

PEDIATRIC CARDIAC GENOMICS CONSORTIUM

CONGENITAL HEART DISEASE GENETIC NETWORK STUDY (CHD GENES)

Funded by the
National Heart, Lung, and Blood Institute,
NIH/DHHS

Version 2: December 16, 2024

Summary Table

Title	Congenital Heart Disease GEnetic NEtwork Study (CHD GENES)
Study Objectives	Investigate relationships between genetic factors and phenotypic and clinical outcomes in patients with congenital heart defects (CHD).
Study Design	Multi-center, prospective observational cohort study of those with CHD and those without CHD who have a genotype strongly associated with CHD. Acquisition of both phenotypic data and source DNA from participant, parents, and family of interest.
Primary Aim	Through genomic analyses, genome-wide association studies, whole exome sequencing, and whole genome sequencing discover gene(s) responsible for CHD.
Secondary Aims	• Identification of mutations responsible for CHD in large numbers of participants. • Genotype/Phenotype correlation including long-term clinical follow-up of enrolled participants to determine how genetics influences the clinical outcome in CHD. • To generate induced pluripotent stem cells (iPSCs) from subjects with defined genetic lesions that are likely tied to the etiology of CHD in those individuals. This work includes characterization of the selected clones as pluripotent and documentation that the relevant cell types (e.g., cardiomyocytes, vascular cells, non-cardiac cells and others) can be differentiated from the iPSCs. • To characterize the iPSC-derived cells with respect to functional, biochemical, biophysical and other properties as relevant for the genomic lesion and the specific form of CHD or a genotype strongly associated with CHD in the subject.
Accrual Objective	Accrual of thousands of participants with CHD, parents, and related family of interest and participants without CHD who have a genotype strongly associated with CHD.
Study Duration	15 years (longer, if funding extended)

Inclusion Criteria	• One of the following:
	  ○ Affected Individuals with CHD, or
	  ○ Participants without CHD who have a genetic syndrome strongly associated with CHD (such as Trisomy or Chr22q deletion), or
	  ○ Pregnant women who have a fetus diagnosed with CHD
	• Age 0 – 99 years of age
	• Males and females

Exclusion Criteria	• Isolated patent foramen ovale (except in confirmed cases of trisomy 21) • Isolated prematurity-associated patent ductus arteriosus • Lack of consent • Isolated pulmonary stenosis secondary to twin-twin transfusion syndrome

INVESTIGATOR SIGNATURE PAGE

Date: 05 April, 2023
Protocol Version: 1.0

Title: Congenital Heart Disease GEnetic NEtwork Study (CHD GENES)

Study Sponsor: National Heart, Lung and Blood Institute; Division of Cardiovascular Sciences

INSTRUCTIONS: The Principal Investigator must print, sign, and date this form indicating he/she is in full agreement with the statement being signed. A copy should be kept in the study records and the original signature page sent to the ACC.

After signature, please return a copy of this form to Cincinnati Children's Hospital B2B ACC at B2BRegulatory@cchmc.org. Retain the original signature page with your study records.

I confirm that I have read the above protocol in the latest version. I understand it, and I will work according to the principles of Good Clinical Practice (GCP) as described in the United States Code of Federal Regulations (CFR) – 21 CFR Parts 45, 50, 56, and 312, and the International Conference on Harmonization (ICH) document "Guidance for Industry: E6 Good Clinical Practice: Consolidated Guidance." Further, I will conduct the study in keeping with local, legal, and regulatory requirements.

- **As the Site Principal Investigator, I agree to conduct [**Congenital Heart Disease GEnetic NEtwork Study (CHD GENES)**.**
- **I agree to carry out the study by the criteria written in the protocol and understand that no changes can be made to this protocol without written permission of the NHLBI.**

_____
Site Principal Investigator Name (Print)

_____ _____
Site Principal Investigator (Signature) **Date**

TABLE OF CONTENTS

CONGENITAL HEART DISEASE GENETIC NETWORK STUDY (CHD GENES) 1
Congenital Heart Disease GEnetic NEtwork Study (CHD GENES) 2
A. Background 7
B. Study Aims and Hypothesis 11
B.1 Study Aims 11
B.2 Hypothesis 12
C. Study Design 13
C.1 Overview 13
C.2 Participant Recruitment and Screening 13
C.3 Genetic and Genomic Analyses 14
C.4 Return of Genetic Testing Results 15
C.5 Study Timeline 15
C.6 Stopping Rules 16
D. Collection of Samples and Data 16
D.1 Sample Collection 16
D.1.1 Proband samples 16
D.1.2 Family samples 17
D.2 Clinical and Phenotypic Data 18
D.2.1 Proband clinical data collection 18
D.2.2 Phenotypic data collection techniques 19
D.3 Family Data Collection 19
D.3.1 Parents data collection 19
D.3.2 Extended family data collection 19
D.3.3 Clinical data collection techniques 19
E. Study Subjects 20
E.1 Inclusion Criteria 20
E.2 Exclusion Criteria 20
E.3 Source of Subjects 21
E.4 Subject Recruitment and Consent 21
E.6 Voluntary Subject Withdrawal 23

E.7 Early Termination of a Subject's Participation 24
E.8 Study Termination 24
F. Statistical and Analytical Approaches 24
F.1 Genome-Wide Association Studies (GWAS) 25
Table 1: 25
F.2 Copy Number Variant (CNV) and Structural Variant (SV) Analyses 26
Table 2: 26
F.3 Candidate Gene, Whole-Exome, And Whole-Genome Sequence Analyses 27
F.3.1 Analysis and annotation of the exome or MIPS sequencing data 27
F.4 Geomarker Acquisition and Analysis 28
G. DATA MANAGEMENT 29
G.1 Data Entry 29
G.2 Data Validation and Monitoring 30
G.3 Specimen Tracking 30
G.4 Data and Specimen Security and Integrity 31
G.4.1 Phenotype security and integrity 31
G.4.2 Genotype data security and integrity 31
G.4.3 Future Use 31
H. Quality Assurance and Quality Control Procedures 32
I. Human Subjects Considerations 33
I.1 Potential risks 33
I.1.1 Psychological distress 33
I.1.2 Blood draw 33
I.1.3 Saliva collection 33
I.1.4 Tissue collection 33
I.1.5 Leftover or discarded samples [OPTIONAL, according to local institutional guidelines] 33
I.1.6 Clinical data collection 33
I.2 Confidentiality, Protection against Risks 34
I.3 Potential benefits 36
I.4 Risk/Benefit Ratio and Importance of Information to Be Obtained 36

I.5 Safety Monitoring 36
I.6 Inclusion of Women and Minorities 37
J. Compensation and Costs 37
J.1 Compensation 37
J.2 Cost 37
K. References 38

A. Background

Congenital heart defects (CHD) are the most common major human birth malformation, affecting ~8 per 1,000 live births [1-4]. CHD are associated with significant morbidity and mortality, and are second only to infectious diseases in contributing to the infant mortality rate [1, 2, 5-7]. Mortality and morbidity rates in patients with CHD have decreased due to the advent of surgical and medical intervention; however, these individuals often require lifelong care, including repeat interventions, resulting in enormous burdens on families, and a large financial burden on the health care system [8]. Children with arrhythmias and cardiomyopathies similarly require extensive medical care throughout life.

Current understanding of the etiology of pediatric cardiovascular disorders is limited. The etiology of congenital heart disease is multifactorial and consists of a combination of environmental, teratogenic, and genetic causes [9-18]. Furthermore, both genetic and environmental factors have been proposed to act as disease modifiers, accounting for the wide variation in phenotypic expression and clinical outcomes of these disorders [11-18]. To date however, few specific genetic or environmental causative factors have been identified. It is estimated that ~11% of CHD patients have some form of chromosomal abnormality causing cardiac malformation [1]. This includes patients with aneuploidy (such as trisomy 13, 18, and 21, as well as Turner Syndrome) who have an elevated prevalence of CHD as compared to the general population [19, 20]. Chromosomal structural abnormalities, such as monosomy 22q11, are a second significant genetic contributor to CHD [9]. More recently, disease-related single-gene defects have been identified in a small number of cases, including mutations in the transcription factors *NKX2.5*, *ZIC3*, *JAG1* and *NOTCH1* [11, 21-25]. Despite these recent discoveries, the cause of the majority of cases of CHD remains largely unknown despite strong evidence of a significant genetic component [9].

The analysis of CHD genetics is complicated by the limited correlation between the genetic defect and specific CHD phenotype. In particular:

1) Similar cardiac defects may be caused by more than one genetic alteration (genetic heterogeneity) [26].
2) One gene may cause more than one CHD phenotype (variable expressivity) [26].
3) A person without CHD may carry a disease-causing gene mutation (reduced penetrance).
4) One or more genes may be causal for CHD associated with chromosomal alterations [9, 27].

5) Single gene defects can cause CHD with or without being recognized as a clinical syndrome [23].

Analyses are further complicated by the likelihood that phenotypic variability results from genetic modifiers, the interaction of genes and environment, and/or stochastic effects.

Due to both the genetic heterogeneity of CHD and the investigators' clinical inability to discriminate subtle anatomic variants, it is likely that within each cohort of patients with anatomically similar CHD, causative genes affecting distinct developmental pathways will be found. Clinical outcome may therefore at least in part depend upon which specific biological pathway is affected, and genotype variants may have distinct, clinically relevant physiologic phenotypes.

Although human genetic analyses are increasingly identifying mutations that cause congenital heart disease (CHD), the molecular events triggered by a CHD gene mutation that leads to the malformation remain poorly defined. For example, 60-70% of patients with deletion of chromosome 22q11.2 (DiGeorge syndrome) have CHD, most often tetralogy of Fallot. Similarly, CHD occurs in 40% of Down syndrome patients, most often atrioventricular septal defects that occur in over 40% of Down syndrome patients with CHD. By contrast, CHD occurs in only ~1% in the general population. While mutations such as chromosome 22q11.2 (DiGeorge syndrome) and Trisomy 21 (Down syndrome) are the unequivocal causes of each syndrome, the reason why some individuals with these predisposing genotypes do not have CHD is unknown. To address this gap in knowledge, and to understand the absence of CHD in some subjects who have normal heart structure and function despite having genotypes that are strongly associated with CHD, we will also enroll subjects without CHD who have a genotype strongly associated with CHD (such as Trisomy, or Chr22q deletion).

Because of low recurrence rates, small sample size, and high phenotypic variability, traditional linkage analysis has met with limited success in defining the genetics of CHD. This study will use state-of-the-art genomic technology coupled with large-scale, multi-center participant recruitment to comprehensively elucidate the genetic causes of CHD. A large number of participants will be necessary to capture the highly variable phenotype and genotype of CHD. Developing a cardiac genetic registry with a DNA repository will aid in rapidly advancing the understanding, diagnosis, and treatment of congenital heart disease and serve as a foundation on which the investigators plan to build a clinical research program in the molecular genetics of congenital heart disease.

Identification of the genetic basis for human disease provides unparalleled opportunities to understand pathogenic mechanisms and to define novel therapies to attenuate disease manifestations. The traditional approach to understanding how gene mutations cause disease is to engineer animal models to carry a specific human gene mutation. While this strategy has merit, it remains limited due to considerable cost, and important differences between human and other species that can limit the relevance of new knowledge that is gained through this approach. Recent strategies to develop and differentiate induced pluripotent stem cells (iPSCs) into specified cell lineages provides the opportunity to directly study the consequences of genetic variants on cell biology of human cell lines, that may more closely resemble human disease states. In addition to understanding mechanisms, these cells also provide a novel and important reagent for screening potential novel therapeutic agents.

The significant breakthrough in stem cell biology that allows the potential to engineer patient-specific cell models of disease requires genetic manipulation of somatic cells. Enforced expression of key transcription factors and/or micro-RNAs can reprogram adult somatic cells (skin fibroblasts and blood cells) to produce pluripotential stem cells. iPSCs can then be differentiated *in vitro* into a wide variety of cell types including cardiomyocytes and endothelial cells. This technology has also been extended to produce iPSCs lines from patients with specific diseases of interest such as Amyotrophic Lateral Sclerosis, Fanconi Anemia, Familial Dysautonomia, long QT syndrome, hypertrophic cardiomyopathy and others. Alternatively, the genotype of a standard iPSC line can be manipulated so that it contains a variant found in a patient with a CHD or any other disorder.

The Pediatric Cardiac Genomics Consortium (PCGC) is a National Heart, Lung and Blood Institute (NHLBI) sponsored research network that has undertaken extensive genomic analyses of subjects with CHD. In so doing, novel genetic defects that are likely to contribute to the etiology of CHD are being uncovered. While modeling in animal systems such as transgenic mice provides opportunities to elucidate CHD pathogenesis, studies with cardiomyocytes and other cell types differentiated from iPSCs that were generated from individuals with CHD genotypes or introduced into a standard iPSC line provide unique opportunities to study pathogenesis in a human cell context.

To address the gap in knowledge as to why some individuals with predisposing genotypes (e.g. trisomy 21, 22q11.2qdel or others) do not have CHD, we will derive iPSCs from subjects with trisomy 21 (Down syndrome) or 22q11.2qdel or others with and without CHD that we will differentiate into cardiomyocytes (iPSC-CMs). We plan to analyze cardiomyocyte gene expression

between subjects with the same genotype (e.g., trisomy 21, 22q11.2qdel or others) who are discordant for CHD. We predict that the expression of genes not encoded in these mutated regions will be different. We expect that the identity of these modifiers and the pathways in which these participate may improve our understanding of the molecular mechanisms for normal and aberrant heart development.

To accomplish these goals, the consortia centers will work collaboratively as part of the PCGC. The consortium is supported by an administrative and data coordinating center, (Children's Hospital Medical Center (CHMC)). Clinical data will be entered into the Medidata Rave clinical data management system and will be stored within Medidata's clinical cloud computing solution. All data extracted from Rave for reporting or analysis will initially be stored at CHMC's secure servers for later ingestion into NHLBI's BioData Catalyst cloud-based environment. In some instances, with prior approval from the PCGC Steering Committee, some data to be used for additional analyses may be entered and stored in a study specific REDCap or other appropriate secure database. Biospecimens will be stored at a central biorepository selected for appropriate expertise and operating standards. Extracted DNA will be sent to core laboratories for genotyping and sequencing. The resulting genetic data will eventually be transferred to and stored in an NIH-governed data repository, chosen by and agreed upon by the PCGC Steering Committee. Genetic data will be linked to the clinical data to allow for detailed correlation between genotype, phenotype, and outcome. The combined data sets will be made available to all investigators through a Consortium-wide request and access process.

The Consortium will also develop resources that will benefit the wider congenital heart disease research community. The de-identified clinical and genetic data will be deposited in publicly accessible databases for use by outside investigators in accordance with NHLBI data sharing policies and an approved plan developed by the PCGC Steering Committee.

ADMINISTRATIVE NOTE: As part of the NIH single IRB requirement for applicable NIH funded studies, the administrative coordinating center will begin the process of converting the PCGC funded studies to a single IRB model as part of the current grant (funding cycle 3.0). As part of this process, the decision was made to combine the two original ongoing studies, Congenital Heart Disease Genetic Network Study (CHD GENES) and iNDUCED PLURIPOTENT STEM CELLS (iPSC) into a single protocol, Congenital Heart Disease Genetic Network Study (CHD GENES). The CHD GENES and iPSC studies previously operated under two separate IRB approved protocols,

(CCHMC HRS IRB 2016-0091 and 2016-0096), and were each also approved locally at the participating sites local IRB of record. All patients enrolled and data collected and entered under the original CHD GENES and iPSC from all active and previously active sites will be rolled over/combined into this new study. Active participating sites will keep their local IRBs opened until they are approved under the new single IRB reliance model (relying on CCHMC IRB). Once active sites are approved to rely on CCHMC as the sIRB, they will be instructed to notify their local IRBs accordingly and close out their current local IRB protocol(s). Sites that were part of the prior funding cycle but who were not renewed for the current grant cycle and are no longer conducting any study activities will be directed to close out their current local IRB protocols.

B. Study Aims and Hypothesis

B.1 Study Aims

The Aims of this study are:

- *Gene discovery* of a comprehensive repertoire of genes responsible for CHD through genomic analyses including (but not limited to) copy-number variation, genome-wide association studies, targeted sequencing, molecular inversion probe sequencing, RNASeq, whole exome sequencing, and whole genome sequencing,
- *Identification of mutations* responsible for CHD in large numbers of participants through sequencing of known CHD candidate genes,
- *Genotype/Phenotype correlation* of enrolled participants to determine how genetics influences the clinical outcome in CHD.
- To generate inducted pluripotent stem cells (iPSCs) from subjects with defined genetic lesions that are likely tied to the etiology of CHD.
- To characterize the iPSC-derived cells with respect to functional, biochemical, biophysical and other properties as relevant for the genomic lesion and the specific form of CHD or a genotype strongly associated with CHD in the subject.

To accomplish these aims, the Consortium will develop and maintain a biorepository of specimens (DNA) and genetic data, along with detailed, phenotypic and clinical outcomes data in order to investigate relationships between genetic factors and phenotypic and clinical outcomes in congenital heart disease. Biospecimens and the derived genetic data will be shared with cardiovascular genetics researchers at participating Centers. Specimen sharing will be performed

following NIH guidelines and federal regulations. Data will also be shared with investigators outside the Consortium consistent with NHLBI data-sharing policies. These biological samples will remain linked to detailed clinical data, including identifiable information as allowed by law(s), and will serve as a resource for long-term investigations into the genetic basis and clinical outcome of congenital cardiovascular disorders beyond the current funding period. Through an increased understanding of the causes and modifiers of congenital heart disease, this initiative's long-term goal is to enhance early detection, treatment and prevention of cardiovascular disease in newborns, children, and adults.

B.2 Hypothesis

Genetic alterations contribute to both the etiology and the outcome of congenital heart disease.

To test this hypothesis, the investigators will recruit large numbers of subjects with sporadic as well as familial congenital cardiovascular disorders (including children and adults with congenital heart disease), as well as subjects with normal heart structure and function who have a genotype strongly associated with CHD (such as Trisomy, or Chr22q deletion) from the participating Centers. Each consented participant and parents (and other family members when available) will donate biological specimens for banking and detailed phenotypic and clinical outcome data will be gathered. Analysis of these biospecimens will permit identification of genetic loci responsible for abnormal cardiovascular development. The identification of such genes and their modifiers will enhance understanding of both normal and abnormal cardiovascular development and will provide the foundation for investigations into the treatment and ultimately the prevention of cardiovascular anomalies. Furthermore, linking genetic information to clinical outcomes is expected to yield information that will facilitate risk stratification and improve treatment through individual genetically tailored regimens.

To address the gap in knowledge as to why some individuals with predisposing genotypes (e.g. trisomy, 22q11.2qdel or others) do not have CHD, we will derive induced pluripotent stem cells (iPSCs) from subjects with trisomy 21 (Down syndrome) or 22q11.2qdel or others with and without CHD that we will differentiate into cardiomyocytes (iPSC-CMs). We plan to analyze cardiomyocyte gene expression between subjects with the same genotype (e.g., trisomy 21, 22q11.2qdel or others) who are discordant for CHD. We predict that the expression of genes not encoded in these mutated regions will be different. We expect that the identity of these modifiers and the pathways in

which these participate may improve our understanding of the molecular mechanisms for normal and aberrant heart development.

C. Study Design

C.1 Overview

This is a multi-center, prospective observational cohort study of those with CHD and those without CHD who have a genotype strongly associated with CHD. Participants will be recruited from the Pediatric Cardiac Genomics Consortium's centers, their satellite facilities, and additional collaborating centers. Participants will undergo detailed phenotyping and biological specimens will be obtained for genetic analyses. Phenotypic information will include data captured during detailed clinical exams at enrollment, chart reviews, structured information extracted from EMR data, and data available from other studies and registries to which some participants may have consented. Noninvasive, clinical imaging data including cardiac magnetic resonance imaging and echocardiography will be collected when available. Genomic technologies will be used to identify common genetic causes of CHD and genetic modifiers of clinical outcome. In a subset of participants, an additional blood sample will be taken to generate iPSCs from subjects with defined genetic lesions that are likely tied to the etiology of CHD in those individuals.

C.2 Participant Recruitment and Screening

Participants with CHD or subjects with normal heart structure and function who have a genotype strongly associated with CHD (such as Trisomy or Chr22q deletion) will be recruited from prenatal and pediatric centers, pediatric cardiology and adult congenital heart centers. Patients with genetic syndromes such as Trisomy or Chr22q deletion will have been diagnosed and well categorized. Participants with isolated prematurity-associated patent ductus arteriosus, isolated patent foramen ovale or isolated pulmonary stenosis secondary to twin-twin transfusion syndrome will be excluded from the study. Family members will be recruited whether or not they have a CHD history. There will be no restrictions to participation based on age, sex or ethnicity. Existing published and preliminary data [28, 29] suggests there will be a slight preponderance of male to female candidates for the study. Whenever possible, the study will recruit 'trios' of participants--children, mothers and fathers--as well as extended family members when appropriate and feasible.

For the iPSC component of the study, all CHD participants will be presented with the option to participate as part of the overall consenting process. Eligibility for this optional part of the study will

occur after enrollment and participation in the main study has begun. Those main study participants who meet the requirements for this optional part of the study and agreed to participation during the consenting process will be asked to provide the additional blood sample as described in the consent form. We will contact currently enrolled affected subjects or subjects with a genotype strongly associated with CHD or parent of affected subjects who have a defined gene mutation and are either still active subjects or are subjects who have given permission for future contact.

Recruitment strategies will include publicizing the study on the Bench to Bassinet website and individual site websites, through brochures and other printed materials, formal presentations by study investigators to their divisions or departments, individual conversations between investigators and patients, and other standard recruitment tools.

C.3 Genetic and Genomic Analyses

The study will use a variety of techniques to identify genetic characteristics associated with the specific CHDs of interest. Microarray analysis will be implemented to detect single nucleotide polymorphisms (SNPs) and copy number variants (CNVs). Genome wide association studies (GWAS) and evaluation of CNVs will seek to identify genetic variants associated with major categories of CHD (e.g., conotruncal defects, left-sided obstructive lesions) and, as numbers allow, within potentially more homogenous subsets (e.g., tetralogy of Fallot, hypoplastic left heart syndrome).

Whole exome sequencing will generate sequence data for all protein coding regions and whole genome sequencing will generate sequence data for the entire genome. Targeted analysis using molecular inversion probe sequencing (MIPS) complements these approaches by providing assessment of select genomic regions. Compared to publically available sequence data, rare variants with potential association to disease will be identified. Based upon data from GWAS, CNV analysis, whole exome sequencing, and whole genome sequencing, candidate genes or loci will be identified. Using capture based techniques and high throughput sequencing, sub genomic libraries consisting of these candidates will be resequenced in large numbers of patients. Such an effort will provide data that supports or refutes their causality in CHD and will define the spectrum of mutation type and location associated with CHD.

Participating sites who collect whole genome sequencing (WGS) as part of their standard clinical care will allow participants that have already had this done clinically, to opt out of providing a

research sample for this purpose. Results and data related to this clinical testing will be collected as detailed within this protocol.

New genetic tools are developed frequently, so other genetic techniques that become available during the study period will be evaluated and used as appropriate to achieve the study aims.

To identify genetic variants that are associated with a specific CHD or clinical outcome, several analytic approaches will be undertaken. Two examples include a familial approach using case-parent triad design and a case-control approach. The case-parent triad design will identify maternal genetic effects and de novo changes by comparing analyses of the proband's DNA with that of the proband's parents. In the case-control design, frequencies of identified genetic variants will be compared between cases and controls. As the technology and science advance, newer analytic approaches are likely to be added.

C.4 Return of Genetic Testing Results

For those participants who requested return of results on their consent, confirmation genetic testing will be arranged with a CLIA certified lab by a central genetic counselor. A genetic counselor will be available to answer questions. The cost for the confirmatory testing will be paid by the study. For cases of deceased minor participants, the research results will be shared as no new sample can be obtained. Each site will have a designated clinician who will receive the confirmation results and, if needed, will arrange for a clinical visit after disclosure of the results.

We will report results for genes (CNVs and SNVs) that are related to CHD and/or incidental findings found within the analyses in this protocol (if family consented to receive them), based on the current ACMG Secondary Findings Gene List. A committee of study PIs will meet periodically to review specific genes and variants that meet criteria to be classified as pathogenic or likely pathogenic in genes that meet ClinGen strong or definitive evidence for association with CHD. This will be an on-going process for the duration of the study as scientific knowledge changes over time.

C.5 Study Timeline

This is a long-term study, defined by the period of funding (and any extensions granted).

C.6 Stopping Rules

There are no planned stopping rules, other than elective withdrawal from the study or cessation of active investigation.

D. Collection of Samples and Data

Data and samples from the below previously IRB approved studies will be incorporated into this database. These studies will be closed with the Cincinnati Children's Hospital IRB once this combined protocol is IRB approved and all data and samples have been confirmed to be transitioned completely.

- 2016-0091 – Congenital Heart Disease Genetic Network Study (CHD Genes)
- 2016-0096 – iNDUCED PLURIPOTENT STEM CELLS (iPSC)

D.1 Sample Collection

The following samples will be collected.

D.1.1 Proband participant samples

1. A sample of whole blood will be obtained from the proband from an indwelling line or in conjunction with a clinically warranted venipuncture whenever possible. Otherwise, trained technicians will perform phlebotomy to obtain a sample of blood from the willing child/family. A sample will be acquired only if drawing said sample will not in any way harm the participant's clinical status or precipitate the need for a blood transfusion. The blood sample will be used to extract DNA and may be used to make a cell pellet and/or immortalized lymphoblastoid cell line to allow for a renewable resource. Approximately 3-10 ml of blood will be drawn from an infant while 10-15 ml will be drawn from anyone older than one year (the volume will not exceed that permitted by weight and will not exacerbate any existing medical conditions). While blood is preferred and should be prioritized, if blood cannot be drawn or the enrollee is unwilling, then a saliva sample (age permitting) will be obtained. In the event that the proband dies following consent, an autopsy specimen may be obtained. We may also request permission to use blood samples collected for clinical procedures that would otherwise be discarded.
2. We will ask permission to use leftover/discarded samples taken during medically indicated procedures for an enrolled proband participant or a mother who is

pregnant with an in-utero diagnosed proband participant during pregnancy including, but not limited to, leftover/ discarded DNA from amniocentesis, chorionic villus samples, or percutaneous umbilical blood sampling. Participation in this collection of discarded samples is optional for all sites. All sites will follow their institutional procedures and guidelines as outlined in IRB reliance and submission documents.

3. If the proband participant is undergoing a medically indicated surgical procedure, then tissue that during the course of surgery would have been discarded (such as discarded atrial tissue resulting from cannulation for cardiopulmonary bypass) may be acquired for research purposes. No additional sample will be collected specifically for this purpose. Participation in this collection of discarded samples is optional for all sites. All sites will follow their institutional procedures and guidelines as outlined in IRB reliance and submission documents.
4. OPTIONAL blood sample for iPSC - Subjects with a defined gene mutation (identified from their initial enrollment in the study) will be identified and will be re-contacted by the study site. If the subject consents, an additional blood sample (3-5ml) will be obtained and used for generating iPSCs.

D.1.2 Family samples

1. Up to 15 ml of whole blood will be drawn from consenting immediate and/or extended family. Blood draws will be done on site or a packet containing instructions for off-site phlebotomy will be sent. If a grandparent, sibling, or distant relative is unwilling or unable to undergo a venipuncture, then a sample of saliva will be acquired. In some instances, samples will be collected by mail. In this event, we will provide a self-addressed stamped return kit for the sputum or blood sample which will be in full compliance with standards for shipping biological materials.
2. We will ask permission to use leftover/discarded samples taken during medically indicated procedures during pregnancy including, but not limited to, leftover/ discarded DNA from amniocentesis, chorionic villus samples, or percutaneous umbilical blood sampling from prenatal cases. Participation in this collection of discarded samples is optional for all sites. All sites will follow their institutional

procedures and guidelines as outlined in IRB reliance and submission documents.

3. If an affected relative is undergoing a surgical procedure, then tissue that during the course of surgery would have been discarded (such as discarded atrial tissue resulting from cannulation for cardiopulmonary bypass) may be acquired for research purposes. No additional sample will be collected specifically for this purpose. Participation in this collection of discarded samples is optional for all sites. All sites will follow their institutional procedures and guidelines as outlined in IRB reliance and submission documents.

D.2 Clinical and Phenotypic Data

Clinical and phenotypic data will be collected as outlined in detail below. In addition, participants will be asked to agree to future contact for the purposes of obtaining additional information. Medical record review and/or programmatic extraction of structured data from the medical record may occur at baseline and longitudinally during the study period for patients who have not been withdrawn.

D.2.1 Proband participant clinical data collection

Proband participant data will include:

1. Cardiac anatomy
2. Additional congenital anomalies or diseases
3. Demographics
4. Medical history, including pregnancy and birth history of the affected proband, information on selected medical conditions and common pregnancy exposures, neurodevelopmental history, and medications
5. Cardiac history, including surgeries and procedures, hospitalizations, complications, and co-morbidities
6. Family history of CHD and other conditions
7. Pedigrees
8. Dysmorphic features
9. Participants may be recontacted during the study period (not more frequently than annually) to document vital status and any changes in medical or cardiac history.
10. Residential addresses, which can be used for geomarker computation. Note: Addresses along with all other PII will remain at the enrollment sites and will not be shared beyond the

sites. Only data that is not directly identifiable will be shared with the study. (see section F4 for details)

D.2.2 Phenotypic data collection techniques

Data collected on the proband participants will be compiled by the following methods:

1. Interviews with the proband participant and parents
2. Review of all medical records, including but not limited to inpatient and outpatients records, images and reports from echocardiograms, electrocardiograms, cardiac MRIs, cardiac catheterizations, and operative reports
3. Programmatic extraction of structured data from the medical record
4. Request for copies of outside records and collection of available pertinent data
5. Physical examination reports by a geneticist, if available

D.3 Family Data Collection

D.3.1 Parents data collection

Data collected on the proband participant's consenting parents may include:

1. Cardiac diagnoses (if any) and cardiac history
2. Relevant medical history including congenital anomalies
3. Demographics
4. Images and reports from any ECGs, cardiac MRIs and/or echocardiograms may be obtained, if the parent has a family history of interest to the research.

D.3.2 Extended family data collection

Data collected on the consenting extended family may include:

1. Cardiac history and diagnoses
2. Relevant medical history including congenital anomalies
3. Demographics
4. Images and reports from any ECGs cardiac MRIs, and/or echocardiograms may be obtained, if the family member has a family history of interest to the research.

D.3.3 Clinical data collection techniques

1. Interviews and survey(s)
2. Medical record review and/or programmatic extraction of structured data from the medical record

3. Queries from the National Death Index (NDI) at the National Center for Health Statistics to identify patients who have died since enrollment. Patient's first name, last name, middle initial, date of birth, father's surname, and sex will be sent via encrypted disc or secure file transfer to NDI to perform the query.

E. Study Subjects

While all patients with pediatric cardiovascular disease and adults with congenital heart disease are of interest, the study will initially focus on four CHD anatomic classifications:

- Atrial septal defects
- Conotruncal abnormalities
- Left-sided obstructive lesions
- Heterotaxy

Additionally, subjects without CHD who have a genetic syndrome strongly associated with CHD (such as Trisomy or Chr22q deletion) will also be recruited.

E.1 Inclusion Criteria

- One of the following:
    - Affected Individuals with CHD, or
    - Subjects without CHD who have a genetic syndrome strongly associated with CHD (such as Trisomy or Chr22q deletion), or
    - Pregnant women who have a fetus with diagnosed CHD
- Age 0 – 99 years of age
- Males and females

E.2. Additional Inclusion Criteria for OPTIONAL iPSC

- Sign separate iPSC consent form
- Participants with a genotype that conveys risk for CHD, regardless of the presence or absence of this birth defect.

E.2 Exclusion Criteria

Patients who meet any of the following criteria will be excluded from the study:

- Isolated patent foramen ovale (except in confirmed cases of trisomy 21)
- Isolated prematurity-associated patent ductus arteriosus
- Lack of consent

- Isolated pulmonary stenosis secondary to twin-twin transfusion syndrome

E.3 Source of Subjects

Patients will be recruited from participating facilities' outpatient clinic and inpatient services and wards that service pediatric patients, pediatric cardiology patients, adult cardiology patients, and women pregnant with fetuses diagnosed with CHD. Sources of potential participants other than the main centers and their auxiliary sites include interested investigators, health care facilities and organizations outside of the PCGC, and self-referral.

Proband participants will be identified through mechanisms including, but not limited to:

- Daily screening of in-patient admission lists and out-patient clinic schedules
- Health care provider referrals
- Screening of medical databases
- Screening rosters of prenatal patients
- Clinician or collaborator referrals from outside institutions
- Self-referral.

At each Center, an ascertainment (screening) log will include minimal data on all screened patients who are known to have an inclusion diagnosis and are approached for consent.
The data will include:

- Date of ascertainment
- Gender
- Age
- Race
- Ethnicity
- Cardiac diagnosis

De-identified data from the screening logs may be used for general study reporting related to demographic reports and report to the OSMB.

E.4 Subject Recruitment and Consent

At each clinical center, the Principal Investigator or other study staff will review cases and conduct subject recruitment. In compliance with institutional policies regarding recruitment and consent activities and good clinical practice, the parent(s) or legal guardian of potential study participants or the potential study participant (if 18 years or older) will be approached. The initial approach may be

as early as during the prenatal period. To capture known patients, an introductory letter and/or brochure describing the study may be sent to families before a scheduled cardiology encounter through an out-patient visit, cardiac catheterization procedure, etc. A web site for the Pediatric Cardiac Genomics Consortium will provide contact information for patients interested in the study.

Patients or parents of patients with CHD or a genotype strongly associated with CHD (such as Trisomy or Chr22q deletion) will be approached for the study. The study will be presented and the consenting process begun. If there is a family history of CHD, the investigators will attempt to recruit other family members into the study. Those families of interest will be approached only after the parent of the minor proband or the adult proband has given approval. Recruitment of other family members will comply with standard procedures and good clinical practice. These procedures may include asking the parents of the proband if they would allow the study team to contact the family member or asking the parents to have the family member contact the study team directly. Approaching extended family may involve sending a letter, sending the study brochure, or a study staff telephone call. The proband's immediate family will be made aware that in contacting extended family, the proband's diagnosis of CHD will most likely be revealed. However, no other information about the proband will be discussed with extended family. Immediate and extended family members will be informed that the study team will **not** reveal each person's decision about participation, medical history, or genetic results. Confidentiality rules restrict sharing of study data. However, each family member will be informed that the family pedigree may be published in scientific journals and reports. Published pedigrees do not list names but would detail clinical and genetic findings.

Consent and assent for newly enrolling participants will be obtained by the PI or designee or the Research Coordinator. The consenting process will comply with good clinical practices and institutional policies. Whenever possible, in-person consent will be obtained; however, in some centers phone consent will be allowed. Documentation for consent (in-person or phone) will be obtained and filed accordingly in the participant research file. The study will be explained to potential subjects using age-appropriate language and easily understood lay terms. To obtain consent for subjects who are minors, the study will be discussed in detail with at least one parent or guardian able to consent for the minor.

Whenever a developmentally appropriate older child or adolescent (aged 11-17) is the potential study subject, assent from the subject will be obtained in addition to parental permission from the

parent or legal guardian. Participants who reach the age of majority during the study period, (i.e. in most states age of majority is 18 but otherwise age of majority as determined by state/local laws), will be re-consented with the adult informed consent form before any further collection of data or study procedures. Data/samples that were collected prior to age of majority may remain in the study for analysis.

All sites will use a study consent and assent that is IRB/Ethics Committee (EC) approved. When remote or electronic consenting is used, institutional policies will be followed and appropriate documentation will be captured. After consenting, a copy of the fully signed consent will be given to participants for their records.

All consents will include language requesting permission to contact the participant at a time in the future to collect additional data that may be useful in the research.

E.4.1 Subject Consent for Currently Enrolled Participants

Currently enrolled participants who were consented under the previous two separate studies (as detailed previously) will not be reconsented to continue participation in this combined protocol. Site approved consent forms allow for the future use of participant's data/specimens.

At the time of site IRB study closure for previous studies, should a site IRB make a determination to require reconsent of individuals who will have ongoing contact with the study team, this requirement will be shared with and documented by the lead site. This determination will be submitted to the sIRB for the individual site that is affected. If, at the time of study closure for previous studies, a site IRB determines notification(s) of currently enrolled participants should occur, this information will be shared with and documented by the lead site. The ACC will provide sIRB approved education and template tools for these notifications and will notify the sIRB of the site requirement.

E.6 Voluntary Subject Withdrawal

If a subject should wish to withdraw from the study, data and samples collected prior to the withdrawal will remain in the database and biorepository, respectively. The subject might wish to have all research being conducted on his/her sample stopped. These requests will result in the sample being destroyed, the subject's name removed from the site-maintained enrollment log, and all data de-identified in the database. Clinical and genetic data already obtained would remain in the databases. If a subject wishes to have all research being conducted on their samples stopped and

their samples destroyed, the participant will be asked to provide this request in writing to the site Investigator.

Data from participants who are unable to give the requested volume of blood will be kept in the study database. All data collected will remain in the database, except as noted above.

E.7 Early Termination of a Subject's Participation

The study PI may choose to terminate a subject's participation if it is felt to be in his/her best interest for clinical or other reasons.

E.8 Study Termination

The study may be terminated only by the Sponsor or other related oversight organizations.

F. Statistical and Analytical Approaches

Various state-of-the-art technologies will be used to generate data on genetic variation that may be related to CHDs. As the technologies and methods for assessing genetic variation are rapidly evolving, it is likely that the most appropriate statistical approaches for analyzing the data will also change with time. However, examples of current approaches are discussed below.

Public databases will be used to provide information on pediatric and population controls. Depending on the available data and the specific hypothesis being evaluated, statistical analyses will be based on data from either case-parent triads or cases and controls.

In general, analyses will be conducted within the overall cohort, major categories of CHD (e.g., conotruncal defects, left-sided lesions) and, as numbers allow, within potentially more homogenous subsets (e.g., tetralogy of Fallot). Additional subgroups defined by clinical outcome may also be analyzed. Analyses of subsets of cases defined by race/ethnicity may also be conducted to ensure that findings based on the full case group are not biased due to population stratification (i.e., systemic differences in the ancestry of cases and controls that are unrelated to disease status). As illustrated by the sample size calculations presented below, very large sample sizes will be required to detect genetic variants that have only a modest impact on disease risk, whereas much smaller sample sizes will be required to detect genes that have a more dramatic impact.

F.1 Genome-Wide Association Studies (GWAS)

Genotype data from GWAS can be analyzed using data from cases and controls or from case-parent triads. Per case, the power to detect associations using case-parent and case-control approaches is similar. Although both approaches have strengths as well as limitations, the triad design has two major advantages over the case-control design:

- Genotype data from case-parent triads can be used to evaluate both maternal and inherited genetic effects, whereas case-control data can be used to evaluate only inherited genetic effects.
- Analyses of the inherited genotype using triad data are immune to bias resulting from population substructure, whereas this is a significant concern in the case-control setting.

Given these advantages, family-based analyses of GWAS data are likely to be preferred over case-control approaches. Hence, power calculations are presented based on analyses of data generated from case-parent triads. However, as noted above, the power of a case-control study with an equivalent number of controls will be comparable to that of a case-parent study.

Sample size and Power: Given the enormous number of statistical tests performed in any GWAS, it is necessary to set the p-value for declaring statistical significance quite low. It is generally agreed that for GWAS, the threshold should be on the order of 10^{-8}. To illustrate the need for large sample sizes for these investigations, minimum sample size requirements to achieve adequate power (>80%) to detect genotype relative risks ranging from 1.2-2.0 under an additive model of inheritance are provided in Table 1. These estimates were obtained using the computer program Quanto (version 1.2.3, Gauderman and Morrison, http://hydra/usc.edu/gxe), and apply to both maternal and inherited genotypes.

Table 1: Minimum sample size requirements for GWAS using $p<10^{-8}$ as the threshold for significance.

Genotype Relative Risk	**Frequency of High Risk Allele**				
	0.10	**0.20**	**0.30**	**0.40**	**0.50**
1.2	13,444	7,711	5,988	5,338	5,219
1.4	3,726	2,177	1,720	1,559	1,548
1.6	1,821	1,082	869	799	804
1.8	1,118	676	550	512	522
2.0	777	477	393	371	381

F.2 Copy Number Variant (CNV) and Structural Variant (SV) Analyses

Evaluation of *de novo* or inherited CNVs and other SVs (herein jointly referred to as SVs) will be conducted using a case-control approach, with controls drawn from publically available databases. In general, analyses of SVs will be based on comparison of the frequency of an identified SV or group of SVs in cases and controls. These comparisons will be made by chi-square analysis or Fisher's exact test and other appropriate statistical tests. Logistic regression analysis will be used to estimate the relative risks associated with groups of SVs. Groups of SVs that will be considered include:

- All SVs
- All SVs within a given gene or region
- All SVs within a given pathway or gene family
- Individual SVs

Where parental data is available, SVs will be classified as inherited or *de novo*.

Sample size and Power: Given the potentially large number of statistical tests that will be performed to assess SVs, it will be necessary to set the p-value for declaring statistical significance quite low. However, unlike GWAS studies, the number of tests to be performed cannot be specified *a priori*. For illustrative purposes, sample size requirements are presented for a p-value threshold of $5x10^{-5}$ (accounting for 1,000 comparisons). However, this threshold will be reconsidered when SV data are available, and can be used to estimate the number of comparisons that will actually be made. Minimum sample size requirements to achieve adequate power (>80%) to detect genotype relative risks ranging from 1.2-2.0 for common variants (minor allele frequency (MAF)$\geq$0.10) and, from 1.2-5.0 for rare variants (MAF=0.001 and 0.01), under an additive model of inheritance, assuming a 1:4 case-control ratio are provided in Table 2. These estimates were obtained using the computer program Quanto (version 1.2.3, Gauderman and Morrison, http://hydra/usc.edu/gxe).

Table 2: Minimum sample size requirements for SV analyses using $p<5x10^{-5}$ as the threshold for significance.

Genotype Relative Risk	**Frequency of High Risk Allele**						
	0.001	**0.01**	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**
1.2	404,600	40,900	4,600	2,600	2,100	1,800	1,800
1.4	108,200	11,000	1,200	740	590	540	540
1.6	51,200	5,200	610	370	280	280	280

1.8	30,500	3,100	310	230	190	180	180
2.0	20,600	2,100	355	160	130	130	130
3.0	6,400	660					
4.0	3,400	350					
5.0	2,200	230					

F.3 Candidate Gene, Whole-Exome, And Whole-Genome Sequence Analyses

Evaluation of sequence variants will be based upon comparison with publically available sequence data. In general, these analyses will focus on identification of rare variants and, in particular, variants that are identified in as few as one case and are unobserved among a large number of control sequences. Power calculations are not provided for the proposed sequence analyses, since the identification of rare variants found only in cases is of interest regardless of the associated p-value.

F.3.1 Analysis and annotation of the exome or MIPS sequencing data

We will use a mix of publicly available and privately developed tools. First, we will map sequence reads to the reference human genome. Reads outside of the targeted sequences will be discarded and statistics on coverage will be collected from the remaining reads. Once the likely mapping locations are identified, the reads will be aligned allowing gaps to the reference genome using BWAmem and comparable software packages in order to identify indels. Potential variants will be called using TERRA, Genome Analysis Toolkit or similar platforms, as well as custom tools already developed in-house at several of the participating centers. We will identify known variants from dbSNP, publicly available genomes and exomes, including the 1000 Genomes Project, TOPMed, UK Biobank, and the Genome Aggregation Database, and exomes previously sequenced at the participating centers.

Sequence variants will be filtered for quality, using criteria such as [30]:
a. Calls with a Phred-like quality score of less than 45 will be excluded
b. Reads supporting the major and minors alleles should have different start and end points
c. Total reads at variant bases should be > 10

For remaining sequence variants lying within exons, their predicted effect on the encoded protein will be determined and deleterious effects predicted, e.g., with Polyphen or predicted premature termination of the protein. The degree of evolutionary conservation will be assessed for

nonsynonymous changes using phyloP or similar scores calculated from multiple alignments of >40 vertebrate species, which are available from the University of California Santa Cruz genome browser (http://genome.ucsc.edu) as well as comparison with invertebrate orthologues.

Rare exonic variants that alter conserved residues and are predicted to have significant impact on protein function in genes will be studied further, such as those in genes that are expressed during development and/or in cardiac tissues. Variants per gene, or per gene class, will be compared with overall or ethnically matched control samples to determine any association with CHD. When appropriate, segregation analyses within the family will be assessed and Lod scores determined. In addition, sequence analyses of that gene in unrelated affected samples will be performed, and newly discovered variants will be examined as above. The significance of rare variants found only in affected cases will be assessed by p-values.

Rare noncoding genome sequences that alter regulatory elements may also alter gene transcription and or translation. These will be identified from whole genome sequencing (WGS) reads, aligned to GRCh38 with the Burrows-Wheeler Aligner (BWA-MEM).Variants will be called using GATK Best Practices recommendations and implemented for base quality score recalibration (QSR), indel realignment, and duplicate removal as described [33]. Standard hard filtering parameters will be used for single nucleotide variants and indel discovery across all PCGC and control WGS samples, followed by N+1 joint genotyping and variant QSR. As the analyses and interpretation of WGS is a rapidly evolving field, we will employ contemporary and emerging strategies to identify candidate variants, and will assess their functional significance in iPSC-CMs and/or CHD tissues.

F.4 Geomarker Acquisition and Analysis

In order to elucidate social determinants that might affect outcomes in CHD patients, we aim to study the relationships between place-based information and health outcomes; for example, air pollution and asthma, neighborhood crime and mental health, or community greenspace and IQ. Study subjects with location information, most commonly a residential mailing address, can be linked to databases of place-based information, or “geomarkers”, in order to conduct these studies. Geocoding is the process of translating a string of text referring to a location (e.g., mailing address) into coordinates on the earth’s surface (e.g., latitude and longitude). These coordinates are required to link participants to their estimated exposures to geomarkers – a process commonly called “geomarker assessment”. Some examples of geomarker assessment commonly performed in health studies using people include distance to the nearest major roadway – a commonly used as a

measure of estimated exposure to traffic related air pollution that is associated with increased risk of asthma – or neighborhood median household income – a commonly used as a measure of community deprivation associated with increased bed days spent in the hospital.
Both geocoding and geomarker assessment involve the use of personal identifiers (addresses or geocodes) and must be conducted in a HIPAA and IRB compliant manner. We will use solutions to this problem that ensure that any HIPAA information strictly stays at each research site and is never shared. One example of such a solution is the DeGAUSS software package [31, 32], which is a standalone, container-based application that can produce geocodes and conduct geomarker assessment. A container is a platform that wraps software into a complete filesystem containing everything it needs to run. For geocoding and geomarker assessment, this includes code, software libraries, and geospatial data. Usable on PC, Mac, or Linux machines, researchers can use DeGAUSS containers to geocode and conduct geomarker assessment without PHI leaving their local machine. After geomarkers are attached to subjects' health information, personal identifiers like address or location coordinates are removed, effectively making the information no longer PHI. This approach can facilitate sharing and collaboration among the PCGC consortium. In addition, the use of containers guarantees the software will always run the same, regardless of its environment, which is a vital requirement for reproducible research.

G. DATA MANAGEMENT

An electronic clinical data management system (CDMS) (Medidata Rave clinical data management system) will be used for the study that is designed to support reliable and secure data entry for clinical research purposes. The system also provides seamless integration of electronic Case Report Forms (eCRF), implementation of protocol amendments, and SAS and XML study data exports. In some instances, with prior approval from the study Steering Committee, some data to be used for additional analyses may be entered and stored in a study specific REDCap or other appropriate database.

G.1 Data Entry

Medidata Rave - Data can be entered directly into Medidata Rave from multiple study sites via a fully validated and 21 CFR Part 11 compliant, secure Web application and stored centrally. Data are entered by subject study identification number; names will not be linked with subject data in the database. Study sites will maintain records in secure areas linking the subject name with the identification number assigned for the study. Study sites will have full access to their own data and

be able to view these data remotely. Study staff will not be able to view subject data associated with other sites.

REDCap (or other appropriate database) – Data entered into any other appropriate database are entered by subject study identification number; names will not be linked with subject data in the database. Study sites will maintain records in secure areas linking the subject name with the identification number assigned for the study. Study sites will have full access to their own data and be able to view these data remotely. Study staff will not be able to view subject data associated with other sites.

G.2 Data Validation and Monitoring

Integrated into the Medidata Rave CDMS are real time validations, including both inter- and intra-instrument data checks. Inconsistent or questionable values are flagged during entry, and a query is automatically generated to the data entry client. These queries provide the information necessary to investigate any data entry errors or resolve questions regarding out-of-range or questionable values. Second-level query tracking allows monitors and data manager's real time access to unresolved queries as well as the date and time of query generation and resolution.

Data entered into REDCap or other appropriate database will be checked by the PCGC Administrative Coordinating Center (ACC) for data entry errors or questionable values and queries issued to the sites to clarify discrepancies.

G.3 Specimen Tracking

Specimen tracking is started from the time of receipt at the site, through shipment to the central biorepository, during handling at the biorepository, and through shipment to various core laboratories. The individual sites and the biorepository will use the ACC's electronic clinical data management system for all specimen tracking including: acquisition, processing, storage, QA/QC, withdrawals, and distribution for utilization.

Each specimen will be labeled with a number consisting of a 3-digit site number followed by a unique specimen number that is different from the subject's unique study ID number. The master list linking specimen IDs to the subject study ID numbers will be maintained under password protection in the CDMS at the ACC. This blinding code system will maintain the confidentiality of the specimens yet allow linkage of the specimens with phenotypic data for analyses by the PCGC investigators.

G.4 Data and Specimen Security and Integrity

G.4.1 Phenotype security and integrity

All data changes are written to an audit trail. The audit trail identifies the data item by table, column and key field. The entry includes the user, date and time, as well as the old value and new value. Data are saved at regular intervals during data entry to prevent loss of information in the event of a disruption of the Internet connection.

Several levels of security are employed to ensure privacy and integrity of the study data, including the following:

- Study access requires use of assigned user names and passwords.
- Individual roles and access levels are assigned by the study data manager.
- Passwords are changed regularly.
- Web-based entry uses secure socket layer (SSL) data encryption.
- Data will not be stored on laptop computers.

G.4.2 Genotype data security and integrity

All genotype data will be transferred to a central data repository, chosen by and agreed upon by the PCGC Steering Committee. All genetic data arising from collected specimens will have standard safeguards in place to ensure participant confidentiality as per good clinical practices, state and local HIPAA rules and IRB/EC guidance.

- The data will be stored in password protected, secure computers and databases at the core laboratories and at the participating consortia centers.
- The data will be identified by study/biospecimen ID number only. There will be no use of patient identifiers.
- Access to the databases will be restricted to PCGC investigators and their analysis teams. Authorizations will recorded upon PCGC investigator approval (via web form) and maintained in a database. Access will be granted to persons based upon consortium guidelines.
- Security, archiving, and backup systems will be in place to ensure data security.
- Access to linked data, which is only maintained at the individual enrollment centers, will be through an approval process at the steering committee level.

G.4.3 Data Sharing/Future Research

During the PCGC funding period, de-identified data (https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard) and specimens (such as blood, tissue, saliva) collected as part of this research will be made available through a Consortium-wide request and access process for ancillary research projects related to CHD and other conditions in accordance to individual participant consents.

Requests for accessing and obtaining stored de-identified data and specimens for any ancillary study will be reviewed to ensure no identifying information is being requested. All proposed projects will be reviewed and data shared following the current Manual of Operations/Manual of Procedures for data and specimen sharing.

At a future point, when funding for the PCGC has ended, the biospecimens will be deposited and maintained in a de-identified manner in an appropriate NIH-funded biorepository to be used as a future resource for the research community. Similarly, information from DNA analyses and the clinical research data sets will be placed into a central data repository, such as the National Center for Biotechnology Information (NCBI) repository and/or the NHLBI's BioData Catalyst database.

H. Quality Assurance and Quality Control Procedures

The ACC has primary responsibility for QC/QA activities of the phenotypic data. The ACC also requires that the sites complete certain QC activities, most of which are monitored by the ACC.

The key QC/QA activities are:

- Data management - oversight including database development and validation, data cleaning and edit checks
- Project management – overall management of site procedures, training for new staff, periodic training and coordinator meetings to ensure protocol adherence and understanding; review of ancillary studies for compliance with B2B policies
- Regulatory – ensuring sIRB oversight for this protocol and reporting for unanticipated problems and protocol violations/deviations; guidance and oversight for regulatory files and essential documents per ICH/GCP guidelines; ensuring adequate IRB approvals and/or exempt status determination is provided prior to releasing stored data or samples

I. Human Subjects Considerations

I.1 Potential risks

I.1.1 Psychological distress

I.1.1.1 Genetic testing: There is a chance that participation in this study could cause psychological distress. Some people involved in genetic studies feel anxious about the possibility of carrying (or their child carrying) an altered gene that places them at risk or that may be passed on to their children. There is a possibility that the testing may detect instances of non-paternity. In such cases, this information will not be shared with the participant, but will be entered into the database.

I.1.1.2 Participating in a repository: There is a chance that participation in a repository may cause psychological stress or long-term anxiety. For this study, all data or specimens entered into a repository will be de-identified.

I.1.2 Blood draw

Minor temporary discomfort may be associated with the removal of blood by venipuncture. There is a risk of bruising and a very small amount of bleeding associated with the blood drawing. There is also a very small risk of infection at the site. Whenever possible, blood samples will be gathered when the participant is scheduled for routine blood testing or procedures.

I.1.3 Saliva collection

There are no known risks to collecting saliva.

I.1.4 Leftover or discarded clinically collected samples

There is no risk associated with the collection of otherwise discarded samples taken during a medically indicated procedure.

I.1.5 Retrospective clinical data collection

There are no significant procedural risks associated with physical examination. The most likely risk for collection of existing data is breach of confidentiality. All study teams will make every effort to ensure the protection of study data during storage and transmission. The least amount of PHI needed to conduct the study will be collected and secured.

I.2 Confidentiality, Protection against Risks

Investigators will take all reasonable measures to protect the confidentiality of subjects and their families.

- Investigators will arrange for counseling, at the participant's expense, if anxious feelings arise in the proband participant or family at any time during the study.
- The results of the genetic tests performed for research purposes will not be placed in the medical record.
- Non-paternity information will be kept in the strictest confidence and will not be divulged to research subjects or their families.
- Each child and parent is assigned a subject identification number (SID). All interview and clinical research data are stripped of direct identifiers and labeled with the study number. The enrollment log with participant identifiers will be maintained at each site in a secured, locked location available only to the study staff.
- The study will follow good clinical practices at all times. Databases will be secured as previously discussed and samples will be stored in secure study freezers.
- All participating laboratories and analysis facilities will follow good clinical practices maintaining data integrity and participant confidentiality.
- Samples for DNA will be stripped of the study SID at the laboratory and assigned specimen numbers that are linked to the SID.
- The risk of breach of subject confidentiality will be minimized by storage of all study materials in a locked file cabinet in a location separate from the laboratory data. The informed consent form states that study data will be made available to the ACC and NIH/NHLBI to ensure study safety and quality control.
- The subject's name and any other identifying information will not appear in any presentation or publication resulting from this study.
- The study team will contact extended family members for recruitment according to institutional policies. Contact will be made with those individuals who have expressed a willingness to at least learn about the research study. Other family members will not be informed of who is and is not participating. The subject will also be warned not to disclose their participation in order to protect their own privacy.
- Results of testing on biological specimens will be shared with the participants who consent to receive them, after confirmation in a clinical lab (CLIA certified lab). Only pathogenic or likely pathogenic CHD genetic results and/or actionable, incidental results, will be reported

to participants. If indicated, a clinical referral will be made for long term care. For cases of deceased minor participants, the research results will be shared as no new sample can be obtained. At the end of the study, the results of the genetic testing may be published for all the subjects as a group. There is a reasonable possibility that no findings will result from this research effort. If findings are detected, it may be years before any utility of these findings are realized.

- In the future, information from DNA analyses and clinical studies or medical records will be placed into an NIH-governed central data repository such as the National Center for Biotechnology Information's database of genotypes and phenotypes (dbGaP). When the results of the genetic tests and other study data are placed in a federal data repository, only doubly de-identified IDs will be transmitted to the repository. Per dbGaP standards a two step removal of identifiers will be used (https://www.ncbi.nlm.nih.gov/gap/docs/submissionguide/). The purpose of a central data repository is to make the study data available for future, yet to be determined, research. The repository has two databases, open access and controlled access. The open access database is available to anyone and includes summarized data regarding the entire collection of study subjects. These summarized data are not linked to medical or personal information. The controlled access database includes de-identified medical information and genomic data that are made available to researchers who have received approval from an NIH Data Access Committee (DAC). The NIH DAC reviews proposed submissions for consistency with appropriate policies to protect the privacy of research participants and confidentiality of their data.
- If an incidental finding is found on a study clinical test such as an ECG or echocardiogram, the PI or other qualified member of the research team will take full responsibility for disclosing the findings to the patients/parents, communicating with their primary care physicians with permission, and making appropriate cardiology referrals as indicated. The subject may choose to seek a second opinion and/or appropriate clinical care. This might change the subject's insurability and employability as it relates to the clinical finding only. The presumption is that detection of a potentially clinically significant finding will prove to be beneficial to the subject in the long run.
- The study is covered under a Certificate of Confidentiality from the National Institutes of Health (NIH). With this Certificate, the researchers of this study cannot be forced to disclose information that may identify a subject, even by a court subpoena, in any federal, state, or local civil, criminal, administrative, legislative, or other proceedings. The Certificate cannot

be used to resist a request for information from the United States government when it is used for evaluating federally funded study projects or for information that must be disclosed to meet the requirements of the Food and Drug Administration (FDA). A Certificate of Confidentiality does not prevent a subject or his/her family from voluntarily releasing information about the subject's involvement in this research. If an insurer, employer, or other person obtains a subject's or family's written consent to receive research information, then the researchers will not use the Certificate to withhold that information.

I.3 Potential benefits

In most cases, the study will not provide direct benefit to individual participants or families. The benefits are those to society as a whole in the improvement of knowledge of the causes of CHD, in the development of new diagnostic tests and ultimately in the improvement of treatment and prognosis.

Research using samples and data may result in inventions or discoveries that could create new tests and medicines that may have commercial value. Although subjects and their families will not receive any compensation now or in the future for their samples or data, income may be derived from future research or commercial use of the grouped data and will be used to support biomedical research.

I.4 Risk/Benefit Ratio and Importance of Information to Be Obtained

The risk/benefit ratio is favorable for this study and adverse events are not anticipated. The baseline risk is minimal because there are no therapeutic interventions. In addition, although an individual subject may not benefit from participation, the results of this study will make important contributions to the improvement of knowledge of the causes of CHD, in the development of new diagnostic tests, and ultimately in the improvement of treatment and prognosis.

I.5 Safety Monitoring

Consistent with NHLBI policy regarding monitoring of clinical research studies, an Observational Study Monitoring Board (OSMB) has been established. This group is independent and serves in an advisory capacity to NHLBI. The OSMB will meet at least yearly (or more frequently if needed) to review data on enrollment, demographics, specimen counts and quality, data completeness and quality, protocol violations, unexpected events and adverse events; these data are prepared for review by the ACC. As there are no interventions in CHD GENES, adverse events are not expected during the conduct of this study. However, if any adverse events related to study

activities are documented, they will be reported to the ACC, OSMB and NHLBI. Adverse events related to study activities should also be reported to IRBs in accordance with institutional policies.

I.6 Inclusion of Women and Minorities

All eligible subjects will be enrolled in this study without regard to gender, race, or ethnicity. The distributions of subjects according to gender, race, and ethnicity will reflect both the epidemiology of congenital heart defects and the referral patterns to participating Centers.

Epidemiologic studies have demonstrated that individuals of varying ethnic and racial backgrounds are affected with CHD with comparable frequency [3]. Participants of all races and ethnic backgrounds will be included in this study. The study will be offered to English- and Spanish-speaking people. Non-English speaking participants will be consented following institutional policies and expectations. If fully translated consent documents are required for any site, these will receive sIRB approval prior to use.

Epidemiologic studies have also demonstrated that approximately equal numbers of males and females are affected with CHD overall, although at birth there is a slight male preponderance. However, left-sided obstructive lesions and conotruncal defects occur significantly more frequently in males (60%) than females for reasons that are not understood [28, 29]. Although this study will approach all eligible subjects regardless of gender, a preponderance of males will likely result as a reflection of the population affected with left-sided obstructive lesions and conotruncal defects. The gender distribution of patients enrolled in this program with other types of CHD will be approximately equal between males and females reflecting the general affected population.

J. Compensation and Costs

J.1 Compensation

Individual centers will determine if, and how, participants will be compensated.

J.2 Cost

There is no cost of participation in this study.

K. References

1. Perry LW, Neill CA, Ferencz C, Rubin JD, Loffredo CA: Infants with congenital heart disease: the cases. In: Epidemiology of Congenital Heart Disease: The Baltimore-Washington Infant Study 19811989. Edited by Ferencz C, Rubin JD, Loffredo CA, Magee CA, vol. 4. Mount Kisco: Future Publishing Company, Inc.; 1993: 33-62.
2. Botto LD, Correa A, Erickson JD: Racial and temporal variations in the prevalence of heart defects. Pediatrics 2001, 107(3):E32.
3. Ferencz C, Rubin JD, McCarter RJ, Brenner JI, Neill CA, Perry LW, Hepner SI, Downing JW: Congenital heart disease: prevalence at livebirth. The Baltimore- Washington Infant Study. Am J Epidemiol 1985; 121(1):31-36.
4. Pierpont, M.E., et al., Genetic basis for congenital heart defects: current knowledge: a scientific statement from the American Heart Association Congenital Cardiac Defects Committee, Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. Circulation, 2007; 115(23): p. 3015-38.
5. Ferencz C, Rubin JD, McCarter RJ, Brenner JI, Neill CA, Perry LW, Hepner SI, Downing JW: Congenital heart disease: prevalence at livebirth. The Baltimore-Washington Infant Study. Am JEpidemiol 1985; 121(1):31-36.
6. Petrini J, Damus K, Johnston RB, Jr.: An overview of infant mortality and birth defects in the United States. Teratology 1997, 56(1-2):8-10.
7. Rosano A, Botto LD, Betting B, Mastroiacovo P: Infant mortality and congenital anomalies from 1950 to 1994: an international perspective. J Epidemiol Community Health 2000, 54(9):660-666.
8. Brown MD, Wernovsky G, Mussatto KA, Berger S: Long-term and developmental outcomes of children with complex congenital heart disease. Clin Perinatol 2005, 32(4):1043-1057.
9. Goldmuntz E, Clark BJ, Mitchell LE, Jawad AF, Cuneo BF, Reed L, McDonald-McGinn D, Chien P, Feuer J, Zackai EH et al: Frequency of 22q11 deletions in patients with conotruncal defects. J Am Coll Card/0/1998, 32(2):492-498.
10. Goldmuntz E, Bamford R, Karkera JD, dela Cruz J, Roessler E, Muenke M: CFC1 mutations in patients with transposition of the great arteries and double-outlet right ventricle. Am J Hum Genet 2002, 70:776-780.Krantz ID, Smith R, Colliton RP, Tinkel H, Zackai EH, Piccoli DA, Goldmuntz E, Spinner NB: Jagged l mutations in patients ascertained with isolated congenital heart defects. Am J Med Genet 1999, 84(1):56-60.
11. Gaynor JW, Gerdes M, Zackai EH, Bernbaum J, Wernovsky G, Clancy RR, Newman MF, Saunders AM, Heagerty PJ, D'Agostino JA et al: Apolipoprotein E genotype and

neurodevelopmental sequelae of infant cardiac surgery. J Thorac Cardlovasc Surg 2003, 126(6): 1736-1745.

12. Mahle WT, Crisalli J, Coleman K, Campbell RM, Tam VK, Vincent RN, Kanter KR: Deletion of chromosome 22q11.2 and outcome in patients with pulmonary atresia and ventricular septal defect, Ann Thorac Surg 2003, 76(2):567-571.
13. Hobbs CA, James SJ, Persian A, Krakowiak PA, Jernigan S, Greenhaw JJ, Lu Y, Cleves MA: Congenital heart defects and genetic variants in the methylenetetrahydroflate reductase gene. J Med Genet 2006, 43(2): 162-166.
14. Michielon G, Marino B, Formigari R, Gargiulo G, Picchio F, Digilio MC, Anaclerio S, Oricchio G, Sanders SP, Di Donate RM: Genetic syndromes and outcome after surgical correction of tetralogy of Fallot. Ann Thorac Surg 2006, 81(3):968-975.
15. Cleves MA, Ghaffar S, Zhao W, Mosley BS, Hobbs CA: First-year survival of infants born with congenital heart defects in Arkansas (1993-1998): a survival analysis using registry data. Birth Defects Res A Clin Mol Teratol 2003, 67(9):662-668.
16. Gaynor JW, Wernovsky G, Jarvik GP, Bernbaum J, Gerdes M, Zackai E, Nord AS, Clancy RR, Nicolson SC, Spray TL: Patient characteristics are important determinants of neurodevelopmental outcome at one year of age after neonatal and infant cardiac surgery. J Thorac Cardlovasc Surg 2007, 133(5):1344-1353, 1353 e1341-1343.
17. Kyburz A, Bauersfeld U, Schinzel A, Riegel M, Hug M, Tomaske M, Valsangiacomo Buchel ER: The fate of children with microdeletion 22q11.2 syndrome and congenital heart defect: clinical course and cardiac outcome. Pediatr Cardiol 200Q, 29(1):76-83.
18. Ziolkowska L, Kawalec W, Turska-Kmiec A, Krajewska-Walasek M, Brzezinska-Rajszys G, Daszkowska J, Maruszewski B, Burczynski P: Chromosome 22q11.2 microdeletion in children with conotruncal heart defects: frequency, associated cardiovascular anomalies, and outcome following cardiac surgery. Eur J Pediatr 2008, 167(10): 1135-1140.
19. Hyett J, Moscoso G, Nicolaides KH. Abnormalities of the heart and great arteries in first trimester chromosomally abnormal fetuses. Am J Med Genet1997; 69:207–16.
20. Mazzanti L, Cacciari E. Congenital heart disease in patients with Turner's syndrome. Italian Study Group for Turner Syndrome (ISGTS). J Pediatr. 1998 Nov; 133(5):688-92.
21. McElhinney DB, Geiger E, Blinder J, Benson DW, Goldmuntz E: NKX2.5 mutations in patients with congenital heart disease. J Am Coll Cardiol 2003, 42(9): 1650-1655.
22. Goldmuntz E, Geiger E, Benson DW: NKX2.5 mutations in patients with tetralogy of fallot. Circulation 2001, 104(21):2565-2568.

23. McElhinney DB, Krantz ID, Bason L, Piccoli DA, Emerick KM, Spinner NB, Goldmuntz EG. Analysis of cardiovascular phenotype and genotype-phenotype correlation in individuals with a JAG1 mutation and/or Alagille syndrome. Circulation 106:2567-2574, 2002.
24. Gebbia M, Ferrero GB, Pilia G, Bassi MT, Aylsworth A, Penman-Splitt M, Bird LM, Bamforth JS, Burn J, Schlessinger D, Nelson DL, Casey B (1997) X-linked situs abnormalities result from mutations in ZIC3. Nat Genet 17:305–308.
25. Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, Grossfeld PD, Srivastava D. Mutations in NOTCH1 cause aortic valve disease. Nature. 2005 Sep 8; 437(7056):270-4. Epub 2005 Jul 17.
26. Benson DW. Advances in cardiovascular genetics and embryology: role of transcription factors in congenital heart disease. Curr Opin Pediatr 2000; 12:497-500.
27. Momma K, Kondo C, Ando M, Matsuoka R, Takao A. Tetralogy of Fallot associated with chromosome 22q11 deletion. Am J Cardiol 1995; 76:618-21.
28. Ferencz C, Loffredo CA, Correa-Villasenor A, Wilson PD. 1997. Malformations of the cardiac outflow tract. Genetic and Environmental Risk Factors of Major Cardiovascular Malformations: The Baltimore-Washington Infant Study: 1981-1989. Armonk: Future Publishing Company, Inc. p 59-102.
29. Ferencz C, Loffredo CA, Correa-Villasenor A, Wilson PD. 1997. Left-sided obstructive lesions. Genetic and Environmental Risk Factors of Major Cardiovascular Malformations: The Baltimore-Washington Infant Study: 1981-1989. Armonk: Futura Publishing Company, Inc. p 166-225.
30. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci 2009, Nov 10; 106(45):19096-101.
31. Brokamp C, Wolfe C, Lingren T, Harley J, Ryan P. Decentralized and Reproducible Geocoding and Characterization of Community and Environmental Exposures for Multi-Site Studies. Journal of American Medical Informatics Association. 25(3). 309-314. 2018.
32. Ryan P, Brokamp C, Blossom J, Lothrop N, Miller R, Beamer P, Visness CM, Zanobetti A, Andrews H, Bacharier LB, Hartert T, Johnson CC, Ownby D, Lemanske RF, Gibson H, Requia W, Coull B, Zoratti EM, Wright AL, Martinez FD, Seroogy CM, Gern J, Gold DR, on behalf of the CREW Consortium. A Distributed Geospatial Approach to Describe Community Characteristics for Multi-Site Studies. Journal of Clinical and Translational Science. 5:e86, 1-8. 2021.
33. Richter F, Morton SU, Kim SW, Kitaygorodsky A, Wasson LK, Chen KM, Zhou J, Qi H, Patel N, DePalma SR, Parfenov M, Homsy J, Gorham JM, Manheimer KB, Velinder M, Farrell A, Marth G, Schadt EE, Kaltman JR, Newburger JW, Giardini A, Goldmuntz E, Brueckner M, Kim R,

Porter Jr GA, Bernstein D, Chung WK, Srivastava D, Tristani-Firouzi M, Troyanskaya OG, Dickel DE, Shen Y, Seidman JG, Seidman CE, Gelb BD. Genomic analyses implicate noncoding de novo variants in congenital heart disease. Nat Genet. 2020 Aug; 52(8):769-777.