The complex SNP and CNV genetic architecture of the increased risk of congenital heart defects in Down syndrome

M. Reza Sailani, Periklis Makrythanasis, Armand Valsesia, et al.

Genome Res. published online June 19, 2013
Access the most recent version at doi:10.1101/gr.147991.112

The complex SNP and CNV genetic architecture of the increased risk of congenital heart defects in Down syndrome

M.Reza Sailani,[1, 2] Periklis Makrythanasis,[1] Armand Valsesia,[14, 15, 16] Federico A. Santoni,[1] Samuel Deutsch,[1] Konstantin Popadin,[1] Christelle Borel,[1] Eugenia Migliavacca,[1] Andrew. J Sharp,[1,20] Genevieve Duriaux Sail,[1] Emilie Falconnet,[1] Kelly Rabionet,[5,17,18] Clara Serra-Juhé,[9,17] Stefano Vicari,[7] Daniela Laux,[19] Yann Grattau,[10] Guy Dembour,[11] Andre Megarbane,[10,12] Renaud Touraine,[13] Samantha Stora,[10] Sofia Kitsiou,[8] Helena Fryssira,[8] Chariklia Chatzisevastou-loukidou,[8] Emmanouel Kanavakis,[8] Giuseppe Merla,[6] Damien Bonnet,[19] Luis A. Pérez-Jurado,[9,17] Xavier Estivill,[5,17,18] Jean M. Delabar[4] and Stylianos E. Antonarakis[1, 2,3] *

1. Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland.

2. NCCR *Frontiers in Genetics,* University of Geneva, Switzerland.

3. iGE3 institute of Genetics and Genomics of Geneva, Switzerland.

4. Unité de Biologie Fonctionnelle et Adaptative (BFA), CNRS EAC 4413, Université Paris Diderot, Sorbonne Paris Cité, 75205 Paris, France.

5. Centre for Genomic Regulation, Biomedical Research Park (PRBB), Barcelona, Spain.

6. Medical Genetics Service, Hospital, IRCCS, San Giovanni Rotondo, Italy.

7. Department of Neuroscience, Children's Hospital Bambino Gesù, Rome, Italy.

8. Department of Medical Genetics, University of Athens, Athens, Greece.

9. Department of Experimental and Health Sciences, Universitat Pompeu Fabra, and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Barcelona, Spain.

10. Institut Jerome Lejeune, Paris, France.

11. Cliniques universitaires Saint-Luc, UCL, Bruxelles, Belgium.

12. Unité de Génétique Médicale et Laboratoire Associé INSERM UMR_S910, Beirut, Lebanon.

13. CHU de Saint-Etienne, hôpital Nord, Saint-Etienne Cedex 2, France.

14. Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland.

15. Ludwig Institute for Cancer Research, UNIL, Lausanne 1005, Switzerland.

16. Swiss Institute of Bioinformatics, UNIL, Lausanne 1005, Switzerland.

17. Hospital del Mar Research Institute (CRG), E-08003 Barcelona, Spain.

18. CIBER in Epidemiology and Public Health (CIBERESP), E-08003 Barcelona, Spain.

19. M3C-Necker, Pediatric Cardiology, University Paris Descartes, Paris, France

20. Present address: Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA.

41 ⍰Correspondence to: stylianos.antonarakis@unige.ch

42 Address for correspondence:

43 Stylianos E. Antonarakis, Department of Genetic Medicine and Development

44 University of Geneva Medical School and University Hospitals of Geneva

45 1 rue Michel-Servet, 1211 Geneva, Switzerland.

46 Telephone number: +41-22-379-5708

47 Fax number: +41-22-379-5706

48

49

50

51

52

53

54

55

56

57

58

59

Abstract

Congenital heart defect (CHD) occurs in 40% of Down syndrome (DS) cases. While carrying three copies of chromosome 21 increases the risk for CHD, trisomy 21 itself is not sufficient to cause CHD. Thus additional genetic variation and/or environmental factors could contribute to the CHD risk. Here we report genomic variations that in concert with trisomy 21, determine the risk for CHD in DS. This case-control GWAS includes 187 DS with CHD (AVSD=69, ASD=53, VSD=65) as cases, and 151 DS without CHD as controls. Chromosome 21 specific association study revealed rs2832616 and rs1943950 as CHD risk alleles (adjusted genotypic *P*-values < 0.05). These signals were confirmed in a replication cohort of 92 DS-CHD cases and 80 DS-without CHD (nominal *P*-value 0.0022). Furthermore, CNV analyses using a customized chromosome 21 aCGH of 135K probes in 55 DS-AVSD and 53 DS-without CHD revealed three CNV regions associated with AVSD risk (FDR ≤0.05). Two of these regions which are located within the previously identified CHD region on chromosome 21 were further confirmed in a replication study of 49 DS-AVSD and 45 DS- without CHD (FDR ≤0.05). One of these CNVs maps near the *RIPK4* gene, and the second includes the *ZNF295* gene, highlighting the potential role of these genes in the pathogenesis of CHD in DS. We propose that the genetic architecture of the CHD risk of DS is complex, and includes trisomy 21, SNP and CNV variations in chromosome 21. In addition, a yet unidentified genetic variation in the rest of the genome may contribute to this complex genetic architecture.

Introduction

Down syndrome (DS) is a common genomic disorder caused by trisomy of human chromosome 21 (Antonarakis 1998; Antonarakis et al. 2002; Antonarakis et al. 2004; Antonarakis and Epstein 2006). Some of its phenotypes (e.g. cognitive impairment) are consistently present in all DS individuals albeit in variable severity, while others show incomplete penetrance (Antonarakis et al. 2002; Antonarakis et al. 2004; Antonarakis and Epstein 2006). Amongst the most notable phenotypes with reduced penetrance are the congenital heart defects (CHD), since approximately 40% of DS individuals present some form of CHD (Ferencz et al. 1989). The most frequent forms of CHD in DS cases are atrioventricular septal defects (AVSD), comprising 43% of the CHD cases, while ventricular septal defects (VSD), atrial septal defects (ASD), and tetralogy of fallot (TOF) comprise 32%, 19%, and 6% of the CHD in DS, respectively. In fact, AVSDs are almost exclusively seen in DS (Ferencz et al. 1989; Roizen and Patterson 2003). The complete underlying genomic or gene expression variation that contributes to the presence of a CHD in DS is unknown. Extensive efforts over the past years to gain a better understanding of the genetic basis of CHD in DS using rare cases of partial trisomy 21 has led to identification of genomic regions on chromosome 21 that when triplicated are consistently associated with CHD. An initial study of rare partial trisomy 21 cases suggested a minimal CHD candidate region on 21q22.3 of approximately 5.27 Mb between markers *D21S3* and *PFKL* (Barlow et al. 2001); this region was later narrowed down to 1.77-Mb (*DSCAM-ZNF295*) (Korbel et al. 2009). In another similar study of partial trisomy 21 cases, the CHD region was mapped to a larger telomeric genomic segment of 15.4 Mb, that overlaps with the segment described above (Lyle et al. 2009). An association study in the mid-90's using a few microsatellite markers on chromosome

21 hinted at a potential association between variation of the *COL6A1* gene region with CHD in Down syndrome (Davies et al. 1995). Also, Grossman et al (Grossman et al. 2011) used *Drosophila*, mouse transgenesis and a cardiac myoblast H9C2 cell model to show that the over-expression of *DSCAM* and *COL6A2* genes cooperatively contribute to ASD in mice, increased abnormalities of heart rhythm and failure in *Drosophila*, and promoted substrate adhesion in H9C2 cell line. Recent studies also suggest the potential contribution of VEGF-A (Ackerman et al. 2012), ciliome and hedgehog (Ripoll et al. 2012), and folate (Locke et al. 2010) pathways to the pathogenicity of CHD in DS. There are also several mouse models for partial or complete trisomy syntenic to human chromosome 21 (Sago et al. 1998; Shinohara et al. 2001; Dunlevy et al. 2010; Yu et al. 2010). CHDs have been observed in the full "trisomy 21" homologous mice Tc1 (Dunlevy et al. 2010) and Dp(10)1Yey/+; Dp(16)1Yey/+; Dp(17)1Yey/+ mice (Yu et al. 2010). In addition, CHDs have been observed in the partial "trisomy 21" model Ts65Dn that is trisomic for 13.4 Mb of the 22.9 Mb chromosome 21 syntenic regions (Moore 2006; Williams et al. 2008). Also, recently engineered duplication of a 5.43 Mb region of Mmu16 from *TIAM1* to *KCNJ6* in the mouse model, Dp(16)2Yey, was reported to cause CHD (Liu et al. 2011).

Candidate non-chromosome 21 genes have been also identified for susceptibility to several CHDs and AVSD in particular (not related to DS). Pathogenic mutations in the *CREDL1* gene (on 3p25) have been found in 6% of individuals with non-trisomy 21-related AVSD (Robinson et al. 2003). Also, *GATA4* mutations (on 8p23) have been found in families with cardiac malformations that included AVSD, VSD, insufficiency of cardiac valves, ASD, and thickening of the pulmonary valve; the data in these families suggest that the same pathogenic mutation could predispose to different types of heart defects

in different individuals (Garg et al. 2003). The creation of transgenic mouse strains using cardiac-specific gene inactivation of hypomorphic *BMP4* alleles resulted in AVSDs that correlate with the level of *BMP4* expression (Jiao et al. 2003). Thus, it is conceivable that gene expression variation of certain loci could contribute to the phenotypic variation of heart defects in DS. Also, CNVs are important component of the overall genomic variability amongst individual genomes (Sharp et al. 2005; Beckmann et al. 2007). Rare and common CNVs have been associated with various phenotypes (Sharp et al. 2005; Redon et al. 2006; Beckmann et al. 2007; Conrad et al. 2010; Craddock et al. 2010; Priest et al. 2012) and possibly they could also be one cause of the CHD risk in DS.

Our current hypothesis for the CHD phenotypes in DS individuals is that three copies of functional genomic elements on chromosome 21 and genetic variation of chromosome 21 and non-chromosome 21 loci predispose to abnormal heart development. Additional variables for these phenotypes could include unknown environmental factors and stochastic events. Thus the CHD phenotypes are likely to be multifactorial, caused both by variation at multiple loci and interactions amongst them and with non-genetic factors. Here we aimed to contribute to the description of the genetic architecture of CHD in DS, and report the results of genome-wide and chromosome 21-specific SNP and CNV association studies using samples from DS individuals with and without CHD. We have also used SNPs that are associated with *cis*-eQTLs as a subset of the SNP space to test the hypothesis that sites controlling the quantitative difference in gene expression contribute to the development of CHD. Furthermore, genome-wide SNP-based interaction studies (GWIS) were attempted to further explore the complex genetic architecture of CHD in DS.

Results

Genome-wide association study of CHD and its sub-phenotypes in DS:

In order to identify genome-wide risk variants for the CHD in DS, we performed a case-control association study using 431,962 non-chromosome 21 SNPs. The cases were 187 samples of DS with various forms of CHD; the controls were 151 samples of DS without detectable CHD (see methods). Fig. 1 shows the "Manhattan plots" of the GWAS *P*-values for all combinations of heart defects (all CHD, AVSD alone, VSD alone, ASD alone). Our intent was to potentially detect GWAS significant signals for the different CHD sub-classes. Supplemental Table S1 summarizes the top *P*-values from all of these association studies. Only one signal remains marginally significant after Bonferroni correction (Bonferroni adjusted *P*-value 0.03), with marker rs160890 on chromosome 5 when only the ASD phenotypic class was considered. The *Q-Q* plots of SNP association test of *P*-values are shown in Supplemental Fig. S1.

Association study using chromosome 21 SNP genotypes

To identify genomic risk variants on chromosome 21 for CHD in DS we performed an association study using 7,238 chromosome 21 trisomic SNPs. For this special trisomic association test, we calculated allelic and genotypic *P*- values (comparing the frequency of the 4 genotypic classes AAA, AAB, ABB, and BBB) in cases versus controls. The Bonferroni correction for multiple testing was utilized based on 7,238 trisomic SNPs tested. The single locus association test for 187 DS with CHD and 151 DS without CHD showed rs2832616 (nominal genotypic $P=3.08\times10^{-6}$) and rs1943950 (nominal genotypic $P=6.83\times10^{-6}$) as CHD risk alleles (Table 1 and Fig. 2). Both SNPs are located in the same LD block ($r^2=1$) in the intergenic region between *GRIK1* and *CLDN17*. Interestingly, both SNPs are *cis*-eQTLs for the *KRTAP7-1* (Dimas et al. 2009; Yang et al. 2010).

In order to validate these results we genotyped the two risk CHD SNPs in a replication sample of 92 DS-CHD and 80 DS without CHD. The nominal genotypic *P*-value was 0.0022 for both SNPs (Table 1). Thus we consider that these two SNPs were validated in the replication sample. When the separate phenotypic classes of CHD were considered, no statistically significant associations were found, except for DS-ASD using allelic *P*-values (Table 1). However for this analysis genotypic *P*-values were not significant (Table 1). This result was not further investigated.

Two-locus interactions:

To test the hypothesis that there exist two loci interactions for CHD risk in DS we used logistic regression to model the interaction between any two SNPs. Three different models were studied: i) interactions among SNPs on the diploid fraction of the genome (i.e. excluding the chromosome 21), ii) interactions among SNPs on the trisomic fraction of the genome, iii) interactions between trisomic SNPs on chromosome 21 and non-chromosome 21 *cis*-eQTLs.

i) Interactions among SNPs on the diploid fraction of the genome:

Since DS is likely to be a disorder of gene expression (Prandini et al. 2007), two-locus interaction study was first performed using only *cis*-eQTL markers that are functionally related to gene expression variation. We used 8,900 SNPs with strong association with gene expression (empirical *P-value* < 1.0×10^{-4}) selected from the Geneva GenCord (Dimas et al. 2009) and HapMap3 CEU datasets (Stranger et al. 2012). After pruning the SNPs as described in the methods, 8,126 SNPs remained. The interaction analysis was done by the fast-epistasis option in PLINK. The results are summarized in Table 2. In

total 33,011,875 SNP×SNP tests were done and an uncorrected *P*-value of 6.0×10^{-9} was used to establish significance in our genome-wide *cis*-eQTL interaction study at the 0.05 level. Correction for multiple testing is done based on $n\times (n-1)/8$ where n is the number of SNPs tested (Becker et al. 2011). This *cis*-eQTL-SNP based interaction study for DS-CHD risk showed a significant interaction signal between rs972372 on chromosome 2 and rs681418 on chromosome 11 (*P*-value = 8.7×10^{-10}, Table 2 and Supplemental Fig. 2). rs972372 is an intronic variant in the *MAP4K4* gene and is associated with the expression of the *CNOT11* gene, while rs681418 is an intronic variant in the *SPA17* gene and is associated with the expression of the *NRGN* gene. This SNP-SNP interaction was tested in a replication study of 83 DS-CHD and 71 DS-without CHD and the nominal *P*-value is 0.047 (Table 2). This marginal significance does not justify a clear declaration of validation of the SNP-SNP interaction result.

Furthermore, two-locus interaction analysis was performed on the entire set of 431,962 SNPs genome-wide. An uncorrected *P*-value of 2.1×10^{-12} (Becker et al. 2011) was used to establish genome-wide significance at the 0.05 level. Supplemental Table S3 shows a summary of these analyses. None of the interacting pairs of SNPs become significant after correcting for multiple testing. Performing the same analysis after SNP LD pruning did not yield any significant signal (data not shown).

ii) Interactions amongst SNPs on the trisomic chromosome 21

For the study of interactions between SNPs on chromosome 21 contributing to CHD, three different tests were performed. First, all 7,238 genotyped SNPs on chromosome 21 were used, performing 26,190,703 interaction tests (*P*-value cutoff for 0.05 level genome-wide significance: 7.5×10^{-9}). Then, after LD SNP pruning, only 2,950 SNPs were used for

interaction analysis, performing 4,349,775 tests (*P*-value cutoff for 0.05 level genome-wide significance: 4.5×10^{-8}). And, finally, all chromosome 21 *cis*-eQTLs (166 *cis*-eQTLs) available from the Geneva GenCord and HapMap3 CEU datasets were used for interaction analysis based on *cis*-eQTLs. A total of 13,695 tests were performed (*P*-value cutoff for 0.05 level significance: 1.4×10^{-5}). None of these tests showed a significant result after correcting for multiple testing (data not shown).

iii) Interactions between SNPs on chromosome 21 and SNPs on the diploid fraction of the genome

An interaction analysis was performed between all chromosome 21 SNPs (7,238 trisomic SNPs) and *cis*-eQTLs present in the non-chromosome 21 fraction of the genome (8,126 SNPs) in order to identify possible interactions between trisomic and disomic regions of the genome contributing to CHD. None of the tested 32,209,100 SNP pair comparisons remained significant after adjusting for multiple testing (an uncorrected *P*-value of 6.2×10^{-9} was used to establish genome-wide significance at the 0.05 level) (data not shown).

Chromosome 21 CNV association analyses:

Copy number variation on the trisomic chromosome 21 could also contribute to the CHD risk in DS and particularly for the AVSD risk (Beckmann et al. 2007; de Smith et al. 2010). After calling CNVs on chromosome 21 (see methods), association tests using 55 DS-AVSD as cases and 53 DS-without CHD as controls were performed on 4,401 CNV tests obtained after performing intersection across all the samples. The *P*-values of association tests are shown in Fig. 3. CNV Association tests with FDR ≤0.05 or *SimpleM P*-value ≤0.05 were considered to be significant. 21 CNV tests out of 4,401 passed this threshold; constituting

three different CNV regions (Table 3). Both FDR and *SimpleM*-based multiple testing correction identified the same CNV regions as significantly associated. These CNV regions, termed here as CNV1, CNV2 and CNV3, are covered by 42, 20, and 62 consecutive probes in our customized NimbleGen CGH array respectively. All these consecutive probes for each CNV region showed the same copy number state.

CNV1 (Chr21: 42,066,443-42,071,313) (*P*-value=2.5×10^{-4}, FDR=0.05) is a 4.9-Kb CNV region located 6 Kb upstream of *RIPK4* (Fig. 3B and Table 3). It contains PBX3 and BCL3 transcription factor binding sites detected in different cell lines (Fig. 3B). Whereas deletions and duplications were found in 25% of DS-AVSDs, no such events were observed in the controls (Table 3). The risk ratio for this CNV is 2.29 (95% *CI*: 1.82-2.82) (Table 3). Moreover, CNV1 overlaps with an inversion reported in the database of genomic variants (Variation_37237) (DGV, http://projects.tcag.ca/variation/) (Fig. 3B). Also for CNV1, there is a SNP probe for rs13048349 which resides within this CNV region and shows the presence of variation in copy number in this region in GenCord cohort (Dimas et al, 2009) that are 215 healthy individuals genotyped by Illumina 550K SNP beadchips, and in our collection of Down syndrome individuals (Supplemental Fig. S3 C and D), while the adjacent SNP probe for rs13048349 outsides of this CNV region did not show any copy number variation (Supplemental Fig. S3 A and B).

CNV2 (Chr21:42,284,480-42,286,300) (*P*-value=1.9×10^{-4}, FDR=0.05) is a 1.8-Kb CNV region located within *ZNF295* including its final exon. Whereas deletions were found in 24% of DS-AVSDs, no deletion was observed in the controls (Table 3). The risk ratio for this CNV region is 1.85 (95% *CI*: 1.33-2.56) (Table 3). Interestingly both CNV1 and CNV2 fall within the 1.7 Mb CHD critical region defined before (Korbel et al. 2009). CNV3 (Chr21:45,541,600-45,555,054) (*P-value*=4.8×10^{-5}, FDR=0.05) is a 13.5-Kb region

located 9 Kb upstream of the *POFUT2* gene. CNV3 includes many transcription factor binding sites identified in a variety of cell lines (P300, CTCF, BCL3, PBX3, TAF1, CEBPB, HEY1) and a long non-coding RNA (*LINC00315*). While the frequency of deletion and duplication in DS-AVSDs is 26%, no such events were observed in the controls (Table 3). The risk ratio for this CNV is 2.26 (95% *CI*: 1.80-2.83) (Table 3). Moreover, CNV3 overlaps with a duplicated fragment (Variation_53268) reported in the DGV. Also all the three detected CNV regions overlap with much larger deletions and duplications reported in the DECIPHER database (https://decipher.sanger.ac.uk) associated with developmental delay (Fig. 3).

To detect the presence of these CNVs in the general (apparently-healthy) population, we used qPCR to see if CNV1, CNV2 and CNV3 are present in the Geneva GenCord cohort (62 samples). We observed duplication of CNV1, CNV2, and CNV3 with 9%, 14%, and 16% frequencies respectively (Supplemental Fig. S4), while no deletions were detected.

Furthermore, we correlated these CNV states detected by qPCR in GenCord cohort and gene expression of nearby genes based on mRNA-seq data available from the same data set (M Gutierrez-Arcelus, T Lappalainen, S Montgomery, A Buil, H Ongen, A Yurovsky, J Bryois, T Giger, L Romano, A Planchon, et al. 2013. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* (In press)) (Supplemental Fig. S5). CNV2, which resides in the final exon of *ZNF295*, correlated with increased expression of this gene compared to copy neutral state (one-sided Mann-Whitney *U*-test P=0.024) (Supplemental Fig. S5). CNV2 is also associated with the expression of *C2CD2* gene (one-sided Mann-Whitney *U*-test P=0.0029) which lies 108 Kb upstream. CNV3 also is associated with an increase in expression of the nearby gene

297 *COL18A1* (one-sided Mann-Whitney *U*-test *P*=0.045), which lies 158 Kb upstream
298 (Supplemental Fig. S5).

299 We subsequently tested the significance of these risk CNVs in a replication study of 49
300 DS-AVSD cases and 45 DS-without CHD controls using NanoString nCounter technology
301 (see methods). The one-sided Mann-Whitney *U*-test was employed to assess the
302 presence of copy number differences between cases and controls for the three detected
303 AVSD-DS risk CNVs. The presence of less copy number state for CNV1 in DS-AVSD cases
304 as compared to DS-without CHD is shown by two probes (FDR ≤0.05) and marginally by
305 additional two probes (FDR<0.1) out of 8 probes tested (Table 4). Also, the presence of
306 less copy number state for CNV2 in DS-AVSD as compared to DS-without CHD is shown
307 by 6 probes (FDR ≤0.05) out of 7 probes tested (Table 4). Thus we conclude that the
308 risk CNV1 and CNV2 were validated in a replication sample. For CNV3, none of the 11
309 probes tested show any significant signal.

310

311 **Discussion**

312 Recently significant progress has been made in the definition of regulatory pathways
313 that control normal and abnormal cardiac valve development that involves well-
314 regulated mechanisms for cell movements and cell–cell interactions amongst distinct
315 populations of heart precursor cells (Gumbiner 1996; Santiago-Martinez et al. 2006;
316 Totong et al. 2011). This suggests that molecules that regulate cell migration and cell-
317 cell adhesion, may contribute to CHD. The fact that the frequency of CHD in DS is much
318 higher than in normal euploid individuals raises the hypothesis that dosage sensitive
319 genes on chromosome 21 greatly increase the risk for CHD. However, the underlying

genomic or gene expression variation, and the allelic architecture that contributes to the manifestation of a CHD in DS are unknown.

In the current study, our hypothesis is that three copies of functional genomic elements on chromosome 21 and genetic variation of chromosome 21 and non-chromosome 21 loci predispose to CHD in DS. Moreover, in order to test the hypothesis that quantitative difference in gene expression contributes to the development of CHD, *cis*-eQTLs were used as a subset of the genomic variation space. Genome-wide interaction studies (GWIS) were attempted to further reveal the complex genetic architecture of CHD in DS. Furthermore, we hypothesized that chromosome 21 CNVs could also contribute to the increased risk of the CHD in DS.

These GWAS and GWIS were performed on a population of 187 DS individuals with CHD and 151 DS individuals without CHD. Although the sample size of this study is not large, power calculations (Purcell et al. 2003) suggest that we could likely identify loci with odds ratio >2. PCA analyses did not reveal any considerable population stratification (cases and controls showed similar distributions in the space defined by the first two principal components) and we did not formally detect any significant inflation factor (lambda=1.01) (Supplemental Figure S7).

The most notable results of this study regarding identifying CHD risk loci in DS that have been verified in our replication study as well were:

i) rs2832616 and rs1943950 within the same LD block on chromosome 21 (both *cis*-eQTLs for *KRTAP7-1* gene) as CHD risk alleles (odd ratio of 2.8 and 2.7 respectively).

ii) A 4.9-Kb CNV upstream of *RIPK4* gene (CNV1) in the previously reported CHD region of chromosome 21 with a risk ratio of 2.29.

iii) A 1.8-Kb CNV within *ZNF295* gene (CNV2) in the previously reported CHD region of chromosome 21 with a risk ratio of 1.85.

iv) A pair of interacting *cis*-eQTLs involving *CNOT11* on chromosome 2 and *NRGN* on chromosome 11 (Bonferroni adjusted *P*-value <0.05).

The chromosome 21 CNV analyses revealed three CNV regions, two of which (CNV1 and CNV2) were verified in an independent replication study and map within the previously reported CHD region of chromosome 21 in DS (Tables 3 and 4, and Fig. 3). The compound effect of trisomy 21 and the presence of less copy number state than the neutral state for CNV1 and CNV2 seem to be associated with the AVSD risk in DS. This highlights the importance of local copy number variation on chromosome 21 in the context of trisomy in the pathogenicity of AVSD in DS. CNV1, which resides 6 Kb upstream of *RIPK4*, contains two transcription factor binding sites (PBX3 and BCL3) and shows enrichment for H3K4Me1 detected in different cell lines (Fig. 3B) suggesting that this CNV region may contain regulatory elements for the nearby gene *RIPK4*. Previously our laboratory has shown an exceptional over-expression of *RIPK4* in the heart tissue of the Ts65Dn DS mouse model (Lyle et al. 2004). This, combined with the fact that *RIPK4* resides in the minimal critical CHD region on chromosome 21 (Korbel et al. 2009) suggests a possible important role in the pathogenicity of CHD in DS. CNV2, which also maps within the critical chromosome 21 CHD region, involves the last and largest exon of *ZNF295,* which encodes the functional portion of the protein (Fig. 3C). *ZNF295* which acts as a transcription repressor (Wang et al. 2005), has been shown to interact with *PPP2R2B* (Glatter et al. 2009) which in *Drosophila* regulates WNT/beta-catenin signaling

pathway. This pathway is required for cardiac differentiation in human embryonic stem cells (Paige et al. 2010).

Furthermore, our genome-wide two-locus interaction analysis based on whole genome *cis*-eQTLs for CHD shows a significant pair of interacting SNPs located in *MAP4K4* and *SPA17* genes (Table 2). Interaction analyses based on whole genome markers revealed no significant interacting pair of SNPs after adjusting for multiple testing (Supplemental Table S3). Analysis of gene ontology (GO) categories (Huang da et al. 2009) on the affected genes showed no significant enrichment (Supplemental Table S7).

In summary, the findings of this study reveal several candidate risk loci for CHD in DS. It is reasonable that many genes or other functional genomic elements may contribute to the development of CHD, since a very large number of genes and signaling pathways regulate heart development, and variation in all of these genes may have a contribution (minor or major) to the risk of CHD in DS. Our results support a multi-factorial model for the development of CHD, and a complex architecture of risk alleles with cumulative effects that explain the developmental phenotypes.

Material and Methods

Sample collection and Genotyping:

In order to identify the genetic components that confer susceptibility to the CHD phenotypes, a cohort of 187 DS patients with well characterized CHD phenotype (through ultrasonography, echocardiography and /or surgery) (AVSD, VSD, ASD), and 151 ethnically matched DS controls (DS without a CHD) was genotyped to perform genetic association studies. Informed consent was obtained for all samples, and the study was approved by the Geneva University Ethics Committee. All DS individuals

(cases and controls) are fully trisomic for the entire chromosome 21. The distribution of phenotypic attributes of 187 DS-CHD individuals is as follows: AVSD=69(37.0%), ASD=51(27.2%), VSD=67(35.8%) (Supplemental Table S2).

Genomic DNA used in our GWAS was extracted from blood. Genotyping of 750ng of genomic DNA was performed using Illumina 550K and 610K BeadChips according to the manufacturer's Infinium II protocol (Illumina, San Diego, USA).

Genomic DNA used in our replication study was extracted from blood, or lymphoblastoid cell lines. The distribution of phenotypic attributes of our replication samples of 92 DS-CHD individuals collected at the Cardiology Department of Hôpital Necker-Enfants malades (Paris, France) after obtaining informed consent is as follows: AVSD=64, ASD=7, VSD=16, and others=5. The 80 DS without CHD controls of replication study come from *Jérôme Lejeune* Foundation (Paris, France).

For quality control analyses, data from samples were used if a minimum of 98% call rate was observed for the sample in GenomeStudio data analysis software (Illumina, San Diego, USA) across all the SNPs common to both HumanHap550 and HumanHap610 BeadChips. Other criteria for SNP filtering were as follows: cluster separation >0.3, MAF> 0.04, HWE *P-value* >0.05, HET -0.2<0>0.2. Due to the complexity of SNP calling in trisomic regions of the genome, we removed chromosome 21 SNPs (trisomic SNPs) from this phase of analysis and called and analyzed them separately (see below). After filtering, subsequent analyses were performed on a dataset of 431,962 SNPs with the overall call rate of 99.997%. For linkage disequilibrium (LD) -based SNP pruning we used the option based on recursively removing SNPs within a sliding window of 50-SNPs

412 that shifts 5-SNPs along with each step, and a variance inflation factor (VIF) threshold of
413 2 (as recommended in the PLINK toolset).

414 *Chromosome 21 SNPs Calling:*

415 Clustering programs such as GenomeStudio assume a diploid state for each SNP. Since
416 there is no standard algorithm for calling triploid genomes, chromosome 21 SNPs were
417 not adequately called by the software; thus they were investigated separately and
418 genotypes called by an in-house procedure. In Genome-Studio software, we used the
419 filter option to choose only chromosome 21 SNPs (trisomic SNPs). Subsequently we
420 used the sample graph option in Genome-Studio to discriminate 4 clusters of SNPs by
421 plotting intensity (R) versus allele ratio (theta) for each individual. These four clusters
422 represented the homozygous AAA genotypes (samples with a low theta value),
423 heterozygous AAB genotypes (samples with an intermediate low theta value), ABB
424 genotypes (samples with an intermediate high theta value) and homozygous BBB
425 genotypes (samples with high theta values) (Supplemental Fig. S6). The boundaries for
426 each cluster were defined visually and we avoid including SNPs located between
427 boundaries of any two adjacent clusters. Then the SNP data for each cluster were
428 imported in an Excel spreadsheet and were assigned to the respective genotype (AAA,
429 AAB, ABB, and BBB) manually. This manual procedure was performed for calling
430 chromosome 21 trisomic SNPs for all the DS samples used in this study. All the DS
431 samples are fully trisomic for chromosome 21 and clearly showed four clusters of
432 genotypes in the GenomeStudio software for chromosome 21 SNPs. To evaluate the
433 quality of trisomic SNPs called by Illumina beadchips, we genotyped rs7282991, and
434 rs2183593 on chromosome 21 by Pyrosequencing for 96 DS samples that were also
435 genotyped by Illumina beadchips. We found 100% concordance for the genotypes called

436 by both methods. To check for trisomic SNPs that violate Hardy-Weinberg equilibrium
437 (HWE) in Down syndrome individuals we used a method previously developed
438 (Kerstann et al. 2004). HWE was tested using two by four Fisher's exact test (HWE *P*-
439 value >0.05). A total of 7238 SNPs out of 8251 chromosome 21 SNPs were retained.

440 Association analysis for non-chromosome 21 SNPs was performed by Fisher's exact test
441 of allele counts in DS individuals with CHD and its different classes (AVSD, ASD, and
442 VSD) and DS individuals without CHD, as implemented in PLINK toolset v0.99q software
443 (Purcell et al. 2007). Quantile-quantile plots (*Q-Q* plots) were generated by WGAViewer
444 (Ge et al. 2008) to detect inflation of statistics due to population stratification.
445 LocusZoom software was used for showing regional association plot (Pruim et al. 2010).
446 A logistic regression model provided in PLINK was applied for genome wide interaction
447 analyses based on the coefficient b3 in $Y \sim b0 + b1.A + b2.B + b3.AB + e$ which is a model
448 based on allele dosage for each SNP, A and B (Purcell et al. 2007).

449 For chromosome 21 trisomic SNPs association and interaction studies, custom scripts
450 implemented in R programming language were used. To test for possible differences in
451 disease susceptibility due to single and two-locus interaction, Fisher's exact test and
452 logistic regression model were applied, respectively. Logistic regression was also used
453 to assess the statistical interaction of trisomic SNPs on chromosome 21 with the disomic
454 SNPs (only *cis*-eQTLs) on the rest of the genome.

455

456 *Population stratification:*

457 Population stratification for the GWAS data was examined by principal component
458 analysis (PCA) using the EIGENSTRAT software v3.0 (Price et al. 2006). We also

459 estimated ancestry proportion by an identity-by-state-based (IBS) clustering method
460 implemented in the PLINK toolset. Briefly, 104,338 SNPs were used by the EIGENSTRAT
461 software and all available SNPs (n=431,962) by the PLINK IBS clustering method to
462 assess an underlying population structure (Supplemental Fig. S7). There was no
463 evidence of population stratification, as the genomic inflation factor calculated by both
464 EIGENSTRAT and PLINK is 1.01 and 1.02 respectively.

465

466 *SNP genotyping by Pyrosequencing and Sanger sequencing:*

467 Pyrosequencing was performed for the genotyping rs2832616, rs2183593, rs7282991
468 and rs9723772. Sanger sequencing was used for genotyping rs1943950 and rs681418
469 that we were not able to genotype them by Pyrosequencing.

470 The PCR primers were designed using Pyrosequencing assay design software ver. 1.0.6
471 (Supplemental Table S5 and S6). The pyrosequencing reactions were automatically
472 performed with a PSQ 96MA system (Pyrosequencing AB) according to the
473 manufacturer's instructions. All reactions were constructed as recommended by the
474 manufacturer's instructions (Pyrosequencing, Westborough, MA, USA). The PCR
475 conditions for pyrosequencing were 94°C for 5 minutes; (94°C for 30 seconds, 60°C for
476 30 seconds; *Tm* reduces ½ °C per cycle, and 72°C for 20 seconds) for 9 cycles; (94°C for
477 30 seconds, 55°C for 30 seconds, and 72°C for 20 seconds) for 29 cycles; 72°C for 5
478 minutes. The PCR conditions for Sanger sequencing were 94°C for 2 minutes; (94°C for
479 30 seconds, 55°C for 30 seconds, and 72°C for 45 seconds) for 26 cycles; 72°C for 5
480 minutes.

481 *Chromosome 21 CNVs analyses:*

To assess whether CNVs on chromosome 21 could predispose to CHD, we comprehensively searched both for known and novel CNVs on chromosome 21 using a custom CGH array. A subset of 55 DS with AVSD and 53 DS without CHD samples were compared in a case-control approach.

Array design: In order to detect CNVs potentially contributing to the CHD in DS, NimbleGen 12x135K custom arrays were used. These arrays were designed to screen human chromosome 21 for both known and de novo CNVs. High-density coverage (a mean of 1 probe per 100bp) was used for a 9 Mb region (chr21:38,000,000-46,944,323) that is known to be associated with a CHD in DS. 89,443 probes were printed for this region. In addition, to genotype 95 known chromosome 21 CNVs (Conrad et al. 2010) , a minimum of 30 probes per known CNV were printed on the arrays. Furthermore, 500 probes on the unique portions of chromosome X and chromosome Y were printed on each sub-array to act as internal controls in sex-mismatched hybridizations, and to allow detection of possible sample mixes in our cohort. The remaining euchromatic portion of chromosome 21 (chr21:13,260,001-37,999,999) was covered by the remaining 42,876 probes available on the Nimblegen 12plex 135K platform. As a reference DNA, a lymphoblastoid cell line derived from a DS individual without CHD was used for all analyses. The reference DNA is fully trisomic for chromosome 21 (Supplemental Fig. S8). Dye swap experiments and technical replicates (between and within arrays and sub-arrays) were performed to measure dye bias effects and to remove random errors introduced in the experiment.

Hybridization signals were extracted using Nimblescan software v2.5 to generate General Feature Format (GFF) files and SignalMap software v1.9.0.05 was used subsequently to visualize and analyze the data. Log_2 ratios were obtained at each probe

position and then were quantile normalized to correct for inter-array variability (Supplemental Fig. S9). Putative CNVs were called as chromosomal segments with unusually high or low $\log_2$ ratios of fluorescent intensity between the test DNA (DS with or without CHD) and reference DNA (a DS individual without CHD) using the genome alteration detection analysis (GADA) algorithm (Pique-Regi et al. 2008) using the options "-M 10 -T 12 -a 0.2".

CNVs calling by GADA were followed by intersection across all the individuals to obtain CNV intersected regions (CNV tests) (Valsesia et al. 2011). A total of 4,401 intersected CNV tests spanning the entire chromosome 21 were identified according to three classes (deletion, duplication and copy neutral). We applied a two by three Fisher's exact test to detect any association signal for 4,401 intersected CNV tests. CNV tests with FDR ≤0.05 were considered to be significant and contiguous CNV tests with FDR ≤0.05 were merged into CNV regions. Moreover, we used *simpleM* (Gao et al. 2008; Gao et al. 2010), a data-driven approach based on PCA that estimates the actual number of effective tests (M_{eff}) in association studies. Briefly, it computes eigenvalues from the pair-wise 4,401×4,401 CNV correlation matrix and then derives the M_{eff} using PCA. M_{eff} is then used in Bonferroni correction on the *P*-values resulted from two by three Fisher's exact test for 4,401 CNV tests.

NanoString CodeSet design

NanoString nCounter technology was used to verify the detected risk CNVs in a replication study of 49 DS-AVSD and 45 DS-without CHD following NanoString's recommended protocol for CNV analyses. The three detected DS-AVSD associated risk CNVs were included in the CodeSet in addition to six control regions selected by

NanoString Technologies (three regions on chromosome 21 and three regions on chromosome 6) and additional two control regions on chromosome 21 which based on our previous Nimblegen CGH arrays did not show any copy number variation between cases and controls. CNV1, CNV2, and CNV3 were covered by 8, 7, and 11 probes respectively. The CodeSet also contains a set of 10 probes called invariant probes designed against autosomal genomic regions that are not copy number variant. The invariant probes were used for normalization purposes. Moreover, the CodeSet includes six positive dsDNA control probes, each targeting a unique DNA sequence present in every assay. The positive control probes were used for technical normalization. Probe sets for each CNV region in the CodeSet were designed and synthesized at NanoString Technologies. All procedures related to DNA quantification, hybridization, detection, and scanning were carried out as recommended by NanoString Technologies at the Genomics platform at the University of Geneva.

NanoString data processing started with a technical normalization of raw NanoString counts for each designed probe using the counts obtained for positive control probe sets prior to normalization using invariant probe set. All normalization procedures were performed by nSolver software v1.1 (NanoString Technologies,Inc). Finally we estimated copy number ratio for each probe relative to the reference sample. One-sided Mann–Whitney *U*-test were employed to compare fold change ratio for each probe between cases and controls. The *P*-value related to each probe was FDR corrected based on the total 35 probes tested.

Quantitative Polymerase Chain Reaction:

Since CNV risk alleles for AVSD in DS should be also common CNVs in normal healthy individuals (healthy non-Down syndrome individuals), quantitative polymerase chain

553 reaction (qPCR) was used to verify their frequency in the normal population. For this
554 purpose we used 62 DNA samples from the Geneva GenCord collection (Dimas et al,
555 2009). The primers designed for each risk CNV plus a control region are shown in Table
556 S4. Primer pairs were tested and efficiencies were measured using standard curves
557 from serial dilutions of genomic DNA. qPCR was performed using the SYBR Green® PCR
558 kit with a standard protocol, in conjunction with the 7900HT Real-time PCR system from
559 Applied Biosystems (ABI, Foster City, California, USA). Each reaction was carried out in
560 triplicate in 384-well plates in a total volume of 10 µl containing 1X ABI CyberGreen
561 Master mix, 0.9 mM each primer and with ~2 ng of DNA. Liquid handling for the 384-
562 well plates was performed with a Biomek robot (Beckman Coulter). The thermal profile
563 recommended by Applied Biosystems was used for amplification (50°C for 2 min, 95°C
564 for 10 min, 40 cycles of 95°C for 15 s and 60°C for 1 min). The generated data were
565 analyzed with SDS 2.2 software (Applied Biosystems). The quantification was carried
566 out using a pair of primers for the *ADARB1* gene on chromosome 21 as a reference gene.
567 For gene copy number assignment, the CT values for each set of triplicates were
568 averaged and normalized against the control primers for the reference gene. The
569 relative copy number for each CNV was calculated as described before (Livak and
570 Schmittgen 2001). We applied a Gaussian Mixture Modeling (GMM) method (Valsesia et
571 al. 2012) to detect copy number variation from the distribution of copy number ratios in
572 an unsupervised approach. We also performed Mann-Whitney *U*-test between these CNV
573 states and gene expression of nearby genes (Supplemental Fig. S5). The gene expression
574 data based on mRNA-Seq for the Geneva GenCord cohort (M Gutierrez-Arcelus, T
575 Lappalainen, S Montgomery, A Buil, H Ongen, A Yurovsky, J Bryois, T Giger, L Romano, A
576 Planchon, et al. 2013. Passive and active DNA methylation and the interplay with genetic

variation in gene regulation. eLife (In press)) are available for primary fibroblast cells, T-cells, and lymphoblastoid cell lines (LCLs).

Genome annotation:

Human genome hg18/NCBI 36 was used as our reference.

Data access

The genotyping and chromosome 21 CNV data from this study have been submitted to the EMBL-EBI European Genome-phenome Archive (http://www.ebi.ac.uk/ega/) under accession no. EGAS00000000129.

Acknowledgments

The authors gratefully acknowledge the individuals and families for their participation. We thank the Genomics Platform of the NCCR program "*Frontiers in Genetics*" for their help in performing qPCR, NanoString and SNP genotyping experiments. We are also grateful to E.T. Dermitzakis and M. Gutierrez-Arcelus for making available the Geneva GenCord data. We also thank F. Bena, S. Gimelli, E. Stathaki, Y. Dupre and M. Gagnebin who assisted in our aCGH experiments. We also thank V. Salle at the Hôpital Necker-Enfants maladies for assisting in our replication sample recruitment. The computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics. Also our special thanks to the CRB-BioJel center (http://www.crb-institutlejeune.com/) for their contribution in our sample collection. The study was supported by grants from the NCCR-*Frontiers in Genetics*, the European AnEuploidy project, the Fondation Child Care, the SNF 144082, and the ERC 249968 to S.E.A. and the Spanish Ministry of Ecomomy and Competitivity to

X.E. P.M was supported by a grant from the Bodossaki foundation. K.P was supported by EMBO long-term fellowship program ALTF 527-2010.

Disclosure declaration

None of the authors have any competing interests.

References

Ackerman C, Locke AE, Feingold E, Reshey B, Espana K, Thusberg J, Mooney S, Bean LJ, Dooley KJ, Cua CL et al. 2012. An excess of deleterious variants in VEGF-A pathway genes in Down-syndrome-associated atrioventricular septal defects. *Am J Hum Genet* **91**(4): 646-659.

Antonarakis SE. 1998. 10 years of Genomics, chromosome 21, and Down syndrome. *Genomics* **51**(1): 1-16.

Antonarakis SE, Epstein CJ. 2006. The challenge of Down syndrome. *Trends Mol Med* **12**(10): 473-479.

Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. 2004. Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat Rev Genet* **5**(10): 725-738.

Antonarakis SE, Lyle R, Deutsch S, Reymond A. 2002. Chromosome 21: a small land of fascinating disorders with unknown pathophysiology. *Int J Dev Biol* **46**(1): 89-96.

Barlow GM, Chen XN, Shi ZY, Lyons GE, Kurnit DM, Celle L, Spinner NB, Zackai E, Pettenati MJ, Van Riper AJ et al. 2001. Down syndrome congenital heart disease: a narrowed region and a candidate gene. *Genet Med* **3**(2): 91-101.

Becker T, Herold C, Meesters C, Mattheisen M, Baur MP. 2011. Significance levels in genome-wide interaction analysis (GWIA). *Ann Hum Genet* **75**(1): 29-35.

Beckmann JS, Estivill X, Antonarakis SE. 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* **8**(8): 639-646.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P et al. 2010. Origins and functional impact of copy number variation in the human genome. *Nature* **464**(7289): 704-712.

Craddock N Hurles ME Cardin N Pearson RD Plagnol V Robson S Vukcevic D Barnes C Conrad DF Giannoulatou E et al. 2010. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**(7289): 713-720.

Davies GE, Howard CM, Farrer MJ, Coleman MM, Bennett LB, Cullen LM, Wyse RK, Burn J, Williamson R, Kessling AM. 1995. Genetic variation in the COL6A1 region is associated with congenital heart defects in trisomy 21 (Down's syndrome). *Ann Hum Genet* **59**(Pt 3): 253-269.

de Smith AJ, Trewick AL, Blakemore AI. 2010. Implications of copy number variation in people with chromosomal abnormalities: potential for greater variation in copy number state may contribute to variability of phenotype. *HUGO J* **4**(1-4): 1-9.

Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, Ingle C, Beazley C, Gutierrez Arcelus M, Sekowska M et al. 2009. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**(5945): 1246-1250.

Dunlevy L, Bennett M, Slender A, Lana-Elola E, Tybulewicz VL, Fisher EM, Mohun T. 2010. Down's syndrome-like cardiac developmental defects in embryos of the transchromosomic Tc1 mouse. *Cardiovasc Res* **88**(2): 287-295.

640 Ferencz C, Neill CA, Boughman JA, Rubin JD, Brenner JI, Perry LW. 1989. Congenital cardiovascular
641 malformations associated with chromosome abnormalities: an epidemiologic study. *J Pediatr*
642 **114**(1): 79-86.
643 Gao X, Becker LC, Becker DM, Starmer JD, Province MA. 2010. Avoiding the high Bonferroni penalty
644 in genome-wide association studies. *Genet Epidemiol* **34**(1): 100-105.
645 Gao X, Starmer J, Martin ER. 2008. A multiple testing correction method for genetic association
646 studies using correlated single nucleotide polymorphisms. *Genet Epidemiol* **32**(4): 361-369.
647 Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-
648 Yamada K, Joo K et al. 2003. GATA4 mutations cause human congenital heart defects and
649 reveal an interaction with TBX5. *Nature* **424**(6947): 443-447.
650 Ge D, Zhang K, Need AC, Martin O, Fellay J, Urban TJ, Telenti A, Goldstein DB. 2008. WGAViewer:
651 software for genomic annotation of whole genome association studies. *Genome Res* **18**(4):
652 640-643.
653 Glatter T, Wepf A, Aebersold R, Gstaiger M. 2009. An integrated workflow for charting the human
654 interaction proteome: insights into the PP2A system. *Molecular systems biology* **5**: 237.
655 Grossman TR, Gamliel A, Wessells RJ, Taghli-Lamallem O, Jepsen K, Ocorr K, Korenberg JR, Peterson
656 KL, Rosenfeld MG, Bodmer R et al. 2011. Over-expression of DSCAM and COL6A2
657 cooperatively generates congenital heart defects. *PLoS Genet* **7**(11): e1002344.
658 Gumbiner BM. 1996. Cell adhesion: the molecular basis of tissue architecture and morphogenesis.
659 *Cell* **84**(3): 345-357.
660 Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists
661 using DAVID bioinformatics resources. *Nat Protoc* **4**(1): 44-57.
662 Jiao K, Kulessa H, Tompkins K, Zhou Y, Batts L, Baldwin HS, Hogan BL. 2003. An essential role of Bmp4
663 in the atrioventricular septation of the mouse heart. *Genes Dev* **17**(19): 2362-2367.
664 Kerstann KF, Feingold E, Freeman SB, Bean LJ, Pyatt R, Tinker S, Jewel AH, Capone G, Sherman SL.
665 2004. Linkage disequilibrium mapping in trisomic populations: analytical approaches and an
666 application to congenital heart defects in Down syndrome. *Genet Epidemiol* **27**(3): 240-251.
667 Korbel JO, Tirosh-Wagner T, Urban AE, Chen XN, Kasowski M, Dai L, Grubert F, Erdman C, Gao MC,
668 Lange K et al. 2009. The genetic architecture of Down syndrome phenotypes revealed by
669 high-resolution analysis of human segmental trisomies. *Proc Natl Acad Sci* **106**(29): 12031-
670 12036.
671 Liu C, Morishima M, Yu T, Matsui S, Zhang L, Fu D, Pao A, Costa AC, Gardiner KJ, Cowell JK et al. 2011.
672 Genetic analysis of Down syndrome-associated heart defects in mice. *Hum Genet* **130**(5):
673 623-632.
674 Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative
675 PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**(4): 402-408.
676 Locke AE, Dooley KJ, Tinker SW, Cheong SY, Feingold E, Allen EG, Freeman SB, Torfs CP, Cua CL,
677 Epstein MP et al. 2010. Variation in folate pathway genes contributes to risk of congenital
678 heart defects among individuals with Down syndrome. *Genet Epidemiol* **34**(6): 613-623.
679 Lyle R, Bena F, Gagos S, Gehrig C, Lopez G, Schinzel A, Lespinasse J, Bottani A, Dahoun S, Taine L et al.
680 2009. Genotype-phenotype correlations in Down syndrome identified by array CGH in 30
681 cases of partial trisomy and partial monosomy chromosome 21. *Eur J Hum Genet* **17**(4): 454-
682 466.
683 Lyle R, Gehrig C, Neergaard-Henrichsen C, Deutsch S, Antonarakis SE. 2004. Gene expression from
684 the aneuploid chromosome in a trisomy mouse model of down syndrome. *Genome Res*
685 **14**(7): 1268-1274.
686 Moore CS. 2006. Postnatal lethality and cardiac anomalies in the Ts65Dn Down syndrome mouse
687 model. *Mamm Genome* **17**(10): 1005-1012.

688 Paige SL, Osugi T, Afanasiev OK, Pabon L, Reinecke H, Murry CE. 2010. Endogenous Wnt/beta-catenin
689 signaling is required for cardiac differentiation in human embryonic stem cells. *PloS one* **5**(6):
690 e11134.
691 Pique-Regi R, Monso-Varona J, Ortega A, Seeger RC, Triche TJ, Asgharzadeh S. 2008. Sparse
692 representation and Bayesian detection of genome copy number alterations from microarray
693 data. *Bioinformatics* **24**(3): 309-318.
694 Prandini P, Deutsch S, Lyle R, Gagnebin M, Delucinge Vivier C, Delorenzi M, Gehrig C, Descombes P,
695 Sherman S, Dagna Bricarelli F et al. 2007. Natural gene-expression variation in Down
696 syndrome modulates the outcome of gene-dosage imbalance. *Am J Hum Genet* **81**(2): 252-
697 263.
698 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components
699 analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**(8): 904-
700 909.
701 Priest JR, Girirajan S, Vu TH, Olson A, Eichler EE, Portman MA. 2012. Rare copy number variants in
702 isolated sporadic and syndromic atrioventricular septal defects. *American journal of medical*
703 *genetics Part A* **158A**(6): 1279-1284.
704 Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ.
705 2010. LocusZoom: regional visualization of genome-wide association scan results.
706 *Bioinformatics* **26**(18): 2336-2337.
707 Purcell S, Cherny SS, Sham PC. 2003. Genetic Power Calculator: design of linkage and association
708 genetic mapping studies of complex traits. *Bioinformatics* **19**(1): 149-150.
709 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI,
710 Daly MJ et al. 2007. PLINK: a tool set for whole-genome association and population-based
711 linkage analyses. *Am J Hum Genet* **81**(3): 559-575.
712 Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen
713 W et al. 2006. Global variation in copy number in the human genome. *Nature* **444**(7118):
714 444-454.
715 Ripoll C, Rivals I, Ait Yahya-Graison E, Dauphinot L, Paly E, Mircher C, Ravel A, Grattau Y, Blehaut H,
716 Megarbane A et al. 2012. Molecular signatures of cardiac defects in Down syndrome
717 lymphoblastoid cell lines suggest altered ciliome and Hedgehog pathways. *PloS one* **7**(8):
718 e41616.
719 Robinson SW, Morris CD, Goldmuntz E, Reller MD, Jones MA, Steiner RD, Maslen CL. 2003. Missense
720 mutations in CRELD1 are associated with cardiac atrioventricular septal defects. *Am J Hum*
721 *Genet* **72**(4): 1047-1052.
722 Roizen NJ, Patterson D. 2003. Down's syndrome. *Lancet* **361**(9365): 1281-1289.
723 Sago H, Carlson EJ, Smith DJ, Kilbridge J, Rubin EM, Mobley WC, Epstein CJ, Huang TT. 1998. Ts1Cje, a
724 partial trisomy 16 mouse model for Down syndrome, exhibits learning and behavioral
725 abnormalities. *Proc Natl Acad Sci* **95**(11): 6256-6261.
726 Santiago-Martinez E, Soplop NH, Kramer SG. 2006. Lateral positioning at the dorsal midline: Slit and
727 Roundabout receptors guide Drosophila heart cell migration. *Proc Natl Acad Sci* **103**(33):
728 12441-12446.
729 Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S,
730 Segraves R et al. 2005. Segmental duplications and copy-number variation in the human
731 genome. *Am J Hum Genet* **77**(1): 78-88.
732 Shinohara T, Tomizuka K, Miyabara S, Takehara S, Kazuki Y, Inoue J, Katoh M, Nakane H, Iino A,
733 Ohguma A et al. 2001. Mice containing a human chromosome 21 model behavioral
734 impairment and cardiac anomalies of Down's syndrome. *Hum Mol Genet* **10**(11): 1163-1175.
735 Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D,
736 Gutierrez-Arcelus M et al. 2012. Patterns of cis regulatory variation in diverse human
737 populations. *PLoS Genet* **8**(4): e1002639.

Totong R, Schell T, Lescroart F, Ryckebusch L, Lin YF, Zygmunt T, Herwig L, Krudewig A, Gershoony D, Belting HG et al. 2011. The novel transmembrane protein Tmem2 is essential for coordination of myocardial and endocardial morphogenesis. *Development* **138**(19): 4199-4205.

Valsesia A, Rimoldi D, Martinet D, Ibberson M, Benaglio P, Quadroni M, Waridel P, Gaillard M, Pidoux M, Rapin B et al. 2011. Network-guided analysis of genes with altered somatic copy number and gene expression reveals pathways commonly perturbed in metastatic melanoma. *PloS one* **6**(4): e18369.

Valsesia A, Stevenson BJ, Waterworth DM, Mooser V, Vollenweider P, Waeber G, Jongeneel CV, Beckmann JS, Kutalik Z, Bergmann S. 2012. Identification and validation of copy number variants using SNP genotyping arrays from a large clinical cohort. *BMC genomics* **13**(1): 241.

Wang J, Kudoh J, Takayanagi A, Shimizu N. 2005. Novel human BTB/POZ domain-containing zinc finger protein ZNF295 is directly associated with ZFP161. *Biochem Biophys Res Commun* **327**(2): 615-627.

Williams AD, Mjaatvedt CH, Moore CS. 2008. Characterization of the cardiac phenotype in neonatal Ts65Dn mice. *Dev Dyn* **237**(2): 426-435.

Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, Deloukas P, Dermitzakis ET. 2010. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* **26**(19): 2474-2476.

Yu T, Li Z, Jia Z, Clapcote SJ, Liu C, Li S, Asrar S, Pao A, Chen R, Fan N et al. 2010. A mouse model of Down syndrome trisomic for all human chromosome 21 syntenic regions. *Hum Mol Genet* **19**(14): 2780-2791.

Figure Legends

Figure 1. Genome-wide Manhattan plots for CHD in DS and its different sub-phenotypes across 431,962 SNPs based on allelic associations. $-\log_{10} P$ values of SNP association tests are plotted relative to their position on each chromosome (alternating black and gray). Chromosome 21 trisomic SNPs are not included here (see text and Fig. 2 for details). The results shown are for **A)** DS-CHD, **B)** DS-AVSD, **C)** DS- ASD, and **D)** DS- VSD.

Figure 2. A) Chromosome 21-wide Manhattan plot and **B)** *Q-Q* plot of SNP genotypic association test *P*-values for 187 DS-CHD and 151 DS-without CHD using 7,238 SNPs on chromosome 21. The red line represents the Bonferroni threshold for chromosome 21-

772 wide α = 0.05. SNPs are plotted in Mb relative to their position on chromosome 21. Two
773 SNPs within the same LD block (r^2=1) reached chromosome 21-wide significance ($P \leq$
774 0.05) (see Table 1 for details). **C)** Regional association plot for the region identified to
775 associate with DS-CHD. The panel shows the recombination rate in the region estimated
776 from HapMap CEU data (http://hapmap.ncbi.nlm.nih.gov/), pairwise LD between SNPs
777 in the region and the SNP identified (labeled in purple), *P*-values for strength of
778 associations and genes in the region. The r^2 values are color coded according to the
779 scale on the panel.

780 **Figure 3. A)** Chromosome 21-wide Manhattan plot of *P*-values for DS-AVSD across 4,401
781 consecutive chromosomal regions for CNV association. The *P*-values are calculated by
782 two by three Fisher's exact test. The blue line represents the FDR threshold for
783 chromosome 21-wide α ≤ 0.05 and CNV tests are plotted in equidistance. Three CNV
784 regions reached genome-wide significance (FDR ≤0.05). **B)** Overview of the CNV1 region
785 (Chr21: 42,066,443-42,071,313) 6 Kbp upstream of *RIPK4* gene. This 4,870 bp CNV
786 region (in gray) is defined by merging 6 contiguous CNV tests (in black, see text for
787 details). CTCF, NSRF and other transcription factors binding sites are present in this
788 region as well as the histone mark H3K4Me1 (data from
789 http://genome.ucsc.edu/ENCODE/). Additionally an inversion (in pink), reported in the
790 database of genomic variation, overlaps with this CNV region. **C)** Overview of the 1,820
791 bp CNV2 region (Chr21:42,284,480-42,286,300) defined by merging two contiguous
792 CNV tests overlapping with the last exon of *ZNF295* gene.

793

Figure S1. *Q-Q* plots of the SNP association test statistic for CHD and its different sub-phenotypes. **A)** For DS-CHD, B**)** for DS-AVSD, **C)** for DS-VSD, and **D)** for DS-ASD. The lower horizontal red line represents the 90 th percentile, while the upper one indicates the observed values lifting off the expected.

Figure S2. *Q-Q* plot of whole-genome *cis*-eQTLs interaction study *P*-values.

Figure S3. Presence of CNV1 in healthy population (GenCord cohort) and Down syndrome individuals shown by Illumine SNP Probe. SNP probe intensity (R) is plotted against Theta value for rs12482855 located outside of the CNV1 and rs13048349 located inside the CNV1 in **A)** and **B)** normal individuals from GenCord cohort and **C)** and **D)** in Down syndrome individuals. Probes for the SNP inside CNV1 clearly aggregate into four clusters in normal individuals and five clusters in Down syndrome individuals. Probes for the nearest SNP outside of the CNV1 form three and four clusters in normal and Down syndrome individuals respectively.

Figure S4. Dosage copy number obtained by the GMM model is plotted against qPCR log2 ratios for **A)** Control region (chr21:41,125,388-41,125,571), **B)** CNV1 (chr21:42,068,711-42,068,995), **C)** CNV2 (chr21:42,284,480-42,284,663), and D) CNV3 (chr21:45,543,908-45,544,146). The copy number states are shown on the top of each plot.

Figure S5. Box-plots of gene expression values (RPKM) and copy number states based on one-sided Mann-Whitney *U*-test. GenCord individuals were divided in two groups based on two or three copy number states for each CNV and RPKM was calculated for each gene of interest. For all the box-plots the width of box is proportional to the amount of data within that group.

Figure S6. Chromosome 21 SNPs calling. We discriminated 4 clusters representing the genotypes AAA, AAB, ABB, and BBB by plotting probe intensity (R) versus allele ratio (theta) for each individual. Similar plots have been prepared for all the DS samples.

Figure S7. **A)** Multi dimensional scaling (MDS) plot (Components 1 and 2) for the DS with and without CHD and comparison with other samples including GenCord dataset (see Dimas et al, 2009), and Hapmap phase2 samples (CHB/JPT and YRI) (http://hapmap.ncbi.nlm.nih.gov/) **B)** Individual loadings on principal components 1 and 2 (PC1, PC2) for EIGENSTRAT analysis for the DS with and without CHD and comparison with other samples including Hapmap phase2 samples (CHB/JPT and YRI). **C)** PC1 and PC2 for EIGENSTRAT analysis for the DS with and without CHD, without out groups. **D)** *Q*-*Q* plot of *P*-values of all DS individuals with CHD versus DS individuals without CHD. Genomic inflation factor calculated by EIGENSTRAT is 1.01 and by PLINK is 1.02.

Figure S8. Array CGH result using a 44K Agilent CGH array illustrating the trisomic state for the entire long arm of chromosome 21 in the lymphoblastoid cell line derived from a DS individual without CHD. The DNA from this cell line is used as the reference DNA in the chromosome 21 CNV experiment.

Figure S9. Box- plots of all 135,679 CNV probes $\log_2$ ratios per sample **A)** before and **B)** after quantile normalization.

Table 1. Significant chromosome 21 trisomic SNPs association test results for DS-CHD and DS -ASD.

SNP	*BP.*	Alleles	Closest gene	*cis*-eQTL	Cohort	*OR (95% CI)*	Genotypic *P*	*Bonf1.*	Allelic *P*	*Bonf2.*
rs2832616	30,395,663	C/**T**	*GRIK1* and *CLDN17*	*KRTAP7-1*	*S*1	2.8(1.9-4.2)	3.1×10^{-6}	0.0224	1.1×10^{-6}	0.009
rs1943950	30,401,723	C/**T**	*GRIK1* and *CLDN17*	*KRTAP7-1*	*S*1	2.7(1.8-4.1)	6.8×10^{-6}	0.0492	3.3×10^{-6}	0.024
rs2832616	30,395,663	C/**T**	*GRIK1* and *CLDN17*	*KRTAP7-1*	*S*2	1.4(0.9-2.1)	0.0022		0.0052	
rs1943950	30,401,723	C/**T**	*GRIK1* and *CLDN17*	*KRTAP7-1*	*S*2	1.4(0.9-2.1)	0.0022		0.0052	
rs2183593	45,616,280	C/**T**	*LINC00315* and *COL18A1*	*ADARB1*	ASD	4.0(2.3-7.0)	2.57×10^{-5}	0.18	1.3×10^{-07}	0.0009
rs7282991	45,605,451	C/**T**	*LINC00315* and *COL18A1*	*ADARB1*	ASD	3.5(2.0-5.8)	9.18×10^{-5}	0.66	1.3×10^{-7}	0.01

BP, base position. Alleles, minor allele is shown in bold. Cohort, *S*1 represents the initial samples of 187 DS-CHD (cases) and 151 DS- without CHD (controls) and *S*2 represents samples from the replication study (92 DS-CHD and 80 DS-without CHD), ASD represents 53 DS-ASD samples and 151 DS- without CHD (controls) . *OR*, odd ratio (95% confidence interval). Genotypic *P*, unadjusted genotypic *P*-value , Allelic *P*, unadjusted allelic *P*-value . *Bonf*1. and 2., Bonferroni corrected *P*-value for the number of SNPs (n = 7,238) tested for genotypic and allelic tests respectively. rs2183593 and rs7282991 (ASD associated risk SNPs) were not validated in the replication study, because of lack of enough ASD samples in the replication cohort.

Table 2. *P*-values of the top three pairs of potentially interacting cis-eQTLs for DS CHD.

SNP1_*Chr.*	SNP location	*cis*-eQTL	SNP2_*Chr.*	SNP location	*cis*-eQTL	Cohort	*P*-value	*Bonf.*
rs972372_2	*MAP4K4*	*CNOT11*	**rs681418_11**	*SPA17*	*NRGN*	*S*1	**8.70×10^{-10}**	**0.007**
rs1815333_13	Intergenic	*ATP8A2*	rs1539871_18	Intergenic	*SMAD2*	*S*1	1.80×10^{-08}	0.149
rs848212_1	*SPEN*	*CLCNKA* and *TMEM51*	rs11191692_10	*PDCD11*	*USMG5* and *FAM26B*	*S*1	9.80×10^{-08}	0.809
rs972372_2	*MAP4K4*	*CNOT11*	**rs681418_11**	*SPA17*	*NRGN*	*S*2	**0.047**	

Cohort, *S*1 represents the initial samples of 187 DS-CHD (cases) and 151 DS-without CHD (controls), and *S*2 represents samples from replication study (83 DS-CHD and 71 DS-without CHD) . *Bonf.* is adjusted *P*-values for multiple testing. Shown in bold are statistically significant pairs. Interaction is based on 8,126 genome-wide *cis*-eQTLs from Geneva GenCord and HapMap3 CEU datasets.

Table 3. Candidate chromosome 21 CNV regions and the frequency of deletion (*Del.*), duplication (*Dup.*) and copy neutral amongst DS-AVSD (n=55) and DS without CHD (n=53).

			DS without CHD(n=53)			DS with AVSD(n=55)						
CNV region	**Size(bp)**	**Location**	***Del.***	***Dup.***	**Neutral**	***Del.***	***Dup.***	**Neutral**	***RR* (95%*CI*)**	***P*-value range**	**FDR range**	***SimpleM***
Chr21:42,066,443 -42,071,313	4870	6 Kb upstream of *RIPK4*	0%	0%	100%	18%	7%	75%	2.29 (1.82-2.82)	0.0002-0.00002	0.03-0.05	0.002-0.02
Chr21:42,284,480 -42,286,300	1820	*ZNF295*	0%	11%	89%	24%	14%	62%	1.85 (1.33-2.56)	0.0001-0.0002	0.04-0.05	0.01-0.02
Chr21:45,541,600 -45,555,054	13454	9 Kb upstream of *POFUT2*	0%	0%	100%	12%	14%	84%	2.26 (1.80-2.83)	0.0002-0.00005	0.03-0.05	0.005-0.02

P-value range, FDR range, and *SimpleM* columns show the lowest and the highest *P*-value, FDR, and *SimpleM* Bonferroni adjusted *P*-value for CNV tests in each CNV region respectively. *RR*, Risk ratio and 95% confidence interval (*CI*) are calculated based on combination of deletions and duplications versus copy neutral. *P*-values are calculated based on two by three Fisher's exact test for each CNV test. Contiguous CNV tests with FDR ≤0.05 were merged into CNV regions (see the text for more details).

886 **Table 4.** NanoString CNV probe association test results in the replication study.

Probe	Probe Coordinate	*P*	FDR_*BH*
Control regions			
Chr6 control 1	chr6:12915005-12915104	0.158	0.34
Chr6 control 2	chr6:47278334-47278433	0.261	0.48
Chr6 control 3	chr6:151086524-151086623	0.040	0.10
Chr21 control 1	chr21:17132182-17132276	0.624	0.78
Chr21 control 2	chr21:24809678-24809777	0.913	0.96
Chr21 control 3	chr21:32970669-32970768	0.805	0.94
Chr21-aCGH control 1	chr21:40187019-40187108	0.440	0.65
Chr21-aCGH control 2	chr21:41125716-41125805	0.848	0.95
CNV 1			
probe 2	chr21:42068170-42068245	0.001	0.02
probe 3	chr21:42068733-42068812	0.017	0.05
probe 6	chr21:42071172-42071242	0.036	0.08
Probe 8	chr21:42071172-42071242	0.039	0.08
CNV 2			
probe 1	chr21:42284095-42284184	0.014	0.05
probe 2	chr21:42284470-42284542	0.018	0.05
probe 4	chr21:42285199-42285288	0.001	0.02
probe 5	chr21:42285633-42285722	0.004	0.02
probe 6	chr21:42286068-42286157	0.016	0.05
probe 7	chr21:42286320-42286409	0.008	0.04

887 *P,* nominal *P*-value . FDR-*BH*, False discovery rate control by the Benjamini–Hochberg
888 procedure. Chr6 control 1,2 and 3, and Chr21 control 1,2 and 3 are Nanostring
889 technologies control regions. Chr21-aCGH control 1 and 2 are control regions that did
890 not show any copy number changes in our initial Nimblegen CGH array experiment
891 between cases and controls.

892

893

894 **Table S1.** Top SNP association results for DS-CHD and its sub-phenotypes.

	Chr.	SNP	*P-value*	*Bonf.*	*FDR*	*OR*
DS-CHD	6	rs2983865	2.43E-06	1	0.3424	0.4729
	18	rs1968470	3.05E-06	1	0.3424	2.099
	18	rs2852749	3.37E-06	1	0.3424	0.3576
	14	rs11621241	3.49E-06	1	0.3424	3.997
	6	rs2983879	3.96E-06	1	0.3424	0.4824
	14	rs17119691	5.24E-06	1	0.3703	3.915
	7	rs1354615	6.00E-06	1	0.3703	2.043
	6	rs2875129	1.09E-05	1	0.4973	1.991
	16	rs9928573	1.26E-05	1	0.4973	2.107
DS-AVSD						
	7	rs1568214	1.72E-06	0.7439	0.3028	2.693
	6	rs2983865	2.08E-06	0.898	0.3028	0.3508
	7	rs7778237	2.43E-06	1	0.3028	2.653
	6	rs3011868	2.80E-06	1	0.3028	2.666
	6	rs2983879	5.06E-06	1	0.3351	0.3711
	6	rs2917461	5.26E-06	1	0.3351	2.594
	3	rs13084784	5.63E-06	1	0.3351	2.623
	6	rs2983882	6.21E-06	1	0.3351	2.554
	7	rs6946669	1.24E-05	1	0.4598	2.472
	7	rs7799573	1.24E-05	1	0.4598	2.472
DS-ASD						
	5	rs160890	7.83E-08	0.03383	0.03383	8.632
	2	rs779355	1.40E-06	0.6054	0.3027	3.076
	9	rs7033354	5.26E-06	1	0.7121	0.2865
	1	rs11102433	8.34E-06	1	0.7121	3.157
	2	rs930655	8.86E-06	1	0.7121	0.3418
	15	rs1672460	9.89E-06	1	0.7121	3.831
	7	rs7789571	1.28E-05	1	0.7551	4.95
	15	rs1704394	1.40E-05	1	0.7551	3.135
	16	rs7498941	2.23E-05	1	0.8094	3.353
	3	rs9850770	2.24E-05	1	0.8094	3.082
DS-VSD						
	10	rs1949058	1.32E-07	0.05703	0.03957	1.284
	10	rs10788451	1.83E-07	0.07913	0.03957	0.5663
	12	rs1847461	4.81E-07	0.2078	0.06925	1.153
	14	rs17119691	1.04E-06	0.4491	0.08982	0.927

14	rs11621241	1.04E-06	0.4491	0.08982	1.038
12	rs10859032	1.40E-06	0.6056	0.1009	1.021
12	rs2393073	1.68E-06	0.7249	0.1036	1.176
12	rs10859030	4.90E-06	1	0.2643	0.824
11	rs545830	6.15E-06	1	0.2952	1.131
2	rs3747517	6.85E-06	1	0.2958	1.016

Chr. , Chromosome. *Bonf*, Bonferroni corrected *P*-value. *FDR*, False discovery rate. *OR*, Odd ratios. Association results were corrected based on the number of SNPs (n = 431,962) tested.

911 **Table S2.** The distribution of phenotypic attributes of 187 DS-CHD individuals and

912 controls.

Country		**Number of Cases**			**Number of Controls (DS without CHD)**		
		Illumina 550K	Illumina 610K	Total	Illumina 550K	Illumina 610K	Total
France	38 AVSD	21	17		102	49	
	39 ASD	17	22				
	45 VSD	22	23				
Italy	9 AVSD	9					
	6 ASD	6					
	5 VSD	5					
Spain	15 AVSD		15				
	2 ASD		2				
	12 VSD		12				
Greece	7 AVSD	7					
	4 ASD	4					
	5 VSD	5					
Total		96	91	**187**	102	49	**151**

913

914

915

916

917

918

919

920

Table S3. *P*-values of top pairs of potentially interacting SNPs from the 431,962 whole genome SNPs for the DS-CHD phenotype and its different sub-phenotypes.

Phenotype	SNP1_Chr.	SNP location	*cis*-eQTL	SNP2_Chr.	SNP location	*cis*-eQTL	*P*-value
CHD	rs4246521_1	*ZNF364*	*ANKRD35/PDE4DIP*	rs2899668_15	Intergenic	*FAM148B/VPS13C*	1.45×10^{-11}
	rs1680662_1	*ZNF364*	*ANKRD35/PDE4DIP*	rs2899668_15	Intergenic	*FAM148B/VPS13C*	2.30×10^{-11}
	rs8004209_14	*TMEM63C*		rs1178778_3	Intergenic	*ROBO1*	5.23×10^{-11}
	rs2966122_16	*CMIP*	*CMC2*	rs9991795_4	*RNF150*	*IL15*	7.90×10^{-11}
AVSD							
	rs7764005_6	Close to *THBS2*		rs12043436_1	*DAB1*	*OMA1*	2.38×10^{-11}
	rs11628040_14	*LINC00520*		rs9898755_17	Intergenic		7.98×10^{-11}
	rs7764005_6	Close to *THBS2*		rs552496_1	Close to *DAB1*		9.74×10^{-11}
	rs2393638_10	*ANK3*	*ANK3*	rs4781070_16	*RMI2*	*RUNDC2A/RPM2*	1.36×10^{-10}
VSD							
	rs7597939_2	*BC042073*		rs6918808_6	Intergenic		3.68×10^{-11}
	rs1379127_4	Intergenic	*PCDH7*	rs12351801_9	*TBC1D2*	*HEMGN*	4.65×10^{-11}
	rs2899668_15	Intergenic	*FAM148B/VPS13C*	rs4246521_1	*ZNF364*	*ANKRD35/PDE4DIP*	7.99×10^{-11}
	rs2899668_15	Intergenic	*FAM148B/VPS13C*	rs1680662_1	*ZNF364*	*ANKRD35/PDE4DIP*	9.10×10^{-11}
ASD							
	rs7724739_5	*ADAMTS12*	*AMACR*	rs6940651_6	*SYNE1*		8.03×10^{-11}
	rs10121424_9	*ZNF618*		rs6444462_3	Intergenic		8.46×10^{-11}
	rs12971916_19	*ZNF321*		rs2497750_1	*PAP2D*	*AGL*	1.16×10^{-08}

The *cis*-eQTLs column shows if any of these interacting SNPs have also *cis*-eQTL effects on gene expression. The threshold for being *cis*-eQTL is set to $p < 0.01$ (see Yang et al. 2010).

930 **Table S4.** The list of primers used for qPCR .

Primer	Sequence
CNV0-F	5′-TGGTAGTCAGTTCAGGACAAATG-3′
CNV0-R	5′-TACCCACTTCAGAGGTGCAA-3′
CNV1-F	5′-CTCTCAGAGGGTTCATGCTG-3′
CNV1-R	5′-CAGCCTCCATAAGACCCTTG-3′
CNV2-F	5′-ACAGAAGCTCCTGGTGACG-3′
CNV2-R	5′-TCCAAGCAACTGAAAATCCA-3′
CNV3-F	5′-CTCAAACACGGTAATGTGTGC-3′
CNV3-R	5′-CCCCTTTAGGAAATGTCTGC-3′

931

932

933 **Table S5.** The list of primers for genotyping rs1943950 and rs681418 by Sanger
934 sequencing.

Primer	Sequence
rs1943950_R	5′-CTCTACCTCTTTTCCTGGGTAAA-3′
rs681418_F	5′-AGGCTGTGTGGAATCAGCATTAT-3′

935

936

937

938

939

940 **Table S6.** The list of primers used in PCR amplification and pyrosequencing genotyping

941 of rs2832616, rs2183593, rs7282991 and rs9723772.

Primer	Function	Sequence
rs2832616_F	Fwd PCR	5´-CTTCAACTTGGGTGGTTCTTTCT-3´
rs2832616_R	Rev PCR	5´-biotin-AGGGGCACATCATTCTGGA-3´
rs2832616_S	Sequencing	5´-GGGTGGTTCTTTCTCTAC-3´
rs2183593_F	Fwd PCR	5´-AGTCCATAACCCACGGCAATT-3´
rs2183593_R	Rev PCR	5´-biotin-CCCATCATGCTTCTCCTGTCAA-3´
rs2183593_S	Sequencing	5´-TCTTCAAGAGTGTGCAG-3´
rs7282991_F	Fwd PCR	5´-AGAGCTGGGTGGCTGGTGA-3´
rs7282991_R	Rev PCR	5´-biotin-GCCCCAGGGACACACAAG-3´
rs7282991_S	Sequencing	5´-GACGCTGGGCCAGCT-3´
rs972372_F	Fwd PCR	5´-biotin-GCGAGTTTAATTCCTCATACGTG-3´
rs972372_R	Rev PCR	5´-CAAGAACTTCTGCGGTCATGT-3´
rs972372_S	Sequencing	5´-GCTACAGTGGAGAACTATTC-3´

942

943

944

945

946

947

948 **Table S7.** DAVID GO term analyses for GWIS top signals.

GO TERM	*P*	Fold Enrichment	FDR-*BH*
Cell adhesion	0.0177	3.74	0.99
Biological adhesion	0.0178	3.73	0.98
Cell-cell adhesion	0.0227	6.32	0.96
Negative regulation of cell proliferation	0.0449	4.83	0.99

949 *P*, nominal *P*-value. FDR-*BH*, False discovery rate control by the *BH* procedure.

950

Figure 1

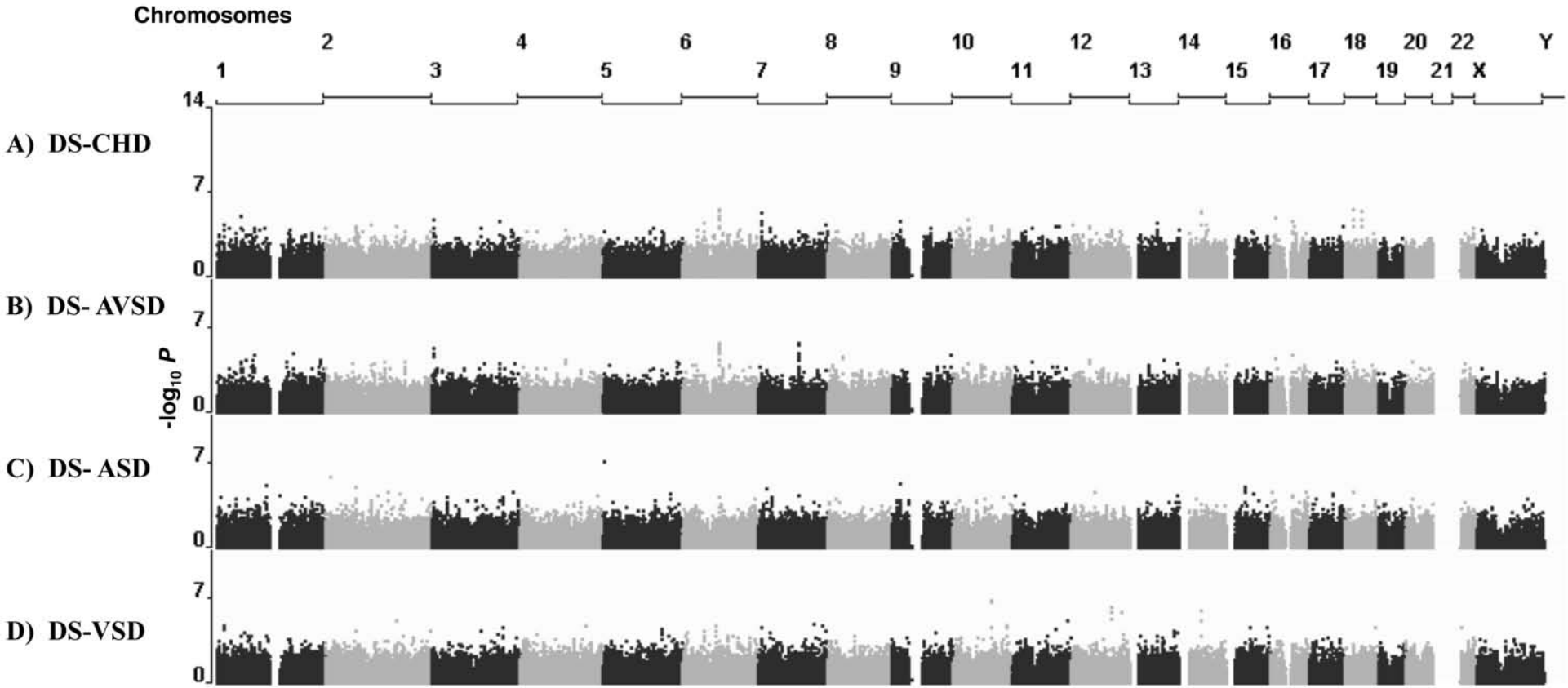
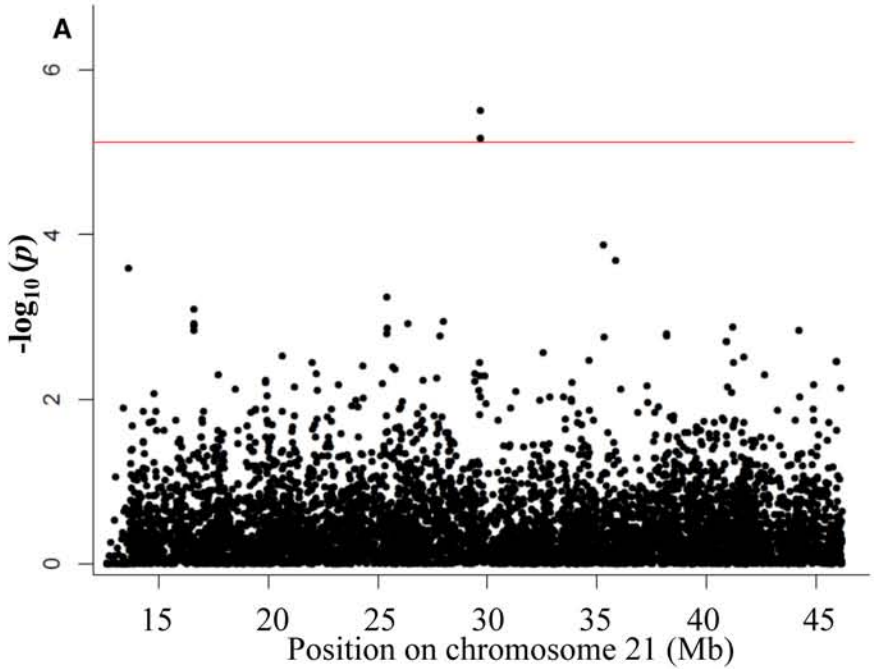
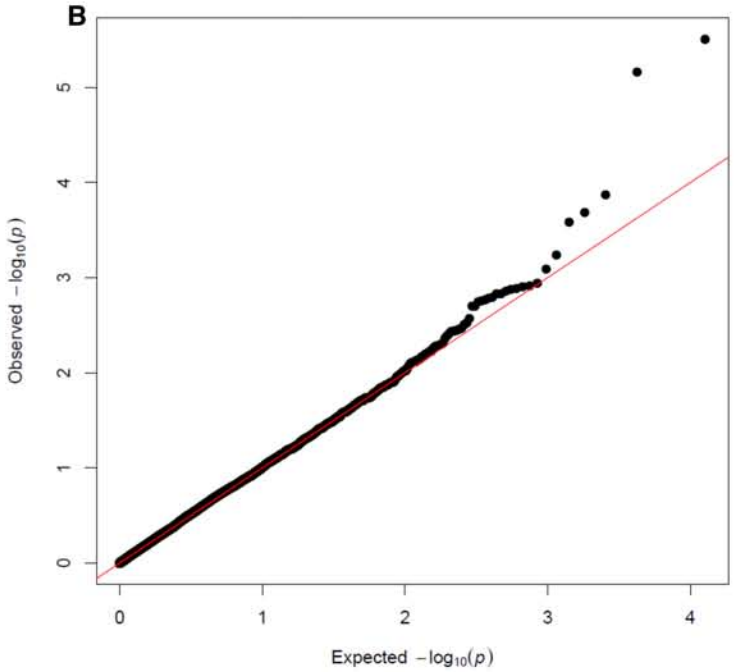

Chromosomes
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
A) DS-CHD
B) DS- AVSD
C) DS- ASD
D) DS-VSD
-log10 P
14
7
0

Figure 2


A
-log10 (p)
0
2
4
6
15
20
25
30
35
40
45
Position on chromosome 21 (Mb)


B
Observed −log10(p)
Expected −log10(p)
0
1
2
3
4
5

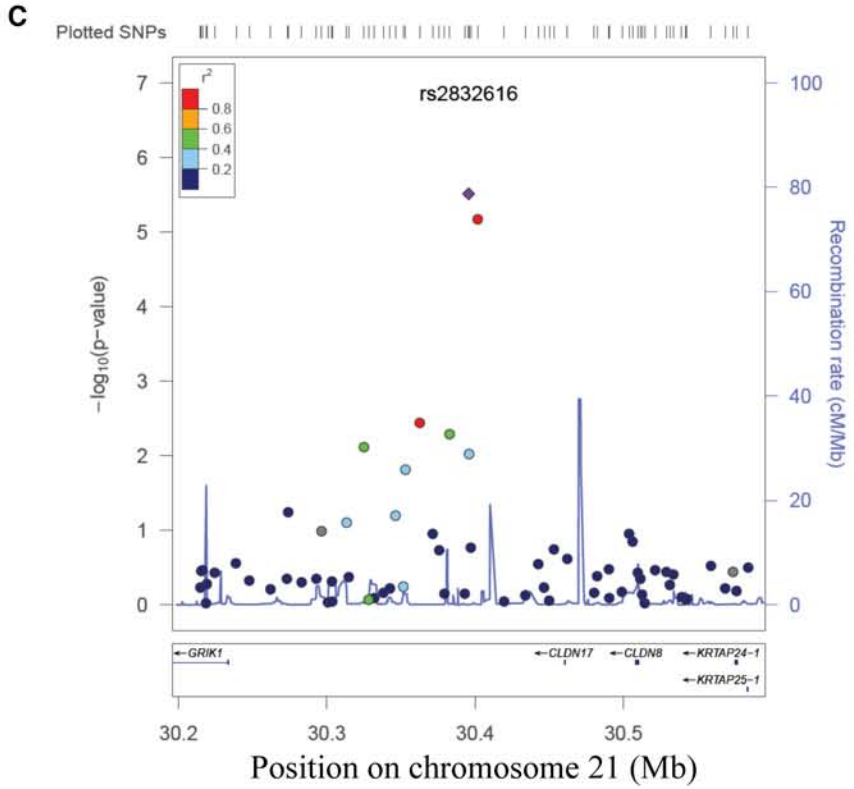

C
Plotted SNPs
rs2832616
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
0
20
40
60
80
100
GRIK1
CLDN17
CLDN8
KRTAP24−1
KRTAP25−1
30.2
30.3
30.4
30.5
Position on chromosome 21 (Mb)

Figure 3

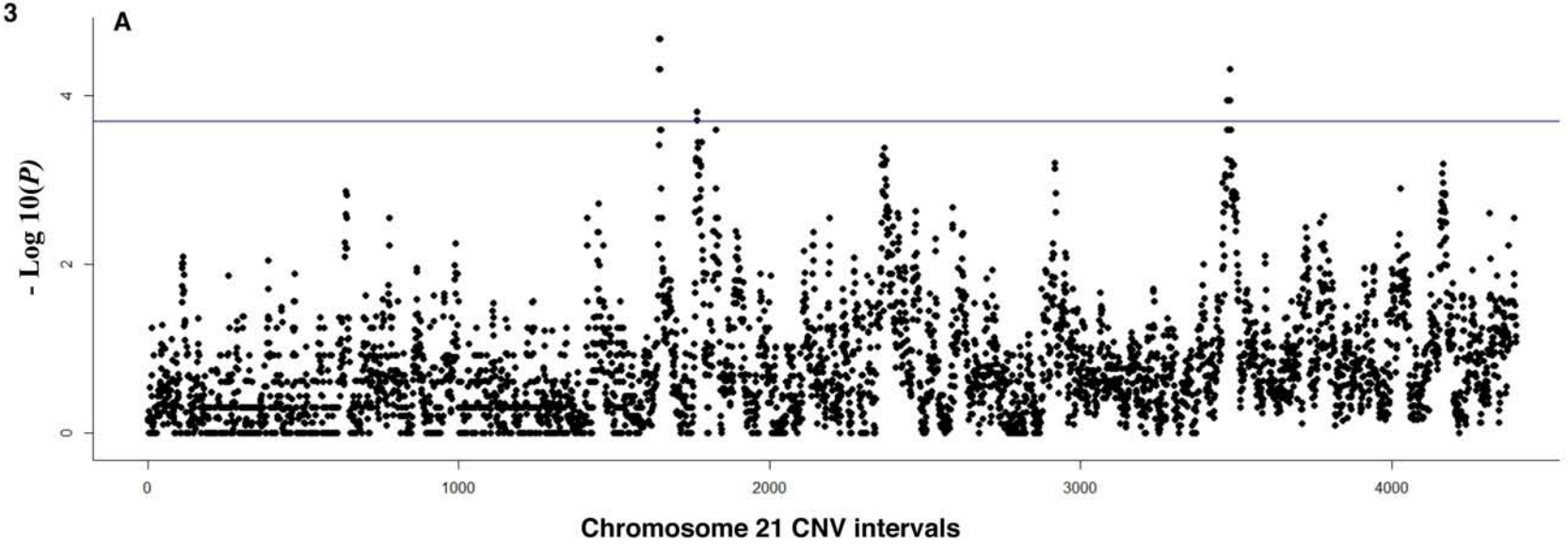

A
- Log 10(P)
0
2
4
0
1000
2000
3000
4000
Chromosome 21 CNV intervals

Figure 3

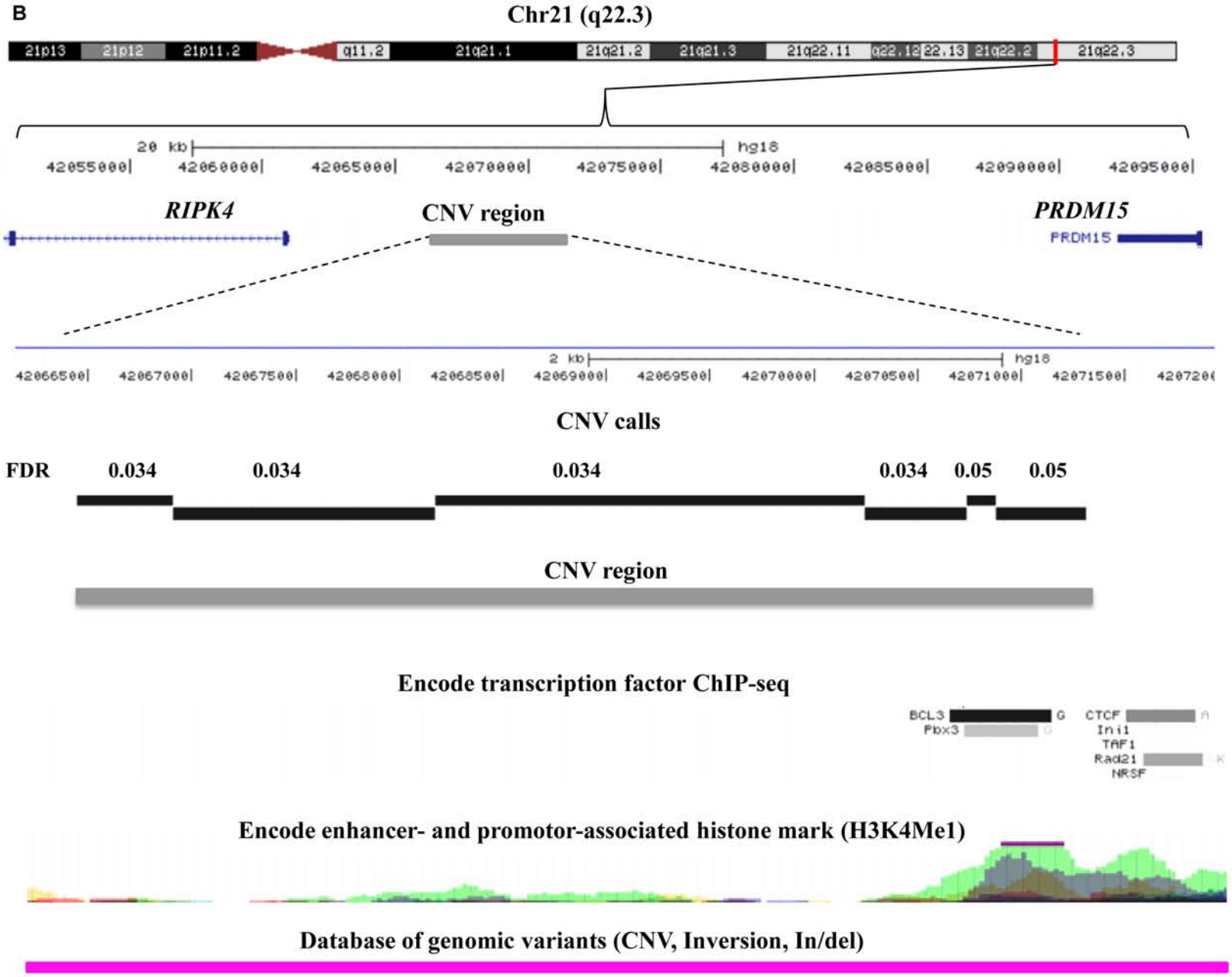

B
Chr21 (q22.3)
21p13
21p12
21p11.2
q11.2
21q21.1
21q21.2
21q21.3
21q22.11
q22.12
22.13
21q22.2
21q22.3
20 kb
hg18
42055000
42060000
42065000
42070000
42075000
42080000
42085000
42090000
42095000
RIPK4
CNV region
PRDM15
PRDM15
2 kb
hg18
42066500
42067000
42067500
42068000
42068500
42069000
42069500
42070000
42070500
42071000
42071500
CNV calls
FDR
0.034
0.034
0.034
0.034
0.05
0.05
CNV region
Encode transcription factor ChIP-seq
BCL3
Pbx3
CTCF
Ini1
TAF1
Rad21
NRSF
Encode enhancer- and promotor-associated histone mark (H3K4Me1)
Database of genomic variants (CNV, Inversion, In/del)

Figure 3

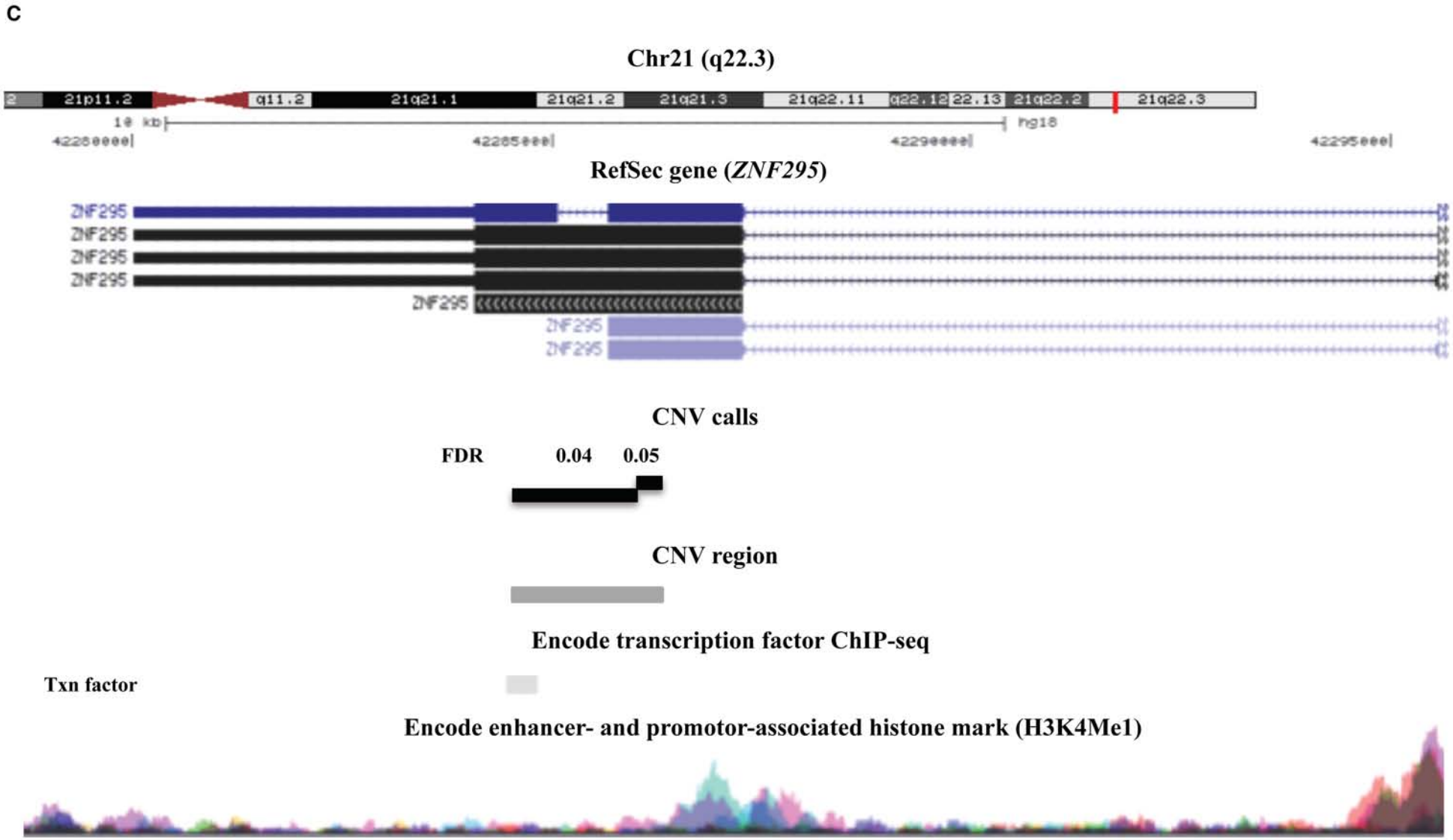

C
Chr21 (q22.3)
21p11.2
q11.2
21q21.1
21q21.2
21q21.3
21q22.11
q22.12
22.13
21q22.2
21q22.3
10 kb
hg18
42280000
42285000
42290000
42295000
RefSec gene (ZNF295)
ZNF295
CNV calls
FDR
0.04
0.05
CNV region
Encode transcription factor ChIP-seq
Txn factor
Encode enhancer- and promotor-associated histone mark (H3K4Me1)

Figure S1

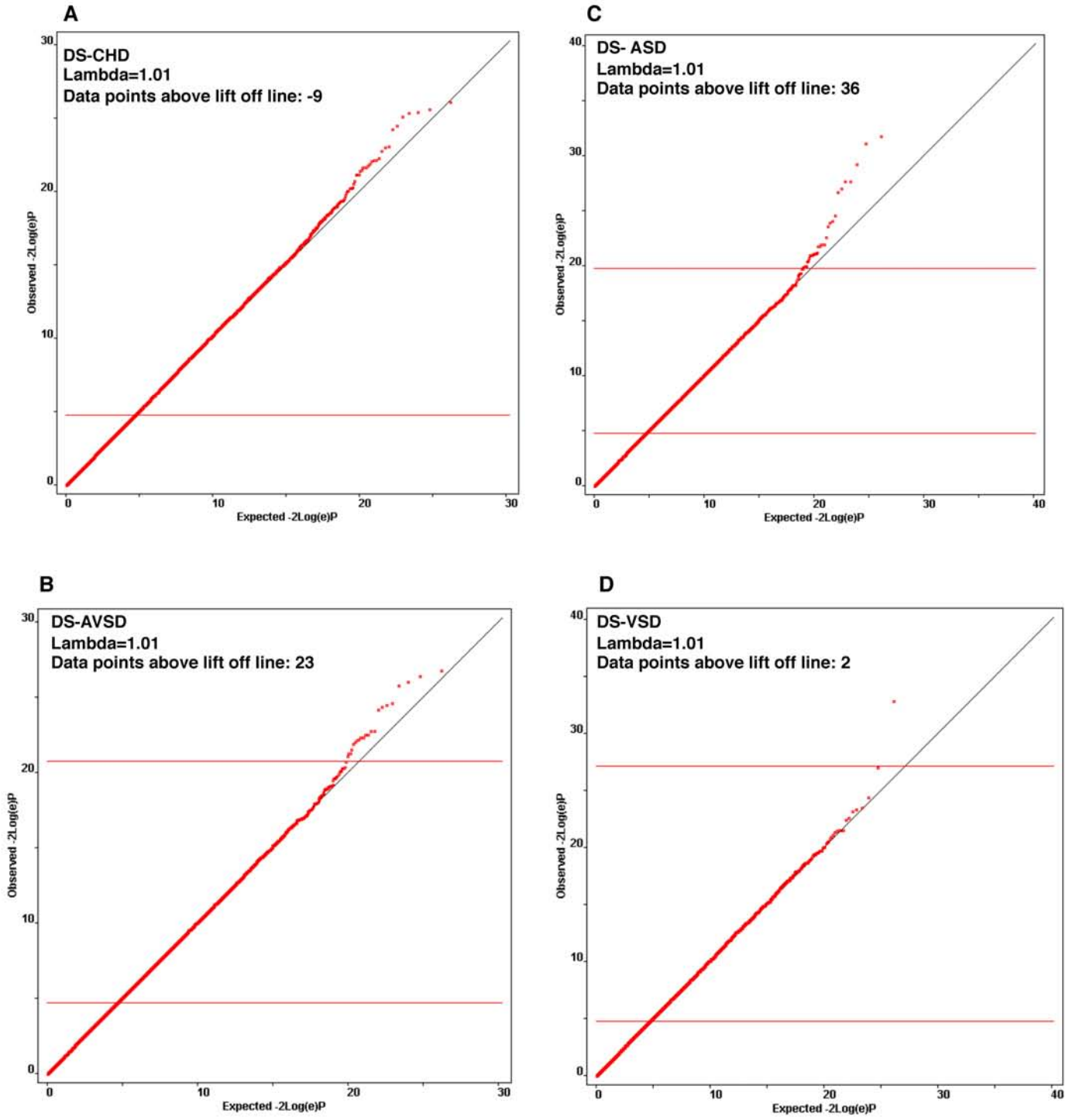

A
DS-CHD
Lambda=1.01
Data points above lift off line: -9
Observed -2Log(e)P
Expected -2Log(e)P
C
DS- ASD
Lambda=1.01
Data points above lift off line: 36
Observed -2Log(e)P
Expected -2Log(e)P
B
DS-AVSD
Lambda=1.01
Data points above lift off line: 23
Observed -2Log(e)P
Expected -2Log(e)P
D
DS-VSD
Lambda=1.01
Data points above lift off line: 2
Observed -2Log(e)P
Expected -2Log(e)P

Figure S2

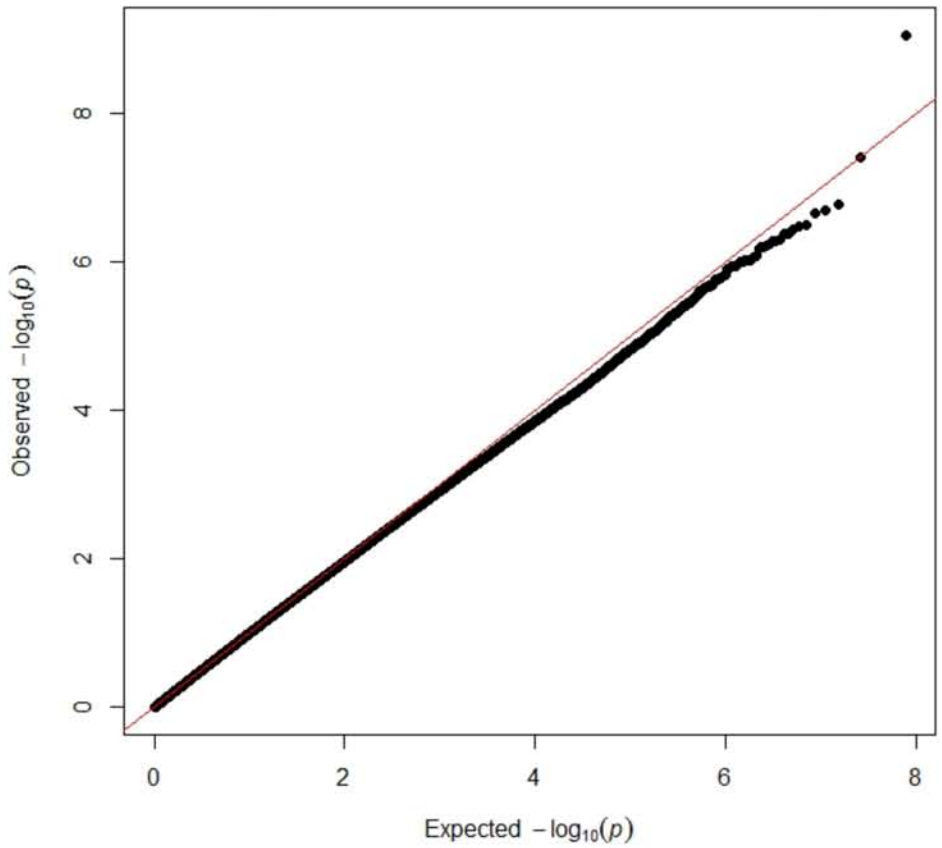


Figure S3

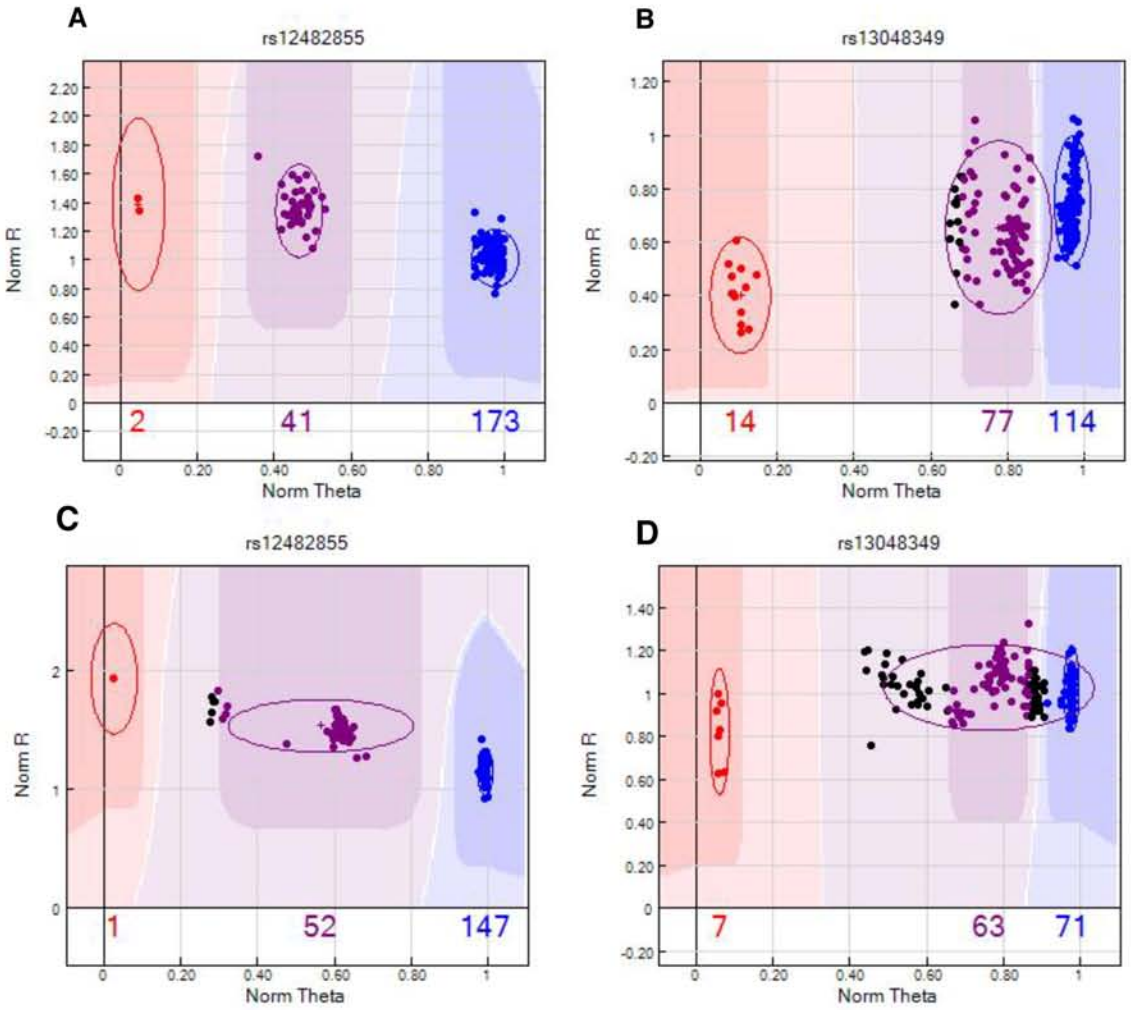

A
rs12482855
Norm R
Norm Theta
2
41
173
B
rs13048349
Norm R
Norm Theta
14
77
114
C
rs12482855
Norm R
Norm Theta
1
52
147
D
rs13048349
Norm R
Norm Theta
7
63
71

Figure S4

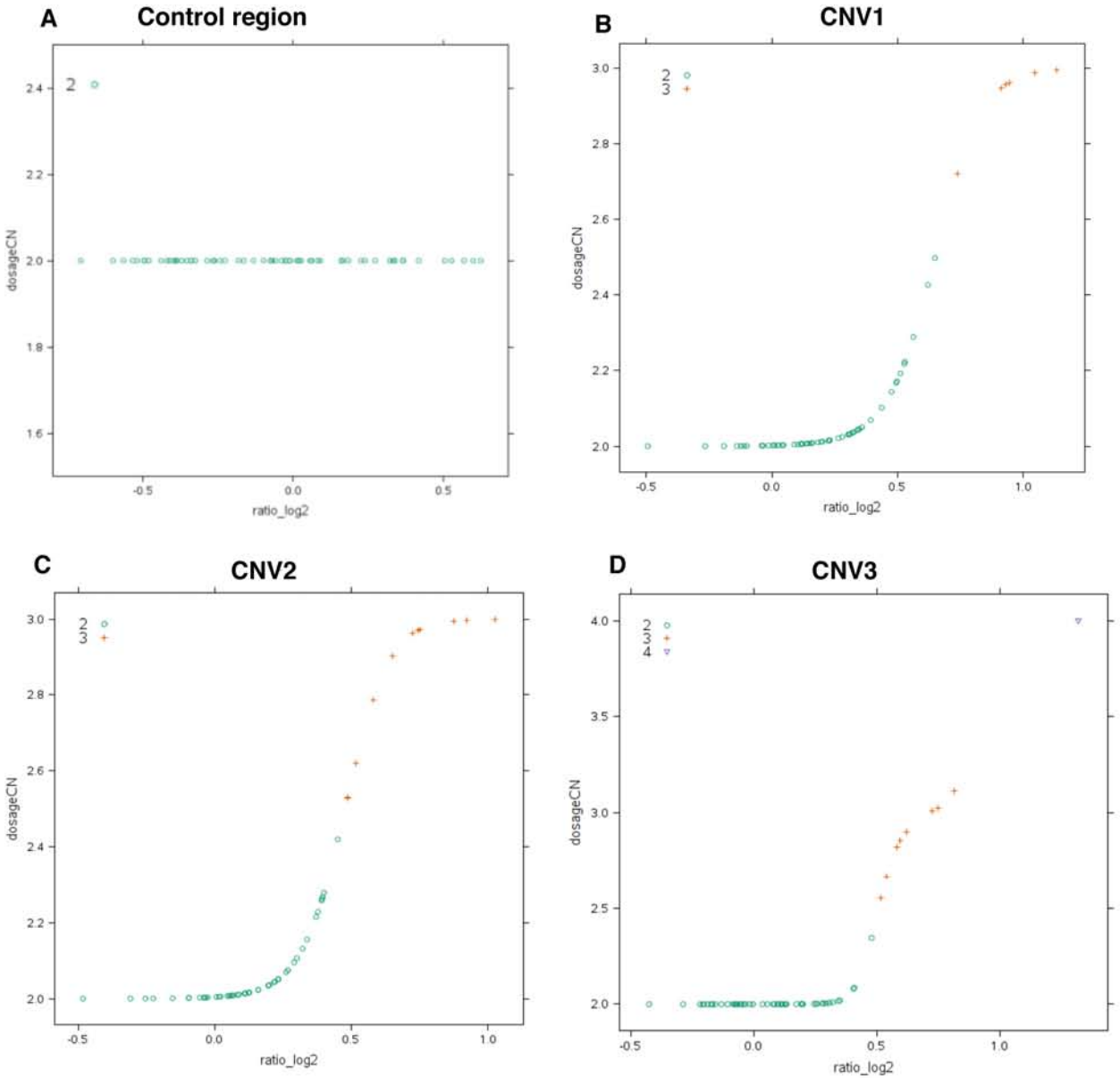


Figure S5

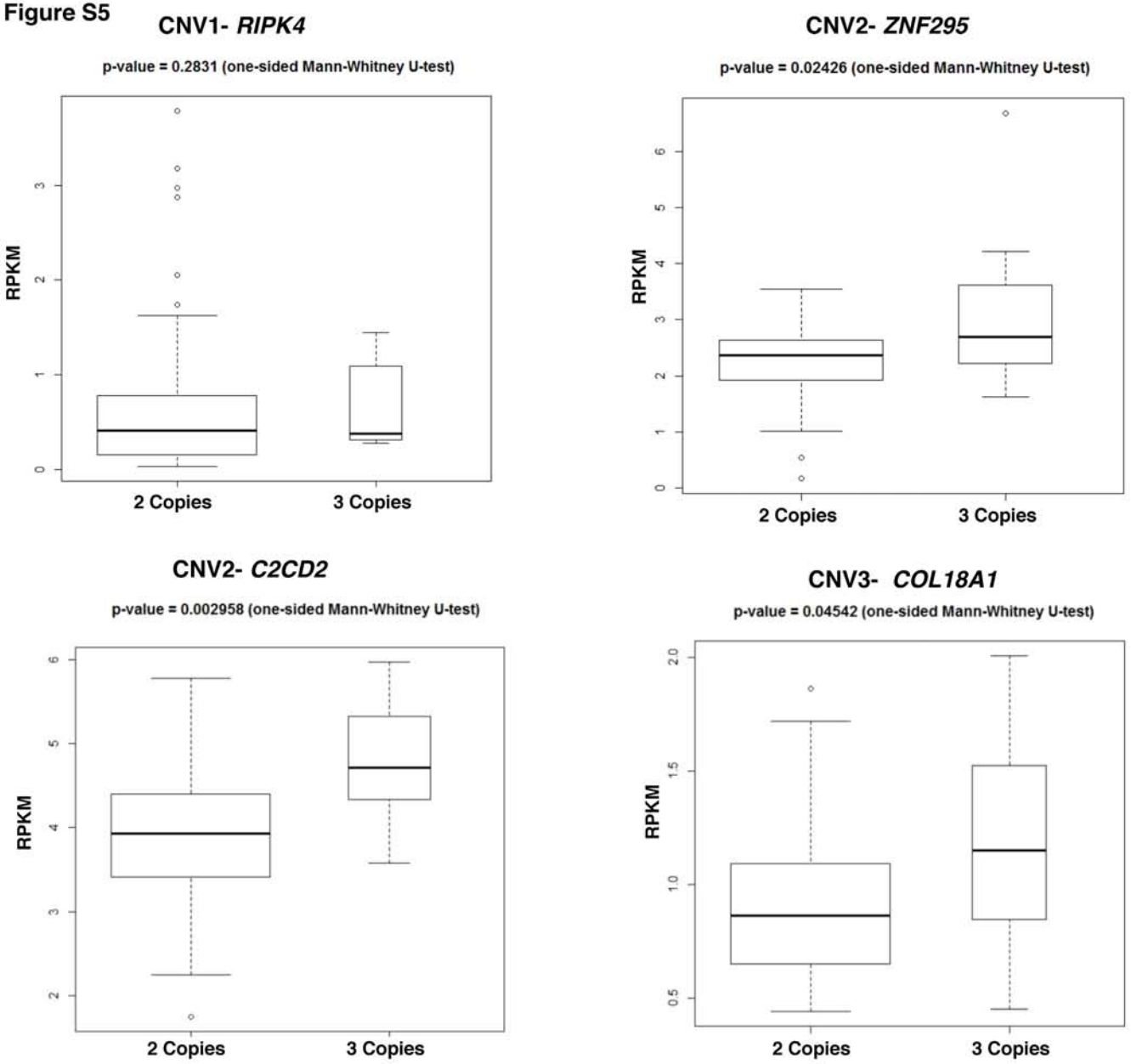

CNV1- RIPK4
p-value = 0.2831 (one-sided Mann-Whitney U-test)
RPKM
0
1
2
3
2 Copies
3 Copies
CNV2- ZNF295
p-value = 0.02426 (one-sided Mann-Whitney U-test)
RPKM
0
1
2
3
4
5
6
2 Copies
3 Copies
CNV2- C2CD2
p-value = 0.002958 (one-sided Mann-Whitney U-test)
RPKM
2
3
4
5
6
2 Copies
3 Copies
CNV3- COL18A1
p-value = 0.04542 (one-sided Mann-Whitney U-test)
RPKM
0.5
1.0
1.5
2.0
2 Copies
3 Copies

Figure S6

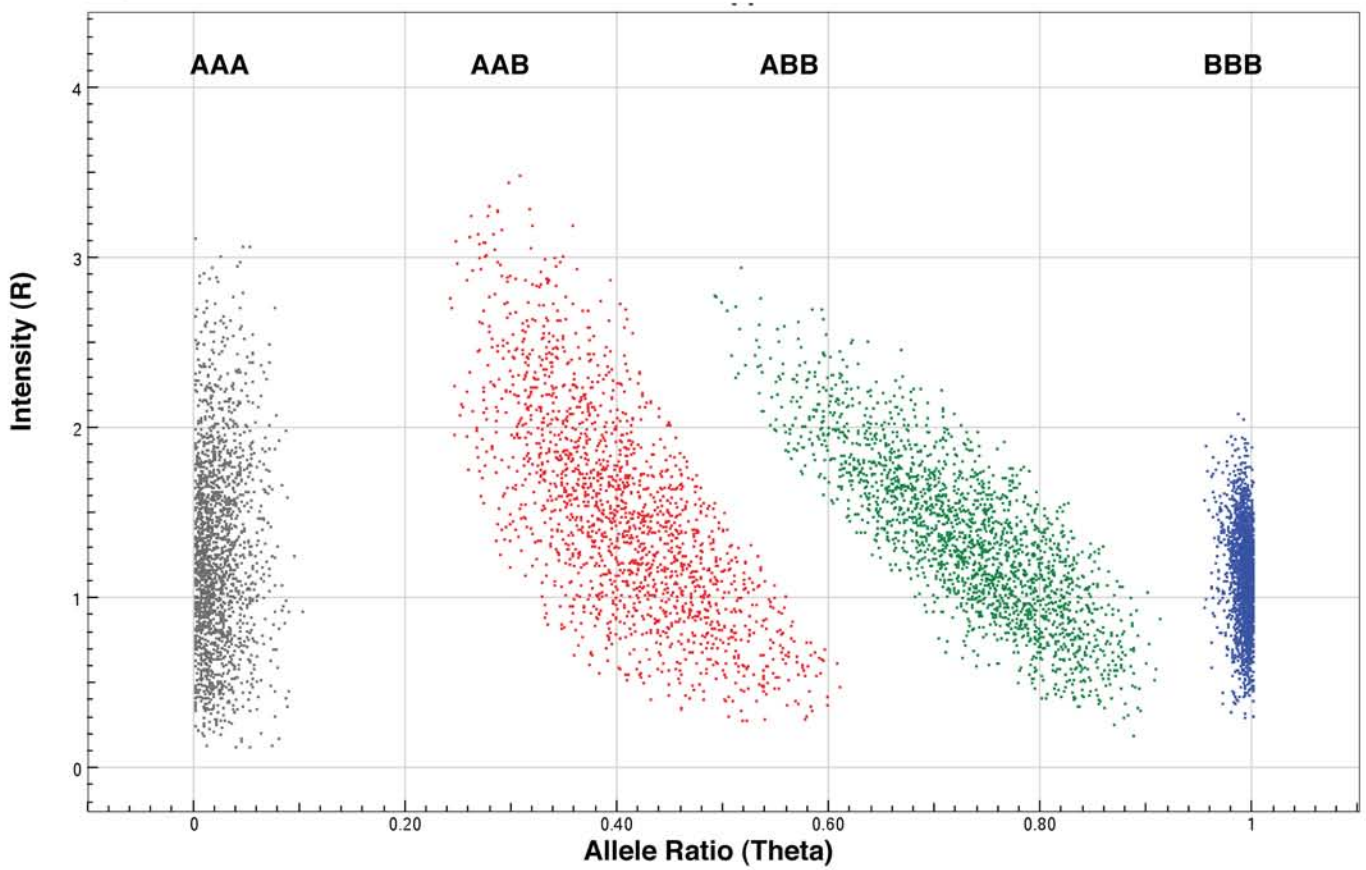
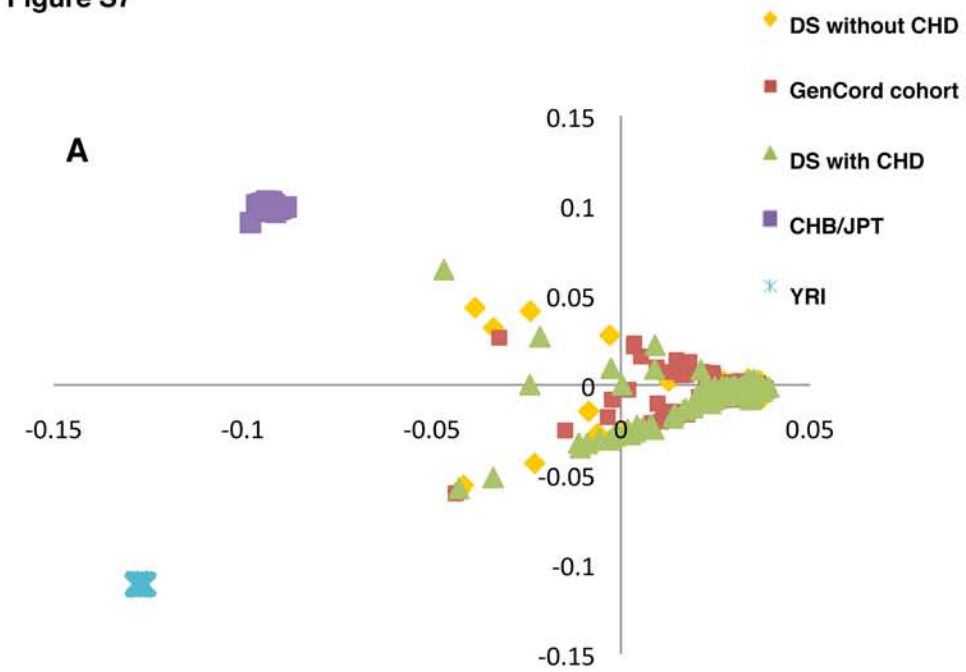
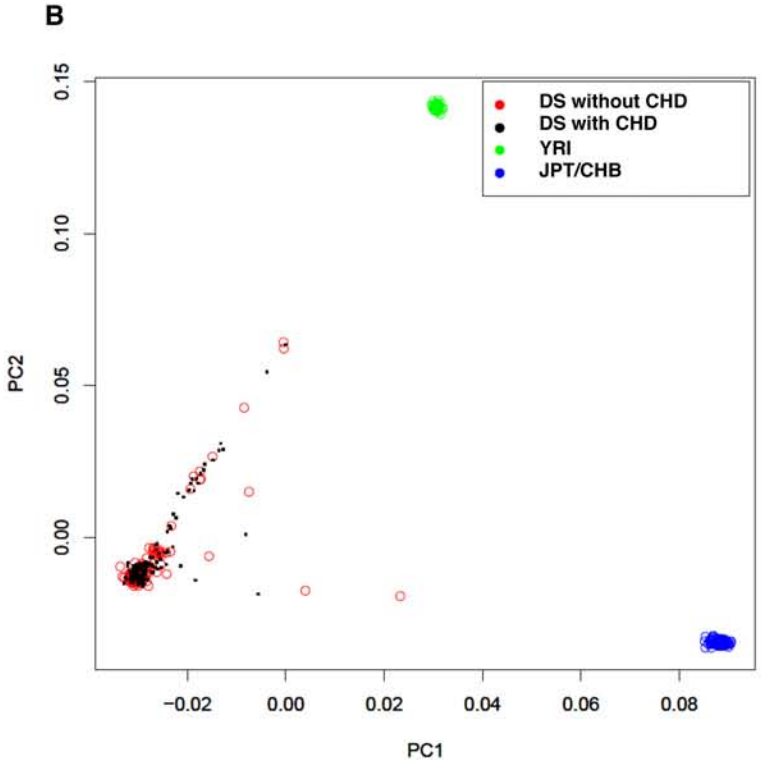
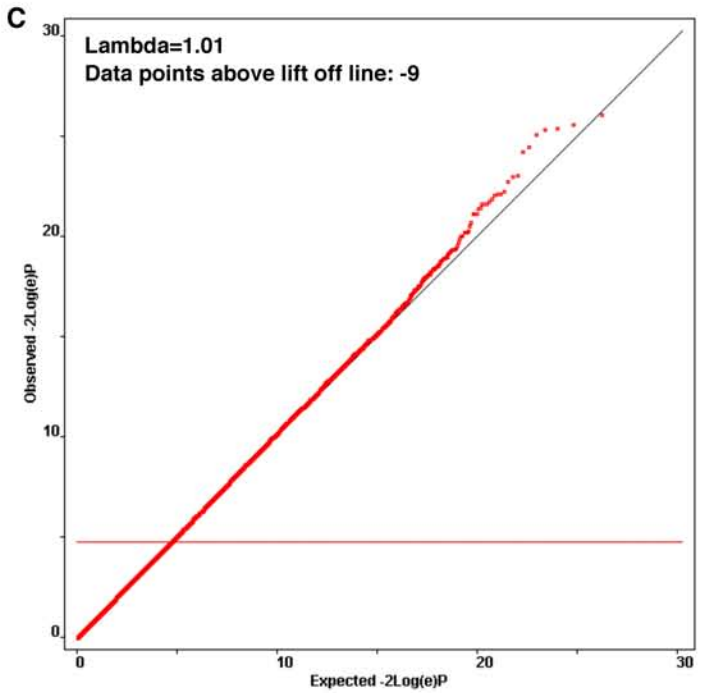

AAA
AAB
ABB
BBB
Intensity (R)
Allele Ratio (Theta)
0
1
2
3
4
0
0.20
0.40
0.60
0.80
1

Figure S7


A
DS without CHD
GenCord cohort
DS with CHD
CHB/JPT
YRI
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.15
-0.1
-0.05
0
0.05


B
DS without CHD
DS with CHD
YRI
JPT/CHB
PC2
PC1
0.15
0.10
0.05
0.00
−0.02
0.00
0.02
0.04
0.06
0.08


C
Lambda=1.01
Data points above lift off line: -9
Observed -2Log(e)P
Expected -2Log(e)P
30
20
10
0
0
10
20
30

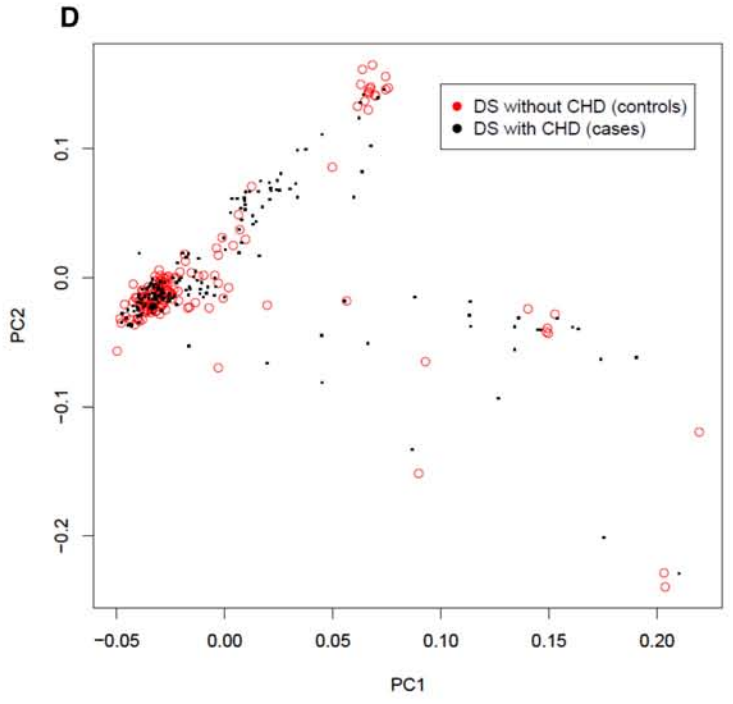

D
DS without CHD (controls)
DS with CHD (cases)
PC2
PC1
0.1
0.0
−0.1
−0.2
−0.05
0.00
0.05
0.10
0.15
0.20

Figure S8

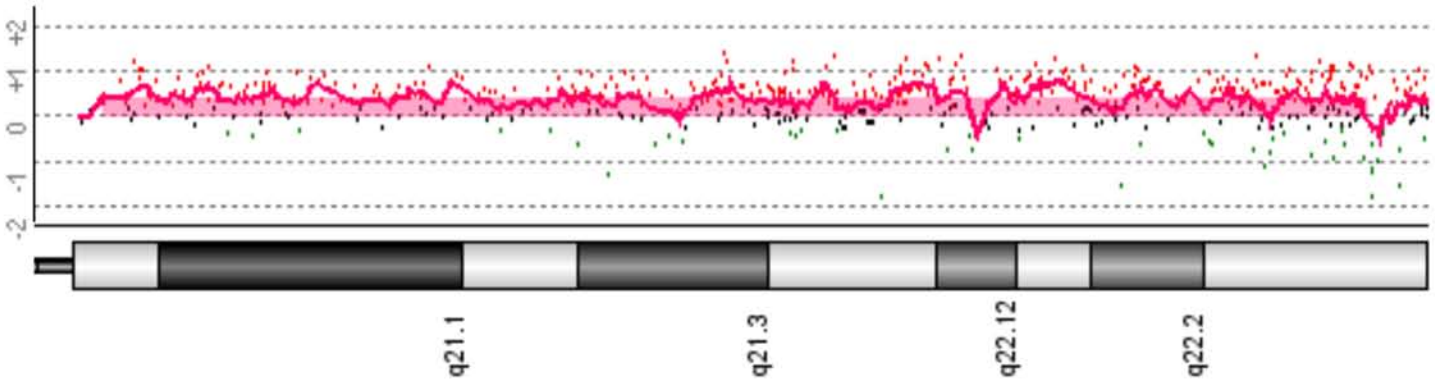
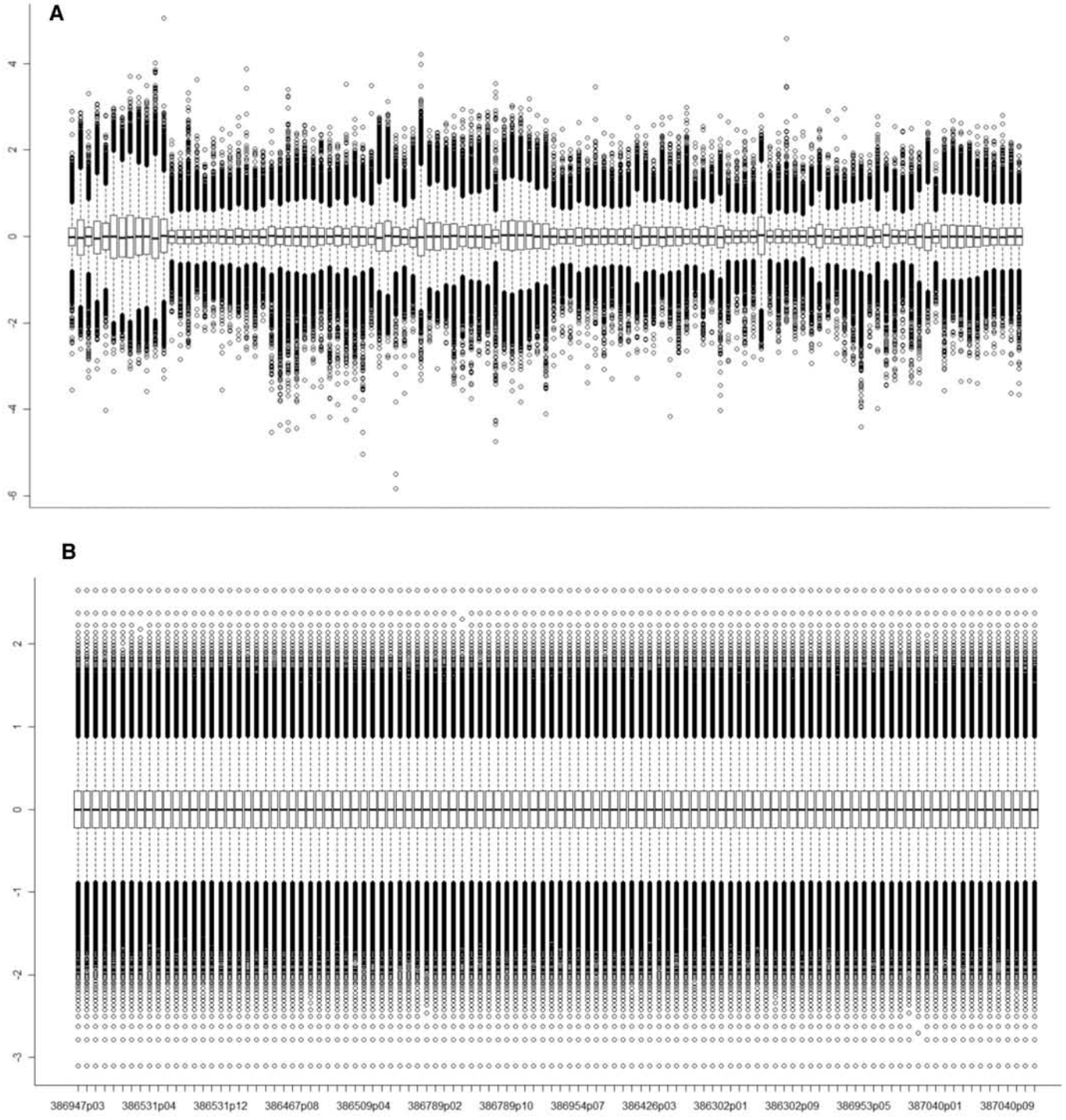


Figure S9


A
4
2
0
-2
-4
-6
B
2
1
0
-1
-2
-3
386947p03
386531p04
386531p12
386467p08
386509p04
386789p02
386789p10
386954p07
386426p03
386302p01
386302p09
386953p05
387040p01
387040p09