

Predicting Complex Disease Risk through Machine Learning and Optimal Polygenic Risk Scores

Christopher B. Cole^{1,2,3}, Majid Nikpay^{1,2}, Ruth McPherson^{1,2}

¹ *Atherogenomics Laboratory, University of Ottawa Heart Institute;*

² *Ruddy Canadian Cardiovascular Genetics Centre, University of Ottawa Heart Institute;*

³ *Statistical Genetics Group, Centre for Addiction and Mental Health*

Abstract

With the recent explosion in genetic sequencing efficiency, personalized medicine is quickly becoming reality. Using high quality genotypic data and novel statistical methods, we can predict complex disease and estimate various environmental contributions. We propose a novel prediction strategy using Best Linear Unbiased Predictors (BLUPs) and Optimal Polygenic Risk Scores (OPRS) to estimate Coronary Artery Disease heritability, pleiotropy, and overall personal risk with validation and comparison to traditional prediction methods. These models may allow researchers to construct an optimally efficient genotyping chip to predict personal risk in a clinical setting and adjust their treatment accordingly.

Keywords: Polygenic Risk Scores, Biostatistics, Optimization

Email address: `chris.c.1221@gmail.com` ()

URL: `chrisbcole.me` ()