Original Article

Design and rationale of a genetic cohort study on congenital cardiac disease: experiences from a multi-institutional platform in Quebec

Marie-Pierre Dubé,[1] Jean-Luc Bigras,[2] Maryse Thibeault,[2] Nathalie Bureau,[2] Philippe Chetaille,[3] Andrea Richter,[2] Jocelyne Mercier,[2] Marc Bellavance,[4] Charles Rohlicek,[5] Rima Rozen,[5] Mona Nemer,[6] Paul Khairy,[1] Roxanne Gendron,[2] Gregor Andelfinger[2]

[1]*Montreal Heart Institute, University of Montreal, 5000, Rue Belanger;* [2]*Department of Pediatrics, Centre Hospitalier Universitaire Sainte Justine, University of Montreal, 3175 Chemin Côte Sainte Catherine, Montréal;* [3]*Cardiology Service, Centre Mère-Enfants, Centre Hospitalier Universitaire de Quebec, 2705, boulevard Laurier, Québec;* [4]*Department of Pediatrics, University of Sherbrooke, 3001, 12e Avenue Nord, Sherbrooke, QC;* [5]*Department of Pediatrics, McGill University, Montréal, 3400, boulevard De Maisonneuve Ouest, Montreal, QC;* [6]*University of Ottawa, 550, rue Cumberland, Ottawa, ON, Canada*

Abstract *Background:* Congenital cardiac disease is the most common malformation, and a substantial source of mortality and morbidity in children and young adults. A role for genetic factors is recognised for these malformations, but overall few predisposing loci have been identified. Here we report the rationale, design, and first results of a multi-institutional congenital cardiac disease cohort, assembled mainly from the French-Canadian population of the province of Quebec and centred on families with multiple affected members afflicted by cardiac malformations. *Methods:* Families were recruited into the study, phenotyped and sampled for DNA in cardiology clinics over the first 3 years of enrolment. We performed segregation analysis and linkage simulations in the subgroup of families with left ventricular outflow tract obstruction (LVOTO). *Results:* A total of 1603 participants from 300 families were recruited, with 169 out of 300 (56.3%) families having more than one affected member. For the LVOTO group, we estimate heritability to be 0.46–0.52 in our cohort. Simulation analysis demonstrated sufficient power to carry out linkage analyses, with an expected mean log-of-odds (LOD) score of 3.8 in 67 pedigrees with LVOTO. *Conclusion:* We show feasibility and usefulness of a population-based biobank for genetic investigations into the causes of congenital cardiac disease. Heritability of LVOTO is high and could be accounted for by multiple loci. This platform is ideally suited for multiple analysis approaches, including linkage analysis and novel gene sequencing approaches, and will allow to establish segregation of risk alleles at family and population levels.

Keywords: Cardiac malformations; population study; biobanking; heritability analysis; power simulations

Received: 9 August 2010; Accepted: 28 March 2011; First published online: 4 July 2011

IN NORTH AMERICA, THE CONSERVATIVE ESTIMATE OF congenital cardiac disease prevalence varies from 4 to 50 per 1000 live births.[1] Congenital cardiac disease accounts for approximately 33% of all birth defects, the single largest category of malformations in Canada, and is the major cause of death among congenital diseases in infants under 1 year of age.[2,3] The impressive progress of the last few decades in the fields of imaging, surgical intervention, and post-operative care for congenital cardiac disease led to a substantial increase in survival, even for complex subtypes. At present, the population of adults with congenital cardiac disease is estimated to exceed that of paediatric patients in absolute numbers.[4] In addition, these survivors may now found their own

Correspondence to: Dr G. Andelfinger, MD, Research Center, Room 2724, CHU Sainte Justine, 3175, Côte Sainte Catherine, Montréal, Québec H3T 1C5, Canada. Tel: 514 345 4931x3244; Fax: 514 345 4896; E-mail: gregor.andel finger@recherche-ste-justine.qc.ca

families in increasing numbers and thus be at risk of passing on genetic predispositions to disease.

In contrast to advances in medical and surgical treatment, relatively little is known about the causes of congenital cardiac disease. To date, aetiology of congenital cardiac disease can be established only in a minority of patients. Epidemiologic studies point to a major role of genetic and epigenetic factors in the pathogenesis of these conditions.[5–7] Evidence is primarily derived from observational studies, estimates of recurrence risks, and formal segregation and heritability analysis. These uniformly suggest that strong genetic factors are implicated in the pathogenesis of congenital cardiac disease.[8–13]

Malformations that cause left ventricular outflow tract obstruction (LVOTO) are the best-studied traits in this regard. Those malformations include aortic stenosis (sub- or supravalvular), bicuspid aortic valve and other morphologic anomalies of the aortic valve, mitral stenosis, coarctation of the aorta, and hypoplastic left heart syndrome. Excluding bicuspid aortic valve, LVOTO still accounts for 10% of all congenital cardiac disease patients in larger population-based studies.[14] This prevalence is lower than that of septal defects; however, the high morbidity and mortality of LVOTO motivates our emphasis on this patient group. Earlier investigations of this disease spectrum did not systematically use echocardiography but instead relied on recurrence of similar conditions among relatives. This pointed to increased recurrence risks and high concordance of phenotypes within families.[15–17] More recently, routine echocardiography screening of family members with LVOTO and heritability analyses have demonstrated a predominant role for genetic factors in this disease spectrum.[12,18,19] These initial studies were successfully followed by linkage analyses identifying significantly linked regions for bicuspid aortic valve, aortic stenosis, hypoplastic left heart syndrome, and other forms of LVOTO.[20–23]

Parallel efforts have been undertaken through candidate gene studies in patient cohorts with selected malformations. Such studies are informed through the discovery of morphogenetic pathways in model systems such as the mouse or zebrafish. As an example, a detailed survey of genes of the *NODAL* pathway, implicated in left–right asymmetry, has shown that combinatorial effects of rare alleles in this cascade lead to congenital cardiac disease.[24]

In addition to the role of genetic and epigenetic factors for a particular trait, understanding population genetics in terms of population size, history, and structure should be taken into consideration in interpreting genetic epidemiological studies.[25] For the French-Canadian population, a major feature is a founder effect: the fact that a relatively small number of individuals gave rise to a major portion of today's large population. In Quebec, slightly more than 8500 individuals of French origin left their descendents, starting in 1608 – with the founding of the New France – to 1759 – the conquest of Quebec by the British. Their descendants now account for 80% of the current population of over 7.5 million. Founder effects are a recognised advantage in tracing the genes responsible for mono- and oligogenic diseases. In addition, recent immigration trends do contribute to intra-population genetic diversity that could influence monogenic and complex traits.[26–28] Several mutations causing known genetic diseases have been described to be specific to French-Canadians and related to their population history.[29–31] To facilitate access to extended French-Canadian family genealogies and population histories, the BALSAC project, a computerised database of French-Canadians from the 17th century up to recent years, is available. This resource serves to automatically reconstruct ascendant and descendant genealogies for the purpose of establishing multi-generational pedigrees dating back to possible common ancestors.[32]

As a necessary prelude to the identification of predisposing alleles, we established a congenital heart disease biobank in 2006. We deliberately concentrated on multiplex families with the initial focus on LVOTO. Here we report the design, rationale, and first results of genetic analysis for this multi-institutional, family-based cohort after 3 years of recruitment.

Methods

Study design

Recruitment of a family-based cohort was set up in four participating tertiary paediatric hospitals throughout the province of Quebec. These centres cover the provincial territory, and all paediatric congenital cardiac disease cases in the province are referred to one of these. The research platform consists of establishing a database of phenotypic and genotypic information for further genetic studies of congenital cardiac disease; a flow chart is shown in Figure 1. Cardiologists in each setting were asked to inform their patients and families of the project. Interested families were referred to the research team that coordinates recruitment and data collection. The latter includes consent, individual medical history, familial history, phenotyping, and DNA sampling. Participation was limited to one visit, without follow-up. However, participants were invited to provide new familial or clinical information deemed relevant to the research project, for example, the birth of an affected individual within a family.

For the purpose of extending small pedigrees to multiplex families, we used an iterative screening
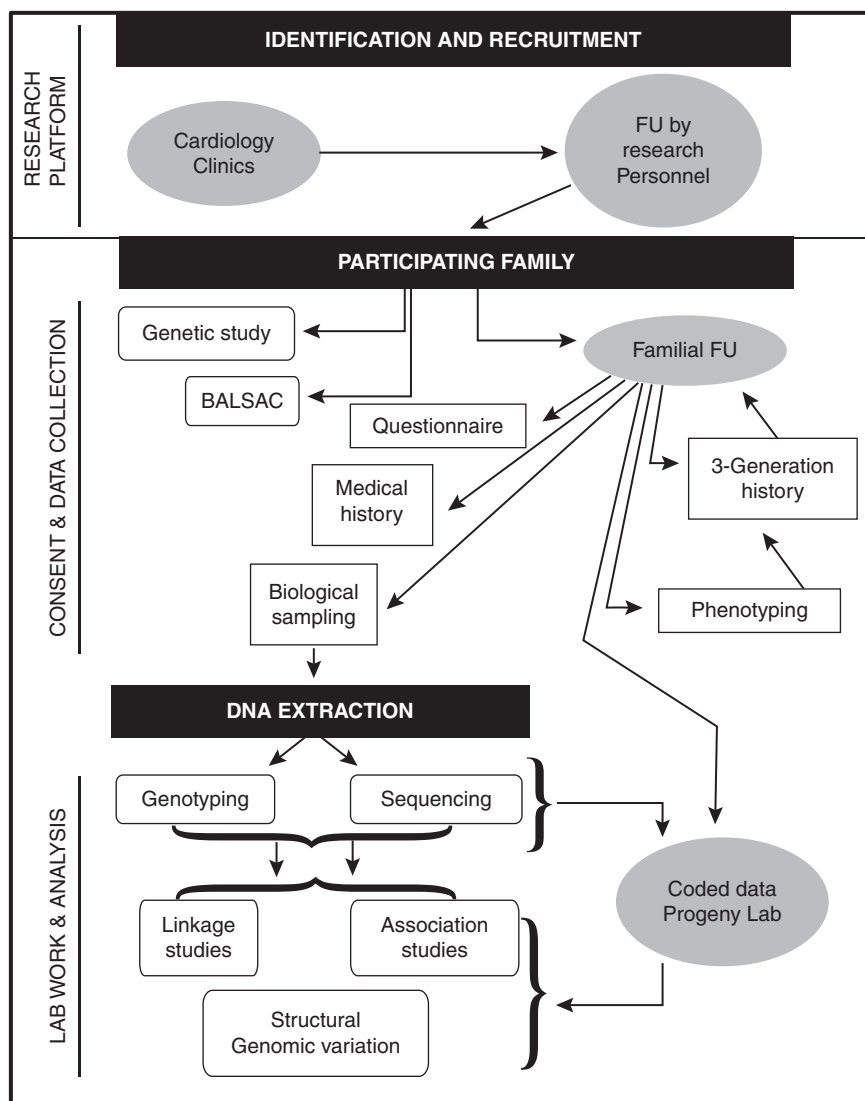
Figure 1.
Schematic depiction of the organisation and workflow of our provincial platform for research into the genetic determinants of congenital heart malformations.

strategy; based on pedigree information in at least three generations, we attempted to recruit all affected family members and their first-degree relatives. For every newly identified proband, we attempted to recruit all living first-degree relatives.

In addition, a cohort of adult controls was recruited to permit case-matched analyses. Considering that research is mainly driven by genetic hypotheses, controls were enrolled in view of *a posteriori*-based matching, for example, sex, age, and ethnicity.

Enrolment of subjects

Subjects were identified through treating cardiologists at participating centres. Inclusion and exclusion criteria are summarised in Table 1. Notably, no age restrictions were imposed, with the exception of a study on genetic

variants of the folate pathway.[33] Recruitment and extension of families remains an ongoing process, with new cases continuously screened and newly identified affected family members reevaluated within the family context.

The study protocol was approved by institutional review boards and is subject to annual review. All subjects provided informed consent. Participants were free to authorise or decline reconstitution of their genealogy using the BALSAC database without compromising participation in the core project.

- *Cases.* Recruitment was mainly targeted towards multiplex families, defined by at least two family members with congenital cardiac defects. To facilitate management of the database, congenital cardiac diseases were grouped arbitrarily into four

Table 1. Inclusion and exclusion criteria.

	Cases	Controls
Inclusion criteria		
Children	Diagnosed congenital cardiac disease	Congenital cardiac disease suspicion invalidated
Adults[*]	Siblings of a case	Healthy siblings of a control
	Diagnosed congenital cardiac disease	First-degree relatives of a control
	First-degree relatives of patients with congenital cardiac disease	Healthy married-in in a congenital cardiac disease family**
Exclusion criteria		
Children	Known genetic disorder	Patent foramen ovale
	Syndromic condition	Other structural cardiac disease
Adults[*]	Known genetic disorder	Patent foramen ovale
	Syndromic condition	Siblings or offspring with congenital cardiac disease

*"Adult" was defined as ⩾18 years of age
**Fully phenotyped, healthy individuals who have married into a family with a Mendelian congenital cardiac disease trait

main categories based on the cardiac malformation of the index case. This subdivision does not reflect any assumptions on possible genetic determinants for those groups. Patients with a LVOTO diagnosis – including sub- and supravalvular aortic stenosis, bicuspid aortic valve or other significant morphologic anomalies of the aortic valve, coarctation of the aorta, mitral stenosis, and hypoplastic left heart syndrome – were grouped together. Patients diagnosed with septal defects, that is, ventricular or atrial, and those with a tetralogy of Fallot are arbitrarily classified in two distinct groups. All other congenital cardiac diseases – for example, transposition of the great arteries, common arterial trunk, pulmonary stenosis, Ebstein's anomaly, cardiomyopathies, and Wolff–Parkinson–White syndrome – are also selected and will be included in specific analyses.

- *Controls.* Healthy paediatric subjects and their first-degree relatives were recruited from a population of children referred for evaluation of an innocent heart murmur. Given that 80% of children will have a clinically insignificant heart murmur at some time or another, we consider children with innocent heart murmurs to be representative of the normal population.[34] However, if a "control subject" or sibling was subsequently diagnosed with congenital cardiac disease, the family was then considered among affected families. Healthy volunteers were recruited as adult controls, predominantly from hospital personnel. Ethnicity of controls reflects that of the overall populations seen at participating hospitals.

Phenotyping

Details of the medical history were derived from an interview as entered on a case report form. The structured questionnaire included information related to pregnancy history of index cases. Family histories included a minimum of three generations. For all participants, weight, height, blood pressure, and heart rate were noted and a standard 12-lead electrocardiogram was obtained and two-dimensional and Doppler echocardiography were performed according to previously described standardised protocols.[18]

Demographic information

We recorded the birthplaces of the proband, their parents, and grandparents. Individuals were categorised as French-Canadian, Caucasians (non-French-Canadian), American Indian, Afro-American or African, Latin American, Asian, or as having mixed origins. French-Canadian individuals were designated in this group if they have at least three of their grandparents of French-Canadian ancestry. This classification, considering grandparental ethnic origin, is an established method and has been validated within Quebec population genetic studies.[35,36] We also applied this classification to the other ethnic groups, and individuals with less than three grandparents belonging to the same ethnic group were considered to have a mixed origin.

Sample collection and biobank procedures

Blood and saliva sampling were performed according to routine procedures for biobanking (Oragene™ DNA self-collection kit, DNA Genotek Inc., Canada). Participants had the option of consenting for sampling of skin fibroblasts or lymphocytes with the purpose of establishing cell lines, or of tissue, which would be discarded during open-heart surgery.

The repository of biological samples and associated data is securely stored at the Research Center of Sainte Justine Hospital. Biological samples were coded before storage and processing with a simple coding scheme to allow for easy extension of the families and for participation in the BALSAC project. Access to the

repository and data decoding is limited to designated authorised research personnel. Sample inventory management allows for easy workflow and following of chains of custody of the samples.

Institutional ethics approval specifically allows sharing of data and samples with other investigators after obtaining project-specific separate approval of the protocols proposed. All information and samples shared with other investigators will be sent coded.

Data storage

All information is recorded in a database set-up with Progeny® Lab 7 (Progeny Software, LLC, South Bend, IN, USA) on a secure server. This database was customised to store and manage clinical, sampling, and genetic data. Files containing pedigree, individual, phenotype, and genotype information can be easily exported in specific formats for analysis. Access to the database by research personnel is limited and controlled.

Segregation analysis

We used the segregation analysis program jPAP v. 1.7.0 to model the inheritance of congenital LVOTO and associated congenital cardiac diseases throughout 67 pedigrees with multiple affected members (88% of all LVOTO multiplex families).[37] Analysis was performed on simple models based on a dichotomous trait (affected/unaffected) without any quantitative co-variables. Nested models were compared by the likelihood ratio test, where the difference in $-2 \ln L$ – that is, natural logarithmic transformation of the likelihood ratio – has a chi-square distribution, with the degrees of freedom equal to the difference in the number of parameters fitted in the most general model. Sporadic (no genetic effect), major gene dominant, recessive and co-dominant, and 2-loci models were compared with the general model. An ascertainment correction for the probands in each family was applied. A chi-square distribution is used to test the fitting of a nested model over a more general one, with a significant chi-square indicating a lack of fit.

Simulation of linkage scenarios

SLINKv. 2.61[38,39] was used to carry out simulations of linkage scenarios and to evaluate the likelihood of obtaining significant linkage peaks (LOD scores). For standard scenarios, the appropriate mode of transmission was chosen, and the disease allele was simulated at a distance of 1.25 centimorgan with a penetrance of 0.85 and a phenocopy rate of 0.025 with an allele frequency of 0.001. Iterative two-point linkage was evaluated for each pedigree and all pedigrees combined assuming varying degrees of genetic homogeneity. The power given by the set of data is defined by the number of LOD scores (>1) reached by each pedigree included in the analysis.

Results

Cohort description

From April, 2006 to March, 2009, we recruited 1603 subjects – including family members – in 300 families (Tables 2a and b). Of all potentially eligible cases identified and approached by the research nurse, 71 individuals in 28 families refused to participate, with an overall recruitment success of 96%. A small percentage (1.7%) of members in participating families refused to participate. Participation was declined for the following reasons: lack of time, no interest in knowing the cause of the malformation, apprehension regarding DNA analyses, and living too far away.

The control group consisted of 28 families, with 87 individuals, and 52 healthy adults. Of individuals specifically approached to participate as controls, 14 declined as individuals or as a family unit (refusal rate 14%).

Some families requested a delay before participation and others postponed study participation due to the stress of clinical interventions for the patient and family. One individual withdrew from the study after consenting. Table 2b and Figure 2 describe the distribution of recruited cases among congenital cardiac disease groups.

Table 3 describes the characteristics of the cohort. Among the probands, defined as being the first affected member seen for medical care, the proportion of non-French-Canadian subjects was 17.6%. From these, Caucasians of other origin and those of mixed origins, which may include at least one parent of French-Canadian origin, represented 5.2% and 8.3% of probands, respectively. Of the recruited families, 169 out of 300 (56%) have more than one affected member.

Segregation analysis

Over 67 pedigrees with LVOTO, models fitting against the general model include the following ones: a trait with a recessive major gene ($\chi^2 = 0.44$, p = 0.9941)

Table 2a. Recruitment among cases and controls.

	Families (n)	Individuals (n)
Cases		
Paediatric	300	1396
Adult	19	106
Controls		
Paediatric	27	80
Adult	14	21
Total	350	1603

Table 2b. Recruitment according to disease entity and possible Mendelian contributions.

	Families (n)*	Individuals (n)	Multiplex families (n/%)	Total affected in multiplex families (n [min–max]/family)**
Left ventricular outflow tract obstruction	109	821	76 (69.7%)	266 [2–30]
Septal defects	103	378	50 (48.5%)	124 [2–5]
Tetralogy of Fallot	28	95	17 (60.7%)	51 [2–8]
Other	60	208	26 (43.3%)	70 [2–7]
Total	300	1502	169	512

*Classified according to primary diagnosis in index case
**Including all structural cardiac disease, excluding coronary artery disease and structural cardiac disease in married-in
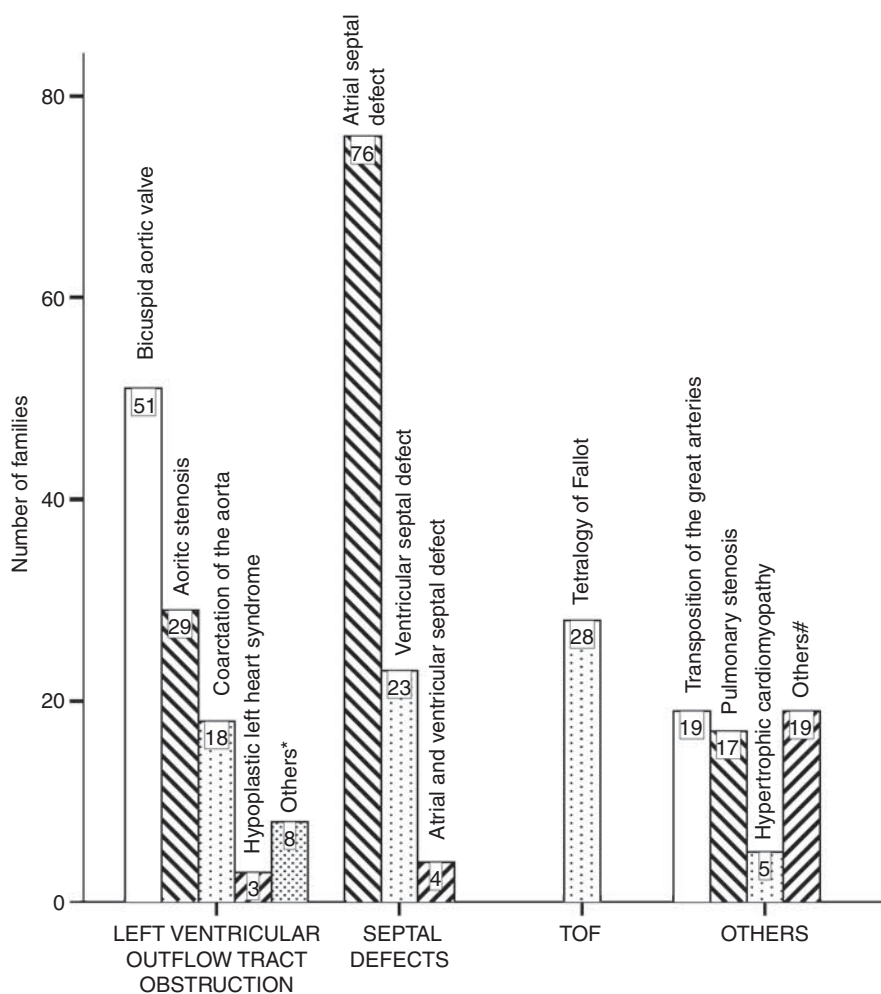

Figure 2.
Distribution of case families based on proband's cardiac affection.

assuming one major gene, or a trait that is an expression of the effect of two (or more) recessive or dominant genes ($\chi^2 = 0.17$–0.90, p $= 0.8254$–0.9822). These results show that a single recessive gene, or two (or more) recessive or dominant genes best account for the inheritance of a LVOTO phenotype with an estimated heritability of 46–52% (Table 4).

Multiplex families and simulation of linkage scenarios

Having established heritability in this model, we carried out linkage simulation in order to evaluate whether our cohort is adequately powered for such analysis. Excluding coronary artery disease, 25.7%

Table 3. Congenital cardiac disease cohort characteristics.

	Probands* (n = 302)	Sibships (n = 1121)	Married-in (n = 79)
Affected (%)	100	24.4	15.2***
Male : Female	154:148	537:584	37:42
Age (years ± SD)	8.4 ± 7.5	31.1 ± 19.4	46.2 ± 13.5 [24–86]
Age range	0–65	0–100	24–86
Ethnicity** (% French-Canadian)	79.8	83.0	93.7

*Two cases have been excluded given the impossibility to recruit the biological parents
**Percentage of French-Canadian. For the probands, 6.0% (18) are non-French-Canadian Caucasian, 1.0% (3) Hispanic, 0.7% (2) American Indian, 1.3% (4) Asian, 8.3% (25) mixed origins, and 5 individuals are of imprecise origin
***Bicuspid aortic valve (n = 3), aortic valve anomaly (n = 1), sub-aortic valve stenosis (n = 1), dilation of the aortic root and/or ascending aorta (n = 4), Wolff–Parkinson–White (n = 1), atrioventricular block (n = 1), pacemaker (n = 1)

Table 4.

Model	Major allele frequency (estimated)	Heritability (estimated)	2lnL	χ^2	p-value
Sporadic	1.0[a]	0[a]	1149.65	79.69 (df = 7)	<0.001
One major gene					
Dominant	0.98	0.36	1112.36	42.40 (df = 5)	<0.001
Recessive	0.78	0.46	1070.40	0.44 (df = 5)	0.9941
Codominant	0.94	0.71	1151.26	81.30 (df = 4)	<0.001
2-loci					
Dom-Dom	0.99/0.78	0.46	1070.86	0.90 (df = 3)	0.8254
Dom-Rec	1.0/1.0	0.52	1070.51	0.55 (df = 3)	0.9077
Rec-Rec	1.0/1.0	0.52	1070.13	0.17 (df = 3)	0.9822
General model	0.98	0.36	1069.96	—	

of related members among all case families recruited had a cardiac defect, as compared with 4.6% in the control pedigrees, that is, one first-degree relative suspected to have a ventricular septal defect, and the three others with distant relatives thought to have congenital cardiac disease. This observation is due to our recruitment bias for multiplex families. Distribution of those affected sibships varied among congenital cardiac disease groups (Table 2b). For families with recurrence of LVOTO, we carried out power calculations to estimate whether the pedigree structure and phenotypic information were sufficient to detect linkage. SLINK program was used on 67 families in the LVOTO group under the assumption of a single point analysis with a biallelic marker. In a cohort context, and assuming that 80% of families share a common genetic factor, linkage analysis could reach an average LOD score of 3.87, with a maximum of 10.93 and a power of 94.2% for LOD > 1. Only one pedigree (102) contains enough information by itself to reach an LOD score over 1. Table 5 lists the pedigrees used for the simulation and their characteristics. Simulation of linkage scenarios for lower penetrances reveals that average LOD scores obtainable in our cohort drop when considering higher genetic heterogeneity. Should only 60% of families share a common genetic factor, linkage analysis could reach average LOD scores of 1.47, with a maximum of 6.47 and a power of 52.1% for LOD > 1; for even higher heterogeneity where only 40% would share a common factor, these values drop to a maximum LOD score of 0.8, maximum of 5.99, and a power of 28.8% for LOD > 1. These results support the assumption that our cohort has the potential to detect genetic factor(s) with a high probability of linkage to disease-causing loci if either a substantial proportion of genetic factors are shared in between families, or if individual families that share those factors are large.

Discussion

The biobank described here was designed to collect a cohort of highly informative families with congenital cardiac disease to be used for the determination of genetic factors predisposing in this disease spectrum. We showed that in as little as 3 years it grew to sufficient size for adequately powered linkage studies. The relatively high number of affected individuals per family has to be interpreted within the context of our recruitment strategy, which does not consist of a cross-sectional study. Ideally, all Quebec patients

Table 5. Pedigrees and their characteristics used for simulation linkage study.

Pedigrees	Individuals (n)	Generations (n)	Phenotyped members (n)	Average LOD score	Max LOD score
1	12	4	4	0.07	0.27
2	15	4	8	0.08	0.28
3	6	3	4	0.07	0.27
4	15	2	15	0.19	1.13
5	9	3	5	0.08	0.29
6	11	4	5	0.03	0.25
7	8	3	4	0.08	0.29
8	25	4	14	0.33	1.01
9	28	5	19	0.20	1.25
10	16	4	4	0.00	0.00
11	12	4	4	0.00	0.00
12	15	5	7	0.19	0.64
13	8	3	5	0.06	0.25
14	9	3	5	0.00	0.00
15	8	3	4	0.00	0.00
16	18	4	8	0.08	0.29
17	18	5	6	0.06	0.24
18	9	3	5	0.00	0.00
20	9	3	5	0.16	0.59
21	3	2	2	0.00	0.00
23	9	3	5	0.06	0.26
24	11	4	4	0.00	0.00
25	12	4	5	0.00	0.00
26	9	3	7	0.06	0.26
27	13	3	5	0.02	0.11
28	9	3	5	0.09	0.29
29	13	5	5	0.05	0.25
30	15	4	5	0.00	0.00
33	14	4	8	0.07	0.28
34	22	5	13	0.05	0.29
35	7	3	3	0.00	0.00
36	8	3	3	0.00	0.00
37	25	4	16	0.34	1.47
38	20	4	8	0.09	0.34
39	8	3	4	0.00	0.00
40	16	5	10	0.12	0.57
41	15	4	7	0.07	0.36
42	16	4	5	0.00	0.00
43	26	4	18	0.40	1.77
44	14	4	4	0.00	0.00
46	26	5	14	0.18	0.74
47	11	4	3	0.00	0.00
48	13	4	6	0.00	0.00
49	23	4	5	0.05	0.21
51	3	2	3	0.00	0.00
52	17	4	6	0.06	0.27
53	17	4	6	0.03	0.25
54	11	4	2	0.00	0.00
55	8	3	4	0.08	0.29
57	20	5	4	0.00	0.00
58	14	4	6	0.04	0.25
59	19	5	6	0.01	0.02
60	16	4	5	0.00	0.00
63	13	4	6	0.07	0.29
65	66	5	50	0.76	2.79
69	16	4	5	0.00	0.00
70	7	3	2	0.00	0.00
72	7	4	3	0.00	0.00
75	8	3	4	0.00	0.00
76	12	4	7	0.00	0.00
77	12	4	4	0.04	0.29
78	14	4	11	0.14	0.70

Table 5. *Continued*

Pedigrees	Individuals (n)	Generations (n)	Phenotyped members (n)	Average LOD score	Max LOD score
79	11	5	5	0.00	0.00
80	11	4	6	0.07	0.29
102	51	5	30	1.32	4.61
2002	14	5	7	0.18	0.59
2505	33	4	20	0.30	1.68
	Total 1009	Median 4	Total 503	Study 3.87	Study 10.93

with congenital cardiac disease should be recruited, regardless of family status (multiplex versus sporadic); however, such an approach is outside the scope of the current study. The study is open-ended, meaning that clinical information and DNA genotyping are continuously updated. Management aspects of the database must deal with migration within the cohort, births, deaths, and changes in the phenotype status, such as surgery or development of arrhythmia or worsening of a pre-existing condition. Our familial database with clinical and genetic data constitutes a specific tool in genetic epidemiology and could support various hypotheses for linkage or association studies. Possible collaborations with other investigators were specifically included in our protocol, with pre-consent from participants, and, therefore, require only the approval of the local ethics committee.

The database is set up to export pedigree information in useful formats for subsequent use in linkage software and other genetic analysis programs. It is also capable of importing information from such analyses, for example, for visualising haplotypes. Such considerations are particularly important in the light of the ever-increasing power of genetic analysis platforms, such as single-nucleotide polymorphism genotyping, high resolution of genomic copy number variants, and other high throughput methods.

The impressive progress made in paediatric cardiology over the last decades has been achieved without *a priori* knowledge of the molecular cascades leading to cardiac malformations. This situation is in great contrast to other areas of cardiovascular medicine, such as coronary artery disease, and other fields, such as oncology. In addition, the limited synergy between human and animal studies further increases the value of family-based human cohorts. As a case in point, mouse models replicating hypoplastic left heart syndrome have yet to be recovered from targeted or N-ethyl-N-nitrosourea mutagenesis screens. Some mouse models replicate semilunar valve disease, one example of which is bicuspid aortic valve as the second most common defect in heterozygous *Nkx2.5* knockout mice.[40] *A priori*, this observation is in contrast with the high heritability and significant linkage

peaks found in this disease spectrum.[20,21,23] This may be due to several reasons: the molecular pathogenesis in humans is not replicated by the current mouse models (more complex genomic variants versus loss-of-function in a single gene); the genes responsible for these malformations have not yet been targeted; the LVOTO disease spectrum is caused by an interplay of different genetic and epigenetic factors not reproduced by routine animal breeding strategies. Linkage gradients in our simulation studies suggest that our cohort has sufficient power to identify novel loci.

The availability of information on pregnancy and environmental influences, as well as the family-based design make our platform ideally suited to assess the biological impact of gene–environment interactions. As an example, we have recently shown that the MTHFD1 p.Arg653Gln variant alters enzyme function and increases risk for congenital cardiac defects, in particular that of tetralogy of Fallot and aortic stenosis.[33] The *in vitro* assays performed in this study suggest that the p.Arg653Gly variant reduces the metabolic activity of this enzyme in the folate pathway and underlies a folate stabilisation effect, which is in turn modulated by folate status.

Despite being set up with a family-based design in mind, this study comprises a control group for several reasons. First, all participants in this arm of the study have been fully phenotyped, including echocardiography. More subtle cardiac malformations such as bicuspid aortic valve and small atrial septal defects may constitute variable expressivity of a genetic trait, yet escape routine physical exam and therefore confound case–control analysis. Second, a family-based control cohort will allow assessing normal and distorted transmission patterns at any loci of interest. Third, the control group will be ethnically matched and thus help to distinguish genetic variants specific to the French-Canadian population *per se* from disease-associated variants.

Recent results from the linkage studies mentioned above and candidate gene analyses in congenital cardiac disease suggest that the most likely underlying genetic model for these conditions is one of rare alleles with major effects.[24] At the same time, genetic heterogeneity of these disorders

can be anticipated to be very important. In order to tease out the molecular basis of variable expressivity and reduced penetrance, our study will allow a multi-pronged approach for the dissection of the genetic architecture of congenital cardiac disease. First, the family-based cohort is well suited for linkage and family-based association analysis and permits to establish segregation patterns of mutant alleles considered to be causative in congenital cardiac disease.[41] The possibility to retrace extended genealogies will make it possible to search for conserved haplotypes as molecular evidence of founder effects. Second, inclusion of sporadic cases will allow case–control designs including genome-wide association studies, which are more powerful to detect common variants causative of common diseases.[42] With the recent progress of sequencing technology, which now allow the use of whole-exome technologies to identify Mendelian factors in rare diseases,[43] our cohort is of great value for complete genome sequencing in multiplex families with congenital cardiac disease. Furthermore, these novel technologies may allow the genetic study of families not tractable by linkage analysis, such as in the setting of considerable genetic heterogeneity. The availability of large-scale sequencing data, as well as additional biobanks including French-Canadian populations, will further expand the applicability of new sequencing technologies for the detection of rare alleles with major effects in congenital cardiac disease.[44,45]

The cohort presented here has inherent limitations, which reflect the initial focus and scope of the project. Our aim was not to provide a cross-sectional sample for genetic epidemiological investigations, but rather to maximise the power of linkage analysis with the goal of dissecting Mendelian components of LVOTO in multiplex families. Therefore, this cohort does not proportionately reflect the disease spectrum in our overall patient population, and at this point may also lack power for investigations of rare disease entities. In addition, current sampling of patients is mostly restricted to peripheral blood, although we have made ethical and infrastructural provisions to extend sampling to myocardial tissues, accessible at surgery or skin fibroblasts and blood for producing lymphoblasts, for cell culture purposes. Such DNA sources can provide insight into somatic mosaicism and provide an inexhaustible supply of DNA.

In conclusion, the cohort recruited with our provincial platform constitutes a unique and open resource for research into the genetics of congenital cardiac disease. Since demographic migration and increasing prevalence of congenital cardiac disease patients of reproductive age have dramatically altered the need for genetic research, the establishment and availability of such resource is timely and will provide a unique opportunity to improve patient counselling and management, and ultimately our understanding of the genetics of cardiac development.

Acknowledgements

References

1. Pierpont ME, Basson CT, Benson Jr DW, et al. Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. Circulation 2007; 115: 3015–3038.
2. Statistics Canada. Deaths, by cause – Chapter XVII: Congenital malformations, deformations and chromosomal abnormalities (Q00–Q99), 2005.
3. Health Canada. Congenital anomalies in Canada – a perinatal health report. Minister of Public Works and Government Services, Ottawa, 2002.
4. Marelli AJ, Mackie AS, Ionescu-Ittu R, Rahme E, Pilote L. Congenital heart disease in the general population: changing prevalence and age distribution. Circulation 2007; 115: 163–172.
5. Allan LD, Crawford DC, Chita SK, Anderson RH, Tynan MJ. Familial recurrence of congenital heart disease in a prospective series of mothers referred for fetal echocardiography. Am J Cardiol 1986; 58: 334–337.
6. Oyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PK, Melbye M. Recurrence of congenital heart defects in families. Circulation 2009; 120: 295–301.
7. Ferencz C, Loffredo CA, Correa-Villasenor A, Wilson PD. Genetic and environmental risk factors of major cardiovascular malformations: the Baltimore–Washington Infant Study: 1981–1989. Future Publishing Co, Armonk, NY, 1997.
8. Whittemore R, Wells JA, Castellsague X. A second-generation study of 427 probands with congenital heart defects and their 837 children. J Am Coll Cardiol 1994; 23: 1459–1467.
9. Brenner JI, Berg KA, Schneider DS, Clark EB, Boughman JA. Cardiac malformations in relatives of infants with hypoplastic left-heart syndrome. Am J Dis Child 1989; 143: 1492–1494.
10. Digilio MC, Marino B, Giannotti A, Toscano A, Dallapiccola B. Recurrence risk figures for isolated tetralogy of Fallot after screening for 22q11 microdeletion. J Med Genet 1997; 34: 188–190.

11. Gill HK, Splitt M, Sharland GK, Simpson JM. Patterns of recurrence of congenital heart disease: an analysis of 6640 consecutive pregnancies evaluated by detailed fetal echocardiography. J Am Coll Cardiol 2003; 42: 923–929.

12. McBride KL, Pignatelli R, Lewin M, et al. Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: segregation, multiplex relative risk, and heritability. Am J Med Genet A 2005; 134: 180–186.

13. Oyen N, Poulsen G, Wohlfahrt J, Boyd HA, Jensen PK, Melbye M. Recurrence of discordant congenital heart defects in families. Circ Cardiovasc Genet 2010; 3: 122–128.

14. Hoffman JI, Kaplan S. The incidence of congenital heart disease. J Am Coll Cardiol 2002; 39: 1890–1900.

15. Pradat P. A case–control study of major congenital heart defects in Sweden – 1981–1986. Eur J Epidemiol 1992; 8: 789–796.

16. Boon AR, Roberts DF. A family study of coarctation of the aorta. J Med Genet 1976; 13: 420–433.

17. Brownell LG, Shokeir MH. Inheritance of hypoplastic left heart syndrome (HLHS): further observations. Clin Genet 1976; 9: 245–249.

18. Cripe L, Andelfinger G, Martin LJ, Shooner K, Benson DW. Bicuspid aortic valve is heritable. J Am Coll Cardiol 2004; 44: 138–143.

19. Hinton Jr RB, Martin LJ, Tabangin ME, Mazwi ML, Cripe LH, Benson DW. Hypoplastic left heart syndrome is heritable. J Am Coll Cardiol 2007; 50: 1590–1595.

20. Hinton RB, Martin LJ, Rame-Gowda S, Tabangin ME, Cripe LH, Benson DW. Hypoplastic left heart syndrome links to chromosomes 10q and 6q and is genetically related to bicuspid aortic valve. J Am Coll Cardiol 2009; 53: 1065–1071.

21. Martin LJ, Ramachandran V, Cripe LH, et al. Evidence in favor of linkage to human chromosomal regions 18q, 5q and 13q for bicuspid aortic valve and associated cardiovascular malformations. Hum Genet 2007; 121: 275–284.

22. Hinton RB, Andelfinger G, Sekar P, et al. Prenatal head growth and white matter injury in hypoplastic left heart syndrome. Pediatr Res 2008; 64: 364–369.

23. McBride KL, Zender GA, Fitzgerald-Butt SM, et al. Linkage analysis of left ventricular outflow tract malformations (aortic valve stenosis, coarctation of the aorta, and hypoplastic left heart syndrome). Eur J Hum Genet 2009; 17: 811–819.

24. Roessler E, Ouspenskaia MV, Karkera JD, et al. Reduced NODAL signaling strength via mutation of several pathway members including FOXH1 is linked to human heart defects and holoprosencephaly. Am J Hum Genet 2008; 83: 18–29.

25. Chapman NH, Thompson EA. Linkage disequilibrium mapping: the role of population history, size, and structure. Adv Genet 2001; 42: 413–437.

26. Moreau C, Vezina H, Labuda D. Founder effects and genetic variability in Quebec. Med Sci (Paris) 2007; 23: 1008–1013.

27. Scriver CR. Human genetics: lessons from Quebec populations. Annu Rev Genomics Hum Genet 2001; 2: 69–101.

28. Yotova V, Labuda D, Zietkiewicz E, et al. Anatomy of a founder effect: myotonic dystrophy in Northeastern Quebec. Hum Genet 2005; 117: 177–187.

29. Laberge AM, Jomphe M, Houde L, et al. A "Fille du Roy" introduced the T14484C Leber hereditary optic neuropathy mutation in French Canadians. Am J Hum Genet 2005; 77: 313–317.

30. Laberge AM, Michaud J, Richter A, et al. Population history and its impact on medical genetics in Quebec. Clin Genet 2005; 68: 287–301.

31. Plante M, Claveau S, Lepage P, et al. Mucolipidosis II: a single causal mutation in the N-acetylglucosamine-1-phosphotransferase gene (GNPTAB) in a French Canadian founder population. Clin Genet 2008; 73: 236–244.

32. Bouchard G Projet BALSAC. www.uqac.ca/balsac 2006.

33. Christensen KE, Rohlicek CV, Andelfinger GU, et al. The MTHFD1 p.Arg653Gln variant alters enzyme function and increases risk for congenital heart defects. Hum Mutat 2009; 30: 212–220.

34. Allen HD, Phillips JR, Chan DP. History and physical examination. In: Allen HD, Clark EB, Gutgesell HP, Driscoll DJ (eds.). Moss' and Adams' Heart Disease in Infants, Children and Adolescents. Lippincott Williams&Wilkins, Philadelphia, 2001, pp. 143–153.

35. Bhopal RS. Ethnicity, Race, and Health in Multicultural Societies: Foundations for Better Epidemiology, Public Health and Health Care, 1st edn. Oxford University Press, Oxford, New York, 2007.

36. Bhat M, Nguyen GC, Pare P, et al. Phenotypic and genotypic characteristics of inflammatory bowel disease in French Canadians: comparison with a large North American repository. Am J Gastroenterol 2009; 104: 2233–2240.

37. Hasstedt SJ. jPAP: document-driven software for genetic analysis. Genet Epidemiol 2005; 29: 255.

38. Ott J. Computer-simulation methods in human linkage analysis. Proc Natl Acad Sci 1989; 86: 4175–4178.

39. Weeks DE, Ott J, Lathrop GM. SLINK: a general simulation program for linkage analysis. Am J Hum Genet 1990; 47: A204.

40. Biben C, Weber R, Kesteven S, et al. Cardiac septal and valvular dysmorphogenesis in mice heterozygous for mutations in the homeobox gene Nkx2-5. Circ Res 2000; 87: 888–895.

41. Laird NM, Lange C. Family-based methods for linkage and association analysis. Adv Genet 2008; 60: 219–252.

42. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273: 1516–1517.

43. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a Mendelian disorder. Nat Genet 2010; 42: 30–35.

44. Durbin RM, Abecasis GR, Altshuler DL, et al. A map of human genome variation from population-scale sequencing. Nature 2010; 467: 1061–1073.

45. Fortin S, Pathmasiri S, Grintuch R, Deschenes M. "Access arrangements" for Biobanks: a fine line between facilitating and hindering collaboration. Public Health Genomics 2010; Epub ahead of print.