Submitter Name: Seon-Kyeong Jang
Submitted email: jangx303@umn.edu
PI Name (if different): Scott Vrieze
PI email (if different): vrieze@umn.edu

# Rare variant heritability of tobacco use: Evidence from deep whole-genome sequencing of up to 26,000 individuals

Seon-Kyeong Jang[1], Luke Evans[2], Allison Fialkowski[1], Donna K. Arnett[3], Diane M. Becker[4], Joshua C[5]. Bis, Jennifer A Brody[5], L. Adrienne Cupples[6], Sean P. David[7,8], Mariza de Andrade[9], Tasha E. Fingerlin[10,11], Sina A. Gharib[5,12], David C Glahn[13], Jeffrey Haessler[14], Susan R. Heckbert[15,16], John E. Hokanson[17], Shih-Jen Hwang[18], Anne E. Justice[19], Wonji Kim[20], Charles Kooperberg[14], Dan Levy[18], Ruth J.F. Loos[21,22], Ani W. Manichaikul[23], Mark T. Gladwin[24], Lisa Warsinger Martin[25], Mehdi Nouraie[24], Olle Melander[26,27], Kari E. North[28], Elizabeth C. Oelsner[29], Anna L. Peljto[30], Michael Preuss[21,22], Bruce M Psaty[5], Dandi Qiao[20], Robert M. Reed[31], Alexander P. Reiner[14], Stephen S. Rich[23], Jerome I. Rotter[33], David A. Schwartz[30,34], Aladdin H. Shadyab[35], Edwin K. Silverman[20], Nicholas L. Smith[15,16], J. Gustav Smith[36,37], Kent D. Taylor[38], Ramachandran S. Vasan[39,40], Victor R. Gordeuk[41], Zhe Wang[21,22], Kerri L. Wiggins[5], Lisa R. Yanek[4], Ivana V. Yang[30], Kendra A. Young[17], Kristin L. Young[28], Yingze Zhang[24], John Blangero[43], Dajiang J. Liu[42], Matthew Keller[2], Scott Vrieze[1]

[1] Department of Psychology, University of Minnesota, Minneapolis, MN, USA
[2] Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA
[3] Dean's Office, University of Kentucky College of Public Health, USA
[4] Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[5] Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA
[6] Department of Biostatistics, Boston University School of Public Health, USA
[7] Department of Family Medicine, Prtizker School of Medicine, University of Chicago, Chicago, IL, USA
[8] NorthShore University HealthSystem, Evanston, IL, USA
[9] Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.
[10] Colorado School of Public Health, University of Colorado Denver - Anschutz Medical Campus, Aurora, CO, USA
[11] Center for Genes Environment and Health, National Jewish Health, Denver, CO, USA.
[12] Center for Lung Biology, Division of Pulmonary, Critical Care and Sleep Medicine, University of Washington, Seattle, WA, USA
[13] Department of Psychaitry, Boston Children's Hosptial and Harvard Medical School, Boston, MA, USA
[14] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 09109, USA
[15] Department of Epidemiology, University of Washington, Seattle WA 98195, USA
[16] Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle WA 98101, USA
[17] Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
[18] Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA
[19] Department of Population Health Sciences, Geisinger Health System, Danville, PA, USA

[20] Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[21] Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[22] The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[23] Center for Public Health Genomics, School of Medicine, University of Virginia, Charlottesville, VA, USA
[24] Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
[25] School of Medicine and Health Sciences, USA
[26] Department of Clinical Sciences, Lund University, Malmö, Sweden
[27] Department of Internal Medicine, Skåne University Hospital, Malmö, Sweden
[28] Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[29] Division of General Medicine, Columbia University Irving Medical Center, Columbia University, New York, NY, USA
[30] Department of Medicine, University of Colorado School of Medicine, Aurora, CO, USA
[31] University of Maryland School of Medicine, USA
[32] Center for Public Health Genomics, School of Medicine, University of Virginia, Charlottesville, VA, USA
[33] The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA
[34] Department of Immunology, School of Medicine, University of Colorado Denver, Aurora, CO, USA
[35] Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA, USA
[36] The Wallenberg Laboratory/Department of Molecular and Clinical Medicine, Institute of Medicine, Gothenburg University, Sweden
[37] Department of Cardiology, Sahlgrenska University Hospital, Gothenburg, Sweden
[38] The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA USA
[39] Sections of Preventive medicine and Epidemiology and cardiovascular medicine, Department of medicine, Boston University School of Medicine, Boston, MA, USA
[40] Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA
[41] Department of Medicine, University of Illinois at Chicago, Chicago, IL, USA
[42] Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA, USA
[43] Department of Human Genetics and South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville TX, USA

While genome-wide association studies have successfully mapped hundreds of common loci to tobacco use, they collectively explain only 5-10% of the variance of tobacco use, compared to ~50% of heritability estimated from close relatives. Here, we examined the contribution of rare variants to heritability of tobacco use in up to 26,000 individuals of European ancestry in the Trans-Omics for Precision Medicine (TOPMed) program with whole genome sequence (WGS data (~30X coverage). Using GREML-LDMS-I, we found that rare variants with MAF 0.1% to 0.01% account for 26% of variation in age of smoking initiation (AgeSmk) and 15% for smoking cessation (SmkCes). Potential confounding was detected due to population structure at rare variants even after adjusting for principal components calculated from common variants,

accounting for about one-third of the estimated rare variant heritability. After further adjustment for this population stratification, we estimated the WGS-based heritability of 0.21 (SE=0.08) for AgeSmk, 0.11 (0.05) for Cigarettes per Day, 0.21 (0.09) for SmkCes, and 0.24 (0.07) for Smoking initiation. These values are 1.4-4.5 times higher than previous SNP-based estimates. We provide a new upper-bound for SNP-based heritability of tobacco use based on estimation in pedigree consisting of close relatives in TOPMed, ranging from 0.18-0.35. The substantial contribution of rare variants over and above common variants for several smoking phenotypes sheds light on the missing heritability and genetic etiology of tobacco use. It also informs fine-mapping strategies since the majority of the rare variant contribution was located in DNA regulatory regions.