Association analysis identifies new risk loci for congenital heart disease in Chinese populations

Yuan Lin[1,2,*], Xuejiang Guo[1,3,*], Bijun Zhao[4,*], Juanjuan Liu[1,3,*], Min Da[5,*], Yang Wen[1,2], Yuanli Hu[5], Bixian Ni[1,2], Kai Zhang[1,2], Shiwei Yang[6], Jing Xu[7], Juncheng Dai[2], Xiaowei Wang[7], Yankai Xia[1,8], Hongxia Ma[2], Guangfu Jin[2], Shiqiang Yu[4], Jiayin Liu[1,9], Bernard D. Keavney[10,11], Judith A. Goodship[10], Heather J. Cordell[10], Xinru Wang[1,8], Hongbing Shen[1,2], Jiahao Sha[1,3], Zuomin Zhou[1,3], Yijiang Chen[7], Xuming Mo[5], Lingfei Luo[12] & Zhibin Hu[1,2]

Our previous genome-wide association study (GWAS) identified two susceptibility loci for congenital heart disease (CHD) in Han Chinese. Here we identify additional loci by testing promising associations in an extended 3-stage validation consisting of 6,053 CHD cases and 7,410 controls. We find GW significant ($P < 5.0 \times 10^{-8}$) evidence of 4 additional CHD susceptibility loci at 4q31.22 (rs1400558, upstream of *EDNRA*, $P_{all} = 1.63 \times 10^{-9}$), 9p24.2 (rs7863990, close to *SMARCA2*, $P_{all} = 3.71 \times 10^{-14}$), 12q24.13 (rs2433752, upstream of *TBX3* and *TBX5*, $P_{all} = 1.04 \times 10^{-10}$) and 20q12 (rs490514, in *PTPRT*, $P_{all} = 1.20 \times 10^{-13}$). Moreover, the data from previous European GWAS supports that rs490514 is associated with the risk of CHD ($P = 3.40 \times 10^{-3}$). These results enhance our understanding of CHD susceptibility.

[1] State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 210029, China. [2] Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China. [3] Department of Histology and Embryology, Nanjing Medical University, Nanjing 210029, China. [4] Department of Cardiovascular Surgery, Xijing hospital, The Fourth Military Medical University, Xi'an 710032, China. [5] Department of Cardiothoracic Surgery, Nanjing Children's Hospital, Nanjing Medical University, Nanjing 210008, China. [6] Department of Cardiology, Nanjing Children's Hospital, Nanjing Medical University, Nanjing 210008, China. [7] Department of Thoracic and Cardiovascular Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China. [8] Department of Toxicology and Key Laboratory of Modern Toxicology of Ministry of Education, School of Public Health, Nanjing Medical University, Nanjing 211166, China. [9] Center of Clinical Reproductive Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing 210029, China. [10] Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, UK. [11] Institute of Cardiovascular Sciences, The University of Manchester, Manchester M13 9NT, UK. [12] Key Laboratory of Freshwater Fish Reproduction and Development, Ministry of Education, Laboratory of Molecular Developmental Biology, School of Life Sciences, Southwest University, Beibei, Chongqing 400715, China. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Z.H. (email: zhibin_hu@njmu.edu.cn) or to L.L. (email: lluo@swu.edu.cn) or to X.M. (email: mohsuming15@sina.com).

Congenital heart disease (CHD) is currently the most common birth defect worldwide. Twenty-eight per cent of all major congenital anomalies consist of heart defects[1]. Epidemiologically, the worldwide prevalence of CHD is estimated to be 8 per 1,000 live births[1–3] and 4 per 1,000 adults[4]. CHD is also a leading cause of perinatal and infant mortality, causing more than 220,000 deaths globally every year[5]. In addition, patients with CHD often need long-term and high-cost medical care[6].

CHD may occur as a part of many recognized chromosomal and Mendelian syndromes[7,8], however, in 75% of cases it manifests as a non-syndromic condition and may result from a multifactorial inheritance model that involves a multitude of susceptibility genes with low- (common variants) or intermediate-penetrance mutations (rare variants)[9]. To date, several genome-wide association studies (GWASs) have achieved success in deciphering the genetic basis of CHD[10–12]. One study on European Caucasians reported that single nucleotide polymorphisms (SNPs) at chromosome 4p16 were associated with the risk of atrial septal defect (ASD), but not other subtypes[11]. Simultaneously, we reported a multistage GWAS of CHD in Han Chinese and identified two CHD susceptibility loci (rs2474937 at 1p12 for *TBX15* and rs1531070 at 4q31.1 for *MAML3*)[10]. Additional genetic factors, particularly those with relatively moderate *P* values in the GWA scan, still remain to be discovered.

Here we conduct an evaluation of the promising associations in an extended 3-stage replication, using 6,053 unrelated CHD cases and 7,410 non-CHD controls, and focus on the SNPs that had *P* values ranging from 10^{-4} to 10^{-5} in the GWA scan (Supplementary Fig. 1, Supplementary Table 1 and Supplementary Data 1). We find evidence of four new CHD susceptibility loci at 4q31.22 (rs1400558, upstream of endothelin receptor type A (*EDNRA*), $P_{all} = 1.63 \times 10^{-9}$), 9p24.2 (rs7863990, close to SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily a, member 2 (*SMARCA2*), $P_{all} = 3.71 \times 10^{-14}$), 12q24.13 (rs2433752, upstream of T-box 3 (*TBX3*) and T-box 5 (*TBX5*), $P_{all} = 1.04 \times 10^{-10}$) and 20q12 (rs490514, in protein tyrosine phosphatase, receptor type, T (*PTPRT*), $P_{all} = 1.20 \times 10^{-13}$). Using data from previous European GWAS, we find that rs490514 is also associated with the risk of CHD ($P = 3.40 \times 10^{-3}$), but no significant association was observed for rs1400558.

Results

Susceptibility loci of CHD in validation studies. For the GWAS scan stage, we analysed association between CHD and 708,275 SNPs in 945 CHD cases and 1,246 control individuals in an additive model using logistic regression analyses with adjustment for the top eigenvector. Forty-five SNPs met the selection criteria for the first-stage validation (Validation I) (Table 1 and Supplementary Data 1). We performed logistic regression analysis under an additive model of the validation samples. In Validation I, 4 SNPs at 4q31.22 (rs1400558), 9p24.2 (rs7863990), 12q24.13 (rs2433752) and 20q12 (rs490514) were consistently replicated in 2,770 ASD, ventricular septal defect (VSD) and ASD combined VSD (the occurrence of an ASD together with a VSD) cases and 3,911 controls from Nanjing. For the second-stage validation (Validation II), additional 1,095 cases with ASD, VSD or ASD combined VSD and 2,379 controls from Xi'an were genotyped to verify the significant associations of the 4 loci, and they were all consistently associated with CHD risk (Table 1 and Supplementary Data 1). Moreover, the associations of the 4 SNPs

---

Table 1 | Associations of 4 SNPs to CHD in the GWA scan and validation studies.

Chr. (cytoband)	SNP	Associated gene	Study	MAF		OR$_{add}$(95% CI)*	P$_{add}$*
				Cases	Controls		
4q31.22	rs1400558 G/A†	*EDNRA*	GWAS	0.39	0.33	1.30(1.14–1.47)	8.07×10^{-5}
			Validation I	0.36	0.34	1.13(1.05–1.21)	1.18×10^{-3}
			Validation II	0.37	0.34	1.15(1.04–1.28)	9.11×10^{-3}
			Combined‡	0.37	0.33	1.17(1.11–1.23)	2.75×10^{-9}
			Validation III§	0.36	0.32	1.18(1.06–1.31)	2.91×10^{-3}
			All combined‖	0.37	0.33	1.15(1.10–1.21)	1.63×10^{-9}
9p24.2	rs7863990 C/T†	*SMARCA2*	GWAS	0.11	0.07	1.54(1.25–1.90)	4.66×10^{-5}
			Validation I	0.10	0.08	1.20(1.06–1.35)	3.00×10^{-3}
			Validation II	0.10	0.08	1.28(1.07–1.52)	6.30×10^{-3}
			Combined‡	0.10	0.08	1.29(1.18–1.40)	7.47×10^{-9}
			Validation III§	0.12	0.08	1.45(1.23–1.72)	1.76×10^{-5}
			All combined‖	0.11	0.08	1.34(1.24–1.44)	3.71×10^{-14}
12q24.13	rs2433752 A/G†	*TBX3–TBX5*	GWAS	0.15	0.20	0.72(0.61–0.84)	4.69×10^{-5}
			Validation I	0.18	0.20	0.86(0.79–0.94)	8.31×10^{-4}
			Validation II	0.17	0.21	0.79(0.69–0.90)	4.68×10^{-4}
			Combined‡	0.17	0.20	0.80(0.75–0.85)	1.80×10^{-11}
			Validation III§	0.18	0.22	0.82(0.72–0.92)	1.23×10^{-3}
			All combined‖	0.17	0.20	0.83(0.78–0.88)	1.04×10^{-10}
20q12	rs490514 A/G†	*PTPRT*	GWAS	0.44	0.38	1.29(1.14–1.46)	5.53×10^{-5}
			Validation I	0.44	0.40	1.15(1.07–1.23)	8.75×10^{-5}
			Validation II	0.42	0.38	1.16(1.05–1.29)	5.33×10^{-3}
			Combined§	0.44	0.39	1.19(1.13–1.25)	3.53×10^{-11}
			Validation III§	0.44	0.39	1.22(1.10–1.35)	1.96×10^{-4}
			All combined§	0.44	0.39	1.19(1.13–1.24)	1.20×10^{-13}

CHD, congenital heart disease; CI, confidence interval; GWAS, genome-wide association study; MAF, minor allele frequency; OR, odds ratio; SNPs, single nucleotide polymorphisms.
*ORs, 95% CIs and *P* values were estimated in an additive model using logistic regression analyses. First principal component were adjusted in GWAS stage.
†Major/minor alleles.
‡GWAS, Validation I and II were pooled together.
§ORs, 95% CIs and *P* values of Validation III were derived from meta-analysis of Validation IIIa and Validation IIIb.
‖GWAS, Validation I, II and III were pooled together.

Table 2 | Stratified analysis by major subtypes of CHD for the 4 identified SNPs.

SNP	Subtypes	MAF		OR$_{add}$(95% CI)*	P$_{add}$*	P†
		Cases	Controls			
4q31.22: rs1400558	ASD	0.36	0.33	1.15(1.06–1.24)	1.01×10^{-3}	
	VSD	0.37	0.33	1.17(1.10–1.25)	1.32×10^{-6}	
	ASD/VSD	0.38	0.33	1.25(1.10–1.41)	4.65×10^{-4}	0.015
	PDA	0.31	0.33	0.88(0.75–1.04)	1.24×10^{-1}	
	TOF	0.37	0.33	1.18(1.01–1.38)	4.04×10^{-2}	
9p24.2: rs7863990	ASD	0.11	0.08	1.36(1.19–1.54)	3.57×10^{-6}	
	VSD	0.10	0.08	1.21(1.09–1.34)	4.25×10^{-4}	
	ASD/VSD	0.12	0.08	1.50(1.25–1.81)	1.93×10^{-5}	0.007
	PDA	0.13	0.08	1.64(1.31–2.05)	1.40×10^{-5}	
	TOF	0.13	0.08	1.76(1.41–2.19)	5.39×10^{-7}	
12q24.13: rs2433752	ASD	0.16	0.20	0.78(0.70–0.86)	1.50×10^{-6}	
	VSD	0.18	0.20	0.84(0.77–0.90)	7.30×10^{-6}	
	ASD/VSD	0.15	0.20	0.70(0.59–0.83)	2.75×10^{-5}	0.249
	PDA	0.18	0.20	0.85(0.70–1.03)	9.82×10^{-2}	
	TOF	0.15	0.20	0.72(0.58–0.89)	1.93×10^{-3}	
20q12: rs490514	ASD	0.42	0.39	1.11(1.03–1.20)	8.75×10^{-3}	
	VSD	0.44	0.39	1.23(1.15–1.30)	6.46×10^{-11}	
	ASD/VSD	0.44	0.39	1.19(1.06–1.35)	3.78×10^{-3}	0.193
	PDA	0.42	0.39	1.12(0.96–1.30)	1.42×10^{-1}	
	TOF	0.41	0.39	1.07(0.92–1.25)	3.64×10^{-1}	

ASD, atrial septal defect; VSD, ventricular septal defect; ASD/VSD, atrial septal defect combined ventricular septal defect; PDA, patent ductus arteriosus; TOF, tetralogy of fallot; MAF, minor allele frequency; OR, odds ratio; SNPs, single nucleotide polymorphisms.
*ORs, 95% CIs and P values were estimated in an additive model using logistic regression analyses.
†P values for heterogeneity test between groups were assessed using the χ^2-based Cochran's Q test.

were evaluated in 1,437 CHD cases from Nanjing and 751 CHD cases from Xi'an that had phenotypes different from those of isolated ASD, isolated VSD or ASD combined VSD, using 773 controls from Nanjing and 347 controls from Xi'an referred to as Validation IIIa and Validation IIIb, respectively. A meta-analysis of Validation IIIa and Validation IIIb was conducted to evaluate the association in a total of 2,188 CHD cases with phenotypes that were different from isolated ASD, isolated VSD or ASD combined VSD, and in 1,120 controls, and the significant associations were still observed (odds ratio (OR) = 1.18, $P = 2.91 \times 10^{-3}$ for rs1400558; OR = 1.45, $P = 1.76 \times 10^{-5}$ for rs7863990; OR = 0.82, $P = 1.23 \times 10^{-3}$ for rs2433752; and OR = 1.22, $P = 1.96 \times 10^{-4}$ for rs490514) (Table 1 and Supplementary Data 1).

Combined analysis of the susceptibility loci. In the combined analysis, additive model of logistic regression analyses were used to estimate the P values of association analyses. We first pooled all of the ASD, VSD and ASD combined VSD cases and non-CHD controls. All 4 SNPs reached GW significance ($P < 5.0 \times 10^{-8}$) for CHD susceptibility (OR = 1.17, $P_{all} = 2.75 \times 10^{-9}$ for rs1400558; OR = 1.29, $P_{all} = 7.47 \times 10^{-9}$ for rs7863990; OR = 0.80, $P_{all} = 1.80 \times 10^{-11}$ for rs2433752; and OR = 1.19, $P_{all} = 3.53 \times 10^{-11}$ for rs490514) (Table 1). In addition, all of the CHD cases and non-CHD controls from the GWA scan and Validations I, II and III were pooled together. Once again, all four SNPs demonstrated significant GW associations (OR = 1.15, $P_{all} = 1.63 \times 10^{-9}$ for rs1400558; OR = 1.34, $P_{all} = 3.71 \times 10^{-14}$ for rs7863990; OR = 0.83, $P_{all} = 1.04 \times 10^{-10}$ for rs2433752; and OR = 1.19, $P_{all} = 1.20 \times 10^{-13}$ for rs490514) without significant heterogeneity between the stages (Table 1).

Stratified analysis by diagnostic groups. In a stratified analysis, we evaluated the associations of the four SNPs using logistic regression analysis under additive model in the major subtypes of ASD, VSD, ASD combined VSD, patent ductus arteriosus (PDA) and tetralogy of fallot (TOF). The heterogeneity between subgroups was assessed with the χ^2-based Cochran's Q test.

Heterogeneities in the ORs for rs1400558 and rs7863990 were observed among the different subtypes ($P = 0.015$ and 0.007, respectively) (Table 2). For the other two SNPs (rs2433752 and rs490514), the association strengths were not significantly different among the five subtypes ($P = 0.249$ and 0.193, respectively) (Table 2). Rs1400558 showed significant associations with all subtypes except PDA (OR = 0.88, $P = 1.24 \times 10^{-1}$). The SNP rs7863990 exhibited significant associations in all of the CHD subtypes, and the strongest association was with the TOF subtype (OR = 1.76, $P = 5.39 \times 10^{-7}$). For rs2433752, although no significant association was observed in PDA subgroup, similar effect sizes were observed for all subtypes including PDA. For rs490514, significant associations were only seen in ASD, VSD and ASD combined VSD groups.

Association analysis in European CHD GWAS. Cordell et al.[11] published two studies of CHD in Europeans, one of them recruited multiple CHD phenotypes and the other one only included TOF cases[12]. Using their existing GWA scan data, we attempted to investigate the associations of the four SNPs we identified and the CHD risk in Europeans. We pooled all samples of these two studies together and evaluated the associations using logistic regression analysis under additive model. Two SNPs (rs1400558 and rs490514) were genotyped in their data. We found that rs490514 is associated with the risk of CHD significantly in Europeans (OR = 1.18, $P = 3.40 \times 10^{-3}$), while no significant association was observed for rs1400558 (OR = 1.02, $P = 6.18 \times 10^{-1}$) (Table 3). We also carried out subsidiary analyses for these two SNPs in six different diagnostic groups: ASD, VSD, transposition of the great arteries (TGA), conotruncal malformations, left-sided malformations and TOF. The association strengths were not significantly different among different subtypes ($P = 0.700$ and 0.908, respectively) (Supplementary Table 2).

Imputation analysis. An imputation analysis in our GWA scan samples identified the associations of the SNPs with CHD risk at

Table 3 | Associations of 2 SNPs to CHD in European GWAS data.

Chr. (cytoband)	SNP	Major/minor alleles	Associated gene	MAF		OR_{add}(95% CI)*	P_{add}*
				Cases	Controls		
20q12	rs490514	A/G	PTPRT	0.12	0.11	1.18(1.06–1.32)	3.40×10^{-3}
4q31.22	rs1400558	G/A	EDNRA	0.29	0.29	1.02(0.94–1.10)	6.18×10^{-1}

CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; SNPs, single nucleotide polymorphisms.
*OR_{add} and P_{add} were estimated in additive model using logistic regression analyses.
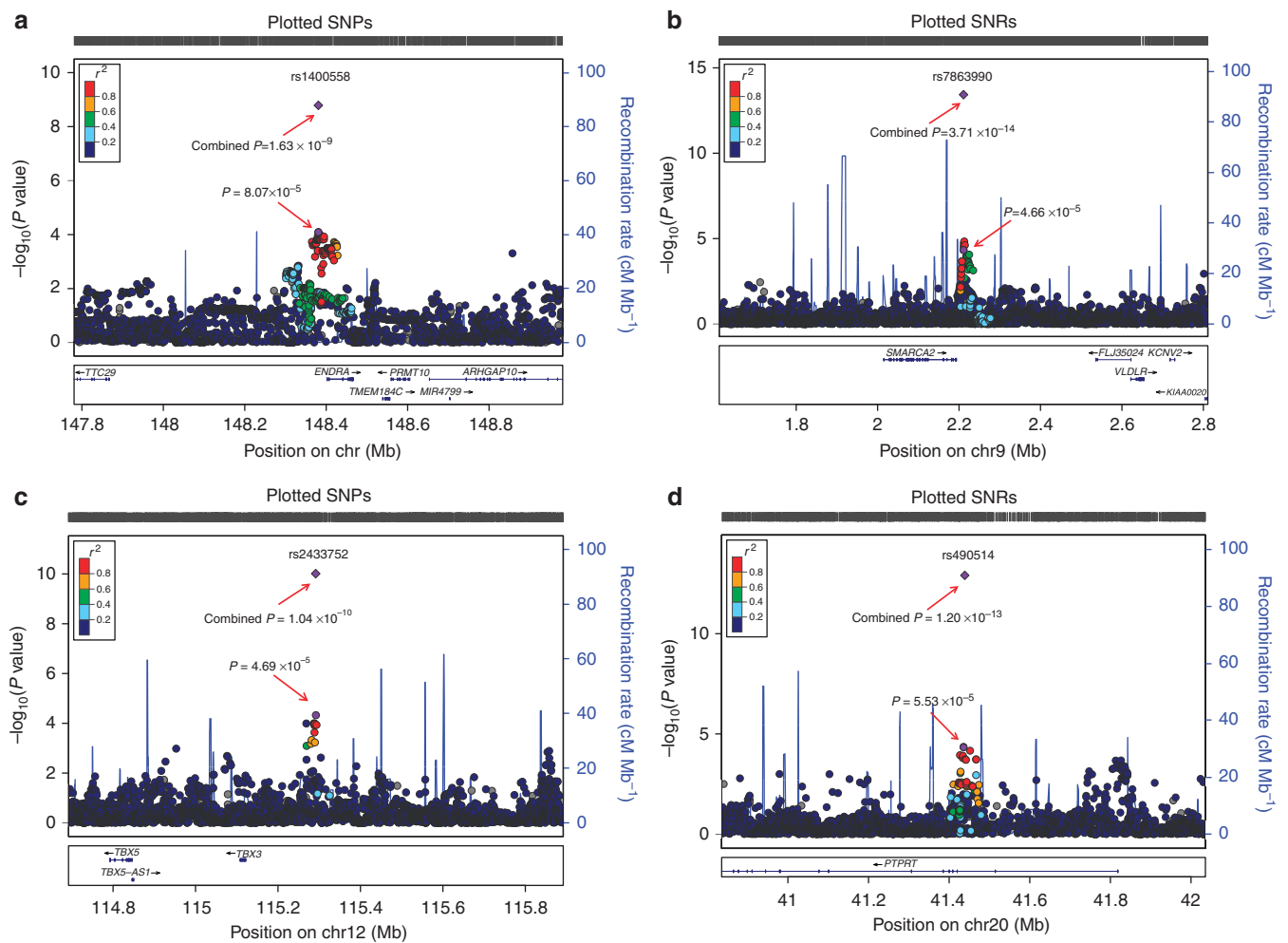
Figure 1 | Regional plot of the four identified lead SNPs. These SNPs are (**a**) rs1400558 at 4q31.22, (**b**) rs7863990 at 9p24.2, (**c**) rs2433752 at 12q24.13 and (**d**) rs490514 at 20q12. The results ($-\log_{10} P$) are shown for the SNPs located in the 600-kb region flanking either side of the marker SNPs in the GWA scan stage. Genetic association testing was performed using logistic regression analyses under an additive model. First principal component was adjusted in GWA scan stage. The marker SNPs are shown as purple circles with their P values in GWA scan stage and as purple diamonds with their P values in the combined stage when all of the CHD cases and non-CHD controls from the GWA scan and Validations I, II and III were pooled together. The r^2 values of the remaining SNPs are indicated by different colours. The genes within the region-of-interest are annotated, with arrows indicating the direction of transcription. For each plot, the recombination rates (right y-axes) of the region are shown according to their chromosomal positions (x-axis).

$P \le 1.0 \times 10^{-4}$ (imputed $r^2 > 0.3$, quality threshold > 0.9, minor allele frequency (MAF) > 0.05, located 500 kb up- or downstream of the four lead SNPs). For 4q31.22, we did not identify any untyped SNPs in high linkage disequilibrium (LD) with the lead SNP that reached $P \le 1.0 \times 10^{-4}$ (Fig. 1a and Supplementary Table 3). For 9p24.2, a series of SNPs reaching $P \le 1.0 \times 10^{-4}$ were in strong LD with rs7863990 (four SNPs: $r^2 = 0.81$–0.90, $P = 1.46 \times 10^{-5}$ to 2.86×10^{-5}, Fig. 1b and Supplementary Table 3). Rs2433748 was in high LD with rs2433752 at 12q24.13 ($r^2 = 0.86$, $P = 9.92 \times 10^{-5}$ and Supplementary Table 3). In

In addition, we found 2 SNPs in high LD with rs490514 at 20q12 ($r^2 = 1.00$ for rs13039258, $P = 5.33 \times 10^{-5}$ and $r^2 = 0.95$ for rs6103054, $P = 6.84 \times 10^{-5}$, Fig. 1d and Supplementary Table 3).

Discussion

Our three-stage validation provides compelling evidence that four new SNPs at 4q31.22 (rs1400558, upstream of *EDNRA*), 9p24.2 (rs7863990, close to *SMARCA2*), 12q24.13 (rs2433752, upstream of *TBX3* and *TBX5*) and 20q12 (rs490514, within the second

intron of *PTPRT*) influence the risk of CHD in Chinese populations. *EDNRA*, *SMARCA2*, *TBX3* and *TBX5* have been implicated to be involved in heart development in multiple previous studies. Moreover, the data of previous European GWAS supports that rs490514 is also associated with the risk of CHD in Europeans.

On chromosome 20, rs490514 is located within the second intron of protein-tyrosine phosphatase receptor type T (*PTPRT*). PTPRT is a member of the PTP family[13]. PTPs are known to be signalling molecules that regulate a variety of cellular processes, including cell growth, differentiation, the mitotic cycle and oncogenic transformation. Their protein domain structures suggest their roles in both signal transduction and cellular adhesion[14,15]. Many members of the PTP family are expressed in various tissues, including the heart[15]. Although Zhao *et al.*[16], found no obvious developmental abnormalities in homozygous *PTPRT* knockout mice, it is possible that it functions in certain genetic background or environmental conditions[17]. Combinations of a heterozygous missense mutation at the conserved PTP catalytic domain of *PTPRT* and a heterozygous intronic deletion of 150 kb removing more than half of intron 1 of *PTPRT* were reported in a Dutch family of three brothers and two sisters affected with intellectual disability, together with congenital cardiac defects[18], indicating possible essential function of *PTPRT* in heart development. In addition, using European GW scan data, we found that rs490514 is also associated with the risk of CHD in Europeans, although the allele frequency of rs490514 in both populations was greatly different (Chinese population: MAF = 0.39 in controls; European population: MAF = 0.11 in controls).

The SNP rs2433752 is located 170 kb and 448 kb upstream of T-box 3 (*TBX3*) and T-box 5 (*TBX5*), respectively. The T-box genes encode important transcription factors involved in the regulation of developmental processes, such as heart development. These genes often act in a combinatorial manner and exhibit an exquisite dose sensitivity when controlling heart developmental processes[19,20]. In the human heart, *TBX3* is expressed in the atrial floor and atrioventricular canal myocardium[21]. Various cardiac defects, such as VSD, double outlet right ventricle and TGA, have been observed in Tbx3-deficient model animals[22–24]. Another nearby gene, *TBX5*, plays important roles in chamber formation, septation and cardiomyocyte differentiation[25]. Both gain- and loss-of-function experiments on *Tbx5* have resulted in abnormal heart phenotypes in animal models, such as looping defects, loss of the ventricular septum and arrested development of the AV cushions, hypoplastic sinoatrial region and left ventricle[26,27]. The polymorphisms in the upstream region potentially influence the transcriptional activity, and the haplo-insufficiency of *TBX3* and *TBX5* has also been proven to be associated with CHD[28,29].

The SNP rs7863990 localizes 17 kb downstream of *SMARCA2*. The encoded protein of *SMARCA2* is part of the Brg1/Brm-associated factor (BAF) chromatin remodelling complex, which is thought to regulate the transcription of certain genes by altering chromatin structure[30]. Each BAF complex utilizes either SMARCA4 (also known as BRG1) or SMARCA2 (also known as BRM) as alternative catalytic subunits[31]. TBX5, Nkx2-5 and GATA4 interact with BAF complexes, and this interaction is important for the *de novo* induction of cardiac differentiation from embryonic mesoderm[32,33]. The disruption of the delicate balance between certain transcription factors and the BAF complexes is likely to be a mechanistic cause of CHD[34]. Some studies have revealed that *Smarca4* is essential for heart development in mouse and zebrafish embryogenesis, whereas *Smarca2* appears to be dispensable[34,35]. In humans, the role of *SMARCA2* in cardiogenesis is unclear; however, several

independent exome sequencing projects have identified *de novo* mutations in *SMARCA2* that are associated with two congenital syndromes that include cardiac defects: Coffin–Siris syndrome and Nicolaides–Baraitser syndrome[36,37].

We also identified associations between rs1400558 and CHD among Chinese. This SNP is located 22 kb upstream of endothelin receptor type A (*EDNRA*). In expression quantitative trait locus (eQTL) analysis using GTEx Portal (http://www.gtexportal.org/home/), we found the risk allele (A) of rs1400558 was significantly associated was lower mRNA expression of *EDNRA* in 86 heart left ventricle tissues ($P = 0.0079$), while no eQTL effects were identified for the other 3 SNPs. In the mouse model, a subpopulation of the first (crescent forming) heart field has been shown to be marked with *Ednra* gene expression[38]. Furthermore, Ednra-mediated signals are involved in myocardial growth, ventricular formation and aortic arch patterning[38,39]. It has been reported that the mice deficient in Ednra signalling exhibit aortic arch malformation and outflow anomalies[39,40]. In addition, *Ednra*-null embryonic hearts often demonstrate hypoplasia of the ventricular wall, low mitotic activity and decreased *Tbx5* expression with reciprocal expansion of *Tbx2* expression[38]. However, in European GWA scan data, no significant association was observed for rs1400558 and the risk of CHD and the allele frequency of this SNP in Chinese and Europeans also showed difference. There might be different lead SNP for Europeans in this region because of underlying genetic heterogeneity, which merits further investigations.

In a stratified analysis, heterogeneity of the association strengths was observed among the different subtypes of CHD. For the SNP rs1400558, similar associations were observed in the ASD, VSD, ASD combined VSD and TOF subtypes but not in PDA, which could be due to the susceptibility gene *EDNRA*, which was reported to be involved in the formation of the ventricles rather than the ductal media. In all of the CHD subtypes, rs7863990 exhibited significant associations, and the strongest association was with the TOF subtype. Studies have reported that BAF complexes interact with many important transcription factors, such as TBX5, Nkx2-5 and GATA4, and may further play essential roles in the formation of the outflow tract. Although no significant association was observed between rs2433752 and PDA, the sample size of this group was relatively small, and the effect size was comparable to VSD. For rs490514, the associations were not different between the subtypes. As the sample size of the PDA and TOF groups was relatively small compared with the other three groups, further studies with larger sample populations are highly recommended.

In conclusion, we identified four new CHD susceptibility loci at 4q31.22, 9p24.2, 12q24.13 and 20q12 in Han Chinese, by performing the extended three-stage validation. These findings further advance our understanding of the susceptibility to CHD.

Methods

Study design. This study was approved by the institutional review boards of Nanjing Medical University and The Fourth Military Medical University. All participants completed the informed consent in writing before taking part in this research. We performed a four-stage case-control analysis. The GWAS phase included 945 CHD cases and 1,246 non-CHD 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-stage validation (Validation I), 2,770 cases and 3,911 controls were recruited from the two hospitals above between March 2006 and June 2013. For the second-stage validation (Validation II), 1,095 cases and 2,379 controls were recruited from Xijing Hospital (Xi'an, China). In addition, 1,437 and 751 CHD cases with phenotypes different from isolated ASD, VSD or ASD combined VSD were recruited from the First Affiliated Hospital of Nanjing Medical University and the Affiliated Nanjing Children's Hospital of Nanjing Medical University (Validation IIIa) and the Xijing Hospital (Validation IIIb), respectively. The Validation IIIa and IIIb included 773 and 347 non-CHD controls from the hospitals in Nanjing and Xi'an above, respectively. We

summarized all subjects in Supplementary Tables 1 and 4. All of the samples used in the validations in the previous study[10] were included in the current study.

Non-syndromic CHD cases were diagnosed on the basis of echocardiography; and some diagnoses were further confirmed by cardiac catheterization and/or surgery (detailed classifications are shown in Supplementary Table 4). CHD cases that manifested additional syndromes, multiple developmental abnormalities or known chromosomal abnormalities were excluded from the study. We next excluded cases if they had positive family history of CHD in their first-degree relatives or if their mothers had maternal diabetes mellitus, phenylketonuria, teratogen exposure or therapeutic drug exposure during pregnancy. All of the controls were non-CHD outpatients who were recruited from the hospitals above over the same time period. Subjects exhibiting congenital anomalies or cardiac disease were excluded from the control group. For each participant, ~2 ml of whole blood was obtained to extract genomic DNA for genotyping analysis.

The discovery cohorts of the two European GWASs comprise CHD cases of multiple phenotypes and TOF cases[11,12]. All cases are self-reported European Caucasian ancestry recruited from multiple centres in the UK and from centres in Leuven, Belgium, Erlangen, Germany and Sydney, Australia. SNP genotyping in the cases was carried out using the Illumina Human660W-Quad array, and genotypes were compared with data for UK population-based controls (genotyped on the Illumina 1.2M chip) obtained from the Wellcome Trust Case Control Consortium 2 (WTCCC2). All diagnoses were established by CHD specialists at the contributing centres, and cases were classified using European Paediatric Cardiac Codes. Cases exhibiting clinical features of recognized malformation syndromes, multiple developmental abnormalities or learning difficulties were excluded from the study[11,12]. After stringent quality control, we analysed the associations of two typed SNPs (rs1400558 and rs490514) and the risk of CHD in the current data set which consisted of 2,111 unrelated CHD cases (ASD: 340; VSD: 191; TGA: 207; conotruncal malformations: 151 and left-sided malformations: 387; TOF: 835) and 5,159 controls.

SNP selection and genotyping for validation. The SNPs for Validation I were selected based on the following criteria: (i) the SNPs had 1.0×10^{-5} < $P \le 1.0 \times 10^{-4}$ in the GWAS stage; (ii) the SNPs were not located in the same chromosomal region/gene of SNPs reported in our previous study[10]; (iii) the SNPs had clear genotyping clusters; (iv) if more than one SNPs were located in the same chromosomal region/gene, only one SNP with the lowest P value was selected; and (v) only the SNP with the lowest P value was selected when multiple SNPs were observed in strong LD ($r^2 \ge 0.8$). A total of 45 SNPs met these criteria (Supplementary Data 1). Significantly associated SNPs ($P < 0.05$) were further genotyped in validation samples.

Genotyping was performed using the iPLEX Sequenom MassARRAY platform (Sequenom, Inc., USA) in Validation I and the TaqMan Allelic Discrimination Assay (Applied Biosystems, Inc., USA) in Validation II and III. The primers of nine SNPs using TaqMan allelic discrimination Assay to perform genotyping were provided in Supplementary Table 5 The details of primers and probes used for TaqMan allelic discrimination assay in validation stage were provided in Supplementary Table 5. We controlled the quality of genotyping based on the following methods: (i) case and control samples were mixture distributed on each 384-well plate; (ii) genotyping was carried out using blind method; (iii) positive and negative controls were included in each plate; and (iv) 5% randomly selected samples were repeat genotyping to calculate the coincidence rate.

Statistical analysis. The association analysis was performed using an additive model in a logistic regression analysis in PLINK 1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/). The heterogeneity between subgroups was assessed with the χ^2-based Cochran's Q test. The meta-analysis was performed in a combined analysis, and a random-effects model was used when heterogeneity between the studies existed (if the P value for the heterogeneity test was <0.05), otherwise, a fixed-effect model was used. We used MACH 1.0 software (http://www.sph.umich.edu/csg/abecasis/MACH/) to impute the ungenotyped SNPs using the LD information from the hg18/1000 Genomes Project database (CHB+JPT as a reference set; June 2010 release). The chromosomal region was plotted using LocusZoom 1.1 (http://csg.sph.umich.edu/locuszoom/). All of the other analyses were performed with R software 2.15.1 (http://www.r-project.org/).

References

1. van der Linde, D. et al. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. J. Am. Coll. Cardiol. **58,** 2241–2247 (2011).
2. Bernier, P. L., Stefanescu, A., Samoukovic, G. & Tchervenkov, C. I. The challenge of congenital heart disease worldwide: epidemiologic and demographic facts. Semin. Thorac. Cardiovasc. Surg. Pediatr. Card. Surg. Annu. **13,** 26–34 (2010).
3. Reller, M. D. et al. Prevalence of congenital heart defects in metropolitan Atlanta, 1998-2005. J. Pediatr. **153,** 807–813 (2008).
4. Marelli, A. J., Mackie, A. S., Ionescu-Ittu, R., Rahme, E. & Pilote, L. Congenital heart disease in the general population: changing prevalence and age distribution. Circulation **115,** 163–172 (2007).
5. Lozano, R. et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet **380,** 2095–2128 (2013).
6. Somerville, J. Grown-up congenital heart disease-medical demands look back, look forward 2000. Thorac. Cardiovasc. Surg. **49,** 21–26 (2001).
7. 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).
8. Weismann, C. G. & Gelb, B. D. The genetics of congenital heart disease: a review of recent developments. Curr. Opin. Cardiol. **22,** 200–206 (2007).
9. Ware, S. M. & Jefferies, J. L. New genetic insights into congenital heart disease. J. Clin. Exp. Cardiol. **S8,** pii: 003 (2012).
10. Hu, Z. et al. A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. Nat. Genet. **45,** 818–821 (2013).
11. Cordell, H. J. et al. Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. Nat. Genet. **45,** 822–824 (2013).
12. Cordell, H. J. Genome-wide association study identifies loci on 12q24 and 13q32 associated with tetralogy of Fallot. Hum. Mol. Genet. **22,** 1473–1481 (2013).
13. Alonso, A. et al. Protein tyrosine phosphatases in the human genome. Cell **117,** 699–711 (2004).
14. Aricescu, A. R. Structure of a tyrosine phosphatase adhesive interaction reveals a spacer-clamp mechanism. Science **317,** 1217–1220 (2007).
15. Paul, S. & Lombroso, P. J. Receptor and nonreceptor protein tyrosine phosphatases in the nervous system. Cell Mol. Life Sci. **60,** 2465–2482 (2003).
16. Zhao, Y. et al. Identification and functional characterization of paxillin as a target of protein tyrosine phosphatase receptor T. Proc. Natl Acad. Sci. USA **107,** 2592–2597 (2010).
17. Barbaric, I., Miller, G. & Dear, T. N. Appearances can be deceiving: phenotypes of knockout mice. Brief. Funct. Genomic. Proteomic. **6,** 91–103 (2007).
18. Schuurs-Hoeijmakers, J. H. et al. Identification of pathogenic gene variants in small families with intellectually disabled siblings by exome sequencing. J. Med. Genet. **50,** 802–811 (2013).
19. Naiche, L. A., Harrelson, Z., Kelly, R. G. & Papaioannou, V. E. T-box genes in vertebrate development. Annu. Rev. Genet. **39,** 219–239 (2005).
20. Hoogaars, W. M., Barnett, P., Moorman, A. F. & Christoffels, V. M. T-box factors determine cardiac design. Cell Mol. Life Sci. **64,** 646–660 (2007).
21. Sizarov, A. Formation of the building plan of the human heart: morphogenesis, growth, and differentiation. Circulation **123,** 1125–1135 (2011).
22. Bakker, M. L. et al. Transcription factor Tbx3 is required for the specification of the atrioventricular conduction system. Circ. Res. **102,** 1340–1349 (2008).
23. Mesbah, K. et al. Tbx3 is required for outflow tract development. Circ. Res. **103,** 743–750 (2008).
24. Chen, D. et al. Genetic analysis of the TBX3 gene promoter in ventricular septal defects. Gene **512,** 185–188 (2013).
25. Greulich, F., Rudat, C. & Kispert, A. Mechanisms of T-box gene function in the developing heart. Cardiovasc. Res. **91,** 212–222 (2011).
26. Takeuchi, J. K. et al. Tbx5 specifies the left/right ventricles and ventricular septum position during cardiogenesis. Development **130,** 5953–5964 (2003).
27. Bruneau, B. G. et al. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. Cell **106,** 709–721 (2001).
28. Frank, D. U. et al. Lethal arrhythmias in Tbx3-deficient mice reveal extreme dosage sensitivity of cardiac conduction system function and homeostasis. Proc. Natl Acad. Sci. USA **109,** E154–E163 (2012).
29. Zhu, Y. et al. Tbx5-dependent pathway regulating diastolic function in congenital heart disease. Proc. Natl Acad. Sci. USA **105,** 5519–5524 (2008).
30. Tsukiyama, T. The in vivo functions of ATP-dependent chromatin-remodelling factors. Nat. Rev. Mol. Cell Biol. **3,** 422–429 (2002).
31. Wu, J. I., Lessard, J. & Crabtree, G. R. Understanding the words of chromatin regulation. Cell **136,** 200–206 (2009).
32. Lickert, H. et al. Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. Nature **432,** 107–112 (2004).
33. Takeuchi, J. K. & Bruneau, B. G. Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. Nature **459,** 708–711 (2009).
34. Takeuchi, J. K. et al. Chromatin remodelling complex dosage modulates transcription factor function in heart development. Nat. Commun. **2,** 187 (2011).
35. Bevilacqua, A., Willis, M. S. & Bultman, S. J. SWI/SNF chromatin-remodeling complexes in cardiovascular development and disease. Cardiovasc. Pathol. **23,** 85–91 (2014).
36. Tsurusaki, Y. et al. Mutations affecting components of the SWI/SNF complex cause Coffin-Siris syndrome. Nat. Genet. **44,** 376–378 (2012).

37. Van Houdt, J. K. *et al.* Heterozygous missense mutations in SMARCA2 cause Nicolaides-Baraitser syndrome. *Nat. Genet.* **44**, 445–449 (2012).

38. Asai, R. *et al.* Endothelin receptor type A expression defines a distinct cardiac subdomain within the heart field and is later implicated in chamber myocardium formation. *Development* **137**, 3823–3833 (2010).

39. Yanagisawa, H. *et al.* Role of Endothelin-1/Endothelin-A receptormediated signaling pathway in the aortic arch patterning in mice. *J. Clin. Invest.* **102**, 22–33 (1998).

40. Kurihara, Y. *et al.* Aortic arch malformations and ventricular septal defect in mice deficient in endothelin-1. *J. Clin. Invest.* **96**, 293–300 (1995).

Author contributions

Z.H., J.S. and H.S. directed the study, obtained financial support and were responsible for study design and interpretation of results. Z.H. performed project management along with Y.L., X.G., X.M. and B.Z., and drafted the initial manuscript along with Y.L. Y.L. and J.D. were responsible for statistical analyses. Y.L., M.D., Y.W., Y.H., B.N. and K.Z. were responsible for sample processing and managed the genotyping data. Ju.L. and L.L. were responsible for functional analysis in zebrafish model. X.M., Y.C., S.Ya., J.X., X.W., Y.X., H.M. and G.J. were responsible for subject recruitment and sample preparation of Nanjing samples. B.Z. and S.Yu were responsible for subject recruitment and sample preparation of Xi'an samples. B.D.K., J.A.G. and H.J.C. provided European CHD GWAS data. Ji.L., X.W., Z.Z. and Y.C. oversaw the study design, and contributed to interpretation of results. All authors approved the final draft of the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Lin, Y. *et al.* Association analysis identifies new risk loci for congenital heart disease in Chinese populations. *Nat. Commun.* 6:8082 doi: 10.1038/ncomms9082 (2015).