Identify low frequency and rare variants associated with blood pressure on chromosome 16 through linkage and association analysis

PDF version of proposal

<table>
<thead>
<tr>
<th>Manuscript id:</th>
<th>4848</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submitted:</td>
<td>Fri, 08/03/2018 - 8:15am</td>
</tr>
</tbody>
</table>

Please specify the original proposal in order to avoid duplicates:

Identify low frequency and rare variants through linkage and association analysis

Provide a short description of changes:
The original proposal includes linkage regions identified from 2 separate studies (CFS and FBPP). We plan to produce 2 separate manuscripts, one for linkage regions identified from each study. Thus, we are revising this proposal and adding a new proposal in order to have a unique paper proposal for each manuscript.

Proposed by: Zhu, Xiaofeng

Proposer's email address: xxz10@case.edu

Please describe the aims of the proposal:
The aim of the proposed research is to identify novel low frequency and rare variants associated with blood pressure traits using whole genome sequencing and family data. For this proposal, only variants within the linkage region on chromosome 16p13 will be examined.

Were all of these aims included in the original study's grant application?: No

Please list any new aims: N/A

Proposal co-authors (Last, First):
Arnett, Donna
Bielak, Lawrence F
Boerwinkle, Eric
Chakravati, Aravinda
Correa, Adolfo
Francheschini, Nora
Guo, Xiuqing
He, Jiang  
He, Karen  
Liang, Jingjing  
Kardia, Sharon  
Kelly, Tanika  
Kooperberg, Charles  
Levy, Dan  
Mathias, Rasika  
Mitchell, Braxton  
Morrison, Alanna C  
Palmas, Walter  
Ramachandran, Vasan  
Rao, DC  
Redline, Susan  
Rich, Stephen  
Rotter, Jerome  
Smith, Jennifer  
Zhu, Xiaofeng

**List phenotypes/traits:**  
SBP, DBP, PP, hypertension status

**Will your proposed analyses include all genomes sequenced through the overall TOPMed program, or a subset?:**  
Subset of genomes

**If subset, please specify:**  
We will include a subset of TOPMed WGS data for the identified linkage region on chromosome 16.

**Brief description of project analysis plans or lookup requested (½-1 ½ page):**  
High blood pressure (BP) is a common condition associated with heart, brain, and kidney diseases. The BP heritability estimated from family data ranges from 30 to 60%. Recent genome-wide association studies (GWAS) have identified over 200 common BP variants that together explain approximately 6% of BP variation. Thus, the heritability of BP remains missing and most of the genetic variants remain to be discovered. It is expected that low frequency and rare variants may be able to explain a portion of the “missing heritability”. Due to lower sensitivity to genetic heterogeneity, linkage analysis of family data can be used to identify low frequency and rare variants. Thus, the aim of the proposal is to identify novel low-frequency and rare variants associated with BP traits using family and whole genome sequencing data. In our previous study, we identified a linkage peak on 16p13 for SBP (MLOD = 2.81) in the Cleveland Family Study European American (CFS-EA) exome array data. Common variants, if associated with BP traits in this region, should have already been discovered by current large GWAS. But there are few GWAS hits within this region. Thus, we hypothesize that low frequency and rare variants in this region contribute to the observed linkage evidence. We will start with TOPMed CFS-EA as the discovery cohort and follow up with additional phase I and I studies. Our analysis plan is as follows: Step 1) We will use the same phenotype harmonization procedure as Dr. Tanika Kelly to harmonize phenotypes...
across participating studies. Among those taking antihypertensive medication, systolic and diastolic blood pressure measures will be imputed by adding 15 mmHg and 10 mmHg, respectively, to observed values. Step 2) Our analysis will be limited to variants segregating within informative families. We will estimate the family-specific LOD (fsLOD) scores for each family and only select variants segregating at least twice in a family with fsLOD >= 0.1. For the observed linkage region, we will test whether a variant is able to explain the observed linkage evidence in the families in which linkage analysis was performed. This will be done by modelling individual variant association and linkage analysis simultaneously. Any substantial reduction of linkage evidence after conditioning on a variant indicates that the variant is a good candidate BP variant. For rare variants, we would not expect substantial reduction in linkage evidence by conditioning on rare variants. We will aggregate rare variants and conduct gene-based association tests. Step 3) For the low frequency and rare variants identified in Step 2, we will divide them into two categories: 1) functional coding variants; 2) non-coding variants. For the non-coding variants, we will predict their functional impacts using functional annotations such as CADD score, fathmm-MKL score, and other information in the WGSA annotation. Only rare non-coding variants that show functional importance will be carried forward in replication analysis. Step 4) The identified coding variants and non-coding variants in Step 3 will be carried forward for replication analysis using the harmonized BP data available across phase I and II TOPMed cohorts. The replication analyses will include both single variant association analysis and gene-based association analysis with kinship adjustments. We will also perform bioinformatics pathway analyses, especially for the pathways involved with the genes harboring variants identified in Step 3.

<p>| How may this analysis overlap other TOPMed projects: | We expect that there may be some overlap with the proposal by Dr. Kelly. However, our proposal focuses on the linkage region on chromosome 16p13 and searches for variants contributing to linkage evidence. This analysis is highly driven by the families that contribute to the linkage evidence. Because of the genetic heterogeneity, especially for low frequency and rare variants, these variants may be missed by Dr. Kelly's approach due to heterogeneity and high statistical penalty. Therefore, this proposal complements Dr. Kelly's proposal. |
|------------------------------------------------------|
| What are your current plans for publication and authorship: | We plan to submit this manuscript for publication in peer-reviewed journals. Authors will include investigators from all studies contributing data for this analysis, as well as other individuals from the blood pressure working group interested in contributing to the design, analysis, interpretation, and/or writing of the manuscript. We plan to actively encourage opportunities for junior investigators to have leadership roles in manuscript development. We will publish under the banner of TOPMed BP Working Group. |
| Will data be shared across institutions? Explain.: | Yes, each study will upload study-specific data to the dbGaP exchange area. For linkage analysis and association analysis, a subset of sequencing data will be needed. Data will be downloaded from the dbGaP exchange area. |
| Will individual-level data be distributed to investigators outside TOPMed? Explain.: | No |</p>
<table>
<thead>
<tr>
<th>Primary Working Group:</th>
<th>Blood Pressure</th>
</tr>
</thead>
</table>
| **Study or Studies associated with this proposal:** | Amish  
CFS  
FHS  
GeneSTAR  
GenSalt  
HyperGEN_GENOA  
JHS  
MESA  
VTE  
WHI |
| **PI Contact:** | I have contacted the PIs of the studies selected above after obtaining Working Group approval (or waiver) and indicated such to those PI. Each PI has either (a) agreed to a collaboration involving use of their study's data in the proposed paper or (b) did not respond to the request within two weeks. Please note that the proposal should be discussed and approved by a Working Group before contacting the study PI for their permission to use their study. I will also select specific consent groups from each study, using the 'Request data sets' form, as they become available. |
| **Additional comments:** | We are splitting our original paper proposal (#1880) into 2 separate proposals as these linkage regions were identified using 2 separate datasets. This revised proposal only focuses on the chr. 16 linkage region identified in CFS. |
| **Scientific review status:** | Approved |
| **Data Request Status:** | Approved |
| **Overall Status:** | Approved |
| **Approval date:** | 08/17/18 |
| **Is this proposal based on a single study/project for which you are an investigator?:** | No |