



Haas, M. E., Aragam, K. G., Emdin, C., Bick, A. G., International Consortium for Blood Pressure Genome-Wide Association Studies, Hemani, G., ... Kathiresan, S. (2018). Genetic association of albuminuria with cardiometabolic disease and blood pressure. *American Journal of Human Genetics*, 461-473. <https://doi.org/10.1016/j.ajhg.2018.08.004>

Peer reviewed version

License (if available):  
CC BY-NC-ND

Link to published version (if available):  
[10.1016/j.ajhg.2018.08.004](https://doi.org/10.1016/j.ajhg.2018.08.004)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Cell Press at <https://www.sciencedirect.com/science/article/pii/S0002929718302702> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms>

## **Genetic association of albuminuria with cardiometabolic disease and blood pressure**

Mary E. Haas,<sup>1,2</sup> Krishna G. Aragam,<sup>1,2,3,4</sup> Connor A. Emdin,<sup>1,2,4</sup> Alexander G. Bick,<sup>1,2,4</sup> International Consortium for Blood Pressure, Gibran Hemani,<sup>5</sup> George Davey Smith,<sup>5</sup> and Sekar Kathiresan<sup>1,2,3,4\*</sup>

<sup>1</sup> Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge MA 02139, USA.

<sup>2</sup> Center for Genomic Medicine and <sup>3</sup> Cardiovascular Research Center, Massachusetts General Hospital, Boston MA 02114, USA.

<sup>4</sup> Department of Medicine, Harvard Medical School, Boston MA 02114, USA.

<sup>5</sup> Medical Research Council Integrative Epidemiology Unit, Population Health Sciences, University of Bristol, Bristol, BS8 2BN, UK

\* Correspondence to:  
Sekar Kathiresan, MD  
Center for Genomic Medicine  
Massachusetts General Hospital  
185 Cambridge Street, CPZN 5.821A  
Boston, MA 02114 USA  
Email: [skathiresan1@mgh.harvard.edu](mailto:skathiresan1@mgh.harvard.edu)  
Phone: 617-724-3091

## **Abstract**

Excretion of albumin in urine, or albuminuria, is associated with the development of multiple cardiovascular and metabolic diseases. However, whether pathways leading to albuminuria are causal for cardiometabolic diseases is unclear. We addressed this question using a Mendelian randomization framework in the UK Biobank, a large population-based cohort. We first performed a genome-wide association study for albuminuria in 382,500 individuals and identified 32 new albuminuria loci. We constructed albuminuria genetic risk scores and tested for association with cardiometabolic diseases. Genetically-elevated albuminuria was strongly associated with increased risk of hypertension (1.38 OR; 95%CI, 1.27-1.50 per 1-SD predicted increase in albuminuria,  $p = 7.01 \times 10^{-14}$ ). We then examined bidirectional associations of albuminuria with blood pressure which suggested that genetically-elevated albuminuria led to higher blood pressure (2.16 mmHg systolic blood pressure; 95%CI, 1.51-2.82 per 1-SD predicted increase in albuminuria,  $p = 1.22 \times 10^{-10}$ ), and that genetically-elevated blood pressure led to more albuminuria (0.005 SD; 95%CI 0.004-0.006 per 1-mmHg predicted increase in systolic blood pressure,  $p = 2.45 \times 10^{-13}$ ). These results support the existence of a feed-forward loop between albuminuria and blood pressure, and imply that albuminuria could increase risk of cardiovascular disease through blood pressure. Moreover, they suggest therapies that target albuminuria-increasing processes could have antihypertensive effects which are amplified through inhibition of this feed-forward loop.

## **Introduction**

In observational epidemiologic studies, albuminuria, or the concentration of albumin excreted in urine, is associated with risk for multiple cardiometabolic diseases: elevations in albuminuria predict development of coronary artery disease, stroke, heart failure, type 2 diabetes, hypertension and all-cause mortality.<sup>1-9</sup> However, whether pathways leading to albuminuria are causally associated with cardiometabolic disease is unclear. Therapies lowering albuminuria are generally associated with reduced cardiovascular disease, for example. However, whether such effects are independent of concomitant reductions in blood pressure is ambiguous.<sup>10-14</sup> Understanding whether associations of albuminuria pathways with disease reflect a

causal relationship or mere correlation may inform whether targeting novel albuminuria-increasing processes could reduce risk for cardiometabolic diseases.

'Mendelian randomization' can provide evidence regarding the hypothesis that a given biomarker-disease relationship is causal.<sup>15</sup> The strengths and limitations of Mendelian randomization can be considered via analogy with a randomized clinical trial. Individuals are assigned to lifelong increase or decrease in a disease risk factor due to the random segregation and independent assortment of genetic polymorphisms at conception, thus minimizing two key limitations of observational epidemiology - reverse causation and confounding. The effect of genetically modifying an exposure (here, albuminuria) can then be tested against increasing or decreasing risk of an outcome (here, cardiometabolic disease). Three assumptions must be met in order for a genetic variant to be a potentially valid instrumental variable in Mendelian randomization: 1) the variant must be strongly associated with the exposure; 2) the variant must not be associated with confounders; and 3) the variant must not be horizontally pleiotropic, i.e., cannot be associated with the outcome independent of the exposure pathway.<sup>16</sup> While the second and third assumptions are hard to prove, many sensitivity analyses have been developed to improve the reliability of Mendelian randomization estimates.<sup>17</sup>

Here, we first identify new genetic variants for albuminuria by conducting a genome-wide association study of albuminuria in 382,500 individuals in the UK Biobank. We subsequently utilized the identified genetic variants as instruments in a Mendelian randomization analysis to test the hypothesis that pathways increasing albuminuria are causal for cardiometabolic diseases.

## **Subjects and Methods**

### **Study Design**

This study had three main components: first, we examined epidemiological associations of baseline albuminuria with incident cardiometabolic disease in UK Biobank. Second, we conducted a genome-wide association study of baseline albuminuria and constructed a polygenic risk score. Finally, we performed a

Mendelian randomization study to test the hypothesis that the associations between processes leading to albuminuria and cardiometabolic diseases are causal.

## **Study Participants**

### *UK Biobank*

Data from 382,500 unrelated individuals of European ancestry with albuminuria measurement in the UK Biobank were used. Samples were excluded for the following reasons: inferred sex did not match reported sex, kinship was not inferred, putative sex chromosome aneuploidy, consent withdrawn, or excessive heterozygosity or missingness, based on centralized sample quality control performed by UK Biobank.<sup>18</sup> Excluded related individuals were defined as one individual in each pair with *KING* coefficient > 0.0884, indicating 2<sup>nd</sup> degree or closer relatedness. European ancestry was determined by self-reported ancestry of British, Irish or Other White, followed by outlier detection using the *R* package *aberrant* with lambda = 40 on genetic principal component (PC)1 and PC2, PC3 and PC4, and PC5 and PC6. Individuals who were outliers for any of the three pairs of PCs were removed from the European ancestry group. Kinship inference and genetic PCs were centrally calculated by UK Biobank.<sup>18</sup>

### Albuminuria and Blood Pressure

Albuminuria was measured at the initial assessment visit (2006-2010); a Beckman Coulter AU5400 clinical chemistry analyzer was used to quantify urine albumin (df-30500, Randox Bioscience, UK; immunoturbidimetric assay, detection range 6.7 – 200 mg/L) and urine creatinine (df-30510, Beckman Coulter (UK), Ltd; enzymatic assay, detection range 88 – 44200  $\mu$ mol/L) concentrations. Urine albumin concentrations below the lower limit of detection (df-30505, n=263654) were set to the lower limit of detection (6.7 mg/L). The resulting urine albumin:creatinine ratio (ACR, mg/g) was natural log-transformed to adjust for right skewedness. Microalbuminuria was defined as urine ACR of 25-355 mg/g in females and 17- 250 mg/g in males; macroalbuminuria >355 mg/g in females and > 250 mg/g in males.<sup>19</sup> Baseline blood pressure was averaged from two measurements taken a few moments apart using an Omron 705 IT electronic blood pressure monitor (df-4079 and df-4080). A sphygmomanometer (df-93 and df-94) was used

if a measurement could not be obtained with the electronic monitor. 381,833 individuals had both blood pressure and albuminuria measurements. Albuminuria and blood pressure can both be decreased by hypertensive medication, but there is no consensus about the magnitude of such effects on albuminuria. Therefore, neither variable was corrected for hypertensive medication use so as not to selectively skew one variable but not the other.

#### Disease Definitions

Prevalent cardiometabolic diseases were defined at study entry through the electronic health record and/or self-report with confirmation via verbal interview by a trained nurse. Detailed definitions for all disease classifications can be found in Table S2. Incident cardiometabolic diseases were ascertained among those not meeting disease criteria at baseline by applying phenotype definitions to longitudinal, in-patient hospital and death registry data linked to the UK Biobank. Participants were censored at the time of disease diagnosis, date of death, or date of last follow-up (i.e., February 9, 2016 for participants enrolled in Wales, February 16, 2016 for participants enrolled in England, and October 31, 2015 for participants enrolled in Scotland), whichever occurred first. Participants were presumed alive at last follow-up if there was no preceding report of death in the death register. UK Biobank was approved by the Research Ethics Committee (reference 16/NW/0274) and informed consent was obtained from all participants. Analysis of UK Biobank data was approved by the Partners HealthCare institutional review board (protocol 2013P001840).

#### *Atherosclerosis Risk in Communities (ARIC)*

10235 unrelated individuals in the Atherosclerosis Risk in Communities study, genotyped using the Affymetrix Genome-wide Human SNP Array 5.0, were imputed to the Haplotype Reference Consortium using the Michigan Imputation Server. Phasing was performed using the Eagle2 algorithm. 4954 variants were removed prior to imputation due to duplication, monomorphism or allele mismatch. Imputation was then performed on 799246 variants using the minimac3 algorithm. 39,235,157 variants in the Haplotype

Reference Consortium were imputed. 6398 individuals were of European ancestry as confirmed by centrally-calculated, European-specific PC analysis and had albuminuria data.

An untimed urine sample was collected during the visit 4 clinical examination. Aliquots were frozen within 12 hours and stored at  $-70^{\circ}\text{C}$ . Albumin and creatinine levels were measured in the University of Minnesota Physicians Outreach Laboratories, Minneapolis, Minnesota, with albumin by a nephelometric method either on the Dade Behring BN100 (assay sensitivity, 2.0 mg/L) or on the Beckman Image Nephelometer, and creatinine using the Jaffe method in order to determine the albumin-to-creatinine ratio (ACR;  $\mu\text{g}/\text{mg}$ ) for participants. Blinded samples ( $n=516$ ) analyzed for quality assurance showed a correlation coefficient ( $r$ ) of the  $\log_e$ -transformed ACR as  $r=0.95$ . ACR was natural log-transformed for association analysis. Genotype and phenotype data were retrieved for analysis from NCBI dbGAP (phs000280.v3.p1) under procedures approved by the Partners HealthCare institutional review board (protocol 2016P002395).

#### *Framingham Heart Study*

8825 individuals from the Offspring and Third Generation cohorts of the Framingham Heart Study, genotyped using the Affymetrix GeneChip Human Mapping 500K Array, were imputed to the Haplotype Reference Consortium using the Michigan Imputation Server. Genetic PCs were calculated on directly-genotyped data using *EIGENSOFT v7.2.1* after removing variants with  $\text{MAF} < 0.01$ , genotype call rate  $< 0.99$  and samples with sample call rate  $< 0.97$  using *PLINK-1.9*. PCs were calculated in unrelated individuals only based on self-reported pedigree, and projected onto related individuals. 6534 individuals were of self-reported white ancestry confirmed by PC analysis. PCs used as covariates for association tests were recalculated in the white subgroup, and 21 individuals were removed as outliers on the basis of this analysis. 6387 of the remaining individuals had albuminuria data available.

Albuminuria was measured at Offspring Exam 8 and Third Generation Exam 1 visits at the Framingham Heart Study Laboratory using a Roche Hitachi 911 Chemistry Analyzer. Urine albumin was quantified by the immunoturbidometric Tina-quant Albumin test (assay sensitivity, 3.0 mg/L); urine creatinine by

colorimetric, modified Jaffe (rate blanked) creatinine test (assay sensitivity, 0.2 mg/100ml). Urine albumin concentrations below the lower limit of detection (n=1682) were set to the lower limit of detection (3.0 mg/L). The resulting urine albumin:creatinine ratio (ACR, mg/g) was natural log-transformed for association analysis. Genotype and phenotype data were retrieved for analysis from NCBI dbGAP (phs000007.v26.p10) under procedures approved by the Partners HealthCare institutional review board (protocol 2016P002395).

#### *International Consortium for Blood Pressure*

The International Consortium for Blood Pressure is a large meta-analysis of study-specific results associating blood pressure with genotypes from the Cardio-MetaboChip SNP Array ( $n_{\max}=201529$ )<sup>20</sup> or imputed to 1000 Genomes Project Phase 1 haplotypes ( $n_{\max}=150134$ ).<sup>21</sup> Summary statistics for SNP effects on systolic and diastolic blood pressure were corrected for anti-hypertensive medication use (+15 mmHg and +10 mmHg for systolic and diastolic blood pressure, respectively) and included body mass index, sex, age, age<sup>2</sup> as covariates, as previously described.<sup>20,21</sup> One limitation of this data set is that adjustment for body mass index and anti-hypertensive medication may lead to associations between genetic variants and adjusted blood pressure being confounded with other factors that influence the adjustment variables (“collider effects”). This could bias Mendelian randomization analyses with albuminuria, which is not adjusted for these factors.<sup>22</sup> SNPs from the Cardio-MetaboChip study<sup>20</sup> were used to construct blood pressure genetic risk scores, whereas association of albuminuria variants with blood pressure was examined using blood pressure effects measured in the 1000G-based study.<sup>21</sup>

## **Statistical Analyses**

### *Observational Epidemiology*

In UK Biobank, Cox proportional hazards regression was used to determine the association of baseline albuminuria with incident cardiometabolic disease (average median follow-up time 7.0 years across diseases). Potential confounding variables were selected per prior epidemiological analyses of albuminuria and cardiometabolic disease,<sup>2,5,6,9,23-25</sup> and a subsequent univariate screen of the selected traits in the UK



Biobank was performed; these criteria yielded age at baseline, sex, current smoking status, body mass index, systolic blood pressure, diastolic blood pressure, baseline diabetes and baseline hyperlipidemia as covariates for inclusion in most Cox proportional hazard models.

### *Genome-Wide Association Study*

UK Biobank samples were genotyped by Affymetrix using either the UK BiLEVE or UK Biobank Axiom arrays. Genotyped variants were then imputed by the UK Biobank central analysis team onto the Haplotype Reference Consortium reference panel.<sup>18</sup> Variant exclusion criteria were Hardy-Weinberg equilibrium  $p \leq 1 \times 10^{-20}$ , QCTOOL INFOscore  $< 0.3$ , variant call rate  $\leq 0.95$  and MAF  $\leq 0.001$  yielding 11,709,857 variants in the analysis. Sex-specific residuals of natural log-transformed urine ACR were analyzed as a continuous trait with age, genotyping array, and the first 10 genetic PCs as covariates via least-squares linear regression under an additive effects model using *Hail v0.1* statistical software (Web Resources).<sup>26</sup> The threshold for statistical significance was empirically determined using permutation testing according to a previous approach.<sup>27</sup> Association of chromosome 21 variants with 1000 random simulated continuous phenotypes were determined using *Hail v0.1*. The necessary significance threshold for a 5% family-wise error rate (FWER) was empirically estimated as the 5<sup>th</sup> percentile of the collection of the minimum variant p-value from each simulated phenotype. The corresponding number of independent tests on chromosome 21 was calculated as  $p = 0.05/\text{threshold}_{5\%FWER}$ , and was scaled to genome-wide using the proportion of the 11,709,857 genome-wide variants located on chromosome 21. This resulted in a genome-wide significance threshold of  $p < 9 \times 10^{-9}$ . Genomic inflation was calculated using the median estimator in the *GenABEL* package in *R*; LD score regression and common (MAF  $> 5\%$ ) SNP genetic correlation and heritability were calculated via *LDSC v1.0.0* using standard variant filtering (MAF  $> 0.01$  & INFOscore  $> 0.9$ ), HapMap3 SNPs and LD scores precomputed from European 1000 Genomes data.<sup>28</sup> Variants were clumped into independent loci using *PLINK-1.9* with  $R^2 > 0.01$  and  $< 1$  MB from the index variant (smallest p-value).

### *Albuminuria Genetic Risk Score*

Up to forty-six SNPs independently associated with albuminuria at a conventional  $p < 5 \times 10^{-8}$  threshold in the genome-wide association study (Table S4) were used to construct weighted polygenic risk scores using *PLINK 2.00a2LM*. Each imputed genotype dosage was multiplied by the effect of the SNP on natural log-transformed urine ACR normalized to 1-SD albuminuria in UK Biobank (0.755 log(mg/g) urine ACR). The resulting weighted dosages were summed to create genetic risk scores. Association of the 46-SNP albuminuria genetic risk score with albuminuria in ARIC and Framingham Heart Study was determined using linear regression with age, sex and the first 10 genetic PCs as covariates. Sensitivity analysis excluding 1 and 10 poorly-imputed variants in ARIC and Framingham Heart Study, respectively, did not substantially affect association results. Variance explained by each score was calculated as the adjusted  $R^2$  from the association of albuminuria with the albuminuria genetic risk score, age, sex and 10 genetic PCs minus the adjusted  $R^2$  from the association of albuminuria with age, sex, and 10 genetic PCs.

#### *Blood Pressure Genetic Risk Scores*

Lead variants of genome-wide significant loci from Cardio-MetaboChip-based ICBP stage 4 meta-analysis<sup>20</sup> were used to construct systolic blood pressure and diastolic blood pressure genetic risk scores. These results did not include UK Biobank. Only variants significantly ( $p < 5 \times 10^{-8}$ ) associated with a specific blood pressure trait were included in that trait's score. rs10164833 was excluded from the systolic blood pressure risk score as it did not replicate in further ICBP meta-analysis. In UK Biobank, each imputed genotype dosage was multiplied by the effect of the SNP on mmHg systolic or diastolic blood pressures from ICBP stage 4 meta-analysis, which were corrected for hypertensive medication use and body mass index,<sup>20</sup> and the resulting weighted dosages were summed. Variance explained by each score in UK Biobank was calculated as the adjusted  $R^2$  from the association of blood pressure corrected for hypertensive medication use with the blood pressure genetic risk score, age, sex and 10 genetic PCs minus the adjusted  $R^2$  from the association of blood pressure corrected for hypertensive medication use with age, sex, and 10 genetic PCs.

#### *Mendelian Randomization*

For individual-level data, association of the albuminuria or blood pressure genetic risk scores with outcomes were assessed using logistic (combined prevalent plus incident disease) or linear (continuous outcomes) two-stage least-squares regression in *Stata v15*. Age at baseline, sex, genotyping array and the first 10 genetic PCs to control for population structure were included as covariates. For summary-level data, the analogous approach is an inverse-variance-weighted (IVW) fixed-effects meta-analysis of the effect of each SNP on the outcome divided by the effect of this SNP on albuminuria.<sup>29,30</sup> Meta-analysis was conducted using the *MendelianRandomization* package<sup>31</sup> in *R*. Effect estimates were normalized to 1-SD albuminuria in UK Biobank (0.755 log(mg/g) urine ACR). Power to detect associations with cardiometabolic disease in UK Biobank were calculated using an online tool (Web Resources, Table S13).

### *Sensitivity Analyses*

We performed the following sensitivity analyses to address several limitations of Mendelian randomization: MR Steiger filtering to remove variants potentially acting through reverse causation,<sup>32,33</sup> calculation of heterogeneity and random-effects IVW meta-analysis to allow for variant effect size heterogeneity (Tables S6-S10, S12),<sup>34,35</sup> median regressions which allow up to 50% of information from variants to violate Mendelian randomization assumptions,<sup>36</sup> MR-Egger regression to detect directional pleiotropy,<sup>37</sup> Cook's distance to detect extreme outliers,<sup>38</sup> and unweighted allele scores to minimize bias from internally-derived weights in individual-level analyses.<sup>39</sup>

### MR Steiger Filtering

The third assumption of Mendelian randomization ('no association independent of the exposure') requires that a variant acts first through the exposure and not the outcome. Observational studies suggest that diseases such as diabetes and hypertension can increase albuminuria.<sup>40,41</sup> Some variants may therefore be associated with albuminuria via first increasing risk of such diseases and secondarily increasing albuminuria. These variants should not be included as instruments for testing the influence of albuminuria on those disease outcomes.

The MR Steiger method strengthens evidence regarding whether a variant acts first through the exposure or outcome under a model of vertical pleiotropy, where the SNP associates with two traits because one trait influences the other. The correlation of a variant with an outcome is a product of both the variant-exposure correlation and the exposure-outcome correlation. The variant-exposure correlation should therefore be greater than the variant-outcome correlation.<sup>33</sup> The MR Steiger method determines the variant-exposure and variant-outcome correlations, and removes variants where the variant-outcome correlation is greater than the variant-exposure correlation. The aim of this approach is to reduce the proportion of variants erroneously included in a Mendelian randomization analysis due to confounding or acting first through the outcome.<sup>32,33</sup> We note that MR Steiger is not designed to distinguish between vertical pleiotropy and horizontal pleiotropy, wherein a SNP influences both traits through independent pathways.

To perform MR Steiger filtering, the correlation of a variant with each exposure and outcome was first determined. For continuous traits from studies with individual-level data, the squared correlation of each variant with a continuous exposure or outcome was calculated as the  $R^2$  from association of the trait with the variant and covariates minus the  $R^2$  from association of the trait with covariates. For continuous traits from studies with summary statistics, the correlation  $R$  was estimated using *get\_r\_from\_pn* in the *TwoSampleMR* package<sup>42</sup> in *R*. Correlation  $R$  of each variant with binary traits was estimated on the logit liability scale<sup>33,43</sup> using *get\_r\_from\_lor* in *TwoSampleMR* modified to use allele frequency measured in UK Biobank. To apply directional MR Steiger filtering, variants with  $R^2_{\text{exposure}} < R^2_{\text{outcome}}$  were removed.

As an additional sensitivity analysis for one-sample Mendelian randomizations with individual-level data, the Steiger test of correlated correlations was used to calculate the probability that the variant-exposure and variant-outcome correlations were different. Variants whose correlation with the exposure was not significantly different from correlation with the outcome, defined as Steiger p-value  $> 0.05$ , were removed. Steiger tests were calculated using the *r.test* in the *psych* package in *R*. This analysis was not performed for two-sample Mendelian randomizations, as correlations of a variant with an exposure or outcome from

summary statistics are only estimated in separate cohorts and therefore may be less appropriate for detecting significant differences between the two measurements.

#### Other Sensitivity Analyses

IVW random-effects, simple median, weighted median and MR-Egger random-effects regression were calculated with normal distributions using the *MendelianRandomization v0.2.0* package in *R*. For these analyses of individual-level data, associations of score SNPs with each outcome were determined via linear or logistic (Wald) regression using age at baseline, sex, genotyping array and the first 10 genetic PCs using *Hail v0.1* statistical software. While MR-Egger can be particularly biased by weak instruments in individual-level or one-sample Mendelian randomization analyses,<sup>44</sup> the fact that MR-Egger regression results were roughly similar between one-sample and two-sample analyses (Tables S8 and S10) suggests that weak instrument bias is not disproportionately affecting these results. Graphs of each variant's effect on exposure vs outcome and IVW-based leave-one-out analyses (Figures S2-S4, S6-S7) were created using the *TwoSampleMR* package in *R*. Variant effect heterogeneity was assessed via Cochran's Q and MR-PRESSO residual sum of squares (RSS), which shows improved false-positive rates.<sup>34</sup> These were calculated using the *MendelianRandomization* and *MRPRESSO*<sup>34</sup> *v1.0* packages, respectively, in *R*. For outlier detection, Cook's distance was calculated on IVW meta-analysis; SNPs with a Cook's distance greater than twice the nominal outlier cutoff  $4/n_{\text{SNPs}}$  were considered for outlier exclusion. Unweighted allele scores were constructed by summing the number of albuminuria- or blood pressure-increasing alleles per individual. Two-stage least-squares regression was used to determine the association of the unweighted allele score with cardiometabolic outcomes as described above. Linkage disequilibrium between variants in albuminuria genetic risk scores and blood pressure genetic risk scores was defined as  $R^2 > 0.2$  and  $< 1$  MB using linkage disequilibrium calculated in the UK Biobank study population via *PLINK-1.9*.

Anti-hypertensive medications reduce albuminuria in addition to lowering blood pressure.<sup>45</sup> We wanted to determine whether not correcting blood pressure and albuminuria for hypertensive medication use confounded albuminuria-blood pressure association results. As a sensitivity analysis, we therefore

excluded individuals on hypertensive medication and re-tested association of an albuminuria genetic risk score with blood pressure. Hypertensive medication use was defined by self-report with confirmation via verbal interview by a trained nurse (df-6177 and df-6153). A genome-wide association study for albuminuria was performed in 302,687 individuals in UK Biobank not on hypertensive medication and who had blood pressure and albuminuria measurements. This yielded 23 independent loci ( $p < 5 \times 10^{-8}$ ,  $R^2 > 0.01$  and  $< 1$  MB from the index variant, Table S11). Effects of the 23 SNPs were used to construct an albuminuria risk score normalized to 1-SD albuminuria in this population (0.713 log (mg/g)). Associations of this albuminuria risk score with blood pressure were determined as above. Directional MR Steiger filtering removed one variant from the risk score for association with both systolic and diastolic blood pressure.

## **Results**

382,500 unrelated individuals of European ancestry in the UK Biobank, a population-based cohort, were used in this study. 54% of participants were female, and the mean age was 56.9 (SD 7.9) years at baseline. Mean baseline systolic blood pressure and diastolic blood pressures were 138.3 (SD 18.6) and 82.3 (SD 10.1) mmHg, respectively; 18940 (5.0%) individuals had diabetes at baseline and 53004 (13.9%) had hyperlipidemia at baseline. The median baseline urine albumin:creatinine ratio (ACR) was 9.8 mg/g (IQR 6.1-16.5). 14.3% had microalbuminuria and 0.4% macroalbuminuria (Tables S1-S2). Baseline urine ACR was natural log-transformed and is referred to as albuminuria (mean 2.3, SD 0.755 log(mg/g)) in subsequent analyses.

### **Association of Albuminuria with Development of Cardiometabolic Diseases**

In UK Biobank, we first examined the association of baseline albuminuria with risk of incident cardiometabolic diseases using Cox proportional hazard regression (average median follow-up time across all diseases, 7.0 years). Baseline albuminuria was strongly associated with subsequent development of cardiometabolic disease (Figure 1): a 1-SD increase in albuminuria was associated with higher hazard of all-cause mortality (1.22 HR; 95%CI 1.20-1.23), coronary artery disease (1.09 HR; 95%CI 1.07-1.11), stroke (1.16 HR; 95%CI 1.12-1.20), peripheral vascular disease (1.23 HR; 95%CI 1.19-1.27), heart failure (1.29

HR; 95%CI 1.26-1.32), type 2 diabetes (1.20 HR; 95%CI 1.18-1.22), chronic kidney disease (1.51 HR; 95%CI 1.48-1.54) and hypertension (1.22 HR; 95%CI 1.20-1.24) but not with other diseases such as skin cancer (1.00 HR; 95%CI 0.98-1.03), even after adjustment for standard metabolic risk factors. Thus, albuminuria measured in UK Biobank is associated with cardiometabolic disease in a manner consistent with previous observational studies.<sup>1,6,7,9,23</sup>

### **Genome-wide Association Study for Albuminuria**

To identify variants to be used as genetic instruments for albuminuria, we conducted a discovery genome-wide association study of albuminuria in the 382,500 UK Biobank participants. Minimal genomic inflation was observed ( $\lambda_{GC} = 1.17$ , LD score regression intercept<sup>28</sup> = 1.02, Figure S1). The common SNP heritability of albuminuria was 0.045 (SE 0.002). In addition to replicating the previously-reported<sup>46</sup> association at the *CUBN* locus (rs10795433: beta -0.024 log(mg/g) for C allele;  $p = 1.37 \times 10^{-24}$ ), we discovered an additional 1247 genome-wide significant ( $p < 9 \times 10^{-9}$ ) associations representing 32 novel independent loci, for a total of 33 genome-wide significant loci (Figure 2, Table 1). Novel associations of potential clinical interest include the *NR3C2* and *COL4A4* loci. *NR3C2* encodes the mineralocorticoid receptor, and mineralocorticoid receptor antagonists such as spironolactone and eplerenone reduce albuminuria when added to other anti-hypertensive medications.<sup>47,48</sup> Mutations in *COL4A4* and neighboring gene *COL4A3* can cause autosomal Alport Syndrome, which is characterized by kidney disease that can include proteinuria.<sup>49</sup> 22 of the 33 loci (or their proxies  $R^2 > 0.8$ ), were available in a smaller previously published genome-wide association study;<sup>46</sup> of these, 20 had a consistent direction of effect and 7 were nominally significant ( $p < 0.05$ , Table S3).

### **Albuminuria Genetic Instrument Strength**

A 46-SNP genetic risk score constructed from the 33 genome-wide significant loci plus an additional 13 loci meeting a conventional significance level of  $p < 5 \times 10^{-8}$  (Table S4) explained 0.7% of the variance in albuminuria in UK Biobank (F-statistic, 2871). The genetic risk score was validated in two additional North American cohorts of European ancestry and non-inflated estimates were obtained. The 46-SNP score was

associated with albuminuria in both the Atherosclerosis Risk in Communities study ( $n=6398$ ,  $p = 6.7 \times 10^{-5}$ , 0.2% variance in albuminuria explained) and the Framingham Heart Study ( $n=6387$ ,  $p = 4.4 \times 10^{-4}$ , 0.2% variance in albuminuria explained; Table S5).

### **Association of Albuminuria Genetic Risk Score with Cardiometabolic Disease**

We examined whether genetically-elevated albuminuria due to the 46-SNP risk score associated with increased risk of cardiometabolic disease in UK Biobank. Genetic predisposition to elevated albuminuria was associated with increased risk of hypertension (1.51 OR; 95%CI, 1.39-1.64 per 1-SD predicted increase in albuminuria due to the 46-SNP score,  $p = 2.68 \times 10^{-22}$ ). However, no significant associations were observed between the albuminuria genetic risk score and risk of all-cause mortality, coronary artery disease, stroke, heart failure, type 2 diabetes, chronic kidney disease or skin cancer (Figure 3).

To remove variants that may act through reverse causation – that is, influence albuminuria through hypertension – from the albuminuria score, we applied MR Steiger filtering.<sup>32,33</sup> This approach removed three variants more directly associated with hypertension. After filtering, the 43-SNP score was still associated with increased risk of hypertension (1.38 OR; 95%CI, 1.27-1.50 per 1-SD predicted increase in albuminuria,  $p = 7.01 \times 10^{-14}$ , Figure 3).

### **Bidirectional Mendelian Randomization of Albuminuria and Blood Pressure**

To further examine the association of albuminuria with hypertension, we investigated the genetic correlations between albuminuria, blood pressure and hypertension. We found significant common SNP genetic correlations between albuminuria and hypertension ( $r_g = 0.16$ ; SE 0.03,  $p = 2.06 \times 10^{-8}$ ), systolic blood pressure ( $r_g = 0.20$ ; SE 0.03,  $p = 3.6 \times 10^{-15}$ ) or diastolic blood pressure ( $r_g = 0.10$ ; SE 0.03,  $p = 3.2 \times 10^{-4}$ ). We performed bidirectional Mendelian randomization between albuminuria and blood pressure to understand the determinants of these correlations. First, we examined association of the albuminuria genetic risk score with blood pressure outcomes. After MR Steiger filtering, albuminuria genetic risk scores remained associated with increased systolic blood pressure (2.16 mmHg; 95%CI, 1.51-2.82 per 1-SD



predicted increase in albuminuria,  $p = 1.22 \times 10^{-10}$ ) and diastolic blood pressure (0.99 mmHg; 0.61-1.36 per 1-SD predicted increase in albuminuria,  $p = 3.40 \times 10^{-7}$ , Figure 4). Next, we investigated the reverse association – blood pressure affecting albuminuria – using a blood pressure risk score as the exposure and albuminuria as the outcome. 47 and 52 variants significantly associated with blood pressure in ICBP,<sup>20</sup> which did not include UK Biobank, explained 1.2% and 1.4% of the variance in systolic and diastolic blood pressure, respectively, in UK Biobank. Both blood pressure risk scores were associated with elevated albuminuria (0.005 SD albuminuria; 95%CI 0.004-0.006 per 1 mmHg predicted increase in systolic blood pressure,  $p = 2.45 \times 10^{-13}$  and 0.007 SD albuminuria; 95%CI 0.005-0.009 per 1 mmHg predicted increase in diastolic blood pressure,  $p = 1.83 \times 10^{-9}$ , Figure 4), validating a previous suggestive report.<sup>20</sup>

### **Sensitivity Analyses**

Seven sets of sensitivity analyses were used to verify the robustness of the associations between albuminuria and hypertension or blood pressure. First, Mendelian randomization results were consistent for a restricted score at the genome-wide significance level of  $p < 9 \times 10^{-9}$  (Table S6). Second, we used several methods to detect and mitigate the effects of pleiotropic variants: 1) MR Egger regression to detect the presence of directional pleiotropy;<sup>37</sup> 2) Cook's distance to detect outlier variants, which can also indicate pleiotropy;<sup>38,50</sup> 3) leave-one-out analyses to determine if associations are biased by a single, potentially-pleiotropic SNP;<sup>42</sup> 4) median-based regressions, which are robust when up to 50% of information comes from invalid variant instruments, including due to pleiotropy.<sup>36</sup> Some directional pleiotropy was observed in the associations between albuminuria and blood pressure but not other associations (Tables S7-S9). MR-Egger regression is especially sensitive to influential points,<sup>51,52</sup> so the observed directional pleiotropy could be due in part to a potential outlier, rs141640975 in the *CUBN* locus (Cook's distance = 0.6-0.7, Figures S2-S3). This variant was the top SNP in the albuminuria GWAS, raising the possibility that it could derive its large effect via aggregating potentially pleiotropic effects of multiple pathways. Excluding this variant reduced directional pleiotropy while maintaining associations between the albuminuria risk score and blood pressure or hypertension (Tables S7-9). Leave-one-out analyses suggested the observed associations were not biased by other single variants (Figures S2-S4, S6-S7). Notably, albuminuria risk scores also

remained associated with hypertension and blood pressure, and blood pressure risk scores with albuminuria, using one or more forms of median regression that allow for many pleiotropic variants (Tables S7-S9).

Third, to be more confident variants were not acting through reverse causation, we used a more stringent MR Steiger filter. This removed variants that were not significantly more associated with albuminuria than outcomes. While effect estimates were slightly attenuated, the associations of albuminuria with blood pressure and hypertension persisted even after this additional filtering (Tables S7-S8).

Fourth, for bidirectional Mendelian randomization it is important that variants in the albuminuria score are not in linkage disequilibrium with variants in the blood pressure score.<sup>22</sup> One pair of variants at the *HOTTIP* locus was in linkage disequilibrium (rs2023844-rs3735533  $R^2 = 0.99$ ). MR Steiger analysis suggests this variant is more directly associated with blood pressure. It was therefore removed from all albuminuria risk scores by directional MR Steiger filtering. Additionally, sensitivity analyses excluding this variant from blood pressure risk scores did not affect association of blood pressure risk scores with albuminuria (Table S9).

Fifth, using the same samples for both discovery of a genetic risk score and analysis of score effects can bias association results towards the observational estimate.<sup>53</sup> Unweighted genetic risk scores can reduce this bias.<sup>39</sup> Unweighted albuminuria risk scores were also associated with increased risk of hypertension and elevated systolic and diastolic blood pressure in UK Biobank (Tables S7-S8).

Sixth, to further mitigate bias from score discovery-analysis overlap in UK Biobank, we examined the effects of albuminuria-associated variants on blood pressure measured in a separate cohort. Blood pressure effects in ICBP were corrected for hypertensive medication use.<sup>21</sup> In this cohort, albuminuria variants were associated with increased systolic blood pressure (2.69 mmHg; 95%CI, 1.18-4.19 per 1-SD predicted increase in albuminuria,  $p = 4.64 \times 10^{-4}$ ) and nominally with increased diastolic blood pressure (1.03 mmHg; 0.10-1.97 per 1-SD predicted increase in albuminuria,  $p = 0.030$ , Figures S5-S6, Table S10).

Finally, hypertensive medications lower both albuminuria and blood pressure. To investigate this source of potential bias, we excluded any individuals in UK Biobank taking hypertensive medication. The effects of the resulting albuminuria genetic risk score on increased blood pressure were largely consistent (Figures S5 & S7, Tables S11-S12) albeit with reduced power in this n=302,687 subset.

## **Discussion**

We used Mendelian randomization to examine if processes leading to elevated albuminuria lead to increased risk of cardiometabolic disease. A genome-wide association study of albuminuria in UK Biobank identified 32 new albuminuria loci. A genetic risk score of up to 46 albuminuria variants was strongly associated with increased risk of hypertension and elevated blood pressure, but showed only weak associations with other cardiometabolic diseases.

These results permit several conclusions. First, processes that increase albuminuria appear to increase risk of hypertension and blood pressure. Although hypertension is commonly thought to increase albuminuria, previous epidemiological studies also suggest that albuminuria predicts development of hypertension.<sup>7,8</sup> Our data add genetic and observational evidence supporting this association. Multiple pathways leading to albuminuria may contribute to hypertension. Albuminuria may arise as a result of generalized endothelial dysfunction,<sup>54-56</sup> which can contribute to development of hypertension.<sup>57,58</sup> Albuminuria can also result from kidney damage. In damaged kidneys, increased blood pressure is thought to help the subfunctional kidney excrete sufficient sodium to maintain sodium homeostasis.<sup>59,60</sup> Consistent with this, severe kidney injury leads to experimental hypertension<sup>61</sup> and mild kidney damage precedes the development of hypertension in multiple experimental models.<sup>59,62</sup> The intrinsic role of the kidney in blood pressure regulation is also supported by the observation that kidney transplantation from hypertensive donors can cause hypertension in previously normotensive recipients.<sup>63,64</sup> Further work is needed to determine the mechanisms via which risk score variants contribute to elevated albuminuria.

Second, application of MR Steiger filtering<sup>32,33</sup> enabled the discovery of evidence for bidirectional effects between albuminuria and blood pressure. The associations of genetically-elevated albuminuria with increased blood pressure, and of genetically-elevated blood pressure with increased albuminuria, suggest that the relationship between albuminuria and blood pressure is bidirectional. This would imply the existence of a feed-forward loop, in which elevated blood pressure leads to increased albuminuria, which in turn would further increase blood pressure. It is important to note that because each Mendelian randomization analysis estimates the effects in one direction, this feedback loop is not formally modelled by such analyses.<sup>53</sup> These results suggest that therapies targeting processes that lower albuminuria could have antihypertensive effects that are further amplified by inhibiting this feed-forward loop. Determining the specific genes and pathways affected by albuminuria variants could assist in rational design of such therapies.

Third, these results imply that processes leading to albuminuria can influence cardiovascular disease through blood pressure. Observational and genetic evidence establishes blood pressure as an important causal risk factor for multiple cardiovascular diseases.<sup>20,65-67</sup> By the principal of two-step Mendelian randomization,<sup>16,68</sup> significant associations between an albuminuria risk score and blood pressure, and between blood pressure risk scores and diseases such as stroke or coronary artery disease,<sup>20</sup> imply that the albuminuria risk score is associated with these diseases at least via blood pressure. This raises the question of why we did not observe significant associations between albuminuria and such diseases. Stronger genetic risk scores may be necessary to detect downstream consequences of a causal relationship between albuminuria and blood pressure: since blood pressure explains only some of the variance in cardiovascular outcomes,<sup>69</sup> albuminuria should have a smaller effect size on cardiovascular disease than on blood pressure. We therefore may have been underpowered to detect such downstream effects of albuminuria-induced hypertension on cardiovascular diseases (Table S13). Larger data sets that generate stronger albuminuria genetic risk scores should help clarify this issue.

A key strength of this study is that albuminuria and genotypes were measured in 382,500 individuals, seven times more than the next largest genome-wide association study,<sup>46</sup> enabling construction of a polygenic risk score that explained 0.2% of the variance in albuminuria in two validation cohorts. We were also able to validate the associations between albuminuria and blood pressure in an outside cohort. Access to individual-level data in UK Biobank allowed us to interrogate whether these associations were confounded by hypertensive medication use. Finally, we used MR Steiger filtering to remove variants that potentially acted through reverse causation, and multiple sensitivity analyses to detect and mitigate pleiotropic variants.

Several limitations should be acknowledged. First, reliance on internally-derived weights in our albuminuria genetic risk score may have biased our results towards the observational associations.<sup>44</sup> To address this, we replicated significant associations using unweighted allele scores and/or in two-sample analyses. Second, an alternate explanation for the bidirectional associations observed is that a shared genetic basis underlies the two traits. If so, SNPs that influence both traits through a shared mechanism could violate the instrument strength independent of direct effect (InSIDE) assumption of standard Mendelian randomization and MR-Egger analyses.<sup>70</sup> Although the associations were consistent using median-based regressions, which do not require the InSIDE assumption,<sup>36,42</sup> we cannot rule out the possibility that associations between albuminuria and blood pressure are due to a shared genetic basis of the two traits rather than causal effects. Third, there was substantial heterogeneity in the causal effect estimates from different variants (Tables S6-S10, S12); i.e., for association of albuminuria variants with hypertension, Cochran's  $Q = 160$  ( $p = 9.5 \times 10^{-16}$ ). This is perhaps not surprising considering the hypothesis under investigation was whether pathways that lead to albuminuria can increase blood pressure and hypertension risk. It is plausible – and quite likely – that multiple albuminuria-inducing pathways exist which could elevate blood pressure to different degrees (i.e., endothelial dysfunction and kidney damage), or not at all (i.e., pathways involved in albumin metabolism or post-renal urine regulation).<sup>51</sup> However, other sources of heterogeneity could nevertheless be present, although these do not necessarily lead to bias.<sup>17,51</sup> Fourth, UK Biobank is a population-based longitudinal cohort. Our study was likely underpowered to detect associations in diseases

less common than hypertension (Table S13); therefore, lack of association of albuminuria with other diseases should not be over-interpreted. Finally, it is important to note that UK Biobank is an older cohort of European ancestry; therefore, results may differ in younger populations or in other ethnic backgrounds.

In conclusion, an albuminuria genetic risk score of up to 46 SNPs was associated with increased risk of hypertension and elevated blood pressure. Application of recently-developed Mendelian randomization methods identified evidence of bidirectional effects from albuminuria-increasing pathways to blood pressure, and from blood pressure to albuminuria. These results provide genetic data to refine and highlight the complex interplay between albuminuria and hypertension.

### **Description of Supplemental Data**

Supplemental Data include seven figures and thirteen tables.

### **Acknowledgements**

This work was funded by the National Institutes of Health R01 HL127564 to S.K. and was conducted using the UK Biobank resource under application 7089. S.K. is supported by a research scholar award from Massachusetts General Hospital, the Donovan Family Foundation, and R01 HL127564; has received grants from Bayer Healthcare, Aegerion Pharmaceuticals, and Regeneron Pharmaceuticals; and has received consulting fees from Merck, Novartis, Sano, AstraZeneca, Alnylam Pharmaceuticals, Leerink Partners, Noble Insights, Quest Diagnostics, Genomics PLC, and Eli Lilly and Company; and holds equity in San Therapeutics and Catabasis Pharmaceuticals. K.G.A. is supported by the American Heart Association Institute for Precision Cardiovascular Medicine (17IFUNP33840012). G.D.S. and G.H. work within the Medical Research Council Integrative Epidemiology Unit, supported by the Medical Research Council (MC\_UU\_00011/1) and the University of Bristol. G.H. is supported by the Wellcome Trust (208806/Z/17/Z). The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by the National Heart, Lung, and Blood Institute (NHLBI, contracts HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C,

HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C); funding for GENEVA was provided by National Human Genome Research Institute grant U01HG004402. The authors thank the staff and participants of the ARIC study for their important contributions. The Framingham Heart Study is conducted and supported by the NHLBI in collaboration with Boston University (contracts N01-HC-25195, N02-HL64278 and HHSN268201500001I). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or NHLBI.

### **Declaration of Interests**

The authors declare no competing interests.

### **Web Resources**

Hail, <https://github.com/hail-is/hail>

Mendelian randomization power calculator, <https://sb452.shinyapps.io/power/>

### **References:**

1. Gerstein, H.C., Mann, J.F.E., Yi, Q., Zinman, B., Dinneen, S.F., Hoogwerf, B., Hallé, J.P., Young, J., Rashkow, A., Joyce, C., et al. (2001). Albuminuria and Risk of Cardiovascular Events, Death, and Heart Failure in Diabetic and Nondiabetic Individuals. *Jama* 286, 421–426.
2. Klausen, K., Borch-Johnsen, K., Feldt-Rasmussen, B., Jensen, G., Clausen, P., Scharling, H., Appleyard, M., and Jensen, J.S. (2004). Very low levels of microalbuminuria are associated with increased risk of coronary heart disease and death independently of renal function, hypertension, and diabetes. *Circulation* 110, 32–35.
3. Hillege, H.L., Fidler, V., Diercks, G.F.H., van Gilst, W.H., de Zeeuw, D., van Veldhuisen, D.J., Gans, R.O.B., Janssen, W.M.T., Grobbee, D.E., de Jong, P.E., et al. (2002). Urinary albumin excretion predicts cardiovascular and noncardiovascular mortality in general population. *Circulation* 106, 1777–1782.
4. Hemmelgarn, B.R., Manns, B.J., Lloyd, A., James, M.T., Klarenbach, S., Quinn, R.R., Wiebe, N., Tonelli, M., Alberta Kidney Disease Network (2010). Relation between kidney function, proteinuria, and adverse outcomes. *Jama* 303, 423–429.
5. Matsushita, K., Coresh, J., Sang, Y., Chalmers, J., Fox, C., Guallar, E., Jafar, T., Jassal, S.K., Landman, G.W.D., Muntner, P., et al. (2015). Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. *The Lancet Diabetes & Endocrinology* 3, 514–525.

6. Brantsma, A.H., Bakker, S.J.L., Hillege, H.L., de Zeeuw, D., de Jong, P.E., Gansevoort, R.T., PREVEND Study Group (2005). Urinary albumin excretion and its relation with C-reactive protein and the metabolic syndrome in the prediction of type 2 diabetes. *Diabetes Care* 28, 2525–2530.
7. Wang, T.J., Evans, J.C., Meigs, J.B., Rifai, N., Fox, C.S., D'Agostino, R.B., Levy, D., and Vasan, R.S. (2005). Low-grade albuminuria and the risks of hypertension and blood pressure progression. *Circulation* 111, 1370–1376.
8. Brantsma, A.H., Bakker, S.J.L., de Zeeuw, D., de Jong, P.E., and Gansevoort, R.T. (2006). Urinary albumin excretion as a predictor of the development of hypertension in the general population. *J. Am. Soc. Nephrol.* 17, 331–335.
9. Scirica, B.M., Mosenzon, O., Bhatt, D.L., Udell, J.A., Steg, P.G., McGuire, D.K., Im, K., Kanevsky, E., Stahre, C., Sjöstrand, M., et al. (2018). Cardiovascular Outcomes According to Urinary Albumin and Kidney Disease in Patients With Type 2 Diabetes at High Cardiovascular Risk: Observations From the SAVOR-TIMI 53 Trial. *JAMA Cardiol* 3, 155–163.
10. Asselbergs, F.W., Diercks, G.F.H., Hillege, H.L., van Boven, A.J., Janssen, W.M.T., Voors, A.A., de Zeeuw, D., de Jong, P.E., van Veldhuisen, D.J., van Gilst, W.H., et al. (2004). Effects of fosinopril and pravastatin on cardiovascular events in subjects with microalbuminuria. *Circulation* 110, 2809–2816.
11. Ibsen, H., Olsen, M.H., Wachtell, K., Borch-Johnsen, K., Lindholm, L.H., Mogensen, C.E., Dahlöf, B., Devereux, R.B., de Faire, U., Fyhrquist, F., et al. (2005). Reduction in albuminuria translates to reduction in cardiovascular events in hypertensive patients: losartan intervention for endpoint reduction in hypertension study. *Hypertension* 45, 198–202.
12. de Zeeuw, D., Remuzzi, G., Parving, H.-H., Keane, W.F., Zhang, Z., Shahinfar, S., Snapinn, S., Cooper, M.E., Mitch, W.E., and Brenner, B.M. (2004). Albuminuria, a therapeutic target for cardiovascular protection in type 2 diabetic patients with nephropathy. *Circulation* 110, 921–927.
13. Holtkamp, F.A., de Zeeuw, D., de Graeff, P.A., Laverman, G.D., Berl, T., Remuzzi, G., Packham, D., Lewis, J.B., Parving, H.H., and Lambers Heerspink, H.J. (2011). Albuminuria and blood pressure, independent targets for cardioprotective therapy in patients with diabetes and nephropathy: a post hoc analysis of the combined RENAAL and IDNT trials. *Eur. Heart J.* 32, 1493–1499.
14. Bakris, G. (2016). Moderately increased albuminuria (microalbuminuria) and cardiovascular disease. UpToDate. <https://www.uptodate.com/contents/moderately-increased-albuminuria-microalbuminuria-and-cardiovascular-disease#H220350>.
15. Davey Smith, G., and Ebrahim, S. (2003). “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* 32, 1–22.
16. Zheng, J., Baird, D., Borges, M.-C., Bowden, J., Hemani, G., Haycock, P., Evans, D.M., and Davey Smith, G. (2017). Recent Developments in Mendelian Randomization Studies. *Curr Epidemiol Rep* 4, 330–345.
17. Hemani, G., Bowden, J., and Davey Smith, G. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* <https://doi.org/10.1093/hmg/ddy163>.
18. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2017). Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv.* <https://doi.org/10.1101/166298>.



19. Mattix, H.J., Hsu, C.-Y., Shaykevich, S., and Curhan, G. (2002). Use of the albumin/creatinine ratio to detect microalbuminuria: implications of sex and race. *J. Am. Soc. Nephrol.* *13*, 1034–1039.
20. Ehret, G.B., Ferreira, T., Chasman, D.I., Jackson, A.U., Schmidt, E.M., Johnson, T., Thorleifsson, G., Luan, J., Donnelly, L.A., Kanoni, S., et al. (2016). The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nature Genetics* *48*, 1171–1184.
21. Wain, L.V., Vaez, A., Jansen, R., Joehanes, R., van der Most, P.J., Erzurumluoglu, A.M., O'Reilly, P.F., Cabrera, C.P., Warren, H.R., Rose, L.M., et al. (2017). Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets From Blood and the Kidney. *Hypertension* *70*, e4–e19.
22. Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* *23*, R89–R98.
23. Matsushita, K., Ballew, S.H., Coresh, J., Arima, H., Ärnlöv, J., Cirillo, M., Ebert, N., Hiramoto, J.S., Kimm, H., Shlipak, M.G., et al. (2017). Measures of chronic kidney disease and risk of incident peripheral artery disease: a collaborative meta-analysis of individual participant data. *The Lancet Diabetes & Endocrinology* *5*, 718–728.
24. Halimi, J.-M., Bonnet, F., Lange, C., Balkau, B., Tichet, J., Marre, M., DESIR Study Group (2008). Urinary albumin excretion is a risk factor for diabetes mellitus in men, independently of initial metabolic profile and development of insulin resistance. The DESIR Study. *Journal of Hypertension* *26*, 2198–2206.
25. Ishani, A., Grandits, G.A., Grimm, R.H., Svendsen, K.H., Collins, A.J., Prineas, R.J., and Neaton, J.D. (2006). Association of single measurements of dipstick proteinuria, estimated glomerular filtration rate, and hematocrit with 25-year incidence of end-stage renal disease in the multiple risk factor intervention trial. *Journal of the American Society of Nephrology* *17*, 1444–1452.
26. Ganna, A., Genovese, G., Howrigan, D.P., Byrnes, A., Kurki, M., Zekavat, S.M., Whelan, C.W., Kals, M., Nivard, M.G., Bloemendal, A., et al. (2016). Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. *Nat. Neurosci.* *19*, 1563–1565.
27. Xu, C., Tachmazidou, I., Walter, K., Ciampi, A., Zeggini, E., Greenwood, C.M.T., UK10K Consortium (2014). Estimating genome-wide significance for whole-genome sequencing studies. *Genet. Epidemiol.* *38*, 281–290.
28. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* *47*, 291–295.
29. Emdin, C.A., Khera, A.V., Natarajan, P., Klarin, D., Zekavat, S.M., Hsiao, A.J., and Kathiresan, S. (2017). Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease. *Jama* *317*, 626–629.
30. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* *37*, 658–665.
31. Yavorska, O.O., and Burgess, S. (2017). MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* *46*, 1734–1739.

32. Hemani, G., Bowden, J., Haycock, P.C., Zheng, J., Davis, O., Flach, P., Gaunt, T.R., and Davey Smith, G. (2017). Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. *bioRxiv*. <https://doi.org/10.1101/173682>.
33. Hemani, G., Tilling, K., and Davey Smith, G. (2017). Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* *13*, e1007081.
34. Verbanck, M., Chen, C.-Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* *50*, 693–698.
35. Burgess, S., and Bowden, J. (2015). Integrating summarized data from multiple genetic variants in Mendelian randomization: bias and coverage properties of inverse-variance weighted methods. *arXiv*. <https://arxiv.org/abs/1512.04486>.
36. Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* *40*, 304–314.
37. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* *44*, 512–525.
38. Richmond, R., Wade, K., Corbin, L., Bowden, J., Hemani, G., Timpson, N., and Davey Smith, G. (2017). Investigating the role of insulin in increased adiposity: Bi-directional Mendelian randomization study. *bioRxiv*. <https://doi.org/10.1101/155739>.
39. Burgess, S., and Thompson, S.G. (2013). Use of allele scores as instrumental variables for Mendelian randomization. *Int. J. Epidemiol.* *42*, 1134–1144.
40. Bianchi, S., Bigazzi, R., and Campese, V.M. (1999). Microalbuminuria in essential hypertension: Significance, pathophysiology, and therapeutic implications. *American Journal of Kidney Diseases* *34*, 973–995.
41. Williams, M.E. (2005). Diabetic Nephropathy: The Proteinuria Hypothesis. *Am J Nephrol* *25*, 77–94.
42. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. *Elife* *7*, 68.
43. Lee, S.H., Goddard, M.E., Wray, N.R., and Visscher, P.M. (2012). A Better Coefficient of Determination for Genetic Profile Analysis. *Genet. Epidemiol.* *36*, 214–224.
44. Hartwig, F.P., and Davies, N.M. (2016). Why internal weights should be avoided (not only) in MR-Egger regression. *Int. J. Epidemiol.* *45*, 1676–1678.
45. Sarafidis, P.A., Khosla, N., and Bakris, G.L. (2007). Antihypertensive therapy in the presence of proteinuria. *Am. J. Kidney Dis.* *49*, 12–26.
46. Teumer, A., Tin, A., Sorice, R., Gorski, M., Yeo, N.C., Chu, A.Y., Li, M., Li, Y., Mijatovic, V., Ko, Y.-A., et al. (2016). Genome-wide Association Studies Identify Genetic Loci Associated With Albuminuria in Diabetes. *Diabetes* *65*, 803–817.

47. Ma, T.K.-W., and Szeto, C.-C. (2012). Mineralocorticoid receptor antagonist for renal protection. *Ren Fail* 34, 810–817.
48. Currie, G., Taylor, A.H.M., Fujita, T., Ohtsu, H., Lindhardt, M., Rossing, P., Boesby, L., Edwards, N.C., Ferro, C.J., Townend, J.N., et al. (2016). Effect of mineralocorticoid receptor antagonists on proteinuria and progression of chronic kidney disease: a systematic review and meta-analysis. *BMC Nephrol* 17, 127.
49. Kruegel, J., Rubel, D., and Gross, O. (2013). Alport syndrome--insights from basic and clinical research. *Nature Reviews Nephrology* 9, 170–178.
50. Corbin, L.J., Richmond, R.C., Wade, K.H., Burgess, S., Bowden, J., Davey Smith, G., and Timpson, N.J. (2016). BMI as a Modifiable Risk Factor for Type 2 Diabetes: Refining and Understanding Causal Estimates Using Mendelian Randomization. *Diabetes* 65, 3002–3007.
51. Burgess, S., Foley, C.N., and Zuber, V. (2018). Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data. *Annu Rev Genomics Hum Genet* 19.
52. Burgess, S., and Thompson, S.G. (2017). Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* 32, 377–389.
53. Haycock, P.C., Burgess, S., Wade, K.H., Bowden, J., Relton, C., and Davey Smith, G. (2016). Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am. J. Clin. Nutr.* 103, 965–978.
54. Deckert, T., Feldt-Rasmussen, B., Borch-Johnsen, K., Jensen, T., and Kofoed-Enevoldsen, A. (1989). Albuminuria reflects widespread vascular damage. The Steno hypothesis. *Diabetologia* 32, 219–226.
55. Clausen, P., Feldt-Rasmussen, B., Jensen, G., and Jensen, J.S. (1999). Endothelial haemostatic factors are associated with progression of urinary albumin excretion in clinically healthy subjects: a 4-year prospective study. *Clinical Science* 97, 37–43.
56. Stehouwer, C.D.A. (2006). Microalbuminuria and Risk for Cardiovascular Disease: Analysis of Potential Mechanisms. *J. Am. Soc. Nephrol.* 17, 2106–2111.
57. Rossi, R., Chiurlia, E., Nuzzo, A., Cioni, E., Origliani, G., and Modena, M.G. (2004). Flow-mediated vasodilation and the risk of developing hypertension in healthy postmenopausal women. *J. Am. Coll. Cardiol.* 44, 1636–1640.
58. Sander, M., Chavoshan, B., and Victor, R.G. (1999). A Large Blood Pressure–Raising Effect of Nitric Oxide Synthase Inhibition in Humans. *Hypertension* 33, 937–942.
59. Cowley, A.W., and Roman, R.J. (1996). The role of the kidney in hypertension. *Jama* 275, 1–25.
60. Crowley, S.D., and Coffman, T.M. (2014). The inextricable role of the kidney in hypertension. *J. Clin. Invest.* 124, 2341–2347.
61. Goldblatt, H., Lynch, J., Hanzal, R.F., and Summerville, W.W. (1934). Studies on Experimental Hypertension: I. the Production of Persistent Elevation of Systolic Blood Pressure by Means of Renal Ischemia. *J. Exp. Med.* 59, 347–379.
62. Johnson, R.J., Herrera-Acosta, J., Schreiner, G.F., and Rodriguez-Iturbe, B. (2002). Subtle acquired renal injury as a mechanism of salt-sensitive hypertension. *N Engl J Med* 346, 913–923.

63. Strandgaard, S., and Hansen, U. (1986). Hypertension in renal allograft recipients may be conveyed by cadaveric kidneys from donors with subarachnoid haemorrhage. *British Medical Journal* 292, 1041–1044.
64. Guidi, E., Menghetti, D., Milani, S., Montagnino, G., Palazzi, P., and Bianchi, G. (1996). Hypertension may be transplanted with the kidney in humans: a long-term historical prospective follow-up of recipients grafted with kidneys coming from donors with or without hypertension in their families. *J. Am. Soc. Nephrol.* 7, 1131–1138.
65. Rapsomaniki, E., Timmis, A., George, J., Pujades-Rodriguez, M., Shah, A.D., Denaxas, S., White, I.R., Caulfield, M.J., Deanfield, J.E., Smeeth, L., et al. (2014). Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people. *Lancet* 383, 1899–1911.
66. Lewington, S., Clarke, R., Qizilbash, N., Peto, R., Collins, R., Prospective Studies Collaboration (2002). Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *The Lancet* 360, 1903–1913.
67. Levy, D., Larson, M.G., Vasan, R.S., Kannel, W.B., and Ho, K.K.L. (1996). The Progression From Hypertension to Congestive Heart Failure. *Jama* 275, 1557–1562.
68. Relton, C.L., and Davey Smith, G. (2012). Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int. J. Epidemiol.* 41, 161–176.
69. Cheng, S., Claggett, B., Correia, A.W., Shah, A.M., Gupta, D.K., Skali, H., Ni, H., Rosamond, W.D., Heiss, G., Folsom, A.R., et al. (2014). Temporal trends in the population attributable risk for cardiovascular disease: the Atherosclerosis Risk in Communities Study. *Circulation* 130, 820–828.
70. Bowden, J., Del Greco M, F., Minelli, C., Davey Smith, G., Sheehan, N., and Thompson, J. (2017). A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 36, 1783–1802.

## **Figure Titles and Legends**

### **Figure 1. Association of albuminuria with incident disease endpoints in UK Biobank.**

Incident disease <sup>a</sup>adjusted for age, sex, current smoking status, body mass index, systolic blood pressure, diastolic blood pressure, baseline diabetes and baseline hyperlipidemia unless otherwise specified. <sup>b</sup>adjusted for age, sex, current smoking status, body mass index, systolic blood pressure, diastolic blood pressure, waist-to-hip ratio and baseline hyperlipidemia. <sup>c</sup>adjusted for age, age<sup>2</sup>, current smoking status, body mass index, baseline diabetes and baseline hyperlipidemia. Bars indicate 95% confidence interval for hazard ratio.

### **Figure 2. Genome-wide association study of albuminuria in UK Biobank identifies 32 new loci.**

33 genome-wide significant loci are indicated by red points. Red line indicates genome-wide significance threshold ( $p = 9 \times 10^{-9}$ ); blue line indicates conventional significance threshold ( $p = 5 \times 10^{-8}$ ).

### **Figure 3. Association of genetic predisposition to increased albuminuria with risk of cardiometabolic disease in UK Biobank.**

Two-stage least-squares regression using albuminuria genetic risk score as instrumental variable; age, sex, genotyping array and 1<sup>st</sup> 10 genetic PCs as covariates. Results are standardized to 1-SD increase in albuminuria due to the genetic risk score. A, genetic risk score composed of all 46 albuminuria variants. B, genetic risk score composed of 43 albuminuria variants after applying directional MR Steiger filtering to remove variants potentially acting in the incorrect direction. Bars indicate 95% confidence interval for odds ratio.

### **Figure 4. Bidirectional Mendelian randomization identifies suggestive causal effects of albuminuria on blood pressure, and of blood pressure on albuminuria.**

A, Mendelian randomization of albuminuria genetic risk scores on blood pressure in UK Biobank (n=381833). Two-stage least-squares regression using albuminuria genetic risk score as instrumental variable on blood pressure outcome; age, sex, genotyping array and 1<sup>st</sup> 10 genetic PCs as covariates.

Results are standardized to 1-SD increase in albuminuria due to the genetic risk score. Genetic risk scores were composed of 44 albuminuria variants after applying directional MR Steiger filtering to remove variants potentially acting in the incorrect direction.

B, Mendelian randomization of blood pressure genetic risk scores on albuminuria. Effects of variants on systolic or diastolic blood pressure were determined in ICBP<sup>20</sup> ( $n_{\max}=201529$ ) and thus corrected for hypertensive medication use and adjusted for body mass index. Two-stage least-squares regression using blood pressure genetic risk score as instrumental variable on albuminuria outcome in UK Biobank ( $n=381833$ ); age, sex, genotyping array and 1<sup>st</sup> 10 genetic PCs as covariates. Results are standardized to 1-mmHg increase in blood pressure due to the genetic risk score. 47 or 52 variants were used to construct scores specific for systolic or diastolic blood pressure, respectively. Directional MR Steiger filtering removed no variants.

SNPs in score, number of SNPs remaining after directional MR Steiger filtering applied. Bars indicate 95% confidence interval for effect on blood pressure (top) or albuminuria (bottom).

## Tables

Lead variant	Nearest Gene(s)	Description	Chr	Position (hg19)	Effect Allele	Noneffect Allele	EAF	Beta (log(mg/g))	SE (log (mg/g))	P value
rs12032996	<i>PHC2-ZSCAN20</i>	Intergenic	1	33920586	G	A	0.838	0.01463	0.00226	9.33E-11
rs10157710	<i>FOXD2-TRABD2B</i>	Intergenic	1	47961691	T	C	0.802	0.01900	0.00209	9.69E-20
rs11264327	<i>EFNA3-EFNA1</i>	Intergenic	1	155095107	A	G	0.399	0.00987	0.00171	7.03E-09
rs4665972	<i>SNX17</i>	Intronic	2	27598097	T	C	0.393	0.01176	0.00172	6.96E-12
rs13394343	<i>SH2D6-MAT2A/PARTICL</i>	Intergenic	2	85754342	C	A	0.570	0.01053	0.00168	3.86E-10
rs10207567	<i>ICA1L</i>	Intronic	2	203714973	C	G	0.813	0.01455	0.00214	1.00E-11
rs1047891	<i>CPS1</i>	Missense	2	211540507	C	A	0.684	0.01205	0.00179	1.71E-11
rs183131780	<i>MIR548AR-LOC646736</i>	Intergenic	2	226684886	T	C	0.002	0.19055	0.01959	2.33E-22
rs35483183	<i>COL4A4</i>	Intronic	2	227876687	A	G	0.123	0.01490	0.00255	5.19E-09
rs35924503	<i>SPHKAP-PID1</i>	Intergenic	2	229131286	C	T	0.001	0.24742	0.02518	8.68E-23
rs112607182	<i>PRKCI</i>	Downstream Variant	3	170027407	T	C	0.077	0.02279	0.00327	3.39E-12
rs7654754	<i>SHROOM3</i>	Intronic	4	77409795	G	A	0.462	0.01020	0.00167	9.96E-10
rs6535594	<i>NR3C2</i>	Intronic	4	149132756	A	G	0.496	0.01146	0.00167	7.12E-12
rs189107782	<i>LINC01262-FRG1</i>	Intergenic	4	190729009	T	C	0.002	0.24502	0.02026	1.12E-33
rs702634	<i>ARL15</i>	Intronic	5	53271420	A	G	0.692	0.01042	0.00181	8.03E-09
rs7731168	<i>CWC27</i>	Intronic	5	64296471	C	G	0.233	0.01253	0.00197	2.19E-10
rs4410790	<i>AGR3-AHR</i>	Intergenic	7	17284577	C	T	0.634	0.01798	0.00173	2.63E-25
rs2023844	<i>HOTTIP</i>	Intronic	7	27243238	A	G	0.926	0.01934	0.00318	1.18E-09
rs17158386	<i>WIPF3-DPY19L2P3</i>	Intergenic	7	29805361	A	G	0.262	0.01330	0.00191	3.65E-12
rs28601761	<i>TRIB1-LINC00861</i>	Intergenic	8	126500031	C	G	0.579	0.01136	0.00171	2.81E-11
rs45551835	<i>CUBN</i>	Missense	10	16932384	A	G	0.014	0.14237	0.00698	2.28E-92
rs144360241	<i>CUBN</i>	Missense	10	16967417	C	T	0.005	0.08186	0.01234	3.31E-11
rs1276720	<i>CUBN</i>	Intronic	10	16971426	T	C	0.745	0.01109	0.00193	8.98E-09
rs141640975	<i>CUBN</i>	Missense	10	16992011	A	G	0.003	0.35876	0.01629	1.75E-107
rs67339103	<i>C10orf11</i>	Intronic	10	77893686	A	G	0.212	0.01522	0.00205	1.07E-13
rs17368443	<i>SBF2</i>	Intronic	11	10296836	C	G	0.061	0.02071	0.00348	2.58E-09
rs2601006	<i>CCT2</i>	5' UTR Variant	12	69979517	C	T	0.657	0.01176	0.00176	2.13E-11
rs4288924	<i>ZFP36L1-ACTN1</i>	Intergenic	14	69302399	G	A	0.480	0.00980	0.00168	5.66E-09
rs8035855	<i>MAPKBP1</i>	Intronic	15	42077961	A	G	0.644	0.01227	0.00174	1.91E-12
rs1145074	<i>SPATA5L1</i>	Intronic	15	45703824	T	A	0.745	0.01140	0.00191	2.41E-09
rs2472297	<i>CYP1A2-CYP1A1</i>	Intergenic	15	75027880	T	C	0.267	0.01812	0.00188	5.31E-22
rs35572189	<i>BAHCC1</i>	Missense	17	79419025	G	A	0.638	0.01051	0.00174	1.44E-09
rs838142	<i>FUT1</i>	3' UTR Variant	19	49252151	A	G	0.723	0.01174	0.00187	3.13E-10

**Table 1. Albuminuria loci from GWAS of 382,500 individuals in UK Biobank.**

Chr, chromosome; EAF, effect allele frequency. For intergenic loci, nearest upstream and downstream RefSeq genes are indicated. Nearest gene should not be taken as evidence of causal gene. Description, most-severe consequence of nearest RefSeq gene.