Significance (1.5 pages)

Importance of the Problem to be Addressed

While there are many statistical methods that calculate polygenic risk scores from GWAS summary statistics, current approaches have limited predictive ability. For example, among psychiatric conditions, polygenic risk scores predict only 2% of the liability variance for major depressive disorder (Wray et al. 2018), 5% for bipolar disorder (Mullins et al. 2021), 3% for neuroticism (Luciano et al. 2018), and 6% for attention deficit hyperactivity disorder (Demontis et al. 2019). The statistical methods that underlie these polygenic risk score calculations all share the assumption that each SNP has only a genetically "additive" effect on the trait. In other words, the methods assume that the trait liability is a weighted sum of minor allele counts at a collection of SNPs, with weights specified by the GWAS estimate of the SNP effect. By making the simplifying assumption that SNP-SNP interactions have no net impact on the trait liability, the investigators restrict the predictive ability of the polygenic risk scores. While the incorporation of SNP-SNP interactions into polygenic risk score calculations doesn't fully resolve the limited predictive ability of polygenic risk scores, including the possibility of SNP-SNP interactions in polygenic risk score calculations will improve predictive ability over methods that neglect SNP-SNP interactions. The reason for this is that the collection of statistical models with possible SNP-SNP interactions contains the collection of statistical models without SNP-SNP interactions.

By improving predictive ability of polygenic risk scores beyond current standards, investigators will more accurately target interventions and preventive measures to those individuals at highest risk of disease. Clinical researchers who use polygenic risk scores to counsel patients at high risk for disease will more accurately identify high-risk patients, while public health researchers will more accurately identify high-risk populations for targeted preventive measures. Together, these efforts will positively impact society by enhancing health, lengthening life, and reducing illness and disability.

Failure to address this issue, by continuing to use current polygenic risk score methods that ignore contributions from SNP-SNP interactions, will result in misclassification of individuals into high-risk and low-risk categories. Misclassification of individuals will attenuate measures of intervention effectiveness, since many of the subjects classified as "high-risk", in fact, will truly be low-risk. Thus, interventions that reduce disease burden and extend healthy lifespan in populations will not be widely implemented and potential gains in lifespan and well-being will not be achieved.

Rigor of the Prior Research Supporting the Aims

Aim 1 (Literature)

Aim 2 (Literature)

Significance of the Expected Research Contribution

Upon successful completion of the proposed research, we expect our contribution to be a computationally efficient and scalable new statistical method for calculating accurate polygenic risk scores. *This contribution is expected to be significant because it will enable accurate identification of individuals at high disease risk and appropriate targeting of preventive interventions.* We will freely share our open source, well documented, and thoroughly tested software implementation of our statistical methods to benefit the research community and those who want to build upon our findings. Society will benefit from our research through the clinical and public health preventive

measures that our work makes possible. Ultimately, our research will promote health, lengthen life, and reduce illness and disability.

Innovation (0.5 pages)

References

Demontis, Ditte, Raymond K Walters, Joanna Martin, Manuel Mattheisen, Thomas D Als, Esben Agerbo, Gísli Baldursson, et al. 2019. "Discovery of the First Genome-Wide Significant Risk Loci for Attention Deficit/Hyperactivity Disorder." *Nature Genetics* 51 (1): 63–75.

Luciano, Michelle, Saskia P Hagenaars, Gail Davies, W David Hill, Toni-Kim Clarke, Masoud Shirali, Sarah E Harris, et al. 2018. "Association Analysis in over 329,000 Individuals Identifies 116 Independent Variants Influencing Neuroticism." *Nature Genetics* 50 (1): 6–11.

Mullins, Niamh, Andreas J Forstner, Kevin S O'Connell, Brandon Coombes, Jonathan RI Coleman, Zhen Qiao, Thomas D Als, et al. 2021. "Genome-Wide Association Study of More Than 40,000 Bipolar Disorder Cases Provides New Insights into the Underlying Biology." *Nature Genetics* 53 (6): 817–29.

Wray, Naomi R, Stephan Ripke, Manuel Mattheisen, Maciej Trzaskowski, Enda M Byrne, Abdel Abdellaoui, Mark J Adams, et al. 2018. "Genome-Wide Association Analyses Identify 44 Risk Variants and Refine the Genetic Architecture of Major Depression." *Nature Genetics* 50 (5): 668–81.