Centers for Common Disease Genomics
Cardiovascular Disease (CCDG CVD) Working Group Plan Addendum 1

Whole Genome Sequencing in Early Onset Atrial Fibrillation

V 09012016

Introduction

The National Heart Lung and Blood Institute, as part of its Trans-Omics for Precision Medicine (TOPMed) program: https://www.nhlbi.nih.gov/research/resources/nhlbi-precision-medicine-initiative/topmed) has provided $10M in co-funding over two years to the NHGRI Centers for Common Disease Genomics (CCDG) program for projects that address the interests and goals of both programs. These funds will be used for a whole genome sequencing project aimed at finding variants influencing risk for early-onset atrial fibrillation (EOAF). This phenotype was considered, but not initially prioritized, in the initial CCDG CVD working group plan.

Disease phenotype: Atrial fibrillation

Atrial fibrillation will affect between 6-12 million individuals in the US by 2050. AF also is associated with increased risks of stroke, dementia, heart failure, death, and high health care costs. Many risk factors for AF have been identified, including advancing age, cardiovascular disease (**CVD**), and CVD risk factors. However, there is little knowledge how to prevent AF. Furthermore, therapies for AF are only partially effective, and are themselves associated with substantial morbidity.

Previously, heritable forms of AF have been considered rare; yet in the last decade, it has been established that AF, and in particular early-onset forms of AF, are heritable. Genome-wide association studies (**GWAS**) provide a powerful tool to identify common variants underlying disease risk. The AFGen Consortium currently consists of investigators from more than 25 studies with >20,000 individuals with AF and >100,000 without AF. In the latest analyses, 14 loci have been identified for AF[1]. Broadly, the loci implicate genes related to cardiopulmonary development, cardiac-expressed ion channels, and cell signaling molecules.

Early-onset atrial fibrillation
We have chosen to focus on individuals with EOAF since such individuals A) have higher disease heritability, B) demonstrate similar AF susceptibility loci as compared to all-comers with AF, yet

[1] Christopherson, I.E. and Ellinor, P.T. Genetics of atrial fibrillation: from families to genomes. Journal of Human Genetics (2016) 61, 61–70 (2016)

exhibit significantly larger effect sizes for common variants, and C) are likely enriched for AF risk variants. We anticipate that these factors will maximize the power for discovery of clinically relevant causal genetic variants for AF.

Collaboration between NHGRI and NHLBI

This collaboration between NHGRI and NHLBI is meant to address the goals of both the TOPMed and CCDG programs. WGS data from these samples will add substantially to NHLBI's ongoing efforts to provide genomic data to support studeis of high interest to its research community. This AFib plan will add WGS data from approximately 6,100 individuals to the 2,900 cases already sequenced through TOPMed. From the CCDG point of view, this project represents a substantial WGS data set from another example disease phenotype that has potential to inform generally how best to discover variants influencing common disease phenotypes.

The collaboration is an opportunity to explore synergies between the two programs. For example, it is highly likely that data from other samples sequenced in the course of other CCDG projects can serve as additional control samples for the AFib study, and similarly other TOPMed WGS data may also serve as common controls for this and other CCDG studies; developing common controls is one of the goals of the CCDG program. Moreover, the CCDG program is coordinating with the TOPMed Infromatics Research Center (IRC) to harmonize data processing between CCDG and TOPMed to facilitate the ability to compare and aggregate data between studies. It is anticipated that all EOAF WGS generated under this plan will be transferred to the TOPMed IRC for variant calling.

Both CCDG investigators, and also the associated NHGRI Genome Sequencing Program Analysis Centers (GSP AC), are expected to include these data in analyses relevant to the wider NHGRI CCDG program goals (see RFA-HG-15-001). We expect that opportunities will arise for direct collaboration with TOPMed investigators.

Samples

In parallel work over the last year, ~20K AF cases were genotyped at the Broad, and ~7,000 of these individuals have EOAF and have not been whole genome sequenced. For all AF cases from 21 participating studies, phenotypic data has been collected centrally and undergone quality control checks, and all samples have been imported, and genotyped on a GWAS array. A DUA has been requested from all of the study PIs for submission of WGS data to dbGaP.

Summary of proposed sequencing

Year 1

Baylor	450	European ancestry EOAF cases from Vanderbilt BioVU

Broad	1540	European ancestry EOAF cases from MGH, Vanderbilt Ablation Registry, UMass, Australia, Penn, Duke, Hopkins, and Texas Heart
MGI	1000	African American samples from Vanderbilt BioVU, Penn & Duke, including 300 AF cases and 700 controls.

Year 2

Baylor	1400	European ancestry cases from Vanderbilt BioVU
Broad	250	Completion of European ancestry EOAF cases MGH, Vanderbilt Ablation Registry, UMass, Australia, Penn, Duke, Hopkins, and Texas Heart
MGI	1500	African American samples from Vanderbilt BioVU, Penn, Duke & Mt. Sinai, including 1367 cases and 133 controls.

Totals	By center:	Baylor 1850 WGS at 30x, Broad 1850 WGS at 30x, MGI 2500 WGS at 30x
	By ancestry:	African American 2500 (1667 cases, 833 controls), European 3700 cases
	By AF type:	European samples will all be EOAF; African American will be a mix of ~800 EOAF and ~900 with AF and other co-morbidities
	Overall:	6140 genomes; 5307 AF cases and 833 controls