The Journal of Infectious Diseases

MAJOR ARTICLE

Investigation of Causal Effects of Protein Biomarkers on Cardiovascular Disease in Persons with HIV

Cavan S. Reilly[*1], Álvaro H Borges[2], Jason V. Baker[3], Sandra E Safo[4], Shweta Sharma[5], Mark N Polizzotto[6], James S Pankow[7], Xiaojun Hu[8], Brad T Sherman[9], Abdel G Babiker[10], Jens D Lundgren[11], H Clifford Lane[12]

[1] Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA; [2] Immunology, Statens Serum Institut, Copenhagen, Denmark; [3] HIV Medicine, Infectious Diseases, Hennepin County Medical Center, Minneapolis, MN, USA; [4] Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA; [5] Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA; [6] Department of Medicine, Australian National University, Canberra, Australia; [7] Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA; [8] Animal and Plant Inspection Service, USDA, Beltsville, MD, USA; [9] Laboratory of Human Retrovirology and Immunoinformatics, Frederick National Laboratories, Frederick, MD, USA; [10] Epidemiology and Medical Statistics, University College London, London, United Kingdom; [11] Department of Infectious Diseases, University of Copenhagen, Copenhagen, Denmark; [12] Division of Clinical Research, National Institutes of Allergy and Infectious Diseases, USA

Background: There is an incompletely understood increased risk for cardiovascular disease (CVD) among people living with HIV (PLWH). We investigated if a collection of biomarkers were associated with CVD among PLWH. Mendelian randomization (MR) was used to identify potentially causal associations.

Methods: Data from follow-up in 4 large trials among PLWH were used to identify 131 incident CVD cases and they were matched to 259 participants without incident CVD (controls). Tests of associations between 460 baseline protein levels and case status were conducted.

Corresponding author contact information: Cavan Reilly, 612-624-9644, cavanr@biostat.umn.edu, 2221 University Ave SE, Suite 200, Minneapolis, MN, 55414, USA

DOI: 10.1093/infdis/jiac496

Results: Univariate analysis found CLEC6A, HGF, IL6, IL10RB, and IGFBP7 as being associated with case status and a multivariate model identified 3 of these: CLEC6A (OR=1.48, p=0.037), HGF (OR=1.83, p=0.012) and IL6 (OR=1.45, p=0.016). MR methods identified 5 significantly associated proteins: AXL, CHI3L1, GAS6, IL6RA, and SCGB3A2.

Conclusions: These results implicate inflammatory and fibrotic processes as contributing to CVD. While some of these biomarkers are well established in the general population and in PLWH (IL6 and its receptor), some are novel to PLWH (HGF, AXL and GAS6) and some are novel overall (CLEC6A). Further investigation into; 1.) the uniqueness of these biomarkers in PLWH and 2.) the role of these biomarkers as targets among PLWH, is warranted.

Keywords: Mendelian randomization, inflammation, fibrosis

INTRODUCTION

People living with HIV (PLWH) are at increased risk of cardiovascular disease (CVD) when compared to the general population [

1,2,3]. The reasons for this increase in risk remain unclear. In addition to a higher prevalence of traditional risk factors [4], patients with HIV infection have evidence of persistent abnormalities in inflammation and coagulation that might be related to duration of infection [5,6,7,8], incomplete immune recovery [9,10], ongoing viral replication [11], transcription of defective proviruses [12], and/or long-term ART toxicity [13]. Recent meta-analyses have estimated this increased risk to be a factor of about 2 after adjustment for potential confounders [14,15]. Recent reviews have summarized these and other mechanisms [16,17]. Identifying mechanisms of this increased risk has the potential to suggest molecular treatment targets.

Recent advances in proteomics have enabled more precise quantification of protein levels in human specimens. While many of these advances have utilized mass spectrometry-based approaches, assays that rely on other molecular techniques have also seen widespread adoption. In particular, the proximity extension assay, made commercially available by Olink, has been used for protein quantification in multiple large cohort studies focused on identification of risk factors for CVD in the general population [18,19,20].

While investigations of associations between potential protein biomarkers and the development of CVD have the potential to elucidate mechanisms, these associations may be attributable to a variety of unmeasured and unknown confounding factors. One established method for overcoming this shortcoming is the use of Mendelian randomization (MR) [21,22]. Given several key assumptions, MR is a statistical technique that can be used to identify causal associations between risk factors and outcomes when one has genotypic data. In MR analysis, one tries to overcome the impact of confounders by identifying a genetic variant that is related to the risk factor but is unlikely to be related to confounders, and then one uses a technique called

instrumental variable analysis to test for an association between the risk factor and the outcome using the identified genetic variant. MR analysis has been used to investigate the causal role of a wide variety of risk factors for CVD, including lower levels of HDL cholesterol [23] and higher levels of the IL6 receptor [24].

Here we investigate the predictive utility and apply MR to these novel protein assays in a group of geographically and racially diverse HIV positive individuals who were recruited into clinical trials conducted by the International Network for Strategic Initiatives in Global HIV Trials (INSIGHT). A total of 15,665 individuals participated in these trials, many contributing specimens, and were followed for up to 10 years with clinical evaluations. Detailed clinical outcomes were documented and assessed by endpoint review committees. These proteomic assays, specimens, statistical techniques, and clinical outcomes were used to investigate the roles of a range of candidate proteins on the development of CVD in PLWH.

METHODS

Participants

Data from participants in four clinical trials (FIRST [25], ESPRIT [26], SMART [27] and START [28]) who consented to studies of genomics (8,428 of 15,665) were combined and a case definition comprising a composite of death, AIDS, or serious non-AIDS conditions was used to identify 500 cases. (Some of these individuals could have experienced a previous event that met the criteria for case status used here.) Cases who had experienced at least one of these outcomes were individually matched on a 1 to 2 basis with controls who had not experienced any of these outcomes over follow-up. Matching was performed within treatment arms of each study and used randomization date and age. This resulted in a nested case-control dataset involving 1493 individuals.

Here for analysis of CVD we focus on the 131 cases who experienced a composite CVD outcome, defined as stroke, MI, or coronary revascularization and 259 matched controls. For analyses not explicitly related to CVD we used the entire cohort of 1493 individuals; the latter analyses were focused on identification of genetic based instruments and by using the larger cohort these analyses had greater power.

Assays

Plasma, mostly obtained during a fasting state, from the baseline study visit for these study participants was used to generate data for 460 proteins from 5 Olink protein panels (cardiometabolic, CVD2, CVD3, immune response, and inflammation). Some proteins were represented on more than one panel. Proteins were excluded from further consideration if more than 10% of the data was below the limit of detection. Observations that were below the limit of detection for retained proteins were imputed at one half the limit of detection on the linear scale.

Genotypes for all 1493 individuals with protein level data were obtained using an Axion Affymetrix array with 770,558 probesets.

Statistical methods

Tests for association between protein levels and the CVD composite outcome controlled for sex, age, self-declared black race, diabetes, hypertension, and treatment assignment in each trial. Generalized estimating equations (GEE) were used to account for matching among cases and controls: both linear and logistic regression results are reported due to the widespread use of linear regression in MR analysis with binary outcomes. These analyses were restricted to CVD cases and their matched controls.

Construction of predictive models for the CVD composite outcome using collections of potential biomarkers started with all proteins meeting marginal statistical significance based on Bonferroni criteria given the number of proteins and all covariates listed for the CVD composite outcome analysis. The protein with the smallest test statistic in absolute value was successively dropped according to a backwards model selection strategy. Once all remaining proteins were significant, each protein meeting the Bonferroni threshold that was dropped was re-entered into the model and was retained if the effect was statistically significant. These analyses were restricted to CVD cases and their matched controls and also used GEE versions of linear regression to be consistent with MR analyses but were supplemented with logistic regression.

One difficulty with using MR analysis in the current context is control of the type I error due to the large number of proteins and genetic variants. To best make use of the large number of proteins and genetic variants while controlling the number of false positives, we devised a sequential testing method to control the family-wise error rate of our collection of MR tests. The basic idea is that there is little power for finding a causal association unless there is an association between a genetic variant and a protein on the one hand and an association between the CVD outcome and a protein on the other. This suggests that we screen these pairs of associations and only conduct the MR test when both members of a pair of these tests reject the null hypothesis. If we have a collection of p proteins and we let $\alpha_{\text{protein,CVD}}$ represent the significance level we require for rejecting the null hypothesis of no association between protein levels and the CVD composite outcome, $\alpha_{\text{protein,SNP}}$ represent the significance level we require for rejecting the null hypothesis of no association between protein levels and SNPs and α_{MR} represent the significance level we require for rejecting the null hypothesis of no association in our MR tests then it transpires that the probability we reject 1 of more MR tests under the null hypothesis of no association between any of the factors involved is approximately $p\,\alpha_{\text{MR}}\alpha_{\text{protein,SNP}}\alpha_{\text{protein,CVD}}$. This is true under the set of assumptions employed in justifying the Bonferroni correction: details are provided in the Supplementary Materials.

Genetic variants were selected by finding SNPs near the coding region of each protein that were significantly associated with the protein's level, then haplotypes were estimated for each

participant using this set of SNPs. The most common haplotype was determined, and individuals were coded as having 0-2 copies of this haplotype. This haplotype was used as a genetic variant in MR tests.

Additional methodological details can be found in the Supplementary Materials.

RESULTS

Table 1 provides a summary of baseline participant characteristics for CVD cases and controls. A similar summary for all cases and controls is presented in Supplementary Table 1. Participants in two of the trials (FIRST and START) were ART naïve at baseline and there is substantial variation among participants across the trials.

Biomarkers associated with clinical outcomes

Summaries of the protein data (along with the Uniprot ID for each protein symbol used here) can be found in Supplementary Tables 2-11. Overall, 86 of the 385 distinct proteins meeting quality control criteria were significantly associated with the CVD composite outcome using a significance level of 0.05. The number of significant proteins found using the individual panels (ignoring overlap) was as follows: cardiometabolic (8), cardiovascular II (14), cardiovascular III (32), immune response (15), and inflammation (24). Tables 2 and 3 present results for all significant proteins using linear and logistic regression models while Supplementary Tables 12-16 present similar results for the proteins which failed to reach significance for each panel. Using a Bonferroni adjustment across all panels simultaneously, 5 proteins differed significantly between cases and controls while controlling for relevant covariates. These were CLEC6A, HGF, IL6, IL10RB, and IGFBP7.

To better understand the interplay between these biomarkers, models were fit that included all of them in addition to relevant covariates. Data from cardiovascular panel III (IGFBP7), the immune response panel (CLEC6A) and the inflammation panel (HGF, IL6, IL10RB) were merged for this analysis. Proteins were dropped according to the backwards selection scheme described above. This resulted in a final model with CLEC6A (OR=1.48, p=0.037, 95% CI: 1.02, 2.13), HGF (OR=1.83, p=0.012, 95% CI: 1.14, 2.95) and IL6 (OR=1.45, p=0.016, 95% CI: 1.07, 1.96) increasing the risk for the CVD composite outcome. Since proteins are on the $\log_2$ scale the OR corresponds to the change in the odds ratio associated with a doubling of the protein level. The final model also indicated differences in CVD risk across the study arms.

Mendelian randomization analysis

With 385 proteins, we can take all significance levels in the sequential testing strategy to be 0.05 and still control the family-wise error rate at 5%. This is what is reported here: Figure 1 provides a graphical presentation of the following results. As noted above, 86 proteins were associated

with CVD at a significance level of 0.05, hence these 86 proteins were selected during the first stage of the sequential procedure. Of these 86 proteins, 10 had no SNPs in the region of its corresponding structural gene and 29 were associated with at least 1 SNP in the corresponding gene. The numbers of such SNPs from the proteins identified as being associated with CVD outcomes in the previous section are as follows (in parentheses): CLEC6A (41), HGF (0), IL6 (0), IL10RB (0) and IGFBP7 (4). Supplementary Table 17 provides a summary of the variants that pass various levels of filtering for each protein associated with the CVD composite outcome with at least 1 variant in its coding region. The filtering displayed in this table is cumulative showing the resulting number of variants after each application of the sequential filtering strategy. Of the 29 proteins that were associated with at least 1 SNP in their coding region, the sequential procedure detected significant associations between 5 proteins and CVD outcomes: these were AXL (p=0.021), CHI3L1 (p=0.045), GAS6 (p=0.015), IL6RA (p=0.049) and SCGB3A2 (p=0.038). Table 4 presents results of all MR tests conducted. All *F*-statistics used for diagnosis of weak instruments exceeded 10 indicating the appropriateness of the approximations used for inference. While only 1 SNP served as an instrument for AXL, all other proteins used a haplotype involving multiple SNPs as an instrument with the numbers for SNPs for each protein as follows (numbers of SNPs in parentheses); CHI3L1 (4), GAS6 (2), IL6RA (68) and SCGB3A2 (27). While all these proteins exhibited a positive association with the composite CVD outcome in the previously described linear and logistic models, CHI3L1 and GAS6 have negative associations with the composite CVD outcome in the MR analysis.

Additional detailed results can be found in the Supplementary Materials.

DISCUSSION

Strong evidence was detected for an association between multiple protein biomarkers and adjudicated CVD outcomes in a global population of PLWH. There were associations between IGFBP7, CLEC6A, IL10RB, HGF, and IL6 as individual biomarkers and CVD outcomes. Since biomarker levels are generally correlated, investigation of joint models examined how the collection of these proteins impact CVD risk. Joint models identified IL6, HGF, and CLEC6A as each independently contributing to CVD risk given the effect of the other proteins. Higher levels of all 3 of these proteins are associated with increased risk for CVD. These proteins were at elevated levels over 6 years prior to the development of CVD in some individuals. This suggests that there may be an association between protein levels and subclinical atherosclerotic disease long before overt CVD.

A strategy of constructing haplotypes and using these haplotypes as instruments in MR analysis provided evidence for causal effects of 5 proteins on CVD outcomes: AXL, IL6RA, CHI3L1, SCGB3A2, and GAS6. Our strategy was unable to find suitable genetic instruments for IL6, IL10RB and HGF, thus we could not effectively test for causal effects for these biomarkers which appear to at least be associated with CVD outcomes. One of the biomarkers with evidence

for a causal role, IL6RA has been demonstrated to have a causal effect on CVD in the general population, hence finding this in an HIV infected population lends support to the hypothesis that at least some mechanisms of CVD development are common among PLWH and the general population. GAS6 has been described as a ligand for AXL so finding both as being involved in the causal pathway for CVD implicates processes that involve this pathway in CVD development. It is intriguing that the direction of the effects for CHI3L1 and GAS6 differed in the MR analysis compared to what we found in the linear model. The most likely explanation is that the models that detect the positive associations are excluding critical covariates (perhaps the levels of other proteins or metabolites), and if these critical covariates had been measured and included the direction of the association would change in the linear models.

There is support in the literature for the proteins identified as being associated with CVD outcomes. The role of IL6 in inflammatory processes relating to CVD among PLWH is well established [6,8] and there are reviews that describe its role in inflammatory processes in the general population [28]. There are also a number of publications that link HGF to various aspects of CVD in the general population [29,30], but this appears to be the first such association described among PLWH. This protein is thought to play a role in cardiovascular remodeling in response to endothelial injury. More specifically it is thought to modulate inflammatory responses in immune cells and to prevent fibrosis mediated by fibroblasts via $TGF\beta_1$ dependent mechanisms (these have been reviewed [31]). The association between CLEC6A (also known as DECTIN-2) levels and CVD outcomes appears to be novel. CLEC6A is a C-type lectin receptor expressed by macrophages, dendritic cells and monocytes with ligands derived from pathogens that induce inflammatory signals upon binding [32].

The literature on the proteins identified using the MR approach is mixed and not entirely consistent with the findings presented here. AXL and its primary ligand GAS6 have been found to be associated with CVD previously and they have been described as being upregulated in response to inflammation and also characterized as being involved in the resolution of inflammatory signals [33,34]. Similarly, CHI3L1 has been found at elevated levels in patients with a variety of inflammatory conditions, including CVD and has been variously described as having pro- and anti-inflammatory roles [35,36]. IL6RA has been well documented to have a critical role in inflammation [6,8]. SCGB3A2 has been described as having anti-fibrotic and anti-inflammatory properties [37,38].

Our findings have some limitations. We do not have comprehensive data on established CVD risk factors across all trials. In particular, we do not have data on smoking or cholesterol levels, and this may induce some residual confounding in our tests for associations between protein levels and the composite CVD outcome. Moreover, these variables could act as effect modifiers, and we would not be able to detect this. However, our models did control for a diagnosis of diabetes and hypertension so some of the potential confounding was accounted for. Since we do not know that the CVD events described here were necessarily the first such event experienced by study participants there is some potential for reverse causality. On the other hand, many of the

events took place years after the specimens examined here were obtained. In addition, we do not have data on the type of stroke or MI across all trials, so further mechanistic understanding of the impact of these proteins is limited. Finally, it is difficult to assess the exclusion restriction assumption in MR analysis that provides the basis for a causal interpretation of the associations described here. To reduce the potential for this we restricted our consideration of SNPs to those near the coding region of the protein of interest.

Our findings indicate that final common pathways leading to the development of CVD among PLWH are similar to the general population. This is particularly the case for IL-6- and GAS6-dependent pathways. We note, however, that the proteins included on the CVD panels were selected based on risk in an HIV negative setting, thus these panels exclude proteins unique to CVD pathogenesis in PLWH should such proteins exist. We also report for the first-time associations between biomarkers of fibrotic/inflammatory pathways with CVD among PLWH. Further investigations are warranted to 1) define whether this reflects mechanistic differences in the development of CVD between PLWH and the general population; and 2) assess the clinical relevance of using these pathways as potential targets for new CVD therapies.

FOOTNOTE PAGE

Conflict of interest: No authors have a conflict of interest.

Funding: This work was supported by the National Institutes of Health [HHSN261200800001E].

Previous presentation: A presentation of some of this material was presented at the Conference on Retroviruses and Opportunistic Infections in 2020.

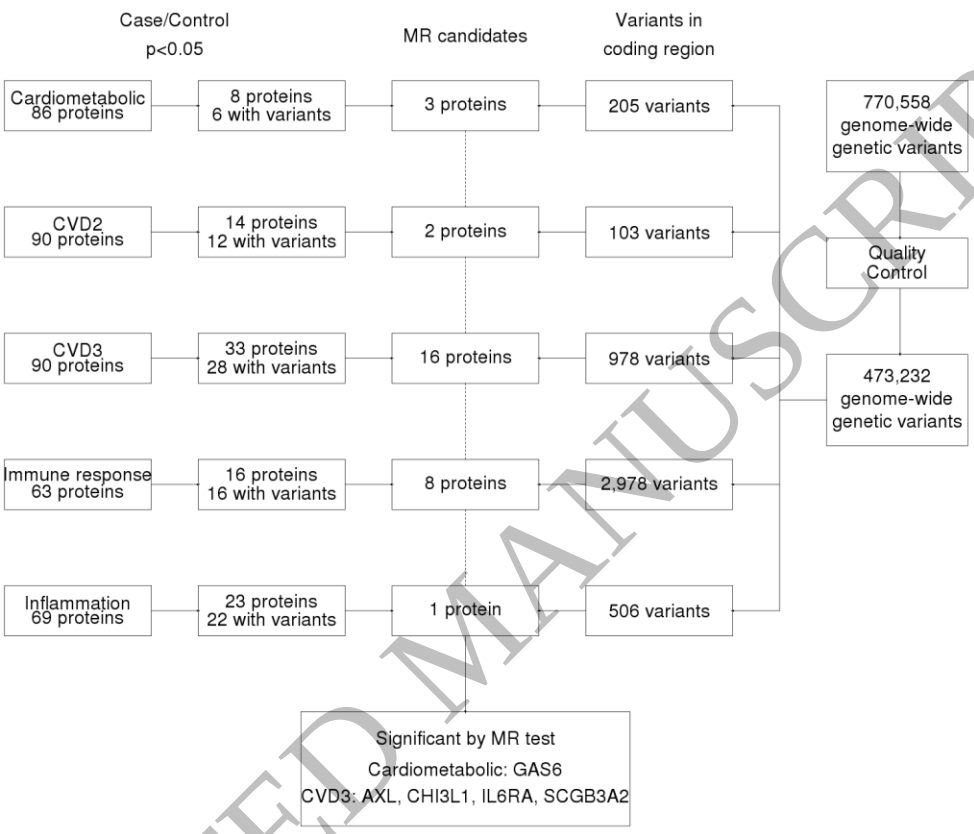

Figure 1
165x165 mm (5.2 x DPI)

References

1 Freiberg MS, Chang CC, Kuller LH, *et al.* HIV infection and the risk of acute myocardial infarction. JAMA Intern Med, **2013**; 173: 614-622.

2 Baker JV, Lundgren JD. Cardiovascular implications from untreated human immunodeficiency virus infection. Eur Heart J, 2011; 32: 945-951.

3 Hsue PY, Deeks SG, Hunt PW. Immunologic basis of cardiovascular disease in HIV-infected adults. J Infect Dis, 2012; 205 Suppl 3: S375-382.

4 Rasmussen LD, Helleberg M, May MT, *et al.* Myocardial infarction among Danish HIV-infected individuals: population-attributable fractions associated with smoking. Clin Infect Dis, **2015**; 60: 1415-1423.

5 Duprez DA, Neuhaus J, Kuller LH, *et al.* Inflammation, coagulation and cardiovascular disease in HIV-infected individuals. PLoS One, **2012**; 7: e44454.

6 Nordell AD, McKenna M, Borges ÁH, Duprez D, Neuhaus J, Neaton JD; INSIGHT SMART, ESPRIT Study Groups.; SILCAAT Scientific Committee. Severity of cardiovascular disease outcomes among patients with HIV is related to markers of inflammation and coagulation. J Am Heart Assoc, **2014**; 3: e000844.

7 Grund B, Baker JV, Deeks SG, *et al.* Relevance of Interleukin-6 and D-Dimer for serious non-AIDS morbidity and death among HIV-positive adults on suppressive antiretroviral therapy. PLoS One, **2016**; 11: e0155100.

8 Borges ÁH, O'Connor JL, Phillips AN, *et al.* Interleukin 6 Is a stronger predictor of clinical events than high-sensitivity C-reactive protein or D-Dimer during HIV infection. J Infect Dis, **2016**; 214: 408-416.

9 Baker JV, Peng G, Rapkin J, *et al.* CD4+ count and risk of non-AIDS diseases following initial treatment for HIV infection. AIDS, **2008**; 22: 841-848.

10 Baker JV, Peng G, Rapkin J, *et al.* Poor initial CD4+ recovery with antiretroviral therapy prolongs immune depletion and increases risk for AIDS and non-AIDS diseases. J Acquir Immune Defic Syndr, **2008**; 48: 541-546.

11 Phillips AN, Neaton JD, Lundgren JD. The role of HIV in serious diseases other than AIDS. AIDS, **2008**; 22: 2409–2418.

12 Imamichi H, Smith M, Adelsberger JW, *et al.* Defective HIV-1 proviruses produce viral proteins. Proc Natl Acad Sci U.S.A., **2020**; 117: 3704-3710.

13 Friis-Møller N, Reiss P, Sabin CA, *et al.* Class of antiretroviral drug and myocardial infarction. N Engl J Med, **2007**; 356: 1723-1735.

14 Eyawo O, Brockman G, Goldsmith CH, *et al. Risk of myocardial infarction among people living with HIV: an updated systematic review and meta-analysis. BMJ Open,* **2019**; 9: e025874.

15 Shah ASV, Stelzle D, Lee KK, *et al.* Global burden of atherosclerotic cardiovascular disease in people living with HIV. Circulation, **2018**; 138: 1100-1112.

16 Kearns A, Gordon J, Burdo TH, Qin X. HIV-1-associated atherosclerosis: Unraveling the Missing Link. J Am Coll Cardiol, **2017**; 69: 3084–98.

17 Hsue PY. Mechanisms of cardiovascular disease in the setting of HIV infection. Can J Cardiol, **2019**; 35: 238-248.
18 Lind L, Siegbahn A, Lindahl B, Stenemo M, Sundström J, Ärnlöv J. Discovery of new risk markers for ischemic stroke using a novel targeted proteomics chip. Stroke, **2015**; 46: 3340-3347.
19 Schiopu A, Bengtsson E, Gonçalves I, Nilsson J, Fredrikson GN, Björkbacka H. Associations between macrophage colony-stimulating factor and monocyte chemotactic protein 1 in plasma and first-time coronary events: A nested case-control study. J Am Heart Assoc, **2016**; 5: e002851.
20 Gonçalves I, Hultman K, Dunér P, *et al*. High levels of cathepsin D and cystatin B are associated with increased risk of coronary events. Open Heart, 2016; 3: e000353.
21 Youngman LD, Keavney BD, Palmer A. Plasma fibrinogen and fibrinogen genotypes in 4685 cases of myocardial infarction and in 6002 controls: test of causality by 'Mendelian randomization'. Circulation, **2000**; 102: 31-32.
22 Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat in Med, **2008**; 27: 1133-1163.
23 Voight BF, Peloso GM, Orhu-Melander M, *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a mendelian randomisation study. Lancet, **2012**; 380: 572-580.
24 Swerdlow DI, Holmes MV, Kuchenbaecker KB, *et al.* The interleukin-6 receptor as a target for the prevention of coronary heart disease: a mendelian randomisation analysis. Lancet, **2012**; 379: 1214-1224.
25 MacArthur RD, Novak RM, Peng G, *et al*. A comparison of three highly active antiretroviral treatment strategies consisting of non-nucleoside reverse transcriptase inhibitors, protease inhibitors, or both in the presence of nucleoside reverse transcriptase inhibitors as initial therapy (CPCRA 058 FIRST Study): a long-term randomized trial. Lancet, 2006; 368: 2125-2135.
26Abrams D, Lévy Y, Losso MH, *et al.* Interleukin-2 therapy in patients with HIV infection. N Engl J Med, **2009**; 361: 1548-1559.
27 The Strategies for Management of Antiretroviral Therapy (SMART) Study Group. CD4+ count–guided interruption of antiretroviral treatment. N Engl J Med, 2006; 355: 2283-2296.
28 Lundgren JD, Babiker AG, Gordin F, *et al*. Initiation of antiretroviral therapy in early asymptomatic HIV infection. N Engl J Med, 2015; 373: 795-807.
28 Hunter CA, Jones SA. IL-6 as a keystone cytokine in health and disease. Nat Immunol, **2015**; 16: 448-457.
29 Bell EJ, Larson NB, Decker PA, *et al.* Hepatocyte growth factor is positively associated with risk of stroke: the MESA. Stroke, **2016**; 47: 2689-2694.
30 Bell EJ, Decker PA, Tsai MY, *et al.* Hepatocyte growth factor is associated with progression of atherosclerosis: the Multi-Ethnic Study of Atherosclerosis (MESA). Atherosclerosis, **2018**; 272: 162-167.
31 Gallo S, Sala V, Gatti S, Crepaldi T. Cellular and molecular mechanisms of HGF/Met in the cardiovascular system. Clin Sci, **2015**; 129: 1173-1193.

32 Sancho D, e Sousa CR. Signaling by myeloid C-type lectin receptors in immunity and homeostasis. Annu Rev Immunol, **2012**; 30: 491-529.
33 Zagórska A, Través PG, Lew ED, et al. Diversification of TAM receptor tyrosine kinase function. Nat Immunol, **2014**; 15: 920-928.

34 Carrera-Silva EA, Chan PY, Joannas et al. T cell-derived protein S engages TAM receptor signaling in dendritic cells to control the magnitude of the immune response. Immunity, **2013**: 39: 160-170.
35 Maegdefessel L, Spin JM, Raaz U, *et al.* miR-24 limits aortic vascular inflammation and murine abdominal aneurysm development. Nat. Commun, **2014**; 5: 5214.
36 Y. Zhou, Peng H, Sun H, Peng X, Tang C, Gaun, *et al.* Chitinase 3-like 1 suppresses injury and promotes fibroproliferative responses in mammalian lung fibrosis. Sci Transl Med, **2014**; 6: 240ra76.
37 Cai Y, Kimura S. Secretoglobin 3A2 exhibits anti-fibrotic activity in bleomycin-induced pulmonary fibrosis model mice. PLoS ONE, **2015**; 10: e0142497.
36 Kido T, Yoneda M, Cai Y, Matsubara T, Ward JM, Kimura S. Secretoglobin superfamily protein SCGB3A2 deficiency potentiates ovalbumin-induced allergic pulmonary inflammation. Mediators Inflamm. **2014**; Article ID 216465.

Table 1: Summary of participant characteristics among CVD cases and controls.

	ESPRIT		**SMART**		**FIRST**		**START**		**Total**	
	Cases	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**	**Cases**	**Contro**
No. participants	**77**	**151**	**16**	**32**	**8**	**16**	**30**	**60**	**131**	**259**
Age	48 [40, 54]	46 [40, 52]	52 [45, 57]	50 [46, 58]	44 [35, 53]	41 [35, 50]	50 [44, 56]	48 [43, 55]	49 [41, 55]	47 [41, 53]
Female	4 (5.2)	13 (8.6)	1 (6.3)	8 (25.0)	2 (25.0)	3 (18.8)	4 (13.3)	14 (23.3)	11 (8.4)	38 (14
Black race	6 (7.8)	14 (9.3)	3 (18.8)	7 (21.9)	4 (50.0)	11 (68.8)	9 (30.0)	12 (20.0)	22 (16.8) 44 (17.0)	
Geographic region										
Africa	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (6.7)	4 (6.7)	2 (1.5)	4 (1.5)
Asia	2 (2.6)	10 (6.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.5) 10 (3.9)	
Europe+Israel	39 (50.6)	73 (48.3)	2 (12.5)	3 (9.4)	0 (0.0)	0 (0.0)	14 (46.7)	33 (55.0)	55 (42.0) 109 (42.1)	
Latin America	6 (7.8)	13 (8.6)	3 (18.8)	2 (6.3)	0 (0.0)	0 (0.0)	0 (0.0)	11 (18.3)	9 (6.9) 26 (10.0)	

Oceania	6 (7.8)	14 (9.3)	3 (18.8)	5 (15.6)	0 (0.0)	0 (0.0)	4 (13.3)	1 (1.7)	13 (9.9) 20 (7.7)
HIV RNA (copies/mL)	50 [50, 400]	50 [50, 400]	50 [50, 50]	50 [50, 400]	86833 [42056, 310714]	68043 [38073, 152010]	13791 [3900, 47000]	9895 [2698, 30450]	400 262 [50, 7348] [50, 6424]
Hepatitis B	5 (7.1)	7 (5.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (3.7)	1 (1.7)	6 (5.0) 8 (3.3)
Regimen at entry[a]									
NRTI, any	76 (98.7)	148 (98.0)	16 (100.0)	27 (100.0)					92 (98.9) 175 (98.3)
abacavir	15 (19.5)	36 (23.8)	4 (25.0)	6 (22.2)					19 (20.4) 42 (23.6)
stavudine	32 (41.6)	62 (41.1)	6 (37.5)	5 (18.5)					38 (40.9) 67 (37.6)
didanosine	19 (24.7)	36 (23.8)	1 (6.3)	3 (11.1)					20 (21.5) 39 (21.9)
lamivudine	58 (75.3)	108 (71.5)	13 (81.3)	21 (77.8)					71 (76.3) 129 (72.5)
tenofovir	3 (3.9)	7 (4.6)	3 (18.8)	8 (29.6)					6 (6.5) 15 (8.4)
zidovudine	34 (44.2)	59 (39.1)	4 (25.0)	12 (44.4)					38 (40.9) 71 (39.9)
Other NRTI	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)					0 (0.0) 0 (0.0)
NNRTI, any	36 (46.8)	72 (47.7)	7 (43.8)	12 (44.4)					43 (46.2) 84 (47.2)
efavirenz	17 (22.1)	48 (31.8)	2 (12.5)	5 (18.5)					19 (20.4) 53 (29.8)
nevirapine	19 (24.7)	28 (18.5)	5 (31.3)	6 (22.2)					24 (25.8) 34 (19.1)
Other NNRTI	0 (0.0)	1 (0.7)	0 (0.0)	1 (3.7)					0 (0.0) 2 (1.1)
PI, any	46 (59.7)	76 (50.3)	8 (50.0)	12 (44.4)					54 (58.1) 88 (49.4)
indinavir	12 (15.6)	24 (15.9)	0 (0.0)	1 (3.7)					12 (12.9) 25 (14.0)
lopinavir	12 (15.6)	16 (10.6)	3 (18.8)	6 (22.2)					15 (16.1) 22 (12.4)
nelfinavir	11 (14.3)	21 (13.9)	3 (18.8)	1 (3.7)					14 (15.1) 22 (12.4)
ritonavir	22 (28.6)	36 (23.8)	3 (18.8)	10 (37.0)					25 (26.9) 46 (25.8)

saquinavir	12 (15.6)	17 (11.3)	0 (0.0)	2 (7.4)			12 (12.9)	19 (10.7)
Other PI	1 (1.3)	1 (0.7)	2 (12.5)	0 (0.0)			3 (3.2)	1 (0.6)

[a] Of those on ART at entry
ART=antiretroviral treatment, NRTI=nucleoside reverse transcriptase inhibitor, NNRTI=non-nucleoside reverse transcriptase inhibitor, PI=protease inhibitor

Table 2: Summaries of significant associations between protein levels from the CVD related panels and CVD outcomes for linear and logistic regression models.

	Linear Model				Logistic Model			
	Slope	SE	*p-value*	95% CI	OR	SE	*P-value*	95% CI
Cardiometabolic Panel								
CCL18	0.08	0.023	0.001	(0.034, 0.126)	1.445	0.163	0.001	(1.158, 1.803)
CD59	0.064	0.028	0.025	(0.008, 0.119)	1.326	0.176	0.034	(1.022, 1.721)
CDH1	0.126	0.052	0.017	(0.023, 0.229)	1.773	0.436	0.02	(1.095, 2.873)
CST3	0.108	0.049	0.027	(0.012, 0.203)	1.661	0.395	0.033	(1.042, 2.648)
GAS6	0.12	0.055	0.029	(0.012, 0.228)	1.746	0.459	0.034	(1.044, 2.922)
IGLC2	0.125	0.04	0.002	(0.046, 0.204)	1.818	0.373	0.004	(1.216, 2.719)
PLA2G7	0.142	0.056	0.011	(0.032, 0.251)	1.943	0.514	0.012	(1.157, 3.264)
PRSS2	0.097	0.041	0.017	(0.017, 0.178)	1.572	0.314	0.024	(1.063, 2.326)
Cardiovascular Panel II								
ACE2	0.05	0.023	0.03	(0.005, 0.095)	1.261	0.134	0.029	(1.024, 1.554)
ADAMTS13	-0.116	0.059	0.049	(-0.231, 0)	0.304	0.185	0.05	(0.092, 1.002)
ADM	0.127	0.044	0.004	(0.04, 0.214)	1.867	0.445	0.009	(1.17, 2.979)
CTSL1	0.142	0.045	0.002	(0.053, 0.23)	1.935	0.412	0.002	(1.275, 2.937)
FGF21	0.031	0.013	0.023	(0.004, 0.057)	1.154	0.073	0.023	(1.02, 1.307)
GAL9	0.193	0.062	0.002	(0.071, 0.316)	2.569	0.806	0.003	(1.39, 4.751)
IL1RL2	-0.07	0.031	0.022	(-0.13, -0.01)	0.676	0.127	0.036	(0.468, 0.975)
IL6	0.099	0.027	<0.001	(0.047, 0.152)	1.593	0.214	0.001	(1.223, 2.073)
KIM1	0.075	0.023	0.001	(0.029, 0.12)	1.417	0.162	0.002	(1.133, 1.772)
PGF	0.14	0.051	0.006	(0.041, 0.24)	1.937	0.485	0.008	(1.185, 3.164)
PRSS8	0.182	0.059	0.002	(0.067, 0.297)	2.459	0.745	0.003	(1.358, 4.451)
SPON2	0.315	0.094	0.001	(0.13, 0.499)	4.914	2.506	0.002	(1.808, 13.352)
TRAILR2	0.113	0.039	0.003	(0.037, 0.188)	1.797	0.52	0.043	(1.019, 3.169)

VEGFD	0.108	0.054	0.048	(0.001, 0.214)	1.629	0.421	0.059	(0.982, 2.704)
Cardiovascular Panel III								
AXL	0.119	0.057	0.035	(0.008, 0.23)	1.756	0.471	0.036	(1.038, 2.971)
CCL15	0.085	0.043	0.046	(0.001, 0.169)	1.5	0.304	0.046	(1.008, 2.231)
CHI3L1	0.046	0.022	0.037	(0.003, 0.09)	1.235	0.126	0.038	(1.011, 1.508)
CHIT1	0.034	0.013	0.009	(0.008, 0.06)	1.196	0.099	0.031	(1.017, 1.407)
CPA1	0.076	0.032	0.02	(0.012, 0.139)	1.419	0.221	0.025	(1.045, 1.926)
CPB1	0.071	0.032	0.026	(0.009, 0.133)	1.385	0.207	0.029	(1.034, 1.856)
CSTB	0.073	0.034	0.032	(0.006, 0.14)	1.412	0.228	0.033	(1.029, 1.937)
CTSD	0.088	0.039	0.026	(0.011, 0.165)	1.518	0.281	0.024	(1.056, 2.182)
CTSZ	0.139	0.057	0.015	(0.027, 0.252)	1.92	0.53	0.018	(1.117, 3.298)
CXCL16	0.18	0.064	0.005	(0.055, 0.306)	2.347	0.736	0.006	(1.27, 4.338)
EPHB4	0.196	0.069	0.004	(0.061, 0.331)	2.534	0.893	0.008	(1.27, 5.057)
FAS	0.138	0.069	0.046	(0.003, 0.273)	1.906	0.627	0.05	(1, 3.633)
GAL4	0.108	0.037	0.003	(0.035, 0.18)	1.649	0.295	0.005	(1.162, 2.342)
GDF15	0.078	0.024	0.001	(0.031, 0.126)	1.447	0.172	0.002	(1.147, 1.826)
GRN	0.174	0.069	0.012	(0.038, 0.309)	2.29	0.79	0.016	(1.165, 4.501)
IGFBP7	0.19	0.046	<0.001	(0.099, 0.281)	2.467	0.621	<0.001	(1.507, 4.04)
IL18BP	0.184	0.055	0.001	(0.075, 0.292)	2.396	0.675	0.002	(1.379, 4.161)
IL2RA	0.155	0.043	<0.001	(0.07, 0.24)	2.104	0.469	0.001	(1.36, 3.256)
IL6RA	0.118	0.06	0.049	(0, 0.236)	1.757	0.508	0.051	(0.997, 3.096)
LTBR	0.201	0.056	<0.001	(0.09, 0.311)	2.629	0.819	0.002	(1.428, 4.84)
MCP1	0.141	0.055	0.01	(0.033, 0.25)	1.96	0.522	0.011	(1.164, 3.302)
OPG	0.175	0.064	0.006	(0.05, 0.3)	2.29	0.724	0.009	(1.233, 4.255)
PCSK9	0.143	0.055	0.009	(0.035, 0.25)	1.962	0.525	0.012	(1.161, 3.315)
PLC	0.137	0.057	0.016	(0.025, 0.248)	1.917	0.525	0.017	(1.121, 3.28)
RARRES2	0.15	0.058	0.01	(0.036, 0.264)	2.087	0.625	0.014	(1.16, 3.755)
SCGB3A2	0.066	0.023	0.003	(0.022, 0.111)	1.358	0.145	0.004	(1.102, 1.673)
ST2	0.087	0.037	0.019	(0.014, 0.16)	1.5	0.274	0.027	(1.048, 2.146)
TFF3	0.13	0.053	0.015	(0.026, 0.234)	1.862	0.509	0.023	(1.09, 3.181)
TNFR1	0.179	0.049	<0.001	(0.083, 0.276)	2.385	0.635	0.001	(1.416, 4.018)
TNFR2	0.149	0.046	0.001	(0.059, 0.239)	2.026	0.475	0.003	(1.279, 3.207)
TNFSF13B	0.149	0.05	0.003	(0.051, 0.246)	2.023	0.5	0.004	(1.245, 3.285)
UPAR	0.133	0.051	0.01	(0.032, 0.233)	1.862	0.472	0.014	(1.133, 3.062)

Table 3: Summaries of significant associations between protein levels from the CVD related panels and CVD outcomes for linear and logistic regression models.

	Linear Model				Logistic Model			
	Slope	SE	*p*-value	95% CI	OR	SE	*p*-value	95% CI

Immune Response Panel								
AREG	0.139	0.044	0.001	(0.054, 0.225)	1.922	0.404	0.002	(1.272, 2.903)
CCL11	0.139	0.047	0.003	(0.047, 0.231)	1.927	0.443	0.004	(1.227, 3.025)
CD83	0.117	0.053	0.027	(0.013, 0.222)	1.749	0.45	0.03	(1.056, 2.896)
CKAP4	0.165	0.044	<0.001	(0.079, 0.252)	2.168	0.509	0.001	(1.369, 3.435)
CLEC4C	0.106	0.045	0.02	(0.017, 0.195)	1.67	0.372	0.021	(1.079, 2.584)
CLEC4D	0.078	0.04	0.047	(0.001, 0.156)	1.437	0.261	0.046	(1.006, 2.052)
CLEC6A	0.134	0.031	<0.001	(0.072, 0.195)	2.027	0.38	<0.001	(1.404, 2.928)
DPP10	0.091	0.045	0.045	(0.002, 0.179)	1.501	0.313	0.051	(0.998, 2.259)
FAM3B	0.159	0.044	<0.001	(0.073, 0.245)	2.143	0.481	0.001	(1.381, 3.325)
HNMT	0.076	0.03	0.013	(0.016, 0.135)	1.423	0.199	0.012	(1.082, 1.872)
IL6	0.119	0.027	<0.001	(0.066, 0.171)	1.754	0.238	<0.001	(1.344, 2.288)
ITM2A	0.075	0.033	0.025	(0.009, 0.141)	1.417	0.228	0.03	(1.035, 1.942)
KRT19	0.078	0.027	0.004	(0.024, 0.131)	1.431	0.183	0.005	(1.114, 1.837)
LAMP3	0.086	0.03	0.004	(0.027, 0.144)	1.501	0.22	0.006	(1.126, 2.002)
LILRB4	0.086	0.038	0.023	(0.012, 0.159)	1.493	0.267	0.025	(1.051, 2.121)
Inflammation	Panel							
CCL11	0.13	0.044	0.003	(0.044, 0.216)	1.848	0.399	0.004	(1.211, 2.821)
CCL20	0.066	0.02	0.001	(0.027, 0.105)	1.358	0.128	0.001	(1.13, 1.633)
CCL25	0.098	0.03	0.001	(0.038, 0.157)	1.565	0.228	0.002	(1.175, 2.083)
CDCP1	0.075	0.033	0.023	(0.01, 0.139)	1.411	0.214	0.023	(1.048, 1.899)
CSF1	0.167	0.062	0.007	(0.045, 0.289)	2.653	1.057	0.014	(1.215, 5.791)
CST5	0.108	0.043	0.012	(0.023, 0.192)	1.674	0.346	0.013	(1.116, 2.511)
CX3CL1	0.132	0.047	0.005	(0.039, 0.225)	1.852	0.419	0.006	(1.188, 2.885)
FGF19	0.06	0.026	0.019	(0.01, 0.111)	1.329	0.165	0.022	(1.041, 1.696)
FGF21	0.043	0.016	0.006	(0.012, 0.073)	1.222	0.091	0.007	(1.057, 1.413)
FGF23	0.097	0.036	0.008	(0.026, 0.169)	1.574	0.3	0.018	(1.082, 2.288)
FLT3L	0.164	0.05	0.001	(0.066, 0.263)	2.224	0.564	0.002	(1.352, 3.656)
HGF	0.207	0.036	<0.001	(0.135, 0.278)	2.777	0.603	<0.001	(1.815, 4.249)
IL10	0.067	0.03	0.025	(0.009, 0.126)	1.365	0.196	0.03	(1.031, 1.808)
IL10RB	0.253	0.061	<0.001	(0.133, 0.373)	3.344	1.055	<0.001	(1.803, 6.205)
IL17C	0.084	0.028	0.003	(0.029, 0.14)	1.501	0.218	0.005	(1.129, 1.995)
IL6.2	0.127	0.028	<0.001	(0.073, 0.181)	1.831	0.262	<0.001	(1.383, 2.424)
IL8	0.075	0.028	0.007	(0.02, 0.13)	1.414	0.19	0.01	(1.086, 1.84)
LIFR	0.224	0.069	0.001	(0.088, 0.36)	2.933	1.072	0.003	(1.433, 6.003)
MCP1	0.147	0.049	0.003	(0.051, 0.243)	2.027	0.493	0.004	(1.259, 3.266)
MMP10	0.123	0.037	0.001	(0.052, 0.195)	1.78	0.325	0.002	(1.245, 2.544)
OPG	0.199	0.062	0.001	(0.078, 0.321)	2.595	0.808	0.002	(1.409, 4.777)
TNFRSF9	0.085	0.042	0.041	(0.003, 0.167)	1.493	0.297	0.044	(1.011, 2.204)
UPA	0.154	0.059	0.009	(0.039, 0.269)	2.058	0.579	0.01	(1.185, 3.572)
VEGFA	0.095	0.043	0.027	(0.011, 0.18)	1.545	0.308	0.029	(1.045, 2.283)

Table 4: Summaries of Mendelian randomization tests between protein levels with the CVD composite outcome.

	Estimate	SE	*p*-value	95% CI
ADAMTS13	1.286	1.722	0.455	(-2.088, 4.661)
AXL	1.628	0.704	0.021	(0.248, 3.008)
CCL18	0.182	0.174	0.295	(-0.159, 0.523)
CHI3L1	-0.150	0.075	0.045	(-0.296, -0.003)
CHIT1	-0.160	1.367	0.907	(-2.839, 2.519)
CLEC4C	-0.014	0.303	0.964	(-0.608, 0.581)
CLEC4D	-0.247	0.231	0.286	(-0.7, 0.207)
CLEC6A	0.424	0.230	0.065	(-0.027, 0.875)
CPA1	0.266	0.367	0.469	(-0.454, 0.986)
CPB1	0.004	0.493	0.993	(-0.962, 0.97)
CST3	-0.480	0.345	0.164	(-1.155, 0.196)
CSTB	0.008	0.334	0.981	(-0.647, 0.663)
CTSD	-0.268	0.259	0.301	(-0.777, 0.24)
FAM3B	0.722	0.900	0.422	(-1.042, 2.486)
FAS	-0.288	0.586	0.623	(-1.437, 0.86)
GAL9	-1.018	0.852	0.232	(-2.688, 0.652)
GAS6	-0.766	0.316	0.015	(-1.386, -0.147)
GRN	1.031	0.846	0.223	(-0.626, 2.689)
HNMT	0.211	0.117	0.072	(-0.019, 0.441)
IGFBP7	-0.121	0.417	0.771	(-0.939, 0.696)
IL10	0.243	0.396	0.539	(-0.533, 1.018)
IL2RA	0.567	0.315	0.072	(-0.051, 1.186)
IL6RA	0.177	0.090	0.049	(0.001, 0.353)
LAMP3	-0.109	0.227	0.630	(-0.554, 0.336)
LILRB4	1.463	2.023	0.470	(-2.502, 5.427)
LTBR	0.336	0.477	0.481	(-0.599, 1.272)
PCSK9	0.172	0.389	0.658	(-0.59, 0.935)
RARRES2	-0.792	0.818	0.333	(-2.395, 0.81)
SCGB3A2	0.397	0.191	0.038	(0.023, 0.771)

Figure Legend

Figure 1: Schematic representation of analysis pipeline. All 5 panels had 92 proteins and the number of proteins passing quality control in each panel is presented on the far left. The next column of boxes from the left shows the number of proteins that differ between cases and controls at a significance level of 0.05 and the number of these proteins that have at least 1 variant. On the far right, information on quality control of genetic variants is presented. Following the description of overall quality control, the number of variants in a protein that differs between cases and controls for a panel is provided. For example, there are 8 proteins in the cardiometabolic panel that differ and 6 of these proteins have at least 1 variant. Across these 6 proteins there are 205 variants in their coding region for a mean of 34 variants per protein. The middle column displays the number of proteins from each panel that have SNPs that are significantly associated with protein levels: these are the proteins for which a Mendelian randomization test was conducted. The protein IL10 appears in 2 panels (immune response and inflammation) hence the number of MR candidates in this Figure exceeds the number of rows in Table 4 by 1. Duplication of proteins across panels and differences in analytical results across the duplicated proteins also creates differences between the results in Tables 2 and 3 and this Figure (for the cardiovascular III, immune response, and inflammation panels).