Dissecting genetic pleiotropy between hyperuricemia and chronic kidney disease using local Bayesian regression methods

Alexa S Lupi[1,2], Nick A Sumpter[3], Megan P Leask[4], Gustavo de los Campos[1,2,5], Tony R Merriman[4], Richard J Reynolds[3], Ana I Vazquez[1,2]

1 Department of Epidemiology and Biostatistics, Michigan State University (MSU), East Lansing, Michigan, United States; 2 Institute for Quantitative Health Science and Engineering, Systems Biology, MSU; 3 Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States; 4 Department of Biochemistry, University of Otago, Dunedin, New Zealand; 5 Department of Statistics, MSU

INTRODUCTION

- Chronic kidney disease (marker **eGFR** <60 mL/min/1.73 m2) and hyperuricemia (serum urate (**sU**) >6.5 mg/dL) are closely related comorbidities. However, their shared genetic variants and how those variants co-effect the traits are not well-defined. Elucidating these variants could have important clinical implications.
- Current approaches are limited. Single-marker regressions (comparing GWAS results) only obtain marginal SNP effects, and whole-genome regressions aren't computationally feasible at the scale of biobank data sample sizes.

Goal

- We fit as series of local Bayesian regressions (LBR) within regions in linkage-disequilibrium (LD) to identify pleiotropic regions between eGFR and sU by estimating local genetic covariances with the goal of untangling the genetic relationship between hyperuricemia and CKD.

METHODS

UK Biobank dataset

- n=333,542 distantly related Caucasians ($r = 0.1$)
- 607,490 autosomal genotyped SNPs

Figure 1 describes the methodology of obtaining genetic covariance estimates within LD segments.

1. BGLR

Fit a series of LBR models for each of the phenotypes: Y_1 = SU, Y_2 = eGFR, and Y_3 = SU + eGFR

2. LD segments

511,828 unique LD segments were identified, with the methodology coming from Funkhouser et al. (2020), inspired by Fernando et al. (2017).

3. Variance estimates

Compute posterior variance for each LD block:

$$Var(X\beta) = \frac{\beta' x' x \beta}{n-1}$$

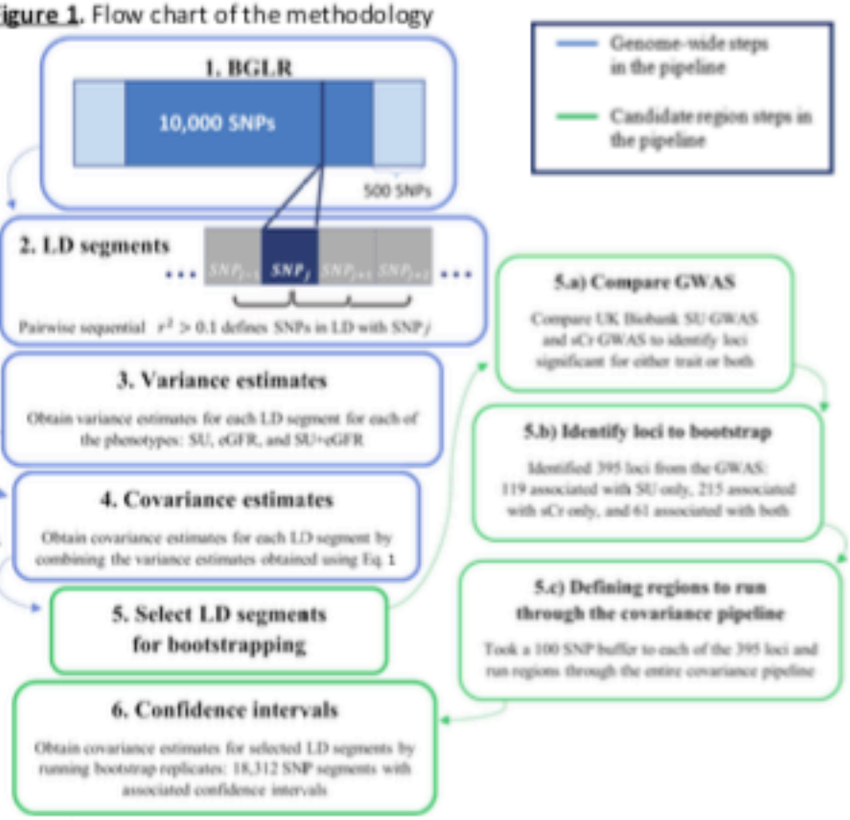
where X is an n x p incidence matrix, β is the est. SNP effects

4. Covariance estimates

Combine variance estimates with Eq. 1:

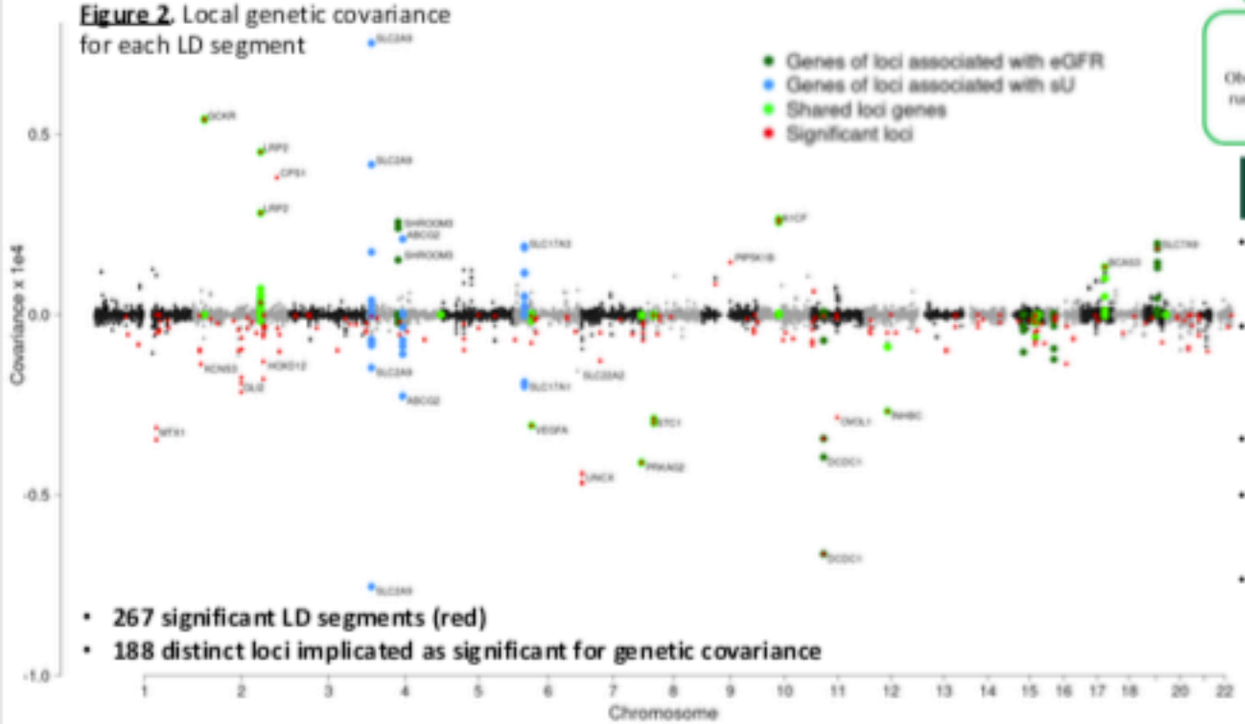
$$Cov(Y_1, Y_2) = \frac{1}{2}[Var(Y_3) - Var(Y_1) - Var(Y_2)]$$

Figure 1. Flow chart of the methodology



RESULTS

Figure 2. Local genetic covariance for each LD segment



- **267 significant LD segments (red)**
- **188 distinct loci implicated as significant for genetic covariance**

DISCUSSION

- Our methodology improves upon existing methodology, as we identified numerous novel loci (including *GLI2, SLC7A9/CEP89, CYP24A1, KCNS3, CHD9, ARL15, PAX8,* and *IGF1R)* in addition to previously hypothesized shared loci.
- We validated all 9 loci identified as shared between the traits by Johnson et al. (2018), including *A1CF*, a lipoprotein synthesis regulator, and *GCKR*, a glucokinase regulator. We also validated 24 of the 35 loci identified by Leask et al. *(accepted)*. Both studies used a comparison of independently run GWAS.
- Our methodology is more robust while still utilizing the well-powered UK Biobank sample size that would be computationally infeasible for a multivariate model.
- We performed a validation of our methodology in the Atherosclerosis Risk in Communities dataset. Due to the small sample size, the results were validated only in the loci with the largest magnitude covariance point estimates.
- Through our novel approach we have provided a robust list of loci significant for local genetic covariance between SU and eGFR. 17 out of 267 (6.4%) of the loci have directionality opposite from the global genetic correlation between eGFR and SU (which is negative). These regions merit detailed investigation as they may involve underlying biological mechanisms with potential clinical implications for CKD and hyperuricemia.

Authors acknowledge support from NIH (P50-AR060772)