Project Singular: Building a Large-Scale Genetics Dataset to Fuel Research Towards Curative Solutions for Single Ventricle Heart Disease

Diane M Pickles, Kirstie E Keller, PhD; Additional Ventures, Palo Alto, CA

No Disclosures

PROJECT SINGULAR

Hope Through Research For Single Ventricle Hearts

BACKGROUND

Congenital heart defects (CHDs) are the most common birth defects affecting nearly 1 in 100 live births (1). Single Ventricle (SV) defects comprise almost 8% of all CHDs and include several disparate clinical diagnoses, including hypoplastic left heart syndrome (HLHS), tricuspid atresia, and pulmonary atresia with an intact ventricular septum (2,3). While considered universally fatal fifty years ago, now most individuals born with SV undergo two to three life-saving, palliative surgeries in the first years of life, culminating in the Fontan operation (2,3). Since the initial intervention, the type, number, and timing of surgeries undergone has evolved considerably, with 80% of SV patients born today living into their 30s because of an improved care paradigm (5), as compared with the initial cohort of patients, whose post-surgery survival at 20 years was less than 50% (6).

There were an estimated 70,000 Fontan patients living worldwide in 2017, and that number is projected to double in the next 20 years (5). Yet, in addition to the invasive surgeries required, most SV patients still experience profound co-morbidities and complications. Physical impacts are daunting and include circulatory failure, arrhythmia, liver fibrosis, and renal dysfunction; equally difficult for patients are the prevalent cognitive, neuropsychological, and behavioral deficits (5). A normal longevity and quality of life are impossible, and the experience of living with an SV defect remains filled with physical and emotional pain and trauma.

5 of every 100,000 babies are born with a single ventricle heart defect

4 genes have been linked to single ventricle heart defects, but none are causal

Right now, there is **NO cure** for single ventricle heart defects

The underlying etiology of SV is poorly understood, with mounting evidence that suggests that the basis is comprised of genetic, epigenetic, mechanobiological, and/or environmental contributions – emerging as a complex disease where there is multiplicity of presentation and outcomes (7-9). To date, four genes have been linked to SV disease (table 1); however, the penetrance is low, and none are causal. This hinders efforts to both predict and prevent the disease and to define the molecular and cellular physiological differences that underpin risk and long-term outcomes.

Project Singular, a study of Additional Ventures, aims to uncover the cause(s) of SV by sequencing at least 5,000 SV patients and immediate family members, in conjunction with collecting phenotypic data, creating the largest genetic dataset in this disease community, and making it available free of charge to researchers.

Candidate Gene	Gene Function	SV Type	References
Connexin43 (GJA1)	Gap-junction protein	HLHS	10
NKX2-5	Cardiac homeobox transcription factor	HLHS, DORV	11
NOTCH1	Membrane ligand-receptor	HLHS	12–15
MYH6	Gap-junction protein	HLHS	15
Other identified potential candidates			
RBFOX2, HAND1, FOX cluster (FOXF1, FOXC2, FOXL1), PTCH1, IRX4, TBX5, BMP2/BMPR2, ETS-1, JAG1, IRX4, MLL2/KMT2D, HUWEI1			6,9,14

Table 1. Genes found to be associated with single ventricle heart defects.

METHODS

ENROLLMENT & SURVEY

Subjects are recruited through a variety of methods including social media outreach, partnerships with existing patient advocacy organizations, and collaborations with clinical sites to provide informational materials regarding the study. Participants enroll virtually through an online portal at projectsingular.org, initially launching in English and later launching in Spanish and Simplified Chinese. First, participants confirm eligibility (individual with SV, parent or sibling of individual with SV, or child of an individual with SV, living in the United States or Canada) and consent to participate in the study; patients sign a medical record release. Both consent and medical release forms are audited by Project Singular study staff.

All participants fill out a brief online survey with patients providing information about their health history, diagnosis, and geographic and demographic information. Survey data is deidentified and loaded into the Terra.Bio platform. Participants receive a biospecimen collection kit (saliva sample for ages 3 and up; buccal sample for under 3 years old) by mail to return to the Broad Institute of MIT and Harvard for sequencing.

RETURN OF RESULTS

While there are no primary results to return, participants can opt in to receive actionable secondary results, which include 73 genes with recommendations for reporting incidental findings in clinical exome and genome sequencing published by the American College of Medicine and Genomics (ACMG). Sequences are run through Fabric, a program developed by the Rare Genomes Project, to identify pathogenic or likely pathogenic variants. If a variant is detected, the participant will be re-identified and staff will confirm the participant's desire for return of results. If confirmed, participants will be referred to Genome Medical, a commercial organization, who will perform confirmatory testing and subsequent genetic counseling, if needed, as well as offering cascade screening for family members, all at no charge to participants. Any errors in variant detection will be reported to Project Singular staff.

MEDICAL RECORD RELEASE & PATIENT DATA

Affected participants have three options to provide their medical records: upload records to the online portal, utilize a clinical research coordinator (CRC) at Boston Children's Hospital for assistance in acquiring and uploading records, or have the CRC reach out to clinical providers for the patient. Once the medical records are acquired, the CRC confirms diagnosis and enters clinical data, which includes detailed cardiac anatomy, medical and surgical history, and prior genetic testing results, into the Broad Institute's Data Study Management (DSM) platform. Affected participant medical records are securely stored in a Google Cloud Bucket. Patient data is deidentified and then transferred into the Terra.Bio platform where it is merged with genomics data. Additional phenotypic patient data will be integrated from other existing SV registries to add power to the dataset.

WHOLE GENOME SEQUENCING

Saliva samples will be processed, quality-controlled, and whole genome-sequenced using Illumina HiSeq at 30X coverage, with methylation arrays conducted on a subset. Sequences will be aligned to the human genome, annotated, deidentified, and securely deposited into the Terra.Bio platform, and merged with the phenotypic information from the participant survey and patient medical record.

DATA ACCESS & USAGE

Deidentified phenotypic and genotypic data will be made available for free to qualified researchers globally after a simplified application process and review by Data Access Committee. Data is available for general research use; commercial use is not prohibited; data use for methods development research irrespective of the specified data use limitations is not prohibited; future use as a control set for any type of health/medical/biomedical study is not prohibited. In addition, utilization of the dataset will be promoted via grant programs and data hackathons sponsored by Additional Ventures.

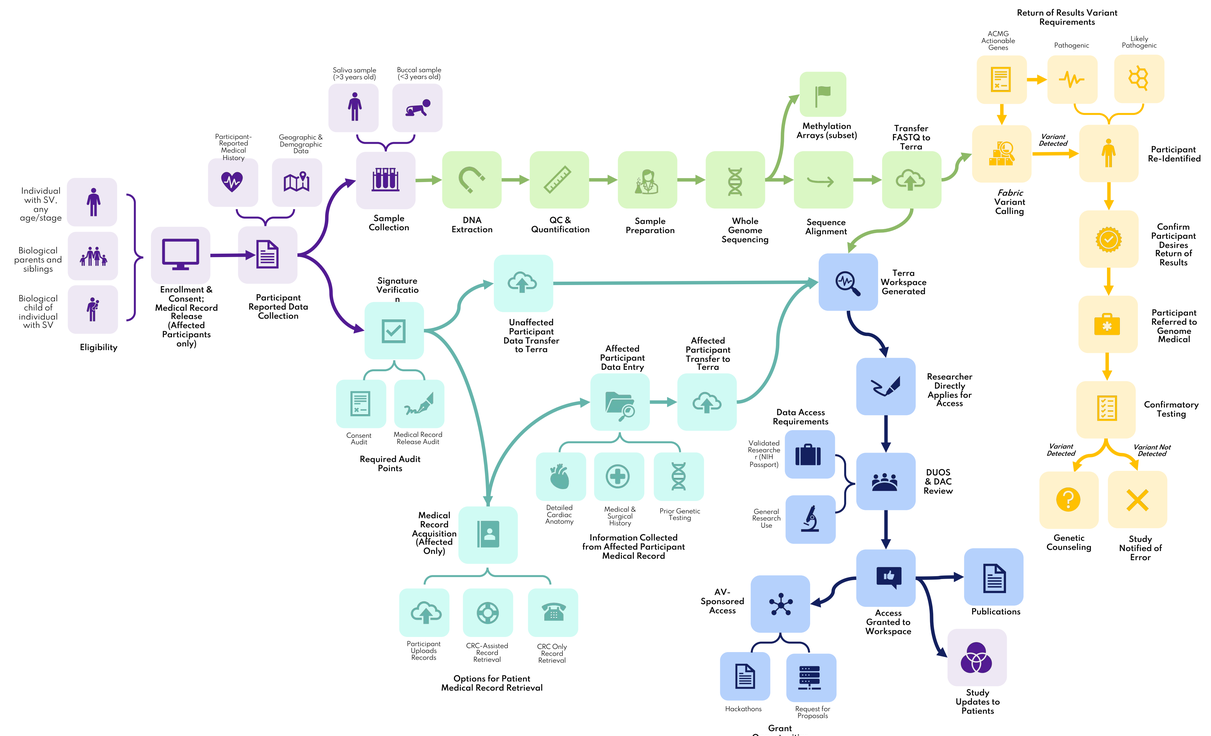

Figure 1. Project Singular Workflow. Schematic of patient, sample, and data processing at Project Singular. Patient Workflow (Purple) describes how eligible participants [individual with single ventricle, any age or stage; biological parents and siblings; and biological children of single ventricle patients – in the United States and Canada] enroll, complete surveys about medical history, demographics, and geographic location, and receive a sample collection kit. Participant Data Workflow (Teal) describes the process by which Project Singular study staff verify enrollees and how medical records are acquired and subsequently entered within the system. Patient information includes cardiac anatomy-based diagnosis, medical and surgical history, and prior genetic testing results. Patient and participant data is transferred into the Terra.Bio platform where it is merged with genomics data. Genomics Workflow (Green) describes the sample processing and ultimate upload into the Terra.Bio platform after alignment and annotation are merged with phenotypic data. Return of Result Workflow (Yellow) Fabric Program is run on all participant sequences to identify variants listed on the American College of Medical Genomics (ACMG) list of recommended incidental findings to be returned as secondary results. Secondary results are returned through a workflow in partnership with the Rare Genomes Project and Genome Medical, who will conduct confirmatory testing and genetic counseling, as needed. Data Usage Workflow (Blue) Researchers apply for data access via the Data Use Operating System with an NIH Commons ID, and requests are reviewed by the Data Access Committee. Additional Ventures incentivizes utilization and promotes awareness of dataset via RFP/grant programs and data-hackathons.

RESULTS

The online portal launched in May 2022 to a small group of potential participants as a pilot study to determine the effectiveness of the enrollment pipeline, medical record acquisition, and sequencing pipeline. To date, 21 individuals have completed enrollment; 9 of these are patients [7 adults, 2 minors] and 12 are family members [8 birth parents, 3 siblings]. There are 5 family cohorts enrolled. Of these, 3 families consist of 2 parents, an affected child, and an unaffected sibling, 1 family consists of 2 parents and an affected child, and 1 family consists of a parent and affected child (figure 2). Medical records are in the process of being acquired and sequencing is underway on returned samples. Feedback from the pilot has led to alterations within the enrollment pipeline, including enrollment workflow instructions and clarifications on the medical record release language.

A comprehensive recruitment and public launch strategy is planned for Fall 2022. Recruitment will include promotional materials distributed via clinics, organic and paid social media, and outreach via patient advocacy and support organizations.

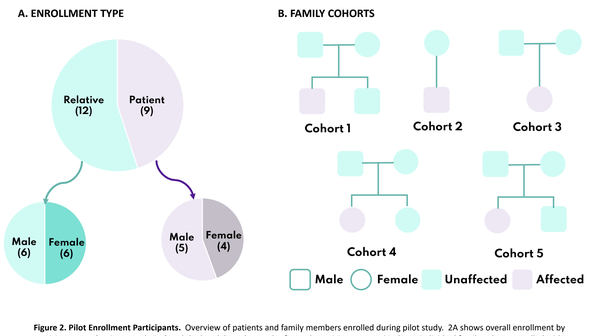

Figure 2. Pilot Enrollment Participants. Overview of patients and family members enrolled during pilot study. 2A shows overall enrollment by participant type (patient or relative) and the breakdown by gender for each participant type. 2B shows individual family cohorts enrolled with detail on gender and affected and unaffected family members.

DISCUSSION

To move toward an understanding of the genetic factors involved in single ventricle heart disease and related sequela, we recognize the need to use whole genome sequencing (WGS) on thousands of samples. We believe that the undiscovered risk variants and casual mutations will be rare and possibly involved in combinations or other complex mechanisms, such as structural alterations. We foresee the possibility that these variants will reside in noncoding regions, which have been largely unexplored by other sequencing efforts. The volume of sequences may help to resolve the role of common variants in congenital heart defects and integrate these data with those on rare variants to understand issues of penetrance, variable expressivity, and pleiotropic effects.

Successful participant enrollment will necessitate overcoming potential trust and privacy concerns around sharing genetics data and medical records, particularly without the potential for return of primary personal results. Working with cardiac centers and local providers will be essential to reach rural, non-English speaking, and low literacy populations.

Together, identifying and understanding the genetic, epigenetic and environmental contribution to SV is critically important. Such knowledge will provide insight into the origin, pathology, mechanisms, and progression of SV and may provide avenues for developing new animal models, predictive measures of disease, genetic counseling tools, clinically-relevant biomarkers, and interventions.

REFERENCES

1. Hoffman, J.I.E.; Kaplan, S. The Incidence of Congenital Heart Disease. J. Am. Coll. Cardiol. 2002, 39, 1890–1900.
2. Heaton, J.; Heller, D. Single Ventricle. In Consults in Obstetric Anesthesiology; Springer: Cham, Switzerland, 2021; pp. 535–536.
3. Rao, P.S. Single Ventricle—A Comprehensive Review. Children 2021, 8, 441.
4. Fontan, F.; Baudet, E. Surgical Repair of Tricuspid Atresia. Thorax 1971, 26, 240–248.
5. Rychik et al. Evaluation and Management of the Child and Adult With Fontan Circulation: A Scientific Statement From the American Heart Association. Circulation 2019, 140, e234–e284.
6. Life after Fontan Palliation—Mayo Clinic. Available online: https://www.mayoclinic.org/
7. Yagi, H. et al. The Genetic Landscape of Hypoplastic Left Heart Syndrome. Pediatr. Cardiol. 39, 1069–1081 (2018).
8. Feinstein, J. A. et al. Hypoplastic left heart syndrome: current considerations and expectations. J. Am. Coll. Cardiol. 59, S1-42 (2012).
9. Saraf, A., Book, W. M., Nelson, T. J. & Xu, C. Hypoplastic left heart syndrome: From bedside to bench and back. J. Mol. Cell. Cardiol. 135, 109–118 (2019).
10. Dasgupta, C. et al. Identification of connexin43 (alpha1) gap junction gene mutations in patients with hypoplastic left heart syndrome by denaturing gradient gel electrophoresis (DGGE). Mutat. Res. 479, 173–186 (2001).
11. Elliott, D. A. et al. Cardiac homeobox gene NKX2-5 mutations and congenital heart disease: associations with atrial septal defect and hypoplastic left heart syndrome. J. Am. Coll. Cardiol. 41, 2072–2076 (2003).
12. Theis, J. L. et al. Compound heterozygous NOTCH1 mutations underlie impaired cardiogenesis in a patient with hypoplastic left heart syndrome. Hum. Genet. 134, 1003–1011 (2015).
13. Iascone, M. et al. Identification of de novo mutations and rare variants in hypoplastic left heart syndrome. Clin. Genet. 81, 542–554 (2012).
14. Garg, V. et al. Mutations in NOTCH1 cause aortic valve disease. Nature 437, 270–274 (2005).
15. Durbin, M. D. et al. Hypoplastic Left Heart Syndrome Sequencing Reveals a Novel NOTCH1 Mutation in a Family with Single Ventricle Defects. Pediatr. Cardiol. 38, 1232–1240 (2017).

Project Singular is funded by Additional Ventures, www.additionalventures.org

ADDITIONAL VENTURES