A framework for detecting noncoding rare variant associations of large-scale whole-genome sequencing studies

Wednesday,2021-08-18 19:01 America/New_York — joshbis

Current Status:
2021-08-23: Revised Penultimate Draft MS received – undergoing expedited review
2021-09-02: Penultimate Draft approved by Steering Committee How to Proceed

BASICS

Penultimate Draft: Cover Note	
Submission not preceded by an approved paper proposal	○ It is a new meta-analysis that derives from a single "project" or "omnibus" proposal as part of a larger consortium of many cohorts in which CHS has agreed to participate and for which no new data are needed Parent Proposal: Omnibus: The NHLBI Trans-Omics for Precision Medicine (TOPMed) Program
First Author	Li, Zilin
CHS Sponsor Name	Psaty, Bruce (University of Washington)
Coauthors	Joshua Bis, Jennifer Brody, Bruce Psaty
Additional authors	Zilin Li, Xihao Li, Hufeng Zhou, Sheila M. Gaynor, Margaret S. Selvaraj,, Theodore Arapoglou, Corbin Quick, Yaowu Liu, Han Chen, Ryan Sun, Rounak Dey, Donna K. Arnett, Lawrence F. Bielak, Joshua C. Bis, Thomas W. Blackwell, John Blangero, Eric Boerwinkle, Donald W. Bowden, Jennifer A. Brody, Brian E. Cade, Matthew P. Conomos, Adolfo Correa, L. Adrienne Cupples, Joanne E. Curran, Paul S. de Vries, Ravindranath Duggirala, Barry I. Freedman, Harald H. H. Göring, Xiuqing Guo, Rita R. Kalyani, Charles Kooperberg, Brian G. Kral, Leslie A. Lange, Ani Manichaikul, Lisa W. Martin, Stephen T. McGarvey, Braxton D.

	Mitchell, May E. Montasser, Alanna C. Morrison, Take Naseri, Jeffrey R. O'Connell, Nicholette D. Palmer, Patricia A. Peyser, Bruce M. Psaty, Laura M. Raffield, Susan Redline,, Alexander P. Reiner, Muagututi`a Sefuiva Reupena, Kenneth M. Rice, Stephen S. Rich, Jennifer A. Smith, Kent D. Taylor, Ramachandran S. Vasan, James G. Wilson, Lisa R. Yanek, Wei Zhao, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Lipids Working Group, The Samoan Obesity, Lifestyle and Genetic Adaptations Study (OLaGA) Group, Jerome I. Rotter, Christen J. Willer,, Pradeep Natarajan,, Gina M. Peloso, and Xihong Lin
Location of analysis	local (not at CC)
Analysts	Zilin Li
CHS Working Group	
Consortium	○ TOPMed
Type of Study	Main |
Data/Analysis type	○ Cross-sectional
Manuscript Keyword(s)	TOPMed | Analysis | STAAR |

1. Assurances

This draft has been reviewed and approved by all coauthors	Yes
Is CHS correctly represented?	Yes
When data are used from ancillary studies, are they included and credited?	Yes
Do you or any member of your Writing Group intend to patent any process, aspect or	No

outcome of these analyses?	
Are these analyses to involve a for-profit Corporation?	No
If genetic data are used, does the paper include the required genetic exclusions statement?	Yes
Is CMS data used?	No
If CMS data were used, does the paper include cell sizes greater than 10?	No CMS data were used in these analyses
I included the current (rev. 2020) CHS grant & contract numbers in the funding acknowledgement	Yes
As part of the CHS data use agreement, I will update this record when my manuscript has been accepted by a journal	Yes
In accordance with the NIH public access policy, I will submit any resulting publication to PubMed Central & update this record with its PMCID within 6 month	Yes

Abstract

Large-scale whole-genome sequencing studies have enabled the analysis of noncoding

rare variants (RVs) associated with complex human traits. Variant set analysis is a powerful approach to study RV association. A key component of this approach is the construction of RV sets that can be used for analysis. However, existing methods have limited scope to define analysis units in the noncoding genome. Here we propose a computationally efficient noncoding RV association detection framework that uses STAAR to provide various strategies for grouping noncoding variants in gene-centric analysis based on functional categories. We also propose SCANG-STAAR in non-gene-centric analysis, which uses dynamic window sizes and incorporates multiple functional annotations. We develop STAARpipeline that performs flexible noncoding RV association analysis, including gene-centric analysis, fixed-window-based and dynamic-window-based non-gene-centric analysis. We apply STAARpipeline to identify noncoding RVs associated with four quantitative lipid traits in 21,015 discovery and 9,123 replication samples from the Trans-Omics for Precision Medicine program. We discover and replicate several noncoding RV associations with lipids. Large-scale whole-genome sequencing studies have enabled the analysis of noncoding rare variants (RVs) associated with complex human traits. Variant set analysis is a powerful approach to study RV association. A key component of this approach is the construction of RV sets that can be used for analysis. However, existing methods have limited scope to define analysis units in the noncoding genome. Here we propose a computationally efficient noncoding RV association detection framework that uses STAAR to provides various strategies for grouping noncoding variants in gene-centric analysis based on functional categories using STAAR. We also propose SCANG-STAAR in non-gene-centric analysis, which uses dynamic window sizes and incorporates multiple functional annotations. We develop STAARpipeline that performs flexible noncoding RV association analysis, including gene-centric analysis, fixed-window-based and dynamic-window-based non-gene-centric analysis. We apply STAARpipeline to identify noncoding RVs associated with four quantitative lipid traits in 21,015 discovery and 9,123 replication samples from the Trans-Omics for Precision Medicine program. We discover and replicate several noncoding RV associations with lipids.

Large-scale whole-genome sequencing studies have enabled the analysis of noncoding rare variants (RVs) associated with complex human traits. Variant set analysis is a powerful approach to study RV association. A key component of this approach is the construction of RV sets that can be used for analysis. However, existing methods have limited scope to define analysis units in the noncoding genome. Here we propose a computationally efficient noncoding RV association detection framework that uses STAAR to provides various strategies for grouping noncoding variants in gene-centric analysis based on functional categories using STAAR. We also propose SCANG-STAAR in non-gene-centric analysis, which uses dynamic window sizes and incorporates multiple functional annotations. We develop STAARpipeline that performs flexible noncoding RV association analysis, including gene-centric analysis, fixed-window-based and dynamic-window-based non-gene-centric analysis. We apply STAARpipeline to identify noncoding RVs associated with four quantitative lipid traits in 21,015 discovery and 9,123

replication samples from the Trans-Omics for Precision Medicine program. We discover and replicate several noncoding RV associations with lipids.

Large-scale whole-genome sequencing studies have enabled the analysis of noncoding rare variants (RVs) associated with complex human traits. Variant set analysis is a powerful approach to study RV association. A key component of this approach is the construction of RV sets that can be used for analysis. However, existing methods have limited scope to define analysis units in the noncoding genome. Here we propose a computationally efficient noncoding RV association detection framework that uses STAAR to provides various strategies for grouping noncoding variants in gene-centric analysis based on functional categories using STAAR. We also propose SCANG-STAAR in non-gene-centric analysis, which uses dynamic window sizes and incorporates multiple functional annotations. We develop STAARpipeline that performs flexible noncoding RV association analysis, including gene-centric analysis, fixed-window-based and dynamic-window-based non-gene-centric analysis. We apply STAARpipeline to identify noncoding RVs associated with four quantitative lipid traits in 21,015 discovery and 9,123 replication samples from the Trans-Omics for Precision Medicine program. We discover and replicate several noncoding RV associations with lipids.

Large-scale whole-genome sequencing studies have enabled the analysis of noncoding rare variants (RVs) associated with complex human traits. Variant set analysis is a powerful approach to study RV association. A key component of this approach is the construction of RV sets that can be used for analysis. However, existing methods have limited scope to define analysis units in the noncoding genome. Here we propose a computationally efficient noncoding RV association detection framework that uses STAAR to provides various strategies for grouping noncoding variants in gene-centric analysis based on functional categories using STAAR. We also propose SCANG-STAAR in non-gene-centric analysis, which uses dynamic window sizes and incorporates multiple functional annotations. We develop STAARpipeline that performs flexible noncoding RV association analysis, including gene-centric analysis, fixed-window-based and dynamic-window-based non-gene-centric analysis. We apply STAARpipeline to identify noncoding RVs associated with four quantitative lipid traits in 21,015 discovery and 9,123 replication samples from the Trans-Omics for Precision Medicine program. We discover and replicate several noncoding RV associations with lipids.

2. Review Method

Paper includes NHLBI author	No

Review method	Expedited – this is a meta-analysis (includes pooled) (1-2 weeks)

3. File

Uploaded Supplemental File(s)	STAARpipeline_v4.docx Description: Manuscript STAARpipeline_Supplementary_Note_v4.docx Description: Suplementary Notes STAARpipeline_Supplementary_Tables_v4.xlsx Description: Supplementary Tables STAARpipeline_Supplementary_Figures_v4.docx Description: Supplementary Figures

PUBLICATION STATUS

Publication	

ABSTRACT & PRESENTATIONS

PP manuscript proposal/penultimate manuscript vote

SC manuscript proposal/penultimate manuscript vote

Add new comment