Human Molecular Genetics, 2012, Vol. 21, No. 7 **1513–1520**
doi:10.1093/hmg/ddr589
Advance Access published on December 22, 2011

Phenotype-specific effect of chromosome 1q21.1 rearrangements and *GJA5* duplications in 2436 congenital heart disease patients and 6760 controls

Rachel Soemedi[1], Ana Topf[1], Ian J. Wilson[1], Rebecca Darlay[1], Thahira Rahman[1], Elise Glen[1], Darroch Hall[1], Ni Huang[2], Jamie Bentham[3], Shoumo Bhattacharya[3], Catherine Cosgrove[3], J. David Brook[4], Javier Granados-Riveron[4], Kerry Setchfield[4], Frances Bu'Lock[5], Chris Thornborough[5], Koenraad Devriendt[6], Jeroen Breckpot[6], Michael Hofbeck[7], Mark Lathrop[8], Anita Rauch[9], Gillian M. Blue[10], David S. Winlaw[10], Matthew Hurles[2], Mauro Santibanez-Koref[1], Heather J. Cordell[1], Judith A. Goodship[1] and Bernard D. Keavney[1,*]

[1]Institute of Genetic Medicine, Newcastle University, Newcastle, UK, [2]Wellcome Trust Sanger Institute, Cambridge, UK, [3]Department of Cardiovascular Medicine, University of Oxford, Oxford, UK, [4]School of Biology, University of Nottingham, Nottingham, UK, [5]East Midlands Congenital Heart Center, Leicester, UK, [6]Center for Human Genetics, University Hospital Leuven, Leuven, Belgium, [7]University Hospital Tubingen, Tubingen, Germany, [8]Commissariat a l'Energie Atomique, Evry Cedex, France, [9]Institute of Medical Genetics, University of Zurich, Zurich-Schwerzenbach, Switzerland and [10]Heart Centre for Children, The Children's Hospital at Westmead, Sydney, Australia

Received October 22, 2011; Revised and Accepted December 10, 2011

Recurrent rearrangements of chromosome 1q21.1 that occur via non-allelic homologous recombination have been associated with variable phenotypes exhibiting incomplete penetrance, including congenital heart disease (CHD). However, the gene or genes within the ~1 Mb critical region responsible for each of the associated phenotypes remains unknown. We examined the 1q21.1 locus in 948 patients with tetralogy of Fallot (TOF), 1488 patients with other forms of CHD and 6760 ethnically matched controls using single nucleotide polymorphism genotyping arrays (Illumina 660W and Affymetrix 6.0) and multiplex ligation-dependent probe amplification. We found that duplication of 1q21.1 was more common in cases of TOF than in controls [odds ratio (OR) 30.9, 95% confidence interval (CI) 8.9–107.6); $P = 2.2 \times 10^{-7}$], but deletion was not. In contrast, deletion of 1q21.1 was more common in cases of non-TOF CHD than in controls [OR 5.5 (95% CI 1.4–22.0); $P = 0.04$] while duplication was not. We also detected rare ($n = 3$) 100–200 kb duplications within the critical region of 1q21.1 in cases of TOF. These small duplications encompassed a single gene in common, *GJA5*, and were enriched in cases of TOF in comparison to controls [OR = 10.7 (95% CI 1.8–64.3), $P = 0.01$]. These findings show that duplication and deletion at chromosome 1q21.1 exhibit a degree of phenotypic specificity in CHD, and implicate *GJA5* as the gene responsible for the CHD phenotypes observed with copy number imbalances at this locus.

INTRODUCTION

Congenital heart disease (CHD) is the most common congenital anomaly, with an incidence of ~7 per 1000 live births (1). CHD may occur as part of recognized chromosomal and Mendelian syndromes (2–4), but in 80% of cases it manifests as a non-syndromic, non-Mendelian condition. Significant familial recurrence risk has been demonstrated in such 'sporadic' CHD

*To whom correspondence should be addressed at: Institute of Genetic Medicine, Newcastle University, Central Parkway, NE1 3BZ Newcastle upon Tyne, UK. Tel: +44 1912418662; Fax: +44 1912418666; Email: b.d.keavney@newcastle.ac.uk

cases, indicating complex genetic contributions to the risk of most CHD (5,6). The most common cyanotic form of CHD is tetralogy of Fallot (TOF), which occurs in ~1 in 2500 live births (1). TOF is characterized by a ventricular septal defect, over-riding of the aortic valve due to anterocephalad deviation of the outlet septum, right ventricular outflow tract obstruction and right ventricular hypertrophy.

Recently, rearrangement hotspots in the human genome which occur via non-allelic homologous recombination (NAHR), and result in recurrent copy number imbalances, have been identified (7,8). One such locus is situated at 1q21.1; copy number variants (CNVs) at this locus have been associated with variable phenotypes exhibiting incomplete penetrance (9–12), including both syndromic (9–11) and non-syndromic (isolated) (11,13) forms of CHD. A recent study by Greenway *et al*. (13) identified 1q21.1 copy number imbalances in 5 out of 512 isolated TOF cases. Although these findings were highly statistically significant when compared with controls ($P = 0.0002$), they as yet await replication for TOF. Moreover, the gene or genes responsible for TOF risk at this locus remain unidentified among the 11 RefSeq genes in the ~1 Mb critical region of 1q21.1. Of these genes, gap junction protein α-5 (*GJA5*, Connexin40, *Cx40*) has previously been proposed as the candidate gene for several cardiac disease phenotypes, including CHD (14–17). Both *Gja5* heterozygous (18%) and homozygous-null (33%) mice exhibit complex heart defects, including conotruncal and endocardial cushion defects (14). However, no *GJA5* point mutation or *GJA5*-specific CNV has been found in CHD patients to date (9,13). We thus examined the 1q21.1 locus in a large cohort of isolated CHD patients ($n = 2436$) in order to estimate more precisely the contribution of 1q21.1 rearrangements to CHD risk, and to identify the causative gene for CHD at this locus.

RESULTS

CNV analyses and concordance between methods of detection

We performed genome-wide CNV analysis on a total of 807 TOF cases and 841 controls using a combination of the PennCNV (18) and QuantiSNP (19) CNV-calling algorithms on data generated using Illumina 660W-Quad chips and found one duplication locus that reached statistical significance for enrichment in TOF versus controls at 1q21.1 (7/807 versus 0/841; $P = 0.007$). A subset of the same TOF patients ($n = 198$) was also analyzed with the Birdseye algorithm on data generated using Affymetrix 6.0 chips. Additionally, we performed targeted multiplex ligation-dependent probe amplification (MLPA) analysis on the 1q21.1 locus in 570 TOF probands, 429 of which were also typed on Illumina 660W, yielding a total of 948 TOF cases examined in this study. A total of 1488 non-TOF CHD cases with clinical diagnoses as shown in Table 1 were analyzed using the PennCNV and QuantiSNP algorithms on Illumina 660W-Quad data; MLPA on the 1q21.1 locus was also performed on 428 of these individuals. When data from the four methods of detection in all cases were merged, we found 100% concordance in CNVs detected in the 1q21.1 locus between the methods used in this study. All CNVs that were detected from the Illumina 660W arrays in individuals that had not also been analyzed with another independent method ($n = 8$) were successfully validated with MLPA.

Table 1. Non-TOF CHD phenotype distribution

Cardiac malformation	*n*
Ventricular septal defect	162
Atrial septal defect	290
Atrioventricular septal defect	63
Transposition of the great arteries	173
Congenitally corrected transposition of the great arteries	37
Common arterial trunk	24
Double-outlet left ventricle	1
PA with ventricular septal defect	18
Double-outlet right ventricle	18
Pulmonary stenosis	79
PA with intact ventricular septum	18
Aortic valve abnormalities	131
Hypoplastic left heart syndrome	15
Mitral valve abnormalities	23
Double inlet left ventricle or right ventricle	23
Ebstein malformation	14
Tricuspid valve abnormalities	32
Aortic arch abnormalities	161
Patent ductus arteriosus	66
Partial anomalous pulmonary venous drainage	19
Total anomalous pulmonary venous drainage	9
Left isomerism	12
Right isomerism	11
Univentricular heart	14
Situs inversus/dextrocardia	5
Heterotaxy	8
Coronary artery anomaly	3
Other	59
Total	1488

Frequency of 1q21.1 rearrangements in control populations

We examined the frequency of NAHR-mediated 1q21.1 rearrangements in 841 individuals from a French population cohort, 5919 WTCCC2 control individuals and 4150 control individuals from previously published studies (8,20,21) that used high-density single nucleotide polymorphism (SNP) platforms comparable to those used in this study. Three duplications and four deletions in 10910 controls were observed (see Supplementary Material, Table S1).

Microduplication of 1q21.1 is strongly associated with sporadic, isolated TOF

We identified NAHR-mediated duplications of 1q21.1 that span the previously reported critical region (9,10) in eight unrelated TOF probands (Fig. 1 and Table 2). The duplications were found to be *de novo* in one proband, inherited from an unaffected mother in three probands and of unknown inheritance (due to unavailability of parental samples) in the remaining four probands. There were no occurrences of 1q21.1 deletion in our TOF cohort. Thus, our findings showed that 1q21.1 microduplications are strongly associated with TOF [8/948 versus 3/10910; $P = 2.2 \times 10^{-7}$; odds ratio (OR) = 30.9, 95% confidence interval (CI) = 8.9–107.6; Table 3], with population attributable risk (PAR) = 0.92%. In contrast,

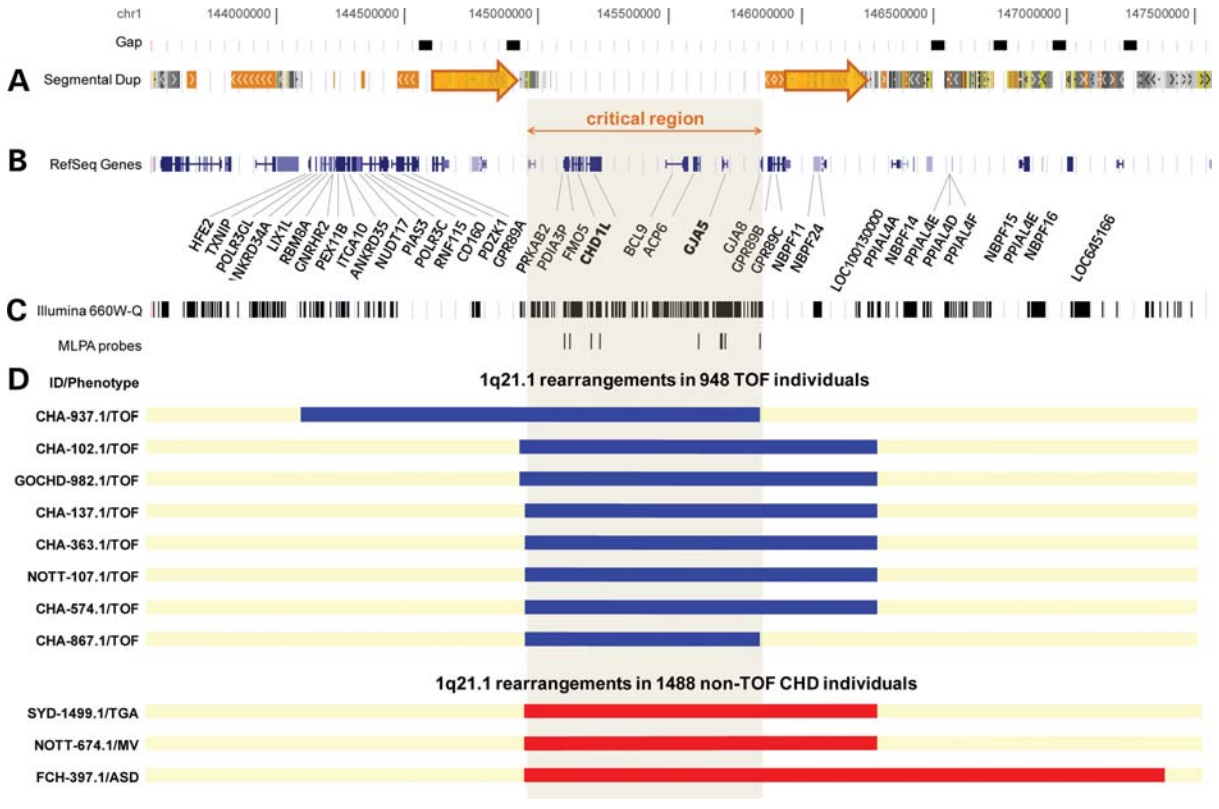

Figure 1. The 1q21.1 region (adaptation from the UCSC hg18 Genome Browser) and the summary of findings in TOF and non-TOF mixed CHD cohort. (**A**) The region of 1q21.1 is complex (143.5–147.5 Mb is shown) due to the presence of extensive segmental duplication blocks and the existing gaps in the reference human genome sequence (NCBI Build 36.1). The largest pair of segmental duplications with >99% homology that mediates most of the rearrangements in this locus is indicated by large orange arrows. (**B**) RefSeq genes in the region. The critical region is indicated by translucent gray block. (**C**) The coverage of the Illumina 660W-Quad (the main platform used in this study) and the location of the custom-designed MLPA probes used in this study are shown. (**D**) Overview of 1q21.1 duplications (blue bars) and deletions (red bars) identified in our study. The cardiac phenotype is shown after the patient identifier. TGA, transposition of the great arteries; MV, mitral valve dysplasia with ventricular septal defect; ASD, atrial septal defect.

we found no evidence that microdeletions of 1q21.1 are associated with TOF (0/948 versus 4/10910; Table 3). Additional phenotypic details of all probands with 1q21.1 CNVs are presented in Supplementary Material, Table S2.

GJA5 duplication is associated with TOF

In addition to NAHR-mediated events, we found 100–200 kb rare duplications within the critical region of 1q21.1 in three patients with TOF. All of these duplications encompass *GJA5*, a strong candidate gene for CHD (14–17). We did not find any deletion within the critical region of 1q21.1 in our TOF cohort. This smaller duplication variant occurred in 2/6760 controls. Thus, small duplications encompassing the *GJA5* gene were also enriched in our TOF cohort compared with controls ($P = 0.01$; OR = 10.7, 95% CI = 1.8–64.3). Additionally, we found a *GJA5* triplication in one patient with pulmonary atresia (PA), like TOF, a cardiac outflow tract phenotype, in the non-TOF CHD cohort (Fig. 2).

Microdeletion of 1q21.1 is associated with isolated congenital heart defects

Examination of the 1q21.1 locus in 1488 cases with other forms of isolated CHD (mixed, non-TOF) revealed three NAHR-mediated deletions and no duplications that spanned the entire critical region (Fig. 1 and Table 2). Thus, 1q21.1 deletion was associated with isolated non-TOF CHD (3/1488 versus 4/10910; $P = 0.04$; OR = 5.5, 95% CI = 1.4–23.2 with PAR = 0.17%, Table 3). In contrast, we found no evidence of association between 1q21.1 duplication and non-TOF CHD (0/1488 versus 3/10910; Table 3). The CHD phenotypes of the cases with 1q21.1 deletion were transposition of the great arteries, atrial septal defect and mitral valve dysplasia with ventricular septal defect (Supplementary Material, Table S2).

DISCUSSION

In 948 isolated TOF cases, we found strong association between duplication at chromosome 1q21.1 and disease risk. We found no association between 1q21.1 deletion and TOF risk. In contrast, among 1488 cases of other isolated CHD phenotypes, we detected association between deletion, rather than duplication, at 1q21.1 and disease risk. Our findings therefore indicate associations between duplication or deletion at this region and CHD that are to a degree specific for particular CHD phenotypes, a novel observation. Furthermore, in the present study, we detected overlapping rare duplication variants of 100–200 kb in size within the critical region of

Table 2. Summary of 1q21.1 CNVs in 2436 CHD patients

Chr	Start	Length (kb)	CN	Patient ID	parental DNA availability	Inheri-tance	Pheno-type	Illumina660 QS	Illumina660 PC	Affy 6.0	MLPA
1	144106312	1742	3	CHA-937.1	–	n/a	TOF	Y	Y	n/a	Y
1	144943150	1350	3	GOCHD-982.1	–	n/a	TOF	Y	Y	n/a	Y
1	144943150	1350	3	NOTT-107.1	P+M	inh-m	TOF	Y	Y	n/a	Y
1	144943150	1350	3	CHA-102.1	–	n/a	TOF	Y	Y	n/a	Y
1	144967972	1325	3	CHA-137.1	P+M	dn	TOF	Y	Y	Y	Y
1	144967972	1325	3	CHA-363.1	P+M	inh-m	TOF	Y	Y	n/a	Y
1	144967972[a]	1321	3	CHA-574.1	M	inh-m	TOF	n/a	n/a	n/a	Y
1	144967972	880	3	CHA-867.1	M	n/a	TOF/PA	Y	Y	n/a	Y
1	145594226	254	3	LEU-30.1	P+M	inh-m	TOF	Y	Y	n/a	Y
1	145594226	254	3	LEU-98.1	–	n/a	TOF	Y	Y	n/a	Y
1	145658466	118	3	CHA-620.1	P+M	inh-m	TOF	Y	Y	Y	Y
1	145658465	118	4	NOTT-319.1	–	n/a	PA	Y	Y	n/a	Y
1	144967972	1419	1	SYD-1499.1	–	n/a	TGA	Y	Y	n/a	Y
1	144967972	1325	1	FCH-397.1	–	n/a	ASD	Y	Y	n/a	Y
1	144967972	1325	1	NOTT-674.1	–	n/a	MV/VSD	Y	Y	n/a	Y

Chr, chromosome; CN, copy number; QS, QuantiSNP; PC, PennCNV; Affy, Affymetrix; Y, yes; n/a, not available; dn, *de novo*; inh-m, inherited from the mother; P, paternal; M, maternal; PA, pulmonary atresia; ASD, atrial septal defect; MV/VSD, mitral valve dysplasia with ventricular septal defect; TGA, transposition of the great arteries.

[a]The proband was typed on the Illumina 660W array but failed SNP QC (low call rate) and thus excluded from the array analyses. However, the mother of the proband was also typed on the Illumina 660W array and passed QC. DNAs from both the proband and mother were analyzed with MLPA, which showed full 1q21.1 duplications with the same breakpoints. This also confirmed the array data from the mother, which passed QC. Thus, the coordinates listed here were inferred from the mother who transmitted the duplication to the respective proband.

Table 3. Phenotypic specificity of 1q21.1 rearrangements in isolated CHD

Patient cohort	1q21.1 microduplications			1q21.1 microdeletions		
	n	*P*-value	OR (95%CI)	*n*	*P*-value	OR (95%CI)
TOF ($n = 948$)	8	2.2×10^{-7}	30.9 (10.2–119.0)	0	NS	–
Non-TOF ($n = 1488$)	0	NS	–	3	0.04	5.5 (1.4–22.0)

NS, not significant.

1q21.1 that are also enriched ($P = 0.01$) in our TOF cohort. These small duplication variants encompass only a single gene in common, i.e. *GJA5*, thus indentifying *GJA5* as a critical CHD gene in this locus.

Chromosome 1q21.1 deletion was first proposed as a cause for CHD by Christiansen *et al*. (11), who found deletions that span the entire critical region of 1q21.1 in one syndromic and two non-syndromic CHD cases among 505 patients referred for clinical genetic evaluation of suspected DiGeorge or Williams' syndrome. All three of the deletion carrying probands had obstruction of the aortic arch as part of their phenotype. However, the specificity of that phenotypic association was likely to have been heavily influenced by selection bias, since aortic arch interruption and supravalvular aortic stenosis are classic cardiovascular manifestations of DiGeorge and Williams' syndromes, respectively. More recently, deletion of 1q21.1 was shown to be present more frequently in patients with variable phenotypes, who had been referred to diagnostic centres principally for mental retardation accompanied by other features, than in controls ($n = 25/5218$ patients; 0/4737 controls; $P = 1.1 \times 10^{-7}$) (9). Twelve of the 25 deletion carriers had CHD as a feature. A second study of 16 557 patients referred to a clinical cytogenetics laboratory who were examined by array comparative genomic hybridization for a range of abnormalities revealed 21 probands with microdeletion and 15 with microduplication. However, only 1 of these 36 patients had CHD without other strong environmental predisposing factors, a patient with a duplication who had a complex heart defect (10). No previous study has estimated the frequency of 1q21.1 rearrangements in patients with mixed CHD phenotypes ascertained on the basis of their heart disease, rather than on the basis of suspected syndromic features; our results demonstrate a modest excess of 1q21.1 deletion in such patients, and no evidence of association with 1q21.1 duplication. In the case of sporadic, isolated TOF, Greenway *et al*. (13) found four 1q21.1 duplications and one 1q21.1 deletion in 512 cases and no occurrence in 2265 controls ($P = 0.007$ and $P = 0.18$ for 1q21.1 duplication and deletion, respectively). Our results confirm that CNV at 1q21.1 is strongly associated with isolated TOF, and in a cohort almost twice as large as that investigated by Greenway *et al*. (13), we demonstrate that duplication is much more strongly associated with TOF than is deletion, for which we found no evidence of association. Interestingly, specificity of 1q21.1 copy number imbalances has been previously described in other associated phenotypes: 1q21.1 duplications but not deletions have been found to be associated with macrocephaly and autism spectrum disorders, while 1q21.1 deletions but not duplications were found to be associated with microcephaly and schizophrenia (10,22).

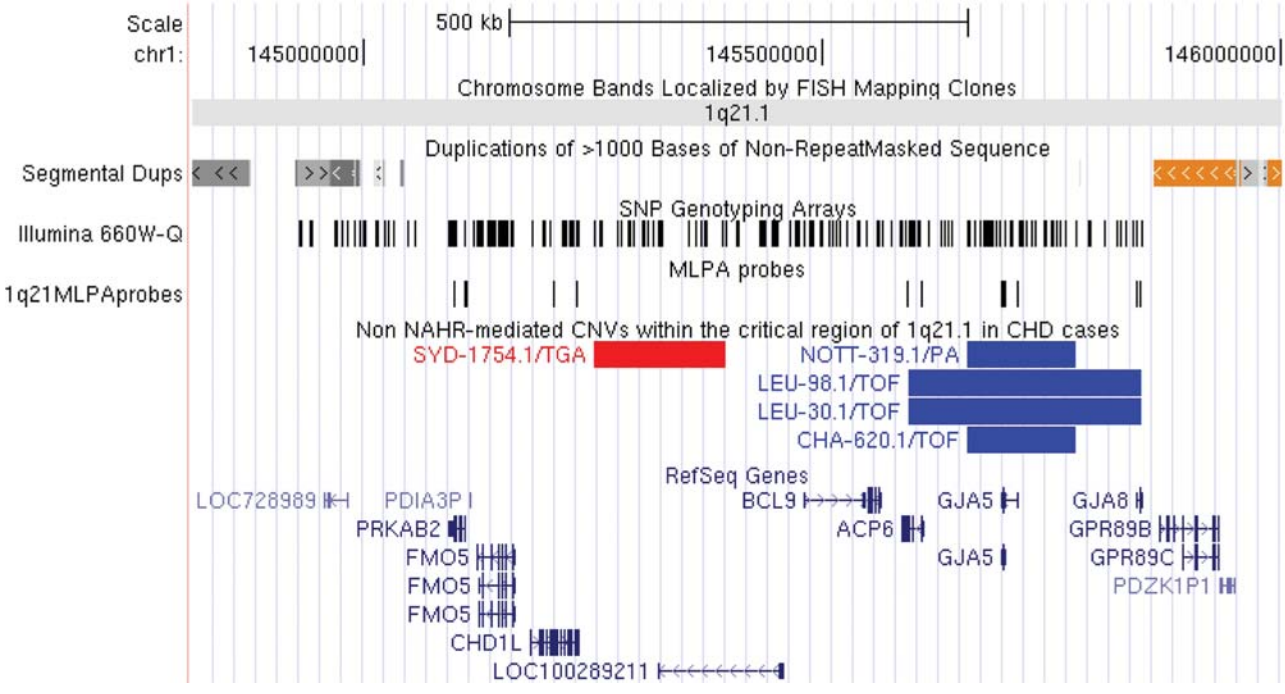

Figure 2. Small duplications encompassing GJA5 within the critical region of 1q21.1. Five CNVs were identified within the critical region of 1q21.1 in 948 TOF and 1488 non-TOF CHD cases. Rare duplications of 100–200 kb in size (shown as blue bars) were found in 3/948 TOF cases encompassing a single gene in common: *GJA5*. Probands LEU-30 and LEU-98 were found to be distantly related, with estimated genome-wide IBD sharing probabilities for sharing (0,1,2) alleles IBD to be (0.8581, 0.1369, 0.0050). However, their estimated IBD-sharing probabilities within the 3 Mb region surrounding *GJA5* are considerably higher (0, 0.64, 0.36), and both of them carry duplications with identical breakpoints. Thus, these two observations are likely to represent one ancestral duplication event. Appropriate correction for the distant relatedness of these two individuals was made in the statistical analyses (see Supplementary Material). In 1488 non-TOF CHD cases, a triplication variant (blue bar) encompassing *GJA5* was found in one patient with PA and one deletion variant (red bar) encompassing the last exon of a non-coding *LOC100289211* gene was found in one patient with transposition of the great arteries (TGA). RefSeq genes in the region, the coverage for the Illumina660W platform and location of the MLPA probes are indicated in the hg18 UCSC Genome Browser (http:// genome.ucsc.edu).

It has been speculated that the 1q21.1 region harbours a single causal gene critical for both cardiovascular and brain development that accounts for both aspects of the rearrangement phenotype, but previous studies had not been able to establish this (9,13). In our TOF cohort, we identified rare smaller duplications within the critical region of 1q21.1, all of which encompass *GJA5*, the strongest candidate gene for the CHD phenotype in this locus. These overlapping small *GJA5* duplications are rare (3/948) in comparison to the NAHR-mediated duplications (9/948) found in TOF cases, but we also found them to be significantly enriched in TOF when compared with controls ($P = 0.01$). With the exception of one patient with PA (a cardiac outflow defect like TOF) who has a *GJA5* triplication, we found no such variant in our non-TOF CHD cohort. Our results therefore suggest duplication of *GJA5* as the most likely mechanism responsible for the association of NAHR-mediated duplication at 1q21.1 and TOF risk. It is not possible to infer directly from our data that *GJA5* deletion is responsible for the association of NAHR-mediated deletion at 1q21.1 and the risk of other forms of isolated CHD, although this seems likely.

GJA5 encodes the cardiac gap junction protein connexin-40, which has key functions in cell adhesion and cell–cell communication. Mice with genetically engineered deletion of *Gja5* have a variety of complex congenital cardiac malformations, in particular of the cardiac outflow tract (14). There are as yet no data from animal models of *Gja5* overexpression, although such data would be of evident interest. However, mice overexpressing *Gja1* (*Cx43*), another connexin isotype, have outflow tract defects (23,24). While the original report of the *Gja5* knockout mouse suggested that *Gja5* was not expressed in neural crest cells in the mouse, more recent work has disputed this finding (14,25). The second heart field plays a critical role in the development of the cardiac outflow tract, and mutations in genes expressed in the second heart field result in outflow tract defects both in mouse models and in humans. *Gja5* was recently shown to be expressed in cells derived from the second heart field during outflow tract development, where it is regulated by the key cardiac transcription factor *Hand2* (25). TOF patients are highly prone to atrial and ventricular arrhythmias in later life which represent a significant source of morbidity and mortality. A number of lines of evidence implicate dysregulation of *GJA5* in atrial arrhythmogenesis (26–29). It would be of interest to determine whether there is differential susceptibility to atrial arrhythmia in TOF patients with and without duplication at 1q21.1 involving *GJA5*.

Our data do not exclude the possibility that other genes in the 1q21.1 region also contribute to CHD risk. Among the possible other candidate genes, *CHD1L* has been shown to be overexpressed in patients with TOF, double-outlet right ventricle and infundibular pulmonary stenosis compared with controls (30). *PRKAB2*, which encodes the β2 subunit of adenosine monophosphate-activated protein kinase, is highly

expressed in the right ventricular outflow tract; mutations in *PRKAG2*, a γ2 subunit of the same protein, have been found to cause some familial forms of hypertrophic cardiomyopathy (31). However, among a total of 2436 CHD cases, we did not find any smaller CNVs within the 1q21.1 region that implicated any gene other than *GJA5,* suggesting that any contribution of such CNVs to CHD risk, while not excluded by our findings, is of small magnitude. Although our findings at *GJA5* are statistically significant ($P = 0.01$), do not require correction for multiple comparisons and are highly biologically plausible, replication of this result in a similarly large and ethnically homogeneous population of TOF patients will be of importance in due course.

As in previous studies, we observed marked variable penetrance of 1q21.1 CNVs. The reasons for this observation remain uncertain. A double-hit model has been previously proposed to explain this variable expressivity (32). However, our power to robustly identify such 'second hits' in the small numbers of cases ($n = 15$) carrying 1q21.1 CNVs in this study is low. Additionally, in all five TOF cases where we observed that the duplication was transmitted from an unaffected parent, transmission was maternal ($P = 0.06$). Although this finding is not significant, it is possible to speculate that parent of origin effects could conceivably in part explain the marked variability in penetrance of cardiac defects with rearrangements in 1q21 that has been observed in several previous studies. A larger study comparing the phenotype when the duplication is paternally or maternally transmitted would be required to address this.

In summary, our study has defined the relationships between 1q21.1 duplication and isolated TOF, and between the reciprocal 1q21.1 deletion and other forms of CHD. Duplication confers an OR for TOF of 31, and accounts for ∼1% of the PAR of TOF, whereas deletion confers an OR and PAR for non-TOF CHD of 6 and 0.2%, respectively. Within the 1q21.1 region, we showed that duplication of *GJA5* alone is associated with an ∼10-fold increase in the risk of TOF, identifying *GJA5* as a critical gene for human heart malformation.

MATERIALS AND METHODS

Study subjects and sample collections

Ethical approval was obtained from the local institutional review boards at each of the participating centres. Informed consent was obtained from all participants (or from a parent/guardian, if the patient was a child too young to themselves consent). Individuals with TOF of Caucasian ancestry together with their parents (when available) were recruited from multiple centres in Newcastle, Leeds, Bristol, Liverpool, Oxford, Nottingham, Leicester (all UK), Leuven (Belgium), Erlangen (Germany) and Sydney (Australia). Individuals with other forms of CHD of Caucasian ancestry were also recruited in Newcastle, Oxford, Nottingham, Leicester, Leuven and Sydney. Patients that exhibited clinical features of syndromic developmental abnormalities or learning difficulties were excluded from the study. Both adult and child patients were recruited. CHD diagnoses and the presence of any associated phenotypes were obtained from the medical record. Parents of cases not reporting a diagnosis of CHD did not undergo echocardiography. All DNAs were further screened for DiGeorge and Williams abnormalities and those found with such anomalies were excluded from the study. Control subjects consisted of 856 individuals of Caucasian ancestry free of reported CHD from a French population cohort and 6000 individuals of Caucasian ancestry free of reported CHD from the Wellcome Trust Case-Control Consortium 2 (WTCCC2) control cohort (3000 individuals from 1958 British Birth Cohort and 3000 individuals from UK Blood Service). Controls did not undergo echocardiography. DNAs from cases were extracted from blood or saliva. DNAs from the French population cohort and UK Blood Service were blood-derived and DNAs from 1958 British Birth Cohort were cell line derived.

Genotyping and quality-control criteria

Genotyping for all CHD probands and the French control cohort was performed at the Centre National de Génotypage (Evry Cedex, France). Normalized intensity and genotype data were obtained from the genotyping centre. TOF patients and family members (907 affected offspring and 747 unaffected parents), 1987 non-TOF CHD patients and 856 unrelated individuals from a French control cohort were genotyped on the Illumina 660W-Quad platform. Additionally, 206 of the same TOF individuals were also typed on Affymetrix 6.0 arrays. Per sample SNP QC analyses were performed to exclude any duplicates, individuals with non Caucasian ancestry and gender mismatches. Samples with low call rate (<98.8%) and high heterozygosity were excluded. Additional intensity quality-control parameters were applied to minimize heterogeneity due to multi-centre variation in DNA source and treatment. Samples were excluded when they failed the following QC criteria: (i) a standard deviation of autosomal log R ratio (LRR) values >3.0; (ii) GC wave factor of the LRR outside the range of $-0.1 < X < 0.1$ (18); (iii) a standard deviation of B allele frequency values >0.15 after GC correction (19). Following these QC measures, 807 TOF patients and 697 of their family members (8 affected and 689 unaffected), 1488 patients with non-TOF CHD and 841 control individuals from the French cohort were included in the final analyses. The phenotype classification summary of non-TOF CHD individuals can be found in Table 1. WTCCC2 controls were typed on Affymetrix 6.0 arrays. Further details on genotyping and QC criteria ($n = 5919$ passed QC) can be found in a previously published study (33) (http://www.wtccc.org.uk/ccc2).

CNV calling algorithms and verification criteria

CNV detection on the Illumina 660W platform was performed using both PennCNV (18) and QuantiSNP (19) CNV-calling algorithms. CNV detection on the Affymetrix 6.0 platform was performed using the Birdseye algorithm from the Birdsuite (34) package. The number of markers used to call CNV within the ∼1 Mb minimal region of NAHR-mediated events in the 1q21.1 locus on the Illumina 660W and Affymetrix 6.0 platforms are 235 and 640, respectively. Within the minimal region (∼100 kb) of small duplications that encompass *GJA5*, 36 and 104 markers were available in the Illumina 660W and Affymetrix 6.0 platforms, respectively. CNV calls

within the 1q21.1 locus that appeared to be artificially split as the result of coverage gaps on the SNP platforms were examined manually and joined (see Supplementary Material, Fig. S1 and S2). All CNV calls with Bayes factor >50 were confirmed with MLPA and Affymetrix 6.0 array (19). CNV coordinates were based on NCBI build 36.1 (hg18).

MLPA design, assay and analysis

We performed MLPA analyses of the critical region of 1q21.1 in 574 TOF probands (433 of which were also typed on the Affymetrix 6.0 and/or Illumina 660W arrays) and 433 non-TOF CHD probands (all were also typed on the Illumina 660W arrays). MLPA probes were designed with MAPD (35) software, using the default settings. The resulting list of candidate probes was subjected to BLAT (36) search in order to ensure specificity and to obtain genomic positions. Candidate probes that overlapped known SNPs (37) and/or segmental duplications (38) [identified by using UCSC Extended DNA utility (39,40)] were excluded. The remaining candidate probes with the highest score given by the MAPD software that met all the MRC Holland guidelines (http://www.mrc-holland.com) were chosen for synthesis. Twenty probes (ranging from 100–140 bp final product size) that targeted *GJA5* (10), *CHD1L* (2), *ACP6* (2), *GJA8* (2), *PRKAB2* (2), *TXNIP* (1) and *ANKRD34A* (1) were synthesized. Nine 1q21 probes were used for each MLPA assay in addition to two control synthetic probes targeting copy number neutral regions. All probe sequences are available upon request.

All MLPA (41) assays were performed with reagents from the MRC Holland P200 kit (Amsterdam, the Netherlands) and custom design synthetic oligonucleotide probes (Integrated DNA Technology, IA, USA). Genomic DNA (100 ng in 5 μl TE) was denatured for 30 min at 95°C and cooled down to 25°C before addition of 35 fmoles of synthetic custom design probes, 1 μl of P200 probe mix and 1.5 μl of MLPA buffer. This was followed by hybridization at 60°C for 16 h. Hybridized probes were ligated with 1 U of Ligase-65 for 15 min at 54°C, followed by ligase deactivation for 5 min at 98°C. Afterwards, 5 μl of the ligated products were added to 15 μl of polymerase chain reaction (PCR) buffer 2:13 dilution in H_2O at 4°C, and the temperature was raised to 60°C before the remaining PCR reagents (2.5 nmoles of dNTPs, 10 pmol fluorescein amidite-labelled universal primers and 2.5 U of SALSA polymerase) were added to make the final reaction volume to 25 μl. PCR reaction was carried out in 33 cycles (95°C for 30 s, 60°C for 30 s and 72°C for 1 min), followed by a final extension at 72°C for 20 min. The final MLPA products were subsequently resolved on ABI 3730xl (Applied Biosystems, CA, USA), using the MRC Holland protocol (http://www.mrc-holland.com). The data were analyzed with the GeneMarker v.1.85 software (SoftGenetics, PA, USA), in which population normalization was applied, and the peak areas were used to calculate relative dosage.

Incorporation of control data from previously published studies

In addition to 841 control individuals from the French cohort and 5919 WTCCC2 controls, we examined the frequency of 1q21.1 rearrangements in 4150 control individuals from previously published studies (8,20,21) that used high-density SNP platforms comparable to the platforms used in this study (with coverage of >200 probes in the critical region of 1q21.1). For details, see Supplementary Material, Table S1.

Statistical analysis

Two-sided Fisher's exact tests were used to compare the frequency of NAHR-mediated 1q21.1 duplications/deletions. Stata 11 was used to compute *P*-values, OR and 95% CI by Cornfield approximation. The frequency of small *GJA5* duplications was compared by maximum likelihood estimation using two binomial distributions (details can be found in Supplementary Material). PAR was calculated using the formula: $100(P(OR-1))/(1+(P(OR-1)))$, in which *P* is the proportion of control population with the 1q21.1 duplication/deletion. Genomewide identity-by-descent (IBD) sharing was calculated using 41692 autosomal SNPs in PLINK (42). IBD sharing for the 1q21.1 locus was calculated using 268 SNPs in a roughly 3 Mb region around *GJA5*.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank all the patients and families who contributed to the study and Sr. Linda Sneddon for assistance with family collection.

Conflict of Interest statement. None declared.

FUNDING

This study was funded by the British Heart Foundation (BHF), European Community's 7th Framework Programme contract ('CHeartED'), Heart Research UK, The Federated Foundation and the Wellcome Trust. B.D.K. is supported by a BHF personal chair. Funding to pay the Open Access publication charges for this article was provided by the British Heart Foundation.

REFERENCES

1. Hoffman, J.I. and Kaplan, S. (2002) The incidence of congenital heart disease. *J. Am. Coll. Cardiol.*, **39**, 1890–1900.
2. Ryan, A.K., Goodship, J.A., Wilson, D.I., Philip, N., Levy, A., Seidel, H., Schuffenhauer, S., Oechsler, H., Belohradsky, B., Prieur, M. *et al.* (1997) Spectrum of clinical features associated with interstitial chromosome 22q11 deletions: a European collaborative study. *J. Med. Genet.*, **34**, 798–804.
3. Freeman, S.B., Taft, L.F., Dooley, K.J., Allran, K., Sherman, S.L., Hassold, T.J., Khoury, M.J. and Saker, D.M. (1998) Population-based study of congenital heart defects in Down syndrome. *Am. J. Med. Genet.*, **80**, 213–217.
4. Wessel, A., Pankau, R., Kececioglu, D., Ruschewski, W. and Bursch, J.H. (1994) Three decades of follow-up of aortic and pulmonary vascular lesions in the Williams-Beuren syndrome. *Am. J. Med. Genet.*, **52**, 297–301.

5. Burn, J., Brennan, P., Little, J., Holloway, S., Coffey, R., Somerville, J., Dennis, N.R., Allan, L., Arnold, R., Deanfield, J.E. *et al.* (1998) Recurrence risks in offspring of adults with major heart defects: results from first cohort of British collaborative study. *Lancet*, **351**, 311–316.
6. Oyen, N., Poulsen, G., Boyd, H.A., Wohlfahrt, J., Jensen, P.K. and Melbye, M. (2009) Recurrence of congenital heart defects in families. *Circulation*, **120**, 295–301.
7. Mefford, H.C. and Eichler, E.E. (2009) Duplication hotspots, rare genomic disorders, and common disease. *Curr. Opin. Genet. Dev.*, **19**, 196–204.
8. Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I. *et al.* (2009) Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.*, **84**, 148–161.
9. Mefford, H.C., Sharp, A.J., Baker, C., Itsara, A., Jiang, Z., Buysse, K., Huang, S., Maloney, V.K., Crolla, J.A., Baralle, D. *et al.* (2008) Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.*, **359**, 1685–1699.
10. Brunetti-Pierri, N., Berg, J.S., Scaglia, F., Belmont, J., Bacino, C.A., Sahoo, T., Lalani, S.R., Graham, B., Lee, B., Shinawi, M. *et al.* (2008) Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat. Genet.*, **40**, 1466–1471.
11. Christiansen, J., Dyck, J.D., Elyas, B.G., Lilley, M., Bamforth, J.S., Hicks, M., Sprysak, K.A., Tomaszewski, R., Haase, S.M., Vicen-Wyhony, L.M. *et al.* (2004) Chromosome 1q21.1 contiguous gene deletion is associated with congenital heart disease. *Circ. Res.*, **94**, 1429–1435.
12. Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.Q., Vincent, J.B., Skaug, J.L., Thompson, A.P., Senman, L. *et al.* (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.*, **39**, 319–328.
13. Greenway, S.C., Pereira, A.C., Lin, J.C., DePalma, S.R., Israel, S.J., Mesquita, S.M., Ergul, E., Conta, J.H., Korn, J.M., McCarroll, S.A. *et al.* (2009) De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat. Genet.*, **41**, 931–935.
14. Gu, H., Smith, F.C., Taffet, S.M. and Delmar, M. (2003) High incidence of cardiac malformations in connexin40-deficient mice. *Circ. Res.*, **93**, 201–206.
15. de Wit, C., Roos, F., Bolz, S.S., Kirchhoff, S., Kruger, O., Willecke, K. and Pohl, U. (2000) Impaired conduction of vasodilation along arterioles in connexin40-deficient mice. *Circ. Res.*, **86**, 649–655.
16. Nao, T., Ohkusa, T., Hisamatsu, Y., Inoue, N., Matsumoto, T., Yamada, J., Shimizu, A., Yoshiga, Y., Yamagata, T., Kobayashi, S. *et al.* (2003) Comparison of expression of connexin in right atrial myocardium in patients with chronic atrial fibrillation versus those in sinus rhythm. *Am. J. Cardiol.*, **91**, 678–683.
17. Lamarche, J., O'Hara, G., Philippon, F., Gilbert, M. and Daleau, P. (2001) Molecular analysis of connexin 40 in the familial form of atrial fibrillation. *Eur. Heart J.*, **22**, 1511–1512.
18. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H. and Bucan, M. (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.*, **17**, 1665–1674.
19. Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C. and Ragoussis, J. (2007) QuantiSNP: an objective Bayes Hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic Acids Res.*, **35**, 2013–2025.
20. Shaikh, T.H., Gai, X., Perin, J.C., Glessner, J.T., Xie, H., Murphy, K., O'Hara, R., Casalunovo, T., Conlin, L.K., D'Arcy, M. *et al.* (2009) High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res.*, **19**, 1682–1690.
21. McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A. *et al.* (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.*, **40**, 1166–1174.
22. Crespi, B., Stead, P. and Elliot, M. (2010) Evolution in health and medicine Sackler colloquium: comparative genomics of autism and schizophrenia. *Proc. Natl Acad. Sci. USA*, **107**(Suppl. 1), 1736–1741.
23. Ewart, J.L., Cohen, M.F., Meyer, R.A., Huang, G.Y., Wessels, A., Gourdie, R.G., Chin, A.J., Park, S.M., Lazatin, B.O., Villabon, S. *et al.* (1997) Heart and neural tube defects in transgenic mice overexpressing the Cx43 gap junction gene. *Development*, **124**, 1281–1292.
24. Kirchhoff, S., Kim, J.S., Hagendorff, A., Thonnissen, E., Kruger, O., Lamers, W.H. and Willecke, K. (2000) Abnormal cardiac conduction and morphogenesis in connexin40 and connexin43 double-deficient mice. *Circ. Res.*, **87**, 399–405.
25. Holler, K.L., Hendershot, T.J., Troy, S.E., Vincentz, J.W., Firulli, A.B. and Howard, M.J. (2010) Targeted deletion of Hand2 in cardiac neural crest-derived cells influences cardiac gene expression and outflow tract development. *Dev. Biol.*, **341**, 291–304.
26. Gollob, M.H., Jones, D.L., Krahn, A.D., Danis, L., Gong, X.Q., Shao, Q., Liu, X., Veinot, J.P., Tang, A.S., Stewart, A.F. *et al.* (2006) Somatic mutations in the connexin 40 gene (GJA5) in atrial fibrillation. *N. Engl. J. Med.*, **354**, 2677–2688.
27. Wirka, R.C., Gore, S., Van Wagoner, D.R., Arking, D.E., Lubitz, S.A., Lunetta, K.L., Benjamin, E.J., Alonso, A., Ellinor, P.T., Barnard, J. *et al.* (2011) A common connexin-40 gene promoter variant affects connexin-40 expression in human atria and is associated with atrial fibrillation. *Circ. Arrhythm. Electrophysiol.*, **4**, 87–93.
28. Liu, X., Wang, F., Knight, A.C., Zhao, J. and Xiao, J. (2011) Common variants for atrial fibrillation: results from genome-wide association studies. *Hum. Genet*. [Epub ahead of print].
29. Sinner, M.F., Ellinor, P.T., Meitinger, T., Benjamin, E.J. and Kaab, S. (2011) Genome-wide association studies of atrial fibrillation: past, present, and future. *Cardiovasc. Res.*, **89**, 701–709.
30. Morano, M., Zacharzowski, U., Maier, M., Lange, P.E., Alexi-Meskishvili, V., Haase, H. and Morano, I. (1996) Regulation of human heart contractility by essential myosin light chain isoforms. *J. Clin. Invest.*, **98**, 467–473.
31. Oliveira, S.M., Ehtisham, J., Redwood, C.S., Ostman-Smith, I., Blair, E.M. and Watkins, H. (2003) Mutation analysis of AMP-activated protein kinase subunits in inherited cardiomyopathies: implications for kinase function and disease pathogenesis. *J. Mol. Cell Cardiol.*, **35**, 1251–1255.
32. Girirajan, S., Rosenfeld, J.A., Cooper, G.M., Antonacci, F., Siswara, P., Itsara, A., Vives, L., Walsh, T., McCarthy, S.E., Baker, C. *et al.* (2010) A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat. Genet.*, **42**, 203–209.
33. Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., 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**, 713–720.
34. Korn, J.M., Kuruvilla, F.G., McCarroll, S.A., Wysoker, A., Nemesh, J., Cawley, S., Hubbell, E., Veitch, J., Collins, P.J., Darvishi, K. *et al.* (2008) Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat. Genet.*, **40**, 1253–1260.
35. Zhi, J. and Hatchwell, E. (2008) Human MLPA Probe Design (H-MAPD): a probe design tool for both electrophoresis-based and bead-coupled human multiplex ligation-dependent probe amplification assays. *BMC Genomics*, **9**, 407.
36. Kent, W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
37. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
38. Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J. and Eichler, E.E. (2001) Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.*, **11**, 1005–1017.
39. Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. and Kent, W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.
40. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
41. Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F. and Pals, G. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.*, **30**, e57.
42. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.