Loss of RNA expression and allele-specific expression associated with congenital heart disease

David M. McKean1,2, Jason Homsy1,2,3, Hiroko Wakimoto1, Neil Patel4, Joshua Gorham1, Steven R. DePalma1,5, James S. Ware1,6,7, Samir Zaidi8, Wenji Ma9, Nihir Patel4, Richard P. Lifton8,10, Wendy K. Chung11, Richard Kim12, Yufeng Shen9,13, Martina Brueckner8, Elizabeth Goldmuntz14, Andrew J. Sharp4,15, Christine E. Seidman1,2,5,*, Bruce D. Gelb4,15,16, & J.G. Seidman1

Congenital heart disease (CHD), a prevalent birth defect occurring in 1% of newborns, likely results from aberrant expression of cardiac developmental genes. Mutations in a variety of cardiac transcription factors, developmental signalling molecules and molecules that modify chromatin cause at least 20% of disease, but most CHD remains unexplained. We employ RNAseq analyses to assess allele-specific expression (ASE) and biallelic loss-of-expression (LOE) in 172 tissue samples from 144 surgically repaired CHD subjects. Here we show that only 5% of known imprinted genes with paternal allele silencing are monoallelic versus 56% with paternal allele expression—this cardiac-specific phenomenon seems unrelated to CHD. Further, compared with control subjects, CHD subjects have a significant burden of both LOE genes and ASE events associated with altered gene expression. These studies identify FGFBP2, LBH, RBFOX2, SGSM1 and ZBTB16 as candidate CHD genes because of significantly altered transcriptional expression.

1 Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. 2 Cardiovascular Division, Brigham and Women’s Hospital, Harvard University, Boston, Massachusetts 02115, USA. 3 Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. 4 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. 5 Howard Hughes Medical Institute, Harvard University, Boston, Massachusetts 02115, USA. 6 National Institute for Health Research Cardiovascular Biomedical Research Unit at Royal Brompton and Harefield National Health Service Foundation Trust and Imperial College London, London SW3 6NP, UK. 7 National Heart and Lung Institute, Imperial College London, London SW3 6NP, UK. 8 Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA. 9 Department of Systems Biology, Columbia University Medical Center, New York, New York 10032, USA. 10 Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06510, USA. 11 Department of Pediatrics and Medicine, Columbia University Medical Center, New York, New York 10032, USA. 12 Section of Cardiothoracic Surgery, University of Southern California Keck School of Medicine, Los Angeles, California 90089, USA. 13 Department of Biomedical Informatics, Columbia University Medical Center, New York, New York 10032, USA. 14 Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. 15 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.G.S. (email: seidman@genetics.med.harvard.edu).
Congenital heart disease (CHD)-causing mutations have been identified in >50 genes including transcription factors, signalling molecules and chromatin modifiers, which direct the temporal and spatial expression of genes during cardiac development. Recent studies have estimated that there are ~400 genes that can harbour loss or gain-of-function mutations that cause CHD (denoted CHD genes). We hypothesized that other CHD genes could be identified by altered expression of one (allele-specific expression (ASE)) or both alleles (loss-of-expression (LOE); Fig. 1).

ASE occurs when transcription from one allele is selectively silenced or enhanced, or when transcripts undergo selective post-transcriptional degradation (for example, nonsense-mediated decay; NMD). ASE occurs physiologically to control dosage effects of chromosome X-encoded genes in females and to silence the maternal or paternal allele of imprinted genes. ASE occurs when one allele can be suppressed by allele-specific chromatin marks, long noncoding RNAs or gene regulatory element mutations. Other ASE studies include all genes where one allele is expressed at a statistically higher level than the other allele, an approach that estimates hundreds of ASE events per tissue and thousands of ASE events per cell; however, this strategy likely results in significant overestimates of ASE-event rates.

We studied ASE in discarded tissues from CHD patients, hypothesizing that ASE events likely to cause CHD should result in substantial allele bias in the expressed transcripts. Hence, we focused on genes that are expressed in fetal heart that (1) are normally biallelically expressed and (2) exhibit extreme ASE (that is, >86% expression of one allele relative to the other). Moreover, we suggest that ASE per se would not be enough to cause a disease phenotype, particularly if dosage compensation resulted in overall normal gene expression. Thus, we focused on extreme ASE events, either with significantly altered gene expression or in which a deleterious mutation was detected in the expressed allele (Fig. 1) as candidate CHD genes.

Biallelic LOE (caused by inadequate trans-acting factors, or combinations of gene regulatory mutations and/or NMD) can cause CHD by either dominant or recessive mechanisms. To identify LOE genes potentially responsible for CHD, we focused on genes (1) that are highly expressed (upper quartile of expressed genes), (2) with tightly regulated cardiovascular expression and (3) with significantly downregulated (>10-fold) expression. We demonstrate that 24% of extreme ASE events in CHD subjects are associated with significantly altered levels of gene expression (compared with 0% of extreme ASE events in control subjects). We identify nine genes in CHD subjects that are functionally null—three due to ASE with a damaging mutation in the expressed allele and six due to biallelic LOE. We propose FGFBP2, LBH, RBFOX2, SGSM1 and ZBTB16 as especially strong candidate CHD genes.

**Results**

**RNaseq expression analyses of CHD subjects and controls.** To identify cardiac gene expression, we studied 144 probands (average age, 2.9 years; range, fetal to 21 years) enrolled in the Pediatric Cardiac Genomics Consortium and performed RNA sequencing (RNaseq) on 172 surgically discarded cardiovascular tissues (Supplementary Data 1 and 2). We also studied gene expression data from ‘normal’ adult (average age, 49.3 years; range, 20–70 years) cardiac tissues (n = 87, left ventricle; n = 26, right atrium) from the Genotype-Tissue Expression Consortium (GTEx, n = 95 subjects; Supplementary Data 3 and 4) and fetal cardiac tissues without CHD (n = 5, gestational age 15–16 weeks).

RNA expression was measured using standard RNaseq procedures (see Methods). The 172 CHD samples were obtained from eight different cardiovascular tissues (aorta, atrial septum, ductus arteriosus, interventricular septum, left ventricle, pulmonary artery, right atrium and right ventricle; Supplementary Data 2 and Supplementary Table 1) with at least six samples per tissue. Approximately 15,300 expressed genes were expressed in each sample (two or more aligned reads per million (r.p.m.); Fig. 2a (grey bars) and Supplementary Table 1), and the average number of genes expressed per subject per tissue ranged from 14,938 (interventricular septum) to 15,883 (pulmonary artery). Expression analyses were performed by comparing single samples to the mean of all other samples of the same tissue type (for example, right atrial CHD sample versus all other right atrial CHD samples).

**Identification of extreme ASE events.** To detect extreme ASE events in genes that are normally biallelically expressed, single-nucleotide polymorphisms (SNPs) were identified from whole exome or genome (WGS, n = 130) or sequencing of CHD probands and available unaffected parents. SNPs were identified in GTEx donors from Illumina Exome and 5M SNP arrays. At each genomic heterozygous SNP (see Methods), an allele ratio (reference RNaseq reads/alternate RNaseq reads) was computed (Supplementary Fig. 1). To focus on ASE events in genes with normal biallelic expression, SNPs with typical biased expression were excluded (see Methods, Supplementary Fig. 1 and Supplementary Table 2). SNPs were then phased using parental genotypes or by assuming that the most highly expressed base at each SNP is encoded by the same allele. Compound allele ratios were then calculated for each gene as the ratio

**Figure 1 | Identification of extreme ASE genes in subjects with CHD.** Shown are both alleles of a gene that differ by the SNP haplgblocks ‘CC’ (blue) and ‘GT’ (red), as identified by WES, WGS or SNP-array genotyping. RNaseq analysis (read counts at heterozygous positions) reveals the expression of both alleles (biallelic RNA expression) or the disproportionate expression of one allele over another (ASE). RNaseq expression analyses (comparing each sample to the average of all other samples within a tissue group) identify relative loss and gain of expression. Variant analysis, in conjunction with RNaseq analysis, can further identify LOF mutations in the expressed allele (*).
of the summed reads corresponding to each allele (Methods). This phasing methodology was 98% accurate in F1 mouse tissues (M. musculus × M. castaneus; Supplementary Table 3) and 100% accurate in three human WGS trios (affected proband and unaffected parents), when we applied a compound allele ratio threshold \( \geq 7 \) (Supplementary Fig. 2c). Transcripts with a compound allele ratio \( \geq 7.2 \) and a binomial \( P < 0.01 \) (Bonferroni-corrected for the number of expressed genes with heterozygous SNPs (Supplementary Data 2 and 4)) were designated as having extreme ASE. Finally, both over-represented ASE genes (that is, > 5% of subjects including at least one control subject have ASE events in the same gene; Supplementary Table 4) and ASE events in genes with low fetal heart expression (Supplementary Table 5) were removed.

DNA-sequencing methodology strongly influenced extreme ASE detection (Fig. 2b). WGS and WES data yielded 5.1 and 1.4 extreme ASE events per CHD subject, respectively; SNP arrays yielded 3.3 extreme ASE events per GTEx donor. In total, we detected 607 extreme ASE events; 491 events in 17 known ASE genes (Supplementary Table 5 and Supplementary Data 6). The set of genes imprinted in non-cardiac tissues were expressed in fetal heart and also contained heterozygous SNPs. These were evaluated for ASE in cardiovascular tissues. Seventeen genes had ASE in at least 50% of CHD and GTEx subjects (Table 1), whereas 31 were predominantly biallelically expressed (Supplementary Table 8). Gene imprinting in mouse cardiovascular tissues was the same as in human cardiovascular tissues except for three genes: Cdkn1c, Igf2r and Sgce. These three genes are biallelically expressed in human hearts, but have ASE in mouse hearts. Biallelic expression of these same genes in other human tissues has been described. After excluding differences that were attributable to genotyping methods, ASE of imprinted genes was indistinguishable in CHD proband tissues and GTEx tissues. ASE events in known imprinted genes were excluded from further analyses.

**Figure 2 | Extreme ASE genes preferentially identified in WGS subjects.** Shown in a are the number of genes with a minimum expression of 2 c.p.m. (grey bars), and the number of expressed genes that contain heterozygous SNPs (black bars) for CHD WGS (n = 30 tissues) and CHD WES probands (n = 142 tissues), GTEx donors (n = 113 tissues) and mouse C57Bl6/Castaneus F1 hybrids (n = 7 tissues). s.d. is indicated. b The distribution of extreme ASE events per subject by genotyping platform. Extreme ASE events were identified in >70% of WGS subjects (n = 14). However, extreme ASE events were identified in only ~20% of WES subjects (n = 130) and ~45% of GTEx donors (n = 95).

**Imprinting in cardiovascular tissues.** Because parental gene imprinting is a major cause of ASE in mammalian tissues, we assessed imprinting in cardiovascular tissues. The set of genes imprinted in fetal and/or neonatal cardiovascular tissues has not been described. Forty-eight genes previously identified as being imprinted in non-cardiac tissues were expressed in fetal heart and also contained heterozygous SNPs. These were evaluated for ASE in cardiovascular tissues. Seventeen genes had ASE in at least 50% of CHD and GTEx subjects (Table 1), whereas 31 were predominantly biallelically expressed (Supplementary Table 8). Gene imprinting in mouse cardiovascular tissues was the same as in human cardiovascular tissues except for three genes: Cdkn1c, Igf2r and Sgce. These three genes are biallelically expressed in human hearts, but have ASE in mouse hearts. Biallelic expression of these same genes in other human tissues has been described. After excluding differences that were attributable to genotyping methods, ASE of imprinted genes was indistinguishable in CHD proband tissues and GTEx tissues. ASE events in known imprinted genes were excluded from further analyses.

**Extreme ASE is attributable to NMD in a minority of cases.** We identified 78 rare, nonsense mutations in genes that were expressed at sufficient levels to evaluate ASE in CHD probands (Supplementary Table 9). Only 14/78 (18%) genes exhibited NMD and had significantly reduced expression of the allele harbouring the LOF mutation, even after employing a less stringent definition of ASE (allele bias > 4). This low percentage of NMD is consistent with previous reports. Unfortunately, this analysis could not be performed on GTEx samples because complete coding sequence data were unavailable for these subjects.

Of the 45 extreme ASE events observed in CHD probands (Supplementary Table 6), seven ASE events (Table 2) resulted from nonsense (ASPN, CTSA, PGMI and RBFOX2), splice site (AARS1D) or frameshift (C7 and RETSAT) variants that caused NMD. In sum, 38/45 (85%) extreme ASE events in CHD probands remain unexplained and could potentially reflect mutations in gene regulatory elements.

**Extreme ASE genes have significant expression changes.** We hypothesized that genes whose expression levels differ between subjects with extreme ASE versus subjects with biallelic expression are candidate CHD genes. After excluding eight CHD and three GTEx tissues that failed quality controls (Methods), we compared gene expression in 37 extreme ASE genes in CHD subjects and 57 extreme ASE genes in GTEx subjects to the mean expression of biallelically expressed samples of the same tissue type (for example, right atrial ASE expression compared with right atrial biallelic expression). In CHD subjects, upregulated ASE genes (fold \( >5 \), \( P < 0.05 \) calculated from z-score)) included Myoz1 and Fgfbp2 (observed in two subjects) and downregulated ASE genes (fold \( <0.65 \), \( P < 0.05 \), \( P \) calculated from z-score) included SGSM1, AARS1D, C5orf46, SDHB, CBRI and RBFOX2 (Table 3). By contrast, no extreme ASE gene had significantly different expression in GTEx subjects.

**Identification of functionally null genes.** We identified CHD gene candidates who had extreme ASE and harboured a deleterious mutation in the expressed allele, and so are unlikely to
make functional protein. Three extreme ASE genes, C17orf97, CRACR2B and FGFBP2, encoded rare, putatively deleterious variants in the expressed allele (Table 2). Although FGFBP2 has relatively common ASE (>5% of subjects (Supplementary Table 6)), its functional null status makes it a candidate recessive CHD gene. These analyses could not be performed in GTEx samples.

Biallelic LOE of genes that are typically both highly expressed and tightly regulated are another class of functionally null genes that may cause CHD. We identified genes with biallelic LOE (folds in six CHD probands (Table 3) but none in GTEx donors (P = 7.8 × 10^{-3}, Fisher Exact test). Five LOE genes, LBH, FRG1B, PHKG1, IRX5 and ZBTB16, have no homozygous LOF variants in the Exome Aggregation Consortium (ExAC) database (exac.broadinstitute.org), while TRMT2B, an X-linked gene, is hemizygous in a significant number of subjects. Significant downregulation of all biallelic LOE genes was confirmed using quantitative PCR (qPCR; Supplementary Table 10), although this analysis identified only an approximately threefold reduction in FRG1B and IRX5. Biallelic LOE of PHKG1 and IRX5 occurred in two subjects both with Kabuki syndrome (MIM147920 (ref. 7)), caused by damaging de novo KMT2D mutations.

### Discussion
Our transcriptome analyses of tissues from CHD patients identified several CHD gene candidates, including RFBOX2, CADD, combined annotation-dependent depletion; CHD, congenital heart disease; ExAC, Exome Aggregation Consortium; FHE, fetal heart expression (reads per million aligned reads); GTEx, Genotype-Tissue Expression Consortium; M, maternal; NE, not expressed; NS, no SNP; P, paternal; SNP, single-nucleotide polymorphism.

| Gene ID | Mutation Predicted | CADD score | ASE effect | MAF | FHE | PCGC cases LOF AF | PCGC controls LOF AF | ExAC LOF AF |
|----------|---------------------|------------|------------|-----|-----|--------------------|---------------------|-------------|
| AARS1D1  | 1-00384 chr17:41102746G>A | Spl Acceptor* | 24.2 | NMD | 4.1 × 10^{-5} | 138 | 0 | 0 | 0 |
| RBF2orf2 | 1-05368 chr22:36155972G>A | Nonsense | 26.9 | M | 0 | 235.0 | 1.4 × 10^{-3} | 0 | 1.8 × 10^{-5} |
| PST1    | 1-01021 chr16:4100580G>T | Nonsense | 40 | M | 0 | 124.2 | 0 | 0 | 6.4 × 10^{-4} |
| CTS1    | 1-01620 chr20:44522702C>A | Nonsense | 36 | M | 4.1 × 10^{-5} | 16.9 | 0 | 1.1 × 10^{-3} | 3.5 × 10^{-4} |
| C7      | 1-00070 chr:340945362TACG TCAGCAGA>T | Frameshift | NA | M | 0 | 156.7 | 2.8 × 10^{-3} | 5.6 × 10^{-4} | 1.4 × 10^{-3} |
| RETSA1  | 1-00024 chr:28557195TCA>T | Frameshift | NA | M | 1.2 × 10^{-3} | 15.8 | 5.6 × 10^{-3} | 5.6 × 10^{-3} | 9.7 × 10^{-3} |

| Gene ID | Mutation Predicted | CADD score | ASE effect | MAF | FHE | PCGC cases LOF AF | PCGC controls LOF AF | ExAC LOF AF |
|----------|---------------------|------------|------------|-----|-----|--------------------|---------------------|-------------|
| FGF2     | 1-01984 chr4:15964134A>C | p.Trp207Gly | 15.41 | Null | 6.3 × 10^{-4} | 0 | 0 | 0 | 1.6 × 10^{-5} |
| C7orf97  | 1-01485 chr17:260239C>T | p.Arg30Trp | 17.77 | Null | 3.5 × 10^{-4} | 16.8 | 0 | 0 | 0 |
| CRACR2B  | 1-04333 chr11:829356G>A | Spl Acceptor* | NA | Null | 4.4 × 10^{-4} | 20.9 | 0 | 0 | 0 |
Table 3 | Extreme ASE and biallelic LOE events with significantly altered gene expression.

| Gene     | ID       | Tissue       | Proband Expr* | Mean Expr ± s.d. (n) | Fold | P value1 | PGCG cases LOF AF | PGCG controls LOF AF | ExAC LOF AF |
|----------|----------|--------------|---------------|----------------------|------|----------|-------------------|----------------------|-------------|
| ASE genes with loss of allele expression |  | | | | | | | | |
| RBFX2 | 1-05368 | DuctArt | 189.1 | 296 ± 44 (15) | 0.64 | 1.6 ± 10−2 | 1.4 ± 10−3 | 0 | 1.8 ± 10−5 |
| SGSML1 | 1-01019 | RA | 60.3 | 133 ± 51 (18) | 0.45 | 4.7 ± 10−2 | 9.4 ± 10−4 | 0 | 9.1 ± 10−5 |
| AARSD1 | 1-00384 | IVS | 65.9 | 108 ± 12 (7) | 0.61 | 3.0 ± 10−4 | 0 | 0 | 0 |
| Csof46 | 1-00713 | LV | 7.7 | 42 ± 17 (9) | 0.18 | 4.4 ± 10−2 | 0 | 0 | 9.1 ± 10−5 |
| SDHB | C417-01 | IVS | 169.6 | 302 ± 25 (7) | 0.56 | <1.0 ± 10−4 | 0 | 0 | 8.2 ± 10−5 |
| CB1 | CHD-1548 | LV | 12.9 | 29 ± 6.0 (10) | 0.44 | 6.1 ± 10−3 | 0 | 2.8 ± 10−4 | 2.6 ± 10−3 |
| ASE genes with gain of allele expression |  | | | | | | | | |
| FGFBP2 | 1-01024 | RA | 5.6 | 1.0 ± 17 (15) | 5.48 | 6.7 ± 10−3 | 9.5 ± 10−4 | 5.6 ± 10−4 | 4.5 ± 10−4 |
| FGFBP2 | 1-01984 | LA | 9.2 | 1.4 ± 0.4 (3) | 6.86 | <1.0 ± 10−4 | 9.5 ± 10−4 | 5.6 ± 10−4 | 4.5 ± 10−4 |
| MVO1 | 1-02697 | RV | 26.9 | 4.2 ± 3.5 (16) | 6.45 | <1.0 ± 10−4 | 0 | 2.7 ± 10−4 | 3.5 ± 10−4 |
| Genes with loss of expression of both alleles |  | | | | | | | | |
| LHC | 1-03051 | AO | 5.4 | 55 ± 15 (6) | 0.10 | 8.0 ± 10−4 | 0 | 0 | 0 |
| ZBTB10 | 1-03316 | AO | 1.2 | 22 ± 5.8 (6) | 0.06 | 3.0 ± 10−4 | 0 | 0 | 0 |
| IKS | 1-03948 | RV | 0.6 | 13 ± 4.1 (20) | 0.05 | 1.9 ± 10−3 | 0 | 0 | 0 |
| PHKG1 | 1-00596 | RV | 2.6 | 30 ± 8.7 (47) | 0.09 | 1.5 ± 10−3 | 0 | 0 | 1.6 ± 10−5 |
| FRSB | 1-04119 | PA | 0.3 | 5.4 ± 11.1 (10) | 0.06 | <1.0 ± 10−4 | 0 | 0 | 0 |
| TRMT1B2 | 1-02921 | RV | 0.0 | 5.5 ± 13.2 (20) | 0 | <1.0 ± 10−4 | 0 | 0 | 2.6 ± 10−4 |
| PHKG4 | 1-03488 | RV | 0.6 | 15 ± 8.1 (20) | 0.04 | 4.0 ± 10−2 | 0 | 1.6 ± 10−5 | 0 |
| FGFBP2 | 1-02697 | RV | 0.5 | 14 ± 3.4 (3) | 0.04 | 1.0 ± 10−4 | 0 | 1.6 ± 10−5 | 0 |

AF, allele frequency; AO, aorta; ASE, allele-specific expression; CHD, congenital heart disease; DuctArt, ductus arteriosus; ExAC, Exome Aggregation Consortium; IVS, interventricular septum; LA, left atrium; LOE, loss-of-expression; LOF, loss-of-function; LV, left ventricle; PCGCC, Pediatric Cardiac Genomics Consortium; PV, pulmonary valve; RA, right atrium; RV, right ventricle.

1Expression in reads per million aligned reads.
2P value calculated from z-score.
3Includes de novo LOF mutations.
4Strong candidate CHD genes (identified in this study).
5Frequency of homozygous (ExAC) or combined homozygous and compound heterozygous (PGCG cases and controls) LOF mutations.
6KMT2D de novo mutations identified in both 1-03948 (LOF) and 1-00596 (damaging-missense).
7Only genes PHKG1 (1-03948, P = 2.7 ± 10−4) and FGFBP2 (1-02697, n = 3) are not significant on their own, but occur in genes significant in other subjects.

a recently discovered definitive CHD gene. Although two LOE genes, FGR1 and TRMT2B, have no known role in cardiac development, the known functions of other candidate genes increase the likelihood that altered expression could cause CHD. In addition, we identified two novel findings related to cardiovascular imprinted genes.

There were no significant differences in imprinting between CHD and GTEx subjects, suggesting that imprinting defects are not a common cause of CHD. However, we note the following unexpected observations. First, the imprinted gene RTL1 has maternal ASE in both human and mouse cardiovascular tissues (Table 1 and Supplementary Table 11) but the opposite (paternal) allele is expressed in other fetal and placental tissues24,25. Temporal and spatial-dependent parental allele switching has been observed in only two other imprinted genes, GRB10 and IGFB2 (refs 26,27). Second, RTL1 is the only protein-coding, imprinted gene with maternal ASE in cardiovascular tissues. Imprinted genes in other tissues do not have a significant maternal or paternal bias. In cardiac tissues, imprinted gene clusters were significantly biased (P = 0.008, Fisher Exact test) for paternal ASE (11 A1; 8 biallelic) versus maternal ASE (1 A1; 12 biallelic). Only 5% (1/21) of maternally expressed, imprinted genes had ASE, while 56% of paternally expressed, imprinted genes (14/25) had ASE (P = 3.1 × 10−4, Fisher Exact test). Together, these data suggest a cardiac-specific mechanism that removes imprinting ‘marks’ that would normally silence maternal allele expression.

Although extreme ASE events were identified in both CHD and GTEx subjects, ASE associated with significantly altered gene expression was observed only in CHD subjects (9/37 CHD subjects; 0/57 GTEx subjects; P = 1.1 × 10−4, Fisher Exact test). As genes with extreme ASE and downregulated expression could phenocopy genes harbouring loss-of-function (LOF) mutations in one allele (ignoring potential dominant-negative effects), we compared the LOF allele frequency (AF) between CHD and control data sets. Five of six ASE genes with reduced expression are constrained and have a low frequency (<0.001) of LOF mutations in the ExAC database (Table 3). Three of these, AARSD1, SDHB and Csof46, have no known functions in the heart. The other two genes, RBFX2 and SGSML1, are strong candidates for contributing to CHD. Prior WES analyses identified de novo RBFX2 LOF variants in four CHD probands (including the subject with ASE), and SGSML1 LOF variants in two CHD probands (one each de novo and inherited), but no RBFX2 or SGSML1 LOF variant in 1,800 controls3,4,8. The LOF AF was significantly higher in CHD probands for both RBFX2 (CHD AF = 1.4 × 10−3; ExAC AF = 1.8 × 10−3; odds ratio (OR) = 78; P = 6.6 × 10−5, Fisher Exact test) and SGSML1 (CHD AF = 9.4 × 10−4; ExAC AF = 9.1 × 10−5; OR = 10.4; P = 0.02, Fisher Exact test) than observed in ∼50,000 control exomes. Moreover, the estimated odds ratio for SGSML1 is conservative as we excluded both the subjects with ASE and another CHD subject with markedly reduced biallelic expression (1-02922: 0.19-fold, P = 0.02, calculated from z-score).

FGFBP2, another strong CHD gene candidate, is predicted to lack all gene functions in two CHD subjects, through two different mechanisms. One subject had extreme ASE with a deleterious mutation in the expressed allele (Table 2), while the other subject is predicted to have complete loss of FGFBP2 gene function due to severely reduced biallelic expression (1-02697,
Table 3). Notably, FGFBP2 is a component of the FGF signalling axis that regulates outflow tract and valve morphogenesis28, and both CHD subjects with abrogated FGFBP2 expression have outflow tract defects.

Two probands with abnormal valve development had biallelic LOE of LBH (limb bud and heart) or ZBTB16. LBH, a cardiac developmental transcriptional co-activator, mediates neural crest migration29, which is required for aortic valve formation30. Overexpression of mouse Lbh produces valvular defects and decreases Nppa (atrial natriuretic hormone) expression31. Consistent with this, LOE of LBH in proband 1-03051 was associated with increased NPPA expression (fold = 12.3, \( P < 0.0001 \), \( P \) calculated from \( z \)-score). However, mice lacking Lbh have no overt cardiovascular defects32. ZBTB16 (also known as PLZF) is a member of the Krueppel (C2H2-type) zinc-finger transcription factor family that regulates expression of GATA4 (ref. 33); GATA transcription factors are important for outflow tract development34. ZBTB16 LOE in proband 1-03316 was associated with reduced GATA4 expression (0.04-fold, \( P = 0.19 \), \( P \) calculated from \( z \)-score).

Kabuki syndrome is a complex developmental disorder including CHD caused by mutation in KMT2D, a histone methyltransferase. Our studies identified markedly reduced expression of PHKG1 in two Kabuki syndrome subjects (Table 3) and decreased expression of IRX5 in one subject. (IRX5 expression is normally low in right atrial tissues, the only sample available from subject 1-00596.) Damaging IRX5 mutations (MIM611714 (ref. 35)) cause CHD with conduction abnormalities, marked frontonasal anomalies and cranial malformations. On the basis of these data, we speculate that KMT2D regulation of IRX5 and PHKG1 contributes to the pathogenesis of Kabuki syndrome.

Although we identified nine ASE events likely related to CHD in 144 probands, our analyses have several limitations. First, most of the CHD tissues were acquired after birth, and genes with aberrant expression that are developmentally downregulated would escape detection. Second, only ~25% of cardiac expressed genes contain heterozygous SNPs per sample, so a large fraction of ASE genes cannot be detected by our methodology. We also detected 3.6-fold more ASE genes in subjects genotyped with WGS than subjects genotyped with WES. WES genotyping does not evaluate untranslated region (UTR) sequences and, hence, does not discover SNPs in these regions, resulting in low efficiency of ASE detection in exome-genotyped samples. Finally, to limit false-positives, we employed stringent definitions of ASE (\( \geq 7.2 \) allele ratio; \( P \leq 0.01 \), Bonferroni-corrected binomial distribution) and LOE (fold < 0.1, \( P < 2.7 \times 10^{-3} \), \( P \) calculated from \( z \)-score) genes, at the expense of under-calling expression differences. We expect that cardiac RNA expression from more CHD tissues will explain a larger proportion of disease.

We found that only a small subset of extreme ASE events was attributable to NMD. Hence, unknown mechanisms accounted for allele gain/loss of gene expression in the majority (85%) of extreme ASE events identified in CHD subjects. While we speculate that mutations in regulatory sequences may lead to altered allele-specific transcription, another contributing factor could be somatic mutations expressed in cardiovascular tissues, but not blood (the DNA source for genomic sequencing), that cause NMD.

In summary, integrated analyses of genomic DNA and RNAseq in CHD cardiac tissues identified preferential silencing of paternally expressed imprinted genes, several extreme ASE and LOE genes relevant to cardiogenesis and potential downstream targets of KMT2D. DNA sequence analyses of 81 trios identified nine de novo mutations likely responsible for disease44. Assessment of RNAseq data from these and 63 singletons identified seven instances (RBFOX2, SGSM1 (\( n = 2 \)), FGFBP2 (\( n = 2 \)), LBH and ZBTB16) with significantly reduced gene expression likely contributing to CHD. These data support the use of RNAseq analyses in identifying disease genes. We expect that further study of WGS will identify damaging mutations in regulatory elements that alter transcription of these CHD genes.

Methods

Patient and control cohorts. CHD probands were recruited from nine centres in the United States and the United Kingdom into the Congenital Heart Disease Genetic Network Study of the Pediatric Cardiac Genomics Consortium (CHD Genes: NCT01196182). The protocol was approved by the Institutional Review Boards of Boston Children’s Hospital, Brigham and Women’s Hospital, Great Ormond St Hospital, Children’s Hospital of Los Angeles, Children’s Hospital of Philadelphia, Columbia University Medical Center, Icahn School of Medicine at Mount Sinai, University of Rochester School of Medicine and Dentistry, Steven and Alexandra Cohen Children’s Medical Center of New York and Yale School of Medicine. Written informed consent was obtained from each participating subject or parent/guardian. Probands with CHD were selected based on availability of cardiovascular tissue and RNA quality (RNA integrity number (RIN)). Cardiac RNA was purified from heart tissue, with the number of reads referred to reads containing the alternate or reference base in a heterozygous SNP) between 0.2 and 0.8.

Whole genomes of 11 probands and three trios were sequenced to a mean read depth of 35.2. gDNA libraries were made from 5 μg of purified DNA and sequenced on an Illumina HiSeq 2000 (101 base paired-end reads). Reads were aligned to the reference genome hg19 using Novoalign (Novocraft) and variants called using HaploTypeCaller (Genotype Analysis Toolkit, GATK)36. Variants were filtered using the hard filters (FisherStrand (FS) < 25, quality by depth (QD) < 40) for passing variants. Identified heterogeneous SNPs had a minimum genotype quality score of 50 and an allele balance (AB, number ALT reads/number REF reads + number ALT reads), where where ALT and REF reads refers to reads containing the alternate or reference base in a heterozygous SNP) between 0.2 and 0.8.

Whole genomes of 11 probands and three trios were sequenced to an average read depth of 35.2. gDNA libraries were made from 5 μg of purified DNA and sequenced on an Illumina HiSeq 2000 (101 base paired-end reads). Reads were aligned to the reference genome hg19 using Novoalign (Novocraft) and variants were called using UnifiedGenotyper and filtered by VQSR (GATK)38. Heterozygous SNPs were identified using the same criteria as for WES.

SNP array genotyping. GTEX subjects were genotyped on both Illumina Exome and Illumina 5M arrays.

Variant annotation and minor allele frequency. Variants were annotated using SnpEff37. Damaging missense variants were predicted using both Polyphen2 (ref. 5) and CADD38. Minor AF (MAF) information for each SNP was extracted from the ExAC database, containing > 55,000 individuals. If AF data were unavailable from ExAC, the maximum MAF reported in dbSNP, ExomeVariant Server, HapMap or 1000 Genomes was chosen for subsequent calculations.

RNAseq and analyses. RNA was purified from RNAlater-treated frozen tissue, using Trizol (Life Technologies). RNA (RIN > 5) was converted into cDNA and into RNAseq libraries as described39. In brief, purified poly-A RNA that had gone through two rounds of oligo-dT selection was converted into cDNA and then made into RNAseq libraries. Libraries were sequenced (Illumina HiSeq 2000 or Illumina HiSeq 2500, 50 base paired-end reads) to a target depth of > 20 million reads (median, 57 million reads; range, 20–250 million reads). Reads were aligned to the hg19 reference genome using TopHat 1.3 (using the following parameters: -m 1 -a 5 --segment-mismatches 3 --segment-length 25 -g 0 --no-novel-juncs), with splice junctions being defined by genes.gtf (Illumina iGenome download). Mitochondrial and duplicate reads were discarded using Samtools and Picard’s MarkDuplicates, respectively. A median of 60% of reads was aligned to the reference genome, hg19, and 36% of reads uniquely aligned to the nuclear genome. Allele-specific reads were tallied using GATK UnifiedGenotyper at each heterozygous position identified by gDNA sequencing (using the following parameters: `--genotyping_mode GENOTYPE_GIVEN_ALLELES --alleles het_snps_only.vcf --output_mode EMIT_ALL_SITES` where a personalized vcf file containing only heterozygous SNP was used as `het_snps_only.vcf`). Gene expression was determined by calculating reads per gene per million aligned reads (r.p.m.).
Quality-control metrics for technologies. Some subjects were excluded from our study. Ignoring SNPs with expected ASE (chromosome X genes, previously reported imprinted genes, and SNPs within alternatively spliced exons), we observed that most heterozygous SNPs (with a minimum of 10 reads) were expressed biallelically (CHD exome (95.7 ± 1.0%), CHD WGS (96.8 ± 0.5%) and GTEx (97.6 ± 0.6%; Supplementary Fig. 2a). Eight CHD and three GTEx subjects with substantially lower biallelic SNP expression (< 75%; data not shown) were removed from this study.

Quality-control metrics for SNPs. To ensure accurate ASE identification, all genotyped SNPs observed in RNAseq data were subjected to quality control. Only SNPs with at least five reads in RNAseq data or both alleles expressed were analysed for ASE.

Low-quality genotype called SNPs. Low-quality SNPs observed in either WES or WGS generally reflected either misaligned DNA sequence reads because of gene orthologues, pseudogenes or other highly similar sequences. As previously described, low-quality SNPs either failed GATK variant filtration or had significantly biased AB. In addition, genes with a single heterozygous SNP, expressed at a high level (> 20 reads), with 100% monoallelic expression were excluded; the reference base of these SNPs was expressed in > 80% of cases (Supplementary Fig. 2b), indicating a suspicious genotype.

SNPs in alternatively spliced exons. Bias expression of SNPs in alternative exons could not be evaluated for ASE because we could not differentiate allele-specific expression in specific splicing. That is, SNPs (Supplementary Data 7, annotated ‘Filter SNP_all splicing’) in exons that were only found in a subset of gene isoforms, as represented in three databases (RefGene.txt (Illumina), UCSC Genes and Basic Gene Annotations Set (ENCODE/Gencode)), were excluded.

SNPs with suspected alignment biases. Sequencing reads containing clustered SNPs (> 100 bp of two other SNPs or SNPs in close proximity to an indel (within 30 bp of an indel) were often not aligned by TopHat/Bowtie, introducing artefacts that appear as allele biases; these SNPs were excluded. In addition, SNPs in repeat regions (Supplementary Data 7, annotated ‘Filter SNP_duplicate_sequence’) were excluded if SNP and flanking sequences (50nt up- and downstream) aligned to multiple genomic locations (identified by BLAT (UCSC Genome Browser)) and if the ALT base was the REF base at one of those multiple locations.

SNPs biased in multiple subjects. Some ‘common’ biased SNPs (biased in > 40% of CHD probands or GTEx subjects, with a minimum of three biased subjects) were present in genes that were bi-allelically expressed. These common biased SNPs (except those likely to contribute to NMD) were filtered out as either not likely to impair cardiac development or as technical artifacts.

Quality-control metrics for genes. Genes expressed in an allele-specific manner in many subjects. Excluded genes included all chromosomes X and Y genes, HLA-genes (that is, HLA-A, HLA-B, HLA-C, HLA-DM, HLA-DOA, HLA-DQB, HLA-DM, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DQA2, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-F, HLA-G, HLA-I, HLA-P and HLA-T) and noncoding genes. Coding genes were identified by SNPPEF EFFECT designations: CODON_CHANGE_PLUS, CODON_CHANGE_MINUS, CODONINSERTION, CODONDELETION, NON_CODON_INSERTION, NON_CODON_INSERTION, TIME_SHIFT, NON_SYNONYMOUS_CODING, COMMISSIONER, STOP_GAINED, NON_SYNONYMOUS_CODING, UTR_3.PRIME, UTR_5.PRIME. In addition, genes with ASE in > 5% of subjects, including at least one GTEx subject, are unlikely to impair cardiac development; these ‘common’ ASE genes are reported in Supplementary Table 4.

RNAs with misaligned reads. Heterozygous SNPs were excluded if > 20% of any other heterozygous SNPs in the same transcript were not expressed, or if a heterozygous coding SNP was not expressed. This pattern reflected misaligned reads.

Genes with low fetal heart expression. RNAs that are unlikely to be involved in cardiac development (normalized fetal heart expression < 2 r.p.m.) were filtered out (Supplementary Table 5). This filter removed five and 24 ASE events from CHD probands and GTEx donors, respectively.

Quality-control confirmation of extreme ASE. Allele bias observed in both aligned and unaligned reads (that is fastq files). As noted above, TopHat/Bowtie alignment can introduce apparent allele bias into aligned RNAseq data. To confirm allele bias was not introduced by TopHat/Bowtie alignment, raw, unaligned sequencing reads containing: 10nt flanking the ALT/REF base, or 20nt either upstream or downstream of the ALT/REF base, or their reverse complements, were counted. The numbers of raw sequencing reads containing ALT and REF sequences were required to be similar to the numbers of ALT and REF aligned reads. Further, allele balances were required to be similar in both unaligned and aligned read counts.

Visual inspection of RNAseq data in Integrative Genomics Viewer (Broad Institute). Biased SNPs (Supplementary Data 7) were filtered out if other SNPs with biallelic expression were observed in the same gene—this was particularly important for CHD exome samples, which are not well genotyped in the 3’ UTR and for GTEx samples (in which genotyping was restricted to common SNPs). In addition, biased SNPs were excluded if visual inspection identified more than two alleles. Thirty-nine SNPs (Supplementary Data 7: Filter SNP_complex_allele_structure) had multiple alleles likely due to misalignment of reads from pseudogenes and/or gene families.

Confirmation of ASE events by Sanger sequencing. At least one biased SNP per extreme ASE event was analysed using Sanger sequencing. PCR products derived from RNA were prepared from 200 ng total RNA by incubation with Superscript III Reverse Transcriptase (Thermo Fisher) and then cDNA was PCR-amplified with Phusion polymerase (New England Biolabs) using gene-specific primers (Supplementary Data 6) flanking the SNPs of interest. PCR products were gel-purified using the QIAquick Gel Extraction Kit (QIAgen) and Sanger sequenced (GENEWIZ, Boston). The relative peak heights of the ALT and REF alleles were measured. Extreme ASE events were confirmed when the relative peak height was > 5.

Allele controls removed REF allele bias. Unless the ALT base causes NMD, there should be no REF base versus ALT base expression bias in monoallelic SNPs. Before removal of ‘low-quality’ genome-wide SNPs, 87.1% of ASE events express the REF base, whereas after quality control, ~ 50% of biased SNPs express the REF base (Supplementary Fig. 2b).

Allele bias and ASE P value calculation. Allele bias and ASE P value were calculated for each SNP that passed quality-control measures. If there were multiple SNPs per gene, we either used phased SNP from (mother (mat) and paternal (pat) alleles) or we made the assumption that if there are multiple SNPs in a gene, the expression bias between these SNPs should be unidirectional; the more strongly expressed SNPs with higher expression are on the same allele. Allele bias was calculated as follows:

- Reads containing heterozygous SNPs were counted and binned into one of four categories, based on inheritance and expression: (1) maternal inheritance, (2) paternal inheritance, (3) unknown inheritance with higher-allele expression and (4) unknown inheritance with lower-allele expression.

- In addition, biased SNPs were excluded if visual inspection identified more than two alleles.

In some cases, the expression bias was not strong enough to produce a significant allele bias.

If allele inheritance can be determined:

- IF SNP-mat sum > SNP-pat sum

- allele bias = (SNP-higher-allele sum + SNP-lower-allele sum)/(SNP-lower-allele sum + SNP-pat sum)

- If SNP-pat sum > SNP-mat sum

- allele bias = (SNP-higher-allele sum + SNP-pat sum)/(SNP-higher-allele sum + SNP-lower-allele sum + SNP-pat sum)

If allele inheritance is unknown:

- allele bias = SNP-higher-allele sum/SNP-lower-allele sum

- ASE P value is calculated using a binomial distribution model, and Bonferroni-corrected by the number of genes containing expressed heterozygous SNPs for each sample (Supplementary Data 2 and 4).

The assumption that the more highly expressed bases at heterozygous SNP positions are all on the same allele has the potential to introduce error into allele bias and statistical assessment of ASE. We directly tested this approach by studying F1 crosses of wild-type C57Bl6 and Castaneus mice. On the basis of parental mouse strain genome DNA sequence (Mouse Genomes Project; Welcome Trust Sanger Institute), we estimated that F1 mice would have 18.5 million heterozygous SNPs, encoded within 19,236 transcripts. To make this more comparable to human WGS data, we only assessed every 18th heterozygous SNP. From RNAseq libraries prepared from left and right atrium, left and right ventricle, pulmonary artery, liver and skin, muscle identified from F1 mouse, on average, 14,678 expressed genes including 9,312 expressed genes with heterozygous SNPs (Fig. 2a). We then determined the false-positive rate of assigning unphased SNPs to alleles by comparing fully phased SNPs to unphased SNPs while varying the minimum allele bias (4 to 7-fold) and minimum read depth of monoallelic SNPs (4, 5, ... 10) in each condition (Supplementary Fig. 2c). We determined that an allele bias of 7.2, with a minimum depth of five reads, yielded one false ASE event and 62 ‘true’ ASE events (1.59% false-positive rate). The false-positive rate associated with assigning SNPs to alleles was also assessed in three human WGS trio, using an allele bias of 7.2 and a minimum read depth of five. Fourteen ASE genes were called regardless of whether phasing was used to assign alleles, or whether phasing was ignored.

Imprinted genes. Human genes previously identified as ‘imprinted’ in any tissue were identified in either of two online databases (http://www.geneimprint.com and www.otago.ac.nz/IGC). Genes with very low human fetal heart expression (< 2 r.p.m.) were excluded because they were unlikely to contribute to normal heart development. Allele bias was calculated for each subject with heterozygous SNPs, and ASE events identified as described above.

NMD analyses. NMD was assessed as described previously22,23. That is, the allele carrying the LOF mutation was at least fourfold less than the normal allele and read-depth sufficient to suggest a statistically significant (P < 0.01) difference in allele ratio. P values were Bonferroni-corrected for the number of heterozygous SNPs mutations per subject (and not the number of expressed genes with heterozygous SNPs), so many more ASE events associated with NMD were identified than in our global ASE analyses.
**Expression analyses.** We calculated both r.p.m. and reads per million aligned reads per kilobase of transcript per gene per sample. On average, 15,479 genes (range: 14,938–17,651) were expressed in cardiovascular tissues (folds < 2 r.p.m.; Fig. 1b and Supplementary Table 1). Each sample was compared with the average expression of all other samples of the same tissue type (fold change), and statistical significance was assessed by z-score. As quality control, tissue groups with more than four samples were included in the analyses. Samples with > 100 highly significant expression differences (fold < 0.05 or fold > 5, P < 0.05, P calculated from z-score) were excluded from the analysis.

RNA expression of ASE genes, which demonstrated significant downregulation (fold < 0.05, P < 0.05) or significant upregulation (fold > 5, P < 0.05), were based on the expected fold change of ~ 0.5 (allele loss of expression) and > 7.2 (gain of expression), and relaxed by 30% because of the variability in gender, age and genotypes.

Fold downregulated: 0.5 + (0.5 × 0.3) = 0.65.
Fold upregulated: 7.2 – (7.2 × 0.3) = 5.

For LOE analyses, we required a more stringent definition of significant downregulation (fold < 0.1, P < 2 × 10−3, P calculated from z-score).

This P value is based on a z-score < 3.

Reported LOE events (Table 3) are limited to polyadenylated transcripts because RNAseq libraries were constructed from polyA-selected mRNA.

**Quantitative RT-PCR.** LOE events in CHD subjects, detected by RNAseq analysis, were confirmed using qPCR. cDNA was prepared from 200 ng RNA from the CHD proband and at least two matched tissue samples using Superscript III Reverse Transcriptase (Thermo-Fisher). CDNA was PCR-amplified using Phusion polymerase (New England Biolabs) with SYBR green and gene-specific primers (Supplementary Table 12) and analysed using Fast Real-Time PCR (Applied Biosystems). Housekeeping gene (ACTB, GAPDH, GUSB and PPIA) expression was used to calculate ΔCt, and ΔΔCt and P value (Student’s t-test) are reported in Supplementary Table 10.

**Data availability.** Clinical and sequence data that support the findings of this study have been deposited in dbGaP under accession phs000571. The data that support the findings of this study are available from the corresponding author upon request.

**References**

1. Gellb, B. D. & Chung, W. K. Complex genetics and the etiology of human congenital heart disease. Cold Spring Harb. Perspect. Med. 4, a013953 (2014).
2. Fahed, A. C., Gellb, B. D., Seidman, J. G. & Seidman, C. E. Genetics of congenital heart disease: the glass half empty. Circ. Res. 112, 707–720 (2013).
3. Glessner, J. T. et al. Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. Circ. Res. 115, 884–896 (2014).
4. Zaidi, S. et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature 498, 220–223 (2013).
5. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249 (2010).
6. Blagitko, N. et al. Human IRX510 is imprinted and expressed from the paternal and maternal allele in a highly tissue- and isoform-specific fashion. Hum. Mol. Genet. 9, 1587–1595 (2000).
7. Zhang, J. et al. The FGF-BMP signaling axis regulates outflow tract valve primordium formation by promoting cushion neural crest cell differentiation. Circ. Res. 107, 1209–1219 (2010).
8. Jais, P. et al. Cardiac microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nat. Genet. 34, 261–262 (2003).
9. Baran, Y. et al. The landscape of genomic imprinting across diverse adult human tissues. Genome Res. 25, 927–936 (2015).
10. Vissers, L. E. et al. De novo mutations in congenital heart disease with parent-of-origin effect database. Nature 526, 517–530 (2015).
11. Morison, I. M., Paton, C. J. & Cleverley, S. D. The imprinted gene and transcription is required for establishment of germline primordial germ cell migration via SDF1. Nature 445, 707–720 (2007).
12. Seitz, H. et al. Imprinted microRNA genes activate transcription of a reciprocally imprinted retrotransposon-like gene. Nat. Genet. 34, 261–262 (2003).
13. Glessner, J. T. et al. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).
14. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).
15. Matsuoka, S. et al. Imprinting of the gene encoding a human cyclin-dependent kinase inhibitor, p57KIP2, on chromosome 11p15 Proc. Natl Acad. Sci. USA 93, 3026–3030 (1996).
16. Homsy, J. et al. De novo dynamic, random monoallelic gene expression in mammalian cells. Science 343, 193–196 (2014).
17. Pediatric Cardiac Genomics Consortium et al. The Congenital Heart Disease Genetic Network Study: rationale, design, and early results. Circ. Res. 112, 698–709 (2013).
18. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).
19. Morison, I. M., Paton, C. J. & Cleverley, S. D. The imprinted gene and parent-of-origin effect database. Nucleic Acids Res. 29, 275–276 (2001).
20. Kalscheuer, V. M., Mariman, E. C., Schepens, M. T., Rehder, H. & Ropers, H. H. The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. Nat. Genet. 5, 74–78 (1993).
21. Müller, B. et al. Evidence that paternally expressed gene of the epsilon-sarcoglycan gene accounts for reduced penetrance in myoclonus-dystonia. Annu. J. Hum. Genet. 71, 1303–1311 (2002).
22. MacArthur, D. G. et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science 335, 823–828 (2012).
23. Moros, L. et al. Genome-wide assessment of imprinted expression in human tissues. Genome Biol. 12, R25 (2011).
24. Seitz, H. et al. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nat. Genet. 34, 261–262 (2003).
25. Phillips, H. M. et al. Neural crest cells are required for correct positioning of the developing outflow cushions and pattern the arterial valve leaflets. Cardiovasc. Res. 99, 453–460 (2013).
26. Bonnard, C. et al. Mutations in IRX5 impair craniofacial development and germ cell migration via SDF1. Nat. Genet. 44, 709–713 (2012).
27. Mckenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1239–1250 (2010).
28. Wang, N. et al. Premyelocytic leukemia zinc finger protein activates GATA4 transcription and mediates cardiac hypertrophic signaling from angiotensin II receptor 2. PLoS ONE 7, e35632 (2012).
29. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. 46, 310–315 (2014).
30. Muelschlegel, J. D. et al. Using next-generation RNA sequencing to examine ischemic changes induced by cold blood cardioplegia on the human left ventricular myocardium transcriptome. Anesthesiology 122, 537–550 (2015).

**Acknowledgements**

We thank Anne Davis, Carolyn Westhoff, Paula Castano, Ana Cepin, Patricia Lanzano, Katrina Celis, Liyong Deng, Kelly Sadamatu and Nhu Tran for assistance with tissue collection. This work was supported by grants from the National Heart, Lung, and Blood Institute to the Pediatric Cardiac Genomics Consortium (U01-HL098188, U01-HL098147, U01-HL098153, U01-HL098163, U01-HL098123 and U01-HL098162) and the Cardiovascular Development Consortium (U01-HL098166) and the Howard Hughes Medical Institute (R.L. and C.E.S.), and the John S. LaDue Cardiovascular Fellowship Fund (R.P.L. and C.E.S.), and an Alan Guttmacher Research Award (J.H.). The views expressed are those of the authors and do not necessarily reflect those of the National Heart, Lung, and Blood Institute or the National Institutes of Health.

**Author contributions**

C.E.S., J.G.S. and D.M.M. conceived the study. J.H., S.R.D., J.S.W., S.Z., Y.S. and D.M.M. analysed exome variants. S.R.D. and D.M.M. analysed whole-genome variants. J.G. and C.E.S., J.G.S. and D.M.M. conceived the study. J.H., S.R.D., J.S.W., S.Z., Y.S. and D.M.M. performed mouse work. Ne.P. and D.M.M. made RNAseq libraries. H.W. and D.M.M. performed mouse work. Ne.P. and...
D.M.M. developed RNAseq, ASE and LOE pipelines and analysed ASE data. W.M. and Ni.P. supported data analysis. W.K.C. and R.K. provided human heart tissues. R.P.L., W.K.C., R.K., Y.S., M.B., E.G., A.J.S., Ne.P. and J.H. edited the manuscript. C.E.S., B.D.G., J.G.S. and D.M.M. wrote 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: McKean, D. M. et al. Loss of RNA expression and allele–specific expression associated with congenital heart disease. Nat. Commun. 7:12824 doi: 10.1038/ncomms12824 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016