UNIVERSITY OF MICHIGAN

Alan Kwong <amkwong@umich.edu>

TOPMed manuscript M6573 : Method paper for Robust HWE

Peloso, Gina Marie <gpeloso@bu.edu> Thu, May 7, 2020 at 12:18 PM
To: Hyun Min Kang <hmkang@umich.edu>, Alan Kwong <amkwong@umich.edu>
Cc: "Cupples, L Adrienne" <adrienne@bu.edu>

Hi Hyun,

I forwarded the manuscript to the FHS P&P. I will pass along the response.

Best,

Gina

From: Hyun Min Kang [mailto:hmkang@umich.edu]
Sent: Thursday, May 07, 2020 12:07 PM
To: Peloso, Gina Marie <gpeloso@bu.edu>; Alan Kwong <amkwong@umich.edu>
Cc: Cupples, L Adrienne <adrienne@bu.edu>
Subject: Re: TOPMed manuscript M6573 : Method paper for Robust HWE

Dear Gina,

Thanks very much for your thoughtful and important comments. We addressed most of the comments in the revised paper. Let me know if you have any additional comments.

Regarding the number of PCs, this was done this way because genotype-based PCs (from sequence-based genotypes) are not available during variant calling, so we used PCs estimated from verifyBamID2 software tool. We have clarified this in the discussion and the method.

Alan is working on the manuscript sent to P&P review. It would be great if you could help with this for FHS P&P.

Thanks,

Hyun.

Hyun Min Kang, Ph.D.
Associate Professor of Biostatistics
University of Michigan, Ann Arbor
Email : hmkang@umich.edu

On Thu, Apr 16, 2020 at 2:48 PM Peloso, Gina Marie <gpeloso@bu.edu> wrote:

> Hi Hyun,
>
> Nicely put together manuscript. A couple of comments to consider:
>
> 1. Why did you use 4 or 2 PCs when TOPMed recommends using 11 to capture the ancestry across the samples? What happens when you use more than 4 PCs?
>
> 2. Somewhere you should make clear the freeze of the TOPMed data that you are using.
>
> 3. I am not able to see the supplemental figures in the file sent. Main figures came through ok.
>
> 4. It might be good to include supplemental tables with the 1000G populations with numbers within each ancestry group that were used and a similar table for TOPMed listing out the freeze 5 cohorts used with sample sizes, ancestries, and the cohort acknowledgements can be included there. I didn't see this included.
>
> 5. Below is the FHS study description and acknowledge. If this needs to go to the FHS P&P, it will be easier if #4 is taken care of before we send off to them.
>
> **Adrienne** – does this manuscript need to go to the FHS P&P?
>
> Best,
>
> Gina
>
> Framingham Heart Study (FHS)
>
> *TOPMed dbGaP accession#: phs000974, Parent dbGaP accession#: phs000007*
>
> The Framingham Heart Study (FHS) is a prospective cohort study of 3 generations of subjects who have been followed up to 65 years to evaluate risk factors for cardiovascular disease.[13-16] Its large sample of ~15,000 men and women who have been extensively phenotyped with repeated examinations make it ideal for the study of genetic associations with cardiovascular disease risk factors and outcomes. DNA samples have been collected and immortalized since the mid-1990s and are available on ~8000 study participants in 1037 families. These samples have been used for collection of GWAS array data and exome chip data in nearly all with DNA samples, and for targeted sequencing, deep exome sequencing and light coverage whole genome sequencing in limited numbers. Additionally, mRNA and miRNA expression data, DNA methylation data, metabolomics and other 'omics data are available on a sizable portion of study participants. This project will focus on deep whole genome sequencing (mean 30X coverage) in ~4100 subjects and imputed to all with GWAS array data to more fully understand the genetic contributions to cardiovascular, lung, blood and sleep disorders.
>
> FHS acknowledges the support of contracts NO1-HC-25195 and HHSN268201500001I from the National Heart, Lung and Blood Institute and grant supplement R01 HL092577-06S1 for this research. WGS for "NHLBI TOPMed: Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study" (phs000974) was performed at the Broad Institute of MIT and Harvard (HHSN268201500014C, 3R01HL092577-06S1, and

3U54HG003067-12S2). We also acknowledge the dedication of the FHS study participants without whom this research would not be possible.

From: Hyun Min Kang [mailto:hmkang@umich.edu]
Sent: Tuesday, April 14, 2020 1:35 PM
To: Peloso, Gina Marie <gpeloso@bu.edu>; Jerome Rotter <jrotter@lundquist.org>; Cupples, L Adrienne <adrienne@bu.edu>
Cc: Alan Kwong <amkwong@umich.edu>
Subject: Fwd: TOPMed manuscript M6573 : Method paper for Robust HWE

Dear Drs. Peloso, Rotter, and Cupples,

I'm forwarding the attachments in a different way to avoid the size limit. I converted the figures into PDF and I hope that it works this time. Please see below for details, and let me know if you have any questions.

Thanks,

Hyun.

Hyun Min Kang, Ph.D.
Associate Professor of Biostatistics
University of Michigan, Ann Arbor
Email : hmkang@umich.edu

---------- Forwarded message ---------
From: **Hyun Min Kang** <hmkang@umich.edu>
Date: Tue, Apr 14, 2020 at 1:29 PM
Subject: TOPMed manuscript M6573 : Method paper for Robust HWE
To: Tom Blackwell <tblackw@umich.edu>, Jonathon LeFaive <lefaivej@umich.edu>, Laura Scott <ljst@umich.edu>, Albert Vernon Smith <albertvs@umich.edu>, Goncalo Abecasis <goncalo@umich.edu>, Michael Boehnke <boehnke@umich.edu>, May Montasser <mmontass@som.umaryland.edu>, Eric Boerwinkle <Eric.Boerwinkle@uth.tmc.edu>, Han Chen <Han.Chen.2@uth.tmc.edu>, Kathleen C. Barnes <Kathleen.Barnes@ucdenver.edu>, Nicholas L. Smith <nlsmith@u.washington.edu>, Dandi Qiao <redaq@channing.harvard.edu>, Wonji Kim <rewki@channing.harvard.edu>, Scott T. Weiss <scott.weiss@channing.harvard.edu>, Jessica Lasky Su <rejas@channing.harvard.edu>, Nicholette D. Palmer <nallred@wakehealth.edu>, L. Adrienne Cupples <adrienne@bu.edu>, Gina M. Peloso <gpeloso@bu.edu>, Angel C. Y. Mak <ChiYan.Mak@ucsf.edu>, Estaban G. Burchard <esteban.burchard@ucsf.edu>, Rasika Mathias <rmathias@jhmi.edu>, Jennifer Smith <smjenn@umich.edu>, Hemant Tiwari <htiwari@uab.edu>, Marguerite Ryan Irvin <irvinr@uab.edu>, Mariza de Andrade <mandrade@mayo.edu>, John Blangero <john.blangero@utrgv.edu>, Courtney Montgomery <Courtney-Montgomery@omrf.org>, Daniel Chasman <DCHASMAN@research.bwh.harvard.edu>, Alexander Reiner <apreiner@uw.edu>, Charles Kooperberg <clk@fredhutch.org>, John Barnard <barnarj@ccf.org>, Mina Chung <chungm@ccf.org>, Brian Cade <bcade@bwh.harvard.edu>, Susan Redline <sredline1@rics.bwh.harvard.edu>, Tanika N. Kelly <tkelly@tulane.edu>, Adolfo Correa <acorrea@umc.edu>, James Wilson <jwilson8@bidmc.harvard.edu>, Jerome Rotter <jrotter@lundquist.org>, Stephen Rich <ssr4n@virginia.edu>, Patrick Ellinor <ellinor@mgh.harvard.edu>, Steven Lubitz <slubitz@mgh.harvard.edu>, Stephen McGarvey <stephen_mcgarvey@brown.edu>, Daniel Weeks <weeks@pitt.edu>, Dan Roden <dan.roden@vumc.org>
Cc: Alan Kwong <amkwong@umich.edu>

Dear Co-authors,

Attached please find a draft of the TOPMed manuscript M6573, entitled 'Robust, flexible, and scalable tests for Hardy-Weinberg Equilibrium across diverse ancestries'. The manuscript is separated into three pieces, (1) main text with references, (2) figures and tables, and (3) supplementary material for your convenience.

We will appreciate it very much if you could review these documents at your convenience and give any comments or suggested revisions (preferably with track changes). In addition, please check the spelling of your name, your institutional affiliation(s). If you want to add study-specific acknowledgements, please let us know. We would like to submit this to the TOPMed DCC in two weeks, April 28th, 2020.

We are considering submission of this manuscript to *Genetics* upon TOPMed approval, but please feel free to give us any suggestions regarding recommended journals.

We included co-authors whom we received responses from our earlier emails in alphabetical order in the middle before the TOPMed banner. We included the study names for those studies we need nomination from co-authors. For the studies whose co-authors are not yet nominated, please let us know the names as soon as possible. Alan Kwong (cc'ed) will follow up with PIs of individual studies regarding this.

We wanted to include as many authors as we can, but given the nature that this is a method paper, we were concerned that the number of authors that the journal expects is somewhat limited. As a result, we set up the following policy in the number of co-authors representing the contributing study in the following ways:

(1) We grouped each study by PIs, and assigned 1-2 co-authors for each group (unless the studies need to nominate different co-authors from each study).

(2) If the study contains fewer than 2,000 contributing samples in total, we requested to nominate one co-author. If more than 2,000, we requested nomination of two co-authors. We understand this strategy is not ideal for everyone, but we had to make a choice as we still expect to have 40 or more authors in this manuscript.

We look forward to receiving feedback and submitting this paper soon. If you have any other questions or comments, please feel free to let me and Alan know.

Sincerely,

Hyun.

Hyun Min Kang, Ph.D.
Associate Professor of Biostatistics
University of Michigan, Ann Arbor
Email : hmkang@umich.edu