856	41/508/1,609	57/712/3,087	0.14	0.11	1.31(1.17–1.47)	1.95×10^{-6}
	All combined	3,097	5,099	63/760/2,274	66/955/4,078	0.14	0.11	1.40(1.27–1.54)	4.99×10^{-12}

[a]Major/minor alleles. [b]Individuals homozygous for the minor allele/heterozygous/homozygous for the major allele. [c]OR_{add} (95% confidence interval (CI)) and P_{add} values were derived from logistic regression analysis in the additive model with adjustment for the top eigenvector.

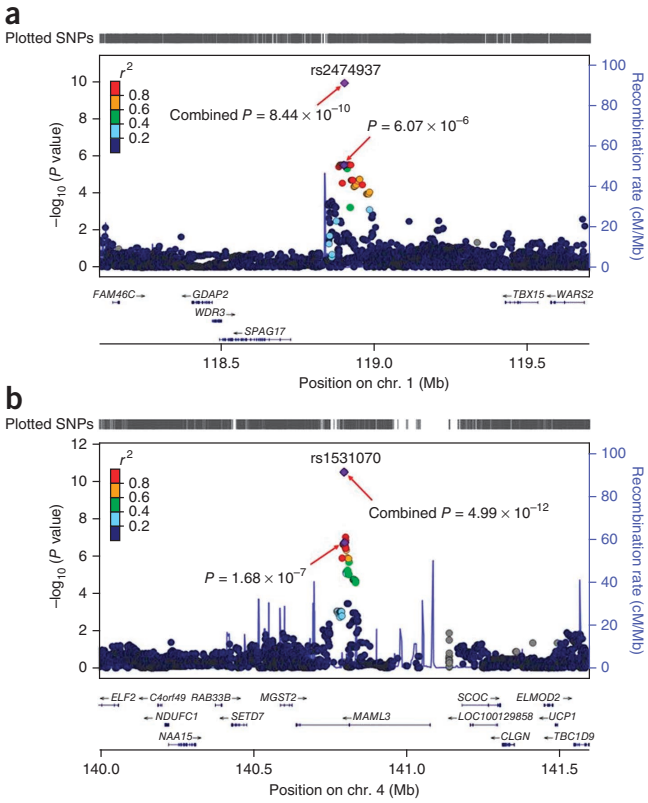

Figure 2 Regional association plots. (**a**,**b**) Plots are shown for the two loci associated with CHM at 1p12 (**a**) and 4q31.1 (**b**). Imputation was performed for each region using 1000 Genomes Project CHB (Han Chinese in Beijing, China) and JPT (Japanese in Tokyo, Japan) data (November 2010 release) as a reference. Results ($-\log_{10}$ (*P* values)) are shown for SNPs in the 1.6-Mb regions centered on the proxy SNPs. Proxy SNPs are shown in purple, and the r^2 values of the other SNPs are indicated by color. The genes within the regions of interest are annotated, and arrows represent the direction of transcription. The right *y* axis shows the recombination rate estimated from the HapMap samples.

the embryonic mesoderm, including of the heart and skeleton. For example, mutations in *TBX5* cause Holt-Oram syndrome, which is characterized by congenital forelimb and cardiac malformations[21–24]. Defects in *TBX1* and *TBX20* are catastrophic for heart development in mice and humans[25–31]. Furthermore, *Tbx2*, *Tbx6* and *Tbx18* are also involved in cardiac chamber formation, regulating left-right patterning and inflow tract development in mice[32–34]. It has been reported that inactivation of the *Tbx15* gene in mice and mutations of *TBX15* in humans may result in severe skeletal malformation[35,36] and that *Tbx15* is also involved in adipocyte differentiation and mitochondrial respiration[37]. However, the target genes of TBX15 have remained elusive so far, and phylogenetic analysis has shown a very close relationship between *Tbx15* and *Tbx18* within the *Tbx1* subfamily[38]. Conservation of the T-box DNA-binding domain allows T-box transcription factors to bind to a core 5′-AGGTGT-3′ sequence[39]. Therefore, when T-box proteins are endogenously coexpressed, there seems to be competition for binding sites within the enhancers of shared target genes. For example, during mouse heart development, Tbx2, a transcriptional repressor, competes with Tbx5, an activator, for binding at the *Nppa* (also known as *Anf*) enhancer[34]. Ectopically, *Tbx6* expression within the segmented paraxial mesoderm could result in *Tbx18* null–like phenotypes, whereas it resulted in *Tbx15* null–like phenotypes within the lateral plate mesoderm[40]. Further studies are warranted to help understand this cross-talk and competition between T-box genes during heart development.

The SNP rs1531070 is in a LD region overlapping the *MAML3* gene (including part of intron 1, all of exon 2 and part of intron 3) (**Supplementary Fig. 4**). Mastermind (Mam) is one of the essential components of the Notch signaling pathway[41]. The MAML1 and MAML3 proteins are the most closely related in the family, with 30% identity in the primary protein sequence. Disruption of *MamL1* in mice causes partial deficiency in Notch signaling *in vivo*[42]. However, *MamL1*-deficient mice did not recapitulate total loss of Notch signaling, suggesting that other mastermind family members could compensate for the loss of *MamL1*. Although *MamL3*-null mice showed no apparent abnormalities, mice null for both *MamL1* and *MamL3* died during the early organogenic period with classic pan-Notch defects[43]; furthermore, expression of the lunatic fringe gene (*Lfng*), which is strictly controlled by Notch signaling in the posterior presomitic mesoderm, was undetectable in this tissue in double-knockout embryos, whereas neither of the single-knockout embryos exhibited any abnormal phenotype. Thus, *MamL1* and *MamL3* have distinct and major roles during the organogenic period in mice as essential components of Notch signaling[43].

We additionally reviewed published studies and summarized previously reported SNPs significantly associated with CHM risk (**Supplementary Table 7**). We did not observe consistent associations in samples from our GWAS stage and those reported previously (**Supplementary Table 8**). Notably, several convincing signals in the GWAS stage were not replicated in subsequent validation stages. We have ruled out the possibility of genotyping error through technical validation using TaqMan assays (with concordance rates of 98.2% and 99.6% for rs9533839 and rs1325324, respectively). To examine potential differences in the frequency of homozygotes with the minor allele between the GWAS and validation stages for the two identified SNPs, we retyped all of the homozygous cases in three stages (GWAS, validation 1 and validation 2), and the concordance rates were 100% for both loci. The difference in frequencies of homozygotes with the minor allele in the GWAS and validation stages for the two identified SNPs may suggest false positive results in the GWAS stage or may be due to the well-known winner's curse[44]. Other explanations include disease heterogeneity, the approaches used in subject recruitment and study design. In this study, we did not perform frequency matching between case and control groups on the basis of age and sex, and differences in frequency may have resulted in unknown effects on the genetic association results, although the genetic variation changed little by age and sex.

In conclusion, this first GWAS of CHM extends our current knowledge of the genetic contributions of low-penetrance variants to CHM in Han Chinese populations and highlights the importance of competition between the T-box transcription factors and of MAML3-Notch signaling in the development of CHM.

URLs. LocusZoom 1.1, http://csg.sph.umich.edu/locuszoom/; MACH 1.0/Minimac, http://www.sph.umich.edu/csg/abecasis/MACH/; PLINK1.07, http://pngu.mgh.harvard.edu/~purcell/plink/; 1000 Genomes Project, http://www.ncbi.nlm.nih.gov/projects/faspftp/1000genomes/; UCSC Genome Browser, http://genome.ucsc.edu/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

The authors wish to thank all the study participants, research staff and students who participated in this work and Q. Wei (University of Texas MD Anderson Cancer Center) for his scientific editing of the manuscript. This work was partly funded by the National Key Basic Research Program Grant (2011CB944304 and 2012CB944902), the 863 Program (2012AA02A515), a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (Public Health and Preventive Medicine), the Natural Science Foundation of China (81000076, 81130022 and 81121001) and the Program for Changjiang Scholars and Innovative Research Team in University (IRT1025).

AUTHOR CONTRIBUTIONS

H.S., J.S., Y.C., Z.Z. and Z.H. directed the study, obtained financial support and were responsible for study design, the interpretation of results and manuscript writing. Y.S. directed the GWAS. J.D. and Y.L. were responsible for statistical analyses. Y.L., S.P., M.D., B.Q., Y.W. and J.W. were responsible for sample processing and managed the genotyping data. X.M., J. Xu, S. Yang, Z.X., Xiaowei Wang, X.G., Y.X., H.M., G.J., S. Yu, J.L. and Xinru Wang were responsible for subject recruitment and sample preparation for the Nanjing samples. B.Z. and J. Xing were responsible for subject recruitment and sample preparation for the Xi'an samples. All authors approved the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Hoffman, J.I. & Kaplan, S. The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* **39**, 1890–1900 (2002).
2. Bruneau, B.G. The developmental genetics of congenital heart disease. *Nature* **451**, 943–948 (2008).
3. Krebs, L.T. *et al.* Notch signaling is essential for vascular morphogenesis in mice. *Genes Dev.* **14**, 1343–1352 (2000).
4. Wang, J., Greene, S.B. & Martin, J.F. BMP signaling in congenital heart disease: new developments and future directions. *Birth Defects Res. A Clin. Mol. Teratol.* **91**, 441–448 (2011).
5. Anderson, L.M. & Gibbons, G.H. Notch: a mastermind of vascular morphogenesis. *J. Clin. Invest.* **117**, 299–302 (2007).
6. Olson, E.N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922–1927 (2006).
7. Srivastava, D. Making or breaking the heart: from lineage determination to morphogenesis. *Cell* **126**, 1037–1048 (2006).
8. Pierpont, M.E. *et al.* Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* **115**, 3015–3038 (2007).
9. Weismann, C.G. & Gelb, B.D. The genetics of congenital heart disease: a review of recent developments. *Curr. Opin. Cardiol.* **22**, 200–206 (2007).
10. Øyen, N. *et al.* Recurrence of congenital heart defects in families. *Circulation* **120**, 295–301 (2009).
11. Burn, J. *et al.* Recurrence risks in offspring of adults with major heart defects: results from first cohort of British collaborative study. *Lancet* **351**, 311–316 (1998).
12. Wessels, M.W. & Willems, P.J. Genetic factors in non-syndromic congenital heart malformations. *Clin. Genet.* **78**, 103–123 (2010).
13. Ware, S.M. & Jefferies, J.L. New genetic insights into congenital heart disease. *J. Clin. Exp. Cardiolog.* **S8**, pii: 003 (2012).
14. Roessler, E. *et al.* Reduced NODAL signaling strength via mutation of several pathway members including *FOXH1* is linked to human heart defects and holoprosencephaly. *Am. J. Hum. Genet.* **83**, 18–29 (2008).
15. Van Driel, L.M. *et al.* Eight-fold increased risk for congenital heart defects in children carrying the nicotinamide N-methyltransferase polymorphism and exposed to medicines and low nicotinamide. *Eur. Heart J.* **29**, 1424–1431 (2008).
16. van Beynum, I.M. *et al.* Common 894G>T single nucleotide polymorphism in the gene coding for endothelial nitric oxide synthase (eNOS) and risk of congenital heart defects. *Clin. Chem. Lab. Med.* **46**, 1369–1375 (2008).
17. Xu, J. *et al.* Functional variant in microRNA-196a2 contributes to the susceptibility of congenital heart disease in a Chinese population. *Hum. Mutat.* **30**, 1231–1236 (2009).
18. Zhao, J. *et al.* The role of height-associated loci identified in genome wide association studies in the determination of pediatric stature. *BMC Med. Genet.* **11**, 96 (2010).
19. Kim, J.J. *et al.* Identification of 15 loci influencing height in a Korean population. *J. Hum. Genet.* **55**, 27–31 (2010).
20. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
21. Li, Q.Y. *et al.* Holt-Oram syndrome is caused by mutations in *TBX5*, a member of the *Brachyury* (*T*) gene family. *Nat. Genet.* **15**, 21–29 (1997).
22. Basson, C.T. *et al.* Mutations in human *TBX5* [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat. Genet.* **15**, 30–35 (1997).
23. Borozdin, W. *et al.* Expanding the spectrum of *TBX5* mutations in Holt-Oram syndrome: detection of two intragenic deletions by quantitative real time PCR, and report of eight novel point mutations. *Hum. Mutat.* **27**, 975–976 (2006).
24. Plageman, T.F. Jr. & Yutzey, K.E. Differential expression and function of Tbx5 and Tbx20 in cardiac development. *J. Biol. Chem.* **279**, 19026–19034 (2004).
25. Jerome, L.A. & Papaioannou, V.E. DiGeorge syndrome phenotype in mice mutant for the T-box gene, *Tbx1*. *Nat. Genet.* **27**, 286–291 (2001).
26. Lindsay, E.A. *et al. Tbx1* haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature* **410**, 97–101 (2001).
27. Merscher, S. *et al. TBX1* is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. *Cell* **104**, 619–629 (2001).
28. Stennard, F.A. & Harvey, R.P. T-box transcription factors and their roles in regulatory hierarchies in the developing heart. *Development* **132**, 4897–4910 (2005).
29. Kirk, E.P. *et al.* Mutations in cardiac T-box factor gene *TBX20* are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *Am. J. Hum. Genet.* **81**, 280–291 (2007).
30. Qian, L. *et al.* Transcription factor neuromancer/TBX20 is required for cardiac function in *Drosophila* with implications for human heart disease. *Proc. Natl. Acad. Sci. USA* **105**, 19833–19838 (2008).
31. Posch, M.G. *et al.* A gain-of-function *TBX20* mutation causes congenital atrial septal defects, patent foramen ovale and cardiac valve defects. *J. Med. Genet.* **47**, 230–235 (2010).
32. Hadjantonakis, A.K., Pisano, E. & Papaioannou, V.E. Tbx6 regulates left/right patterning in mouse embryos through effects on nodal cilia and perinodal signaling. *PLoS ONE* **3**, e2511 (2008).
33. Christoffels, V.M. *et al.* Formation of the venous pole of the heart from an *Nkx2–5*-negative precursor population requires *Tbx18*. *Circ. Res.* **98**, 1555–1563 (2006).
34. Habets, P.E. *et al.* Cooperative action of Tbx2 and Nkx2.5 inhibits ANF expression in the atrioventricular canal: implications for cardiac chamber formation. *Genes Dev.* **16**, 1234–1246 (2002).
35. Singh, M.K. *et al.* The T-box transcription factor Tbx15 is required for skeletal development. *Mech. Dev.* **122**, 131–144 (2005).
36. Lausch, E. *et al. TBX15* mutations cause craniofacial dysmorphism, hypoplasia of scapula and pelvis, and short stature in Cousin syndrome. *Am. J. Hum. Genet.* **83**, 649–655 (2008).
37. Gesta, S. *et al.* Mesodermal developmental gene *Tbx15* impairs adipocyte differentiation and mitochondrial respiration. *Proc. Natl. Acad. Sci. USA* **108**, 2771–2776 (2011).
38. Meins, M., Henderson, D.J., Bhattacharya, S.S. & Sowden, J.C. Characterization of the human *TBX20* gene, a new member of the T-Box gene family closely related to the *Drosophila H15* gene. *Genomics* **67**, 317–332 (2000).
39. Naiche, L.A., Harrelson, Z., Kelly, R.G. & Papaioannou, V.E. T-box genes in vertebrate development. *Annu. Rev. Genet.* **39**, 219–239 (2005).
40. Wehn, A.K. & Chapman, D.L. *Tbx18* and *Tbx15* null-like phenotypes in mouse embryos expressing Tbx6 in somitic and lateral plate mesoderm. *Dev. Biol.* **347**, 404–413 (2010).
41. Artavanis-Tsakonas, S., Matsuno, K. & Fortini, M.E. Notch signaling. *Science* **268**, 225–232 (1995).
42. Oyama, T. *et al.* Mastermind-1 is required for Notch signal-dependent steps in lymphocyte development *in vivo*. *Proc. Natl. Acad. Sci. USA* **104**, 9764–9769 (2007).
43. Oyama, T. *et al.* Mastermind-like 1 (MamL1) and mastermind-like 3 (MamL3) are essential for Notch signaling *in vivo*. *Development* **138**, 5235–5246 (2011).
44. Kraft, P. Curses—winner's and otherwise—in genetic epidemiology. *Epidemiology* **19**, 649–651, discussion 657–658 (2008).

ONLINE METHODS

Study population. The study was approved by the institutional review boards of all participating centers and hospitals. All cases and/or their parents as well as controls provided informed consent for participation in this study. The GWAS stage included 945 sporadic ASD, VSD and ASD/VSD cases and 1,246 controls recruited from the First Affiliated Hospital of Nanjing Medical University and the Affiliated Nanjing Children's Hospital of Nanjing Medical University (Nanjing, China) between March 2006 and March 2009. For the first validation stage (validation 1), 1,578 cases with ASD, VSD or ASD/VSD and 2,301 controls were recruited from the above two hospitals between March 2009 and March 2012. For the second validation stage (validation 2), 582 cases with ASD, VSD or ASD/VSD and 1,565 controls were recruited from Xijing Hospital (Xi'an, China). In addition, 864 and 256 CHM cases with phenotypes different from isolated ASD, isolated VSD or ASD/VSD were recruited from the First Affiliated Hospital of Nanjing Medical University (validation 3a) and the Xijing Hospital (validation 3b), respectively, as the third validation stage (validation 3), using controls shared with validation 1 and validation 2, respectively. Some of the samples were reported in previously published studies[17]. Non-syndromic CHM cases were diagnosed on the basis of echocardiography, with some diagnoses further confirmed by cardiaccatheterization and/or surgery (detailed classifications are shown in **Supplementary Table 9** and ref. 2). Cases who had clinical features of developmental syndromes, multiple major developmental anomalies or known chromosomal abnormalities were excluded. Cases were also excluded if they had a positive family history of CHM in a first-degree relative (parent, sibling or child), maternal diabetes mellitus, phenylketonuria, maternal exposure to teratogens (for example, from pesticides and organic solvents) or maternal exposure to therapeutic drugs during the intrauterine period. Controls were outpatients without CHM from the same geographic areas. They were recruited from the hospitals above during the same time period. Controls with congenital anomalies or cardiac disease were excluded. Subjects were not excluded from the control group if they had a family history of congenital or cardiac anomalies. All subjects were genetically unrelated individuals of Han Chinese ancestry. For each participant, approximately 2 ml of whole blood was obtained to extract genomic DNA for genotyping analysis.

Quality control in the GWAS stage. Samples (individuals) with an overall genotyping rate of <95% were excluded from further analysis (30 subjects); no samples were excluded because of sex discrepancies. Overall, 24 unexpected duplicates or probable relatives were excluded on the basis of pairwise identity by state (PI_HAT > 0.25). SNPs were excluded if (i) they did not map to autosomal chromosomes; (ii) they had a call rate of <95%; (iii) they had a MAF of <0.05 in controls; or (iv) the genotype distribution for the SNP deviated in controls from that expected under Hardy-Weinberg equilibrium ($P < 1 \times 10^{-5}$). We detected population outliers and stratification using a PCA-based method. A set of 55,236 common autosomal SNPs with low LD ($r^2 < 0.1$) was used to classify outliers (20 samples), using the founders of the HapMap trios of the YRI (Yoruba in Ibadan, Nigeria; $n = 90$), CEU (Utah residents of Northern and Western European ancestry; $n = 90$), CHB ($n = 45$) and JPT ($n = 44$) populations as internal controls and our GWAS participants who remained after the removal of samples with low call rate, ambiguous sex and familial relationship (**Supplementary Fig. 1a**). PCA showed that the cases and controls were genetically matched (**Supplementary Fig. 1b**), and the genomic control inflation factor (λ) was 1.065. After these quality control processes, a total of 945 ASD, VSD and ASD/VSD cases and 1,246 controls were included in the final analysis with 708,275 SNPs.

SNP selection and genotyping in the validation phase. SNPs for the first validation stage were selected if they had a P value of $\leq 1.0 \times 10^{-5}$ in the GWAS stage, and only one SNP (with the lowest P value) was selected when multiple SNPs showed strong LD ($r^2 \geq 0.8$). The SNPs significantly associated with CHM risk in the first validation stage were further genotyped in samples from the second validation stage. Genotyping in the two validation stages was performed with TaqMan assays (Applied Biosystems). Genotyping for more than 10% of cases and controls in the GWAS and validation stages was randomly repeated for the two identified SNPs (rs2474937 and rs1531070) using TaqMan assays, and the concordance rates were 99.5% and 99.7%, respectively. In addition, we retyped all cases who were homozygous for the minor allele in three stages (GWAS, validation 1 and validation 2) using TaqMan assays for the two identified SNPs ($n = 47$ and 63 for homozygotes with the rs2474937[C] and rs1531070[A] alleles, respectively), and the concordance rates were 100% for both loci. Technicians who performed genotyping experiments were blinded to the case-control status of the samples.

Quantitative RT-PCR. Quantitative RT-PCR (qRT-PCR) was performed to measure the mRNA expressions of *TBX15*, *MAML3* and *SPAG17* in tissues. RNA was isolated from 13 cardiac tissues, 8 pairs of liver cancer and adjacent non-cancer tissue, and 8 pairs of lung tumor and adjacent non-cancer tissue. TaqMan Gene Expression Master Mix (Applied Biosystems), primers and probes were used to perform the qRT-PCR assays for *TBX15* and *MAML3*, and Power SYBR Green PCR Master Mix (Applied Biosystems) and primers were used to analyze *SPAG17* expression. All RT-PCR samples, including no-template controls, were run using the ABI7900 Real-Time PCR System (Applied Biosystems), with runs performed in triplicate. Expression of the *ACTB* gene (β-actin) was used to normalize expression levels. Relative levels of expression were calculated using the equation $2^{-\Delta CT}$ (C_T, cycle threshold), in which $\Delta C_T = C_{T\ gene} - C_{T\ ACTB}$.

Statistical analysis. Population structure was evaluated by PCA as implemented in the software package EIGENSTRAT 3.1 (ref. 45). The significant ($P < 0.05$) top eigenvector was included in the logistic regression model as a covariate for genetic association analysis. We used PLINK 1.07 for general genetic association analysis[46]. We used MACH 1.0 and Minimac software (see URLs) to impute untyped SNPs using LD information from the 1000 Genomes Project database (with CHB and JPT samples as the reference set; November 2010 release). Chromosome regions were plotted using an online tool, LocusZoom 1.1 (see URLs). P values were two sided, and ORs were estimated with an additive model by logistic regression analyses if not specified otherwise. The X^2-based Cochran's Q statistic was also calculated to test for heterogeneity between groups in stratified analysis. Analyses were also performed using R (2.11.1) or Stata version 9.2 (StataCorp).

45. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
46. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

doi:10.1038/ng.2636