

Name: Sandra Sanchez-Roige

Email: sanchezroige@ucsd.edu

Multi-ancestry meta-analysis of tobacco use disorder based on electronic health record data prioritizes novel candidate risk genes and reveals associations with numerous health outcomes

Sylvanus Toikumo^{1,2*}, Mariela V Jennings^{3*}, Benjamin Pham³, Hyunjoon Lee⁴, Travis T Mallard⁴, Sevim B Bianchi³, John J Meredith³, Laura Vilar-Ribo⁵, Heng Xu³, Alexander Hatoum⁶, Emma C Johnson⁷, Vanessa Pazdernik⁸, Zeal Jinwala¹, Brittany S Leger³, Maria Niarchou⁹, Michael Ehinmowo, Penn Medicine BioBank, Million Veteran Program, psycheMERGE Substance Use Disorder, Greg D Jenkins⁸, Anthony Batzler⁸, Richard Pendegraft⁸, Abraham A Palmer³, Hang Zhou¹⁰, Joanna M Biernacka¹¹, Brandon J Coombes⁸, Joel Gelernter¹², Ke Xu¹², Dana Hancock¹³, Nancy J Cox¹⁴, Jordan W Smoller⁴, Lea K Davis¹⁴, Amy C Justice¹⁵⁻¹⁷, Henry R Kranzler^{1,2}, Rachel Kember^{1,2}, Sandra Sanchez-Roige^{3,14}

¹Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; ²Mental Illness Research, Education and Clinical Center, Crescenz VAMC, Philadelphia, PA, USA; ³Department of Psychiatry, University of California San Diego, San Diego, CA, USA; ⁴Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; ⁵Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain; ⁶Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA; ⁷Department of Psychiatry, Washington University School of Medicine, Saint Louis, Missouri, USA; ⁸Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, USA; ⁹Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; ¹⁰Department of Psychological and Brain Sciences, Washington University in Saint Louis, St. Louis, MO, USA; ¹¹Department of Psychiatry & Psychology, Mayo Clinic, Rochester, MN, USA; ¹²Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; ¹³Behavioral and Urban Health Program, Behavioral Health and Criminal Justice Division, RTI International, Research Triangle Park, NC, USA; ¹⁴Department of Medicine, Division of Genetic Medicine, Vanderbilt University, Nashville, TN, USA; ¹⁵Yale University School of Public Health, New Haven, CT, USA; ¹⁶Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA; ¹⁷Yale University School of Medicine, New Haven, CT, USA.

Tobacco use disorder (TUD) is the most prevalent substance use disorder worldwide. Genetic factors influence smoking behaviors and although strides have been made using genome-wide association studies (GWAS) to identify risk variants, only a small amount of the variance has been accounted for. We leveraged five biobanks to perform a multi-ancestral meta-analysis of TUD (derived via electronic health records, EHR) in 740,937 individuals. We identified 72 independent risk loci; integration with functional genomic tools uncovered 330 potential risk genes, primarily expressed in the brain. TUD was genetically correlated with smoking traits and comorbid psychiatric traits from traditionally ascertained cohorts, externalizing behaviors in children, and hundreds of medical outcomes, including HIV infection, heart disease, and pain. This work furthers our biological understanding of TUD and establishes the use of EHR as sources of phenotypic information for studying the genetics of TUD.