**Supplementary documents**

**MR-TRYX: Exploiting horizontal pleiotropy to infer novel causal pathways**

Yoonsu Cho1, Philip C Haycock1, Tom R Gaunt1, Jie Zheng1, Andrew P Morris2, George Davey Smith1, Gibran Hemani1

1 MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

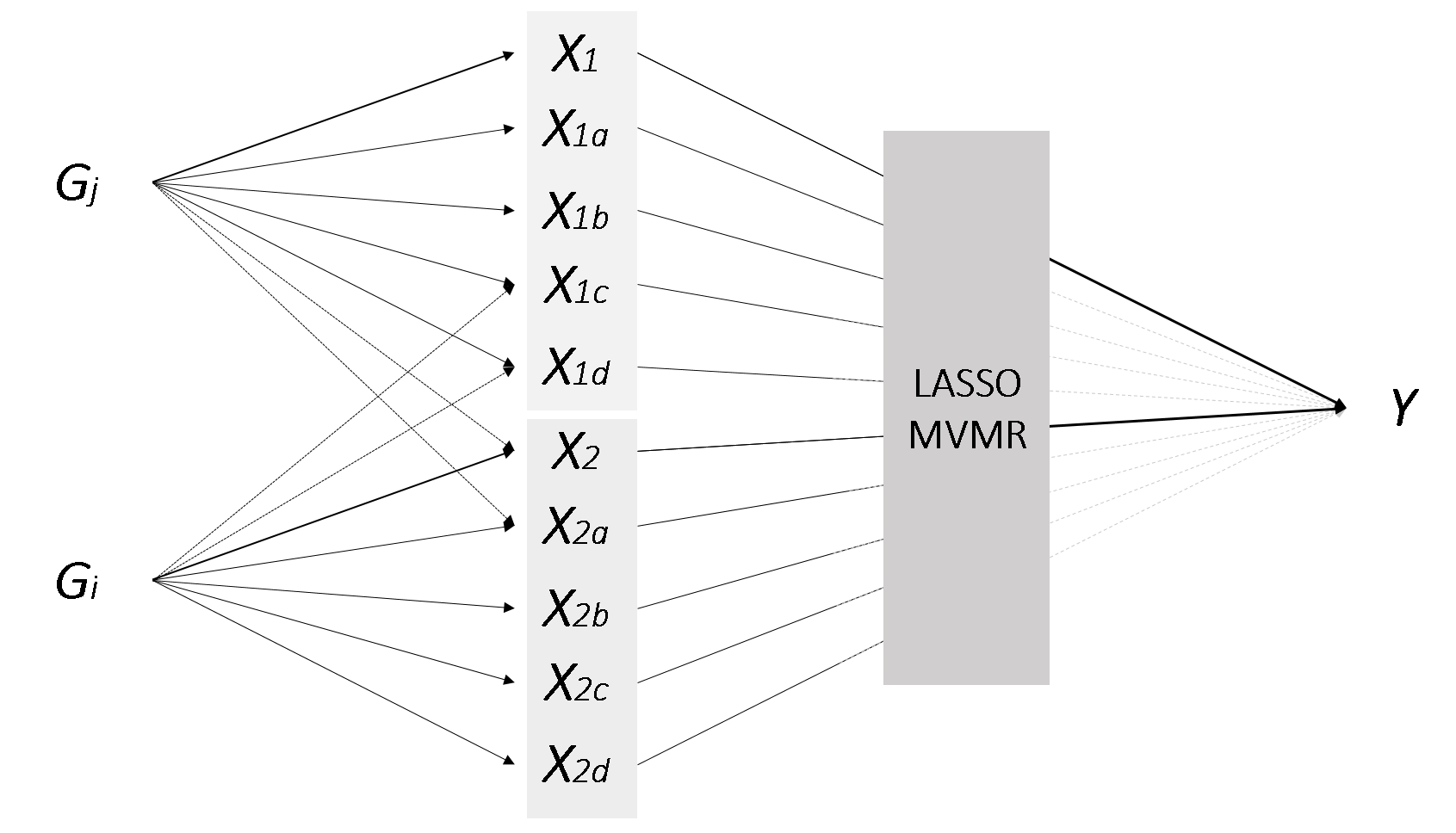
2 Department of Biostatistics, University of Liverpool, Liverpool, UK.

**Supplementary figure legends**

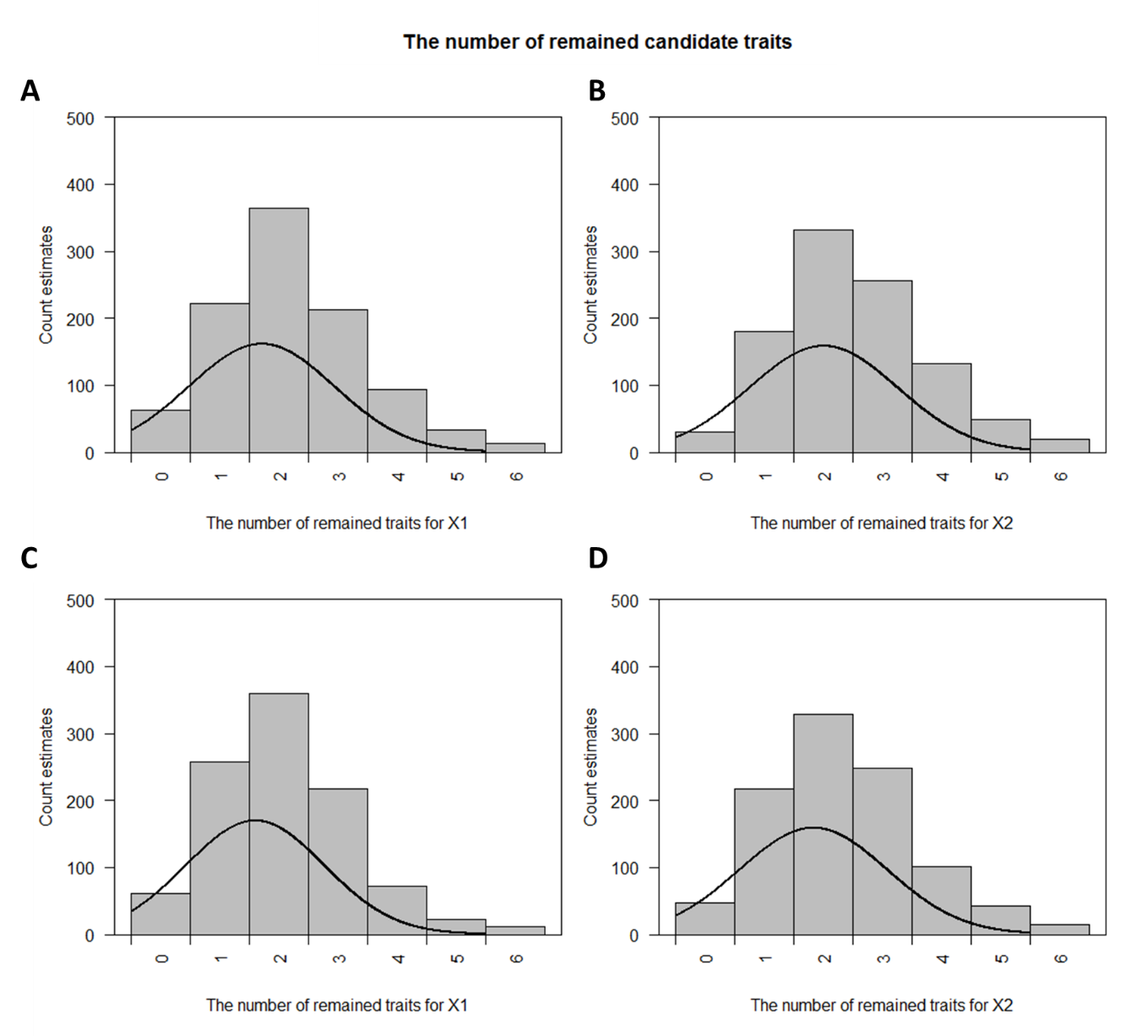
**Figure S1. The design of simulation 1.** Directed acyclic graphs illustrating that potentially account for the causal relationship between the candidate traits including the original trait and redundant traits (e.g. X1b, X2b, etc) and the outcome. The genetic variants Gj and Gi affect the traits. The dotted line between genetic variants and the traits indicates that genetic variant G affects the trait X1s also influences the traits X2s via horizontal pleiotropy. The LASSO step selects the traits that are associated with the outcome and have non-zero effects. The bold line between the traits and the outcome Y indicates the association between selected traits by the LASSO steps and the outcome.

**Figure S2. The number of candidate traits included in LASSO-MVMR.** The graph presents the number of traits (e.g. X1 and X2) and their proxies that were survived after shrinkage step through the LASSO and were included in LASSO-MVMR analysis. (A) and (B) show the number of X1s and X2s remained after LASSO step where there is no pleiotropy whilst (C) and (D) shows the number of X1s and X2s after LASSO where the genetic variants are pleiotropic. The line indicates the normal distribution.

**Figure S1. The design of simulation 1.**

****

**Supplementary figure 2. The number of candidate traits included in LASSO-MVMR.**

****

**Supplementary Table S1. Sources of data used in this study.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Consortium/Studies** | **Use** | **Study population** | **Imputation** | **Model** | **Adjustments** |
| CARDIoGRAMplusC4D1 | SNP-Log odds CHD  SNP-Log odds MI | 22 233 cases and 64 762 controls of European descent  43 676 cases and 128,199 controls of European, South/East Asian, and Hispanic and African American ancestry | MACH, IMPUTE, or BIMBAM  (ref: HapMap II CEU)  IMPUTE, or MINIMAC  (ref: HapMap II/III or 1000 Genomes phase 1 ver. 3) | ADD | Age, sex and genomic control inflation factor |
| HaemGen | SNP-Platelet count2  SNP-Haemoglobin concentration3  SNP-Packed cell volume3  SNP-Red blood cell count3 | 66 867 individuals of European ancestry  71 861 individuals of European or South Asian ancestry  63 511 individuals of European or South Asian ancestry  66 214 individuals of European or South Asian ancestry | IMPUTE, BIMBAM, or MACH  (ref: HapMap II CEU)  MACH, IMPUTE, or BEAGLE  (ref: HapMap II) | ADD | Age, sex and other cohort‐specific covariates (where appropriate)  Cohort‐specific covariates  (where appropriate) |
| GUGC4 | SNP-Urate | 110 347 individuals of European ancestry | MACH, IMPUE, BEAGLE, BIMBAM  (ref: HapMap II CEU) | ADD | Age, sex, and study-specific covariates, if applicable |
| CKDGen5 | SNP-Serum Cystatin C | 133,413 individuals of European ancestry | MACH, IMPUTE, or BIMBAM  (ref: HapMap II CEU) | ADD | Age, sex, study site and genetic principal components (where appropriate). |
| IPSCSG6 | SNP-Primary sclerosing cholangitis | 4 796 cases and 19 955 population controls of European and American ancestry | IMPUTE2  (1000 Genomes Phase I ver. 3) | ADD |  |
| Dubois et al.7 | SNP-Celiac disease | 4 533 cases and 10 750 controls of European ancestry | BEAGLE and CEU, TSI, MEX and GIH  (ref: HapMap III) | ADD | Population stratification |
| GLGC8 | SNP-HDLC  SNP-LDLC  SNP-TC | 188 578 individuals of European, East Asian, South Asian, and African ancestry | MACH  (ref: HapMap II CEU) | ADD | Age, sex, principal components of  genomic ancestry (some studies),  and genomic control inflation factor |
| UK Biobank\* | SNP-SBP  SNP-Treatments  SNP-Anthropometric measures  SNP-Smoking  SNP-Alcohol drinking  SNP-Diagnosed diseases  SNP-Education  SNP-Exercise  SNP-Memory | Approx. 340 000 individuals of European ancestry | IMPUTE2  (ref: Haplotype Reference Consortium and UK10K haplotype resource) | ADD | Population structure (where appropriate). |

CARDIoGRAM, Coronary ARtery DIsease Genome-wide Replication and Meta-analysis; GUGC, Global Urate Genetics Consortium; CKDGen, Chronic Kidney Disease Genetics; IPSCSG, International Primary Sclerosing Cholangitis Study Group; GLGC, Global Lipids Genetics Consortium; CEU, Northern Europeans from Utah; ADD, additive.

\* UK Biobank GWAS results from Neale Lab: http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank.

**Supplementary Table S2. MR of candidate traits on each outcome (Excel).**

**Supplementary Table S3. Results of supplementary simulation 1.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Presence of horizontal pleiotropy** | | | |
|  | **No pleiotropy** | | **With pleiotropy** | |
|  | **Beta estimate** | **P value** | **Beta estimate** | **P value** |
| **LASSO-MVMR** |  |  |  |  |
| X1 | 0.192 (0.004) | < 1.00e-159 | 0.187 (0.003) | < 1.00e-159 |
| X2 | 0.185 (0.004) | < 1.00e-159 | 0.187 (0.003) | < 1.00e-159 |
| **MVMR** |  |  |  |  |
| X1 | 0.397 (0.511) | 1.35e-133 | 0.410 (0.019) | 2.12e-103 |
| X2 | 0.002 (0.506) | 4.49e-1 | -0.008 (0.019) | 6.62e-1 |

S.D., standard deviation; MVMR, multivariable Mendelian randomization. 1 Beta estimates were derived from the LASSO-MVMR and normal MVMR. The simulate causal effect is 0.2.

**Supplementary Table S4. Results of supplementary simulation 2.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Probability to be selected by LASSO1** | | | |
| **Scenario2** | **S1** | **S2** | **S3** | **S4** |
| X1 | 1.000 | 0.467 | 1.000 | 0.225 |
| X2 | 0.473 | 1.000 | 1.000 | 0.217 |

S.D., standard deviation; MVMR, multivariable Mendelian randomization. 1 The values represent the probability of the number of times each trait is selected by LASSO out of 1000 repeated simulations. 2 X1 was instrumented by 100 genetic variants whilst X2 was instrumented by 20 genetic variants. 1 S1: X1 has an effect on Y (β=0.3), where X2 has no effect on Y (β=0.0); S2: X1 has no effect on Y (β=0.0), where X2 has an effect on Y (β=0.3); S3: Both of X1 and X2 have effects on Y (β=0.3); S4: Neither of X1 and X2 have effects on Y (β=0.0).

**References**

1. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015;47(10):1121-30. doi: 10.1038/ng.3396

2. Gieger C, Radhakrishnan A, Cvejic A, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature* 2011;480(7376):201-8. doi: 10.1038/nature10659

3. van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature* 2012;492(7429):369-75. doi: 10.1038/nature11677

4. Huffman JE, Albrecht E, Teumer A, et al. Modulation of genetic associations with serum urate levels by body-mass-index in humans. *Plos One* 2015;10(3):e0119752. doi: 10.1371/journal.pone.0119752

5. Kottgen A, Pattaro C, Boger CA, et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet* 2010;42(5):376-84. doi: 10.1038/ng.568

6. Ji SG, Juran BD, Mucha S, et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat Genet* 2017;49(2):269-73. doi: 10.1038/ng.3745

7. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;42(4):295-302. doi: 10.1038/ng.543

8. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013;45(11):1274-83. doi: 10.1038/ng.2797