

# Distinct pathway-based effects of blood pressure and body mass index on cardiovascular traits: comparison of novel Mendelian Randomization approaches

Maria K Sobczyk<sup>1\*</sup>, Tom G Richardson<sup>1</sup>, Genevieve M Leyden<sup>1</sup>, Tom R Gaunt<sup>1\*</sup>

1) MRC Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Oakfield House, Bristol BS8 2BN, United Kingdom

\* Corresponding authors: [maria.sobczyk@bristol.ac.uk](mailto:maria.sobczyk@bristol.ac.uk), [tom.gaunt@bristol.ac.uk](mailto:tom.gaunt@bristol.ac.uk)

## Abstract

**Background:** Mendelian randomization (MR) leverages genetic variants as instrumental variables to determine causal relationships in epidemiology. However, challenges persist due to heterogeneity arising from horizontal pleiotropy. On the other hand, exploration of the biological underpinnings of such heterogeneity across variants can enhance our understanding of disease mechanisms and inform therapeutic strategies. Here, we introduce a new approach to instrument partitioning based on enrichment of Mendelian disease categories and compare it to a method based on genetic colocalisation in contrasting tissues.

**Methods:** We employed one-sample and two-sample MR methodologies using blood pressure (BP) exposure SNPs grouped by proximity to Mendelian disease genes affecting the renal system or vasculature, or body mass index (BMI) variants related to mental health and metabolic Mendelian disorders. We then compared the causal effects of Mendelian-partitioned SNPs on cardiometabolic outcomes with subsets inferred from gene expression colocalisation in kidney, artery (for BP), adipose, and brain tissues (for BMI). Additionally, we assessed whether effects from these groupings could emerge by chance using random SNP subset sampling.

**Results:** Our findings suggest that the causal relationship between systolic BP and coronary heart disease is predominantly driven by SNPs associated with vessel-related Mendelian diseases over renal. However, kidney-oriented SNPs showed more pronounced effect size in the colocalization-based analysis, hinting at a multifaceted interplay between pathways in the disease aetiology. We consistently identified a dominant role of Mendelian vessel and coloc artery exposures in driving the negative effect of diastolic BP on left ventricular stroke volume and positive effect of systolic BP on type 2 diabetes. We also found higher causal estimates for metabolic versus mental health SNPs when dissecting BMI pathway contribution to atrial fibrillation risk using Mendelian disease. In contrast, brain variants yielded higher causal estimates than adipose in the colocalization method.

**Conclusions:** This study presents a novel approach to dissecting heterogeneity in MR by integrating clinical phenotypes associated with Mendelian disease. Our findings emphasize the importance of understanding tissue-/pathway- specific contributions in interpreting causal relationships in MR. Importantly, we advocate caution in interpreting pathway-partitioned effect size differences without robust statistical validation.

**Short title:** Pathway effects of blood pressure and BMI on cardiovascular traits

**Keywords:** body mass index; blood pressure; cardiovascular disease; Mendelian disease; colocalisation; Mendelian randomization; genetic epidemiology

## Introduction

Mendelian randomization (MR) is a statistical method used in epidemiology to study the causal relationship between a risk factor (exposure) and an outcome (disease or trait) by leveraging genetic variants derived from genome-wide association studies (GWAS) as instrumental variables<sup>1</sup>. The technique is based on the principles of Mendelian inheritance, which states that genetic variants, such as single nucleotide polymorphisms (SNPs), are randomly assigned during meiosis and therefore should be less prone to confounding factors or reverse causation that typically plague observational studies (subject to meeting certain assumptions<sup>2</sup>).

However, causal estimates derived from individual genetic variants used as instrumental variables in MR studies can be highly variable<sup>3</sup>. Heterogeneity in a given single result can arise from two main sources of bias due to violation of key MR assumptions. These are horizontal pleiotropy, which occurs when one or more genetic variants influence the outcome through multiple independent pathways and weak instrument bias which can result in imprecise causal estimates and increase heterogeneity<sup>4</sup>.

Of chief interest in our study, causal effects vary due to endpoint phenotypes being de facto composites, representing divergent underlying biological mechanisms covered by different genetic instruments, which unlike bias can improve understanding of disease aetiology and help design better targeted interventions. Three broad types of approaches have been used so far when studying biological sources of heterogeneity: direct clustering based on SNP associations with exposure and outcome<sup>5,6</sup>, clustering of variant associations across a set of traits<sup>7-9</sup>, or instrument clustering informed by tissue gene expression patterns<sup>10-13</sup>. In particular, a biological hypothesis-driven approach proposed by Leyden et al. (2022)<sup>11</sup> clusters genetic instruments for body mass index (BMI) based on the tissue (brain or adipose) where a given BMI SNP is found to colocalise with an expression quantitative trait locus (eQTL). The Bayesian colocalisation method coloc<sup>14</sup> is employed here to compare the association signals at a specific genomic region for the two traits of interest (gene eQTL and BMI), considering variables such as effect sizes and allele frequencies to determine to what extent these are consistent with a single shared causal variant for both traits. In this way, a given SNP instrument is putatively linked to a particular gene whose expression (either in the adipose or brain tissue, or both) potentially mediates distinct causal effect on a set of cardiometabolic traits.

Although this approach was used to prioritise the putative causal tissue types underlying BMI-associated genes, in general the coloc method as originally implemented has been shown to lack specificity when assigning SNPs to genes on its own, particularly when using eQTL data due to the co-expression of nearby genes<sup>15</sup>. Another approach of prioritizing candidate genes at GWAS loci is to leverage the knowledge of Mendelian monogenic diseases, which are caused by rare mutations with large effects on phenotypes. Several studies have reported an enrichment of Mendelian disease genes near GWAS loci across various phenotypes, suggesting shared genetic basis between complex and Mendelian traits<sup>16–18</sup>. However, not all Mendelian disease genes are equally relevant for a given complex trait, and trait symptom similarity ought to be a key metric for gene prioritization<sup>19</sup>.

In this paper, we introduce a new approach of Mendelian disease category-driven stratification of variants for common exposures used in MR. Blood pressure (BP) is a highly polygenic risk factor for a number of cardiovascular<sup>20–22</sup> and metabolic<sup>23–25</sup> conditions, with both vasculature-<sup>26–28</sup> and kidney-<sup>29–31</sup> expressed genes shown previously to be of key importance. Kidneys control blood pressure by regulating blood volume and electrolyte balance<sup>32</sup>, chiefly through natriuresis response<sup>33</sup> and the renin-angiotensin-aldosterone system (RAAS) hormonal axis. Accordingly, impaired kidney function has long been linked to hypertension<sup>34</sup>. The vasculature regulates blood pressure via modulation of vascular tone. This is achieved through the processes of vasoconstriction and vasodilation, controlled by the smooth muscle cells in the arterial walls<sup>35</sup>. The endothelium lining the inner surface of blood vessels plays a pivotal role by releasing an array of vasoactive substances<sup>36</sup>. Endothelial cells secrete endothelin, a potent vasoconstrictor and nitric oxide, the key vasodilator<sup>37,38</sup> and can also influence blood pressure through inflammatory mechanisms<sup>39</sup>. If instruments acting on BP via these two key mechanisms show different estimates of effect on an outcome in MR, we hypothesize that this will be due to pleiotropy in one or both subsets of instruments.

We begin by contrasting the kidney and vascular components of blood pressure risk factor burden on cardiometabolic disease. To achieve this, we carry out one-sample and two-sample multivariable MR analyses utilising blood pressure (systolic and diastolic) exposure variants grouped by co-sharing genetic loci with Mendelian disease genes whose symptoms affect either the renal system or vasculature (**Figure 1**). We compare our results to the colocalisation-based method proposed by Leyden et al. (2022)<sup>11</sup> by linking blood pressure genetic variants to regulation of gene expression in kidney or arteries. We then return to the BMI exposure reported by Leyden et al.<sup>11</sup> to ask if the effect of variants grouped by link to metabolic and mental health Mendelian disease corresponds to the effect obtained by adipose and brain tissue colocalisation-based subsets. Finally, using random re-sampling of the exposure SNP sets we investigate if the effect size differences observed in coloc- or Mendelian- based SNP subdivisions are likely to arise by chance. Our findings provide valuable insights into tissue-based underpinnings of causal links between BP or BMI and cardiometabolic traits.

## Materials and Methods

### *Exposure and outcome GWAS datasets*

Our exposure datasets consisted of systolic (SBP) and diastolic blood pressure (DBP) GWAS produced by ICBP in 2018<sup>27</sup> and body mass index (BMI) produced by the GIANT consortium in 2018<sup>40</sup> (**Supplementary Table 1**). These large studies, comprising ~1 million and ~0.7 million individuals of European ancestry, respectively, were highly powered and returned a number of top variants (~900) implicated in a variety of biological processes in each GWAS.

Our outcome GWAS datasets included common cardiometabolic diseases: atrial fibrillation (AF)<sup>41</sup>, heart failure (HF)<sup>42</sup>, coronary heart disease (CHD)<sup>43</sup>, myocardial infarction (MI)<sup>43</sup>, stroke<sup>44</sup>, and type 2 diabetes (T2D, see **Supplementary Table 1** for details). GWAS for continuous measurements of cardiac function included: left ventricular end-diastolic volume (LVEDV), left ventricular end-systolic volume (LVESV), left ventricular ejection fraction (LVEF), and left ventricular stroke volume (SV).

As a negative control, we selected an outcome whose incidence is likely not to be causally impacted by BMI or BP as indicated by a previous MR study<sup>45</sup>: age-related macular degeneration (AMD)<sup>46</sup>.

All the GWAS summary statistics, with the exception of left ventricular function traits (downloaded from <https://personal.broadinstitute.org/ryank/>) and early-onset AMD (downloaded from [https://homepages.uni-regensburg.de/~wit59712/earlyamd/winkler\\_et\\_al\\_earlyamd\\_meta.gz](https://homepages.uni-regensburg.de/~wit59712/earlyamd/winkler_et_al_earlyamd_meta.gz)) were accessed via Open GWAS<sup>47</sup>.

We note that we used exposure and outcome GWAS datasets with non-overlapping participants whenever feasible, to limit bias<sup>48</sup>, however a significant proportion of individuals in the AF and early-onset AMD (~20-40%), as well as in the left ventricular function GWASes were obtained based on UK Biobank individuals, who are also included in our exposure GWAS for BMI and BP.

### *Genetic instrument selection*

BMI and BP GWAS summary statistics were obtained in GWAS-VCF<sup>49</sup> format from the OpenGWAS platform, and were subsequently converted to the TwoSampleMR<sup>50</sup> package format using the `gwasvcf_to_TwoSampleMR` function from the `gwasglue` (<https://mrcieu.github.io/gwasglue/>) R package. In order to identify independent genetic instruments for each exposure, we first filtered SNPs that showed strong association at a genome-wide significance level ( $p$ -value  $< 5 \times 10^{-8}$ ). Next, the SNPs were clumped using `ld_clump` wrapper for `plink` ver 1.9.43<sup>51</sup> from the `ieugwasr` R package (<https://mrcieu.github.io/ieugwasr>) to ensure that linkage disequilibrium (LD) as measured by  $r^2$  was  $< 0.001$  within 10 Mbp in 1000 Genomes<sup>52</sup> European panel. We sometimes used SNP proxies showing high genetic correlation ( $r^2 > 0.8$ ) in the instances when the chosen SNP was missing in the outcome dataset.

### *Assignment of genetic instruments to pathways: Mendelian disease*

Having obtained 887 independent SNP instruments for BMI, 914 for diastolic BP and 863 for systolic BP, we used MendelVar<sup>19</sup> to partition the variants into subsets

enriched for Mendelian disease categories. Briefly, we used the MendelVar<sup>19</sup> pipeline to generate LD-based genomic intervals around each input SNP (delimited by most distant SNP with minimum  $r^2 = 0.8$  within 1 Mbp in either direction). Then, we checked for enrichment of phenotype ontology terms linked to Mendelian disease genes (defined as coding region with 1000 bp 5' and 3' flanking regions) present within the interval using the INRICH<sup>53</sup> software contained in the MendelVar platform.

#### *Assignment of genetic instruments to tissues: colocalisation*

We followed the method described in Leyden et al. (2022)<sup>11</sup> to assign SNPs to subsets with eQTL colocalisation evidence in at least one of the two chosen tissue types (kidney and vasculature). For BP traits, we used cis-eQTLs in kidney (NephQTL2<sup>54</sup> tubulointerstitial  $n=311$  and glomerular  $n=240$ ) and arteries (GTEx 8<sup>55</sup> aorta  $n=387$  and coronary  $n=212$ ); similar sample sizes across tissues should result in comparable power. We used intervals of +/- 100 kbp centred on each exposure SNP for colocalisation in the coloc R software<sup>56</sup>, and a stringent posterior probability H4 (PP<sub>H4</sub>, hypothesis 4: shared causal variant between exposure GWAS and cis-eQTL dataset) threshold of 0.9 to partition SNPs into coloc-based tissues.

#### *Enrichment analysis within subsets*

We used ToppFun from ToppGene Suite<sup>57</sup> and over-representation module in ConsensusPathDB<sup>58</sup> (both with default settings) to test for global enrichment of functional terms in gene subsets identified by MendelVar and coloc harnessing popular ontologies, such as GO<sup>59</sup>, Reactome<sup>60</sup> and KEGG<sup>61</sup>.

#### *One sample Mendelian Randomization analyses*

Our methodology for one-sample MR analyses followed the protocol described in Leyden et al. (2022)<sup>11</sup> and utilized individual-level data from the UK Biobank<sup>62</sup> (Application number: 81499). This involved the creation of a genetic risk score (GRS), factoring in all SNPs or coloc/Mendelian SNP subsets, weighted according to their respective effect sizes. The sample comprised 334,398 unrelated people of European descent, and was established after excluding participants with withdrawn consent, genetically related, or those who did not cluster with "white European" group based on K-means clustering (K=4).

Subsequently, we used all exposure (BMI, DBP, SBP) SNPs to derive MR estimates employing either linear or logistic regression using AER and mass R packages. We analysed all available outcomes (**Supplementary Table 2**: AF, HF, CHD, MI, T2D, stroke, AMD, LVEDV, LVESV, LVEF, SV) and adjusted for variables such as age, sex, the leading 10 principal components, and a binary marker for genotype chips. A rank-based inverse normal transformation was applied to continuous cardiac outcomes (LVEDV, LVESV, LVEF, SV) before analysis. We carried out the univariable analysis using each coloc or Mendelian SNP GRS as exposure in turn. Finally, we deployed multivariable models which incorporated either the two coloc- or two Mendelian-based SNP subset GRS to quantify the direct effect of each subset on the outcome.

To estimate the type 1 error rate due to the fraction of shared individuals between our GRS construction GWAS datasets and the cases and controls in the UK Biobank sample, we used the "sample overlap" web app



(<https://sb452.shinyapps.io/overlap/>). The results indicated that, based on the strong F-statistics of our instruments, the overlap of samples is unlikely to cause significant bias in our analyses (type 1 error rate < 0.05).

### *Sensitivity analyses*

We also used GWAS exposure and outcome datasets described above in two-sample MR analysis. Firstly, we carried out univariable MR analysis, using all top SNPs for a given exposure and then Mendelian or coloc SNP subsets separately. Secondly, for coloc SNP subsets we ran multivariable MR including both tissue subsets to obtain mutually adjusted “independent” effect as per Leyden et al. (2022). In this analysis, each instrument’s effect size is weighted by coloc’s  $PP_{H4}$  in its target tissue. We were not able to carry out multivariable MR for Mendelian SNP subsets due to lack of a suitable metric by which to scale individual variants’ effects. We chose the “TwoSampleMR”<sup>50</sup> R package for the standard inverse-variance weighted, mode-weighted, median-weighted and MR-Egger MR analyses as well as multivariable MR analysis. We then calculated instrument’s strength (F-statistics,  $R^2$ ) and heterogeneity (Cochran’s Q and  $I^2$ )<sup>63</sup>.

In addition to including a negative control for the outcome (AMD), we also attempted to provide a negative control SNP partitioning in the Mendelian disease context. We subset the exposure SNPs by a feature which was not expected to biologically influence the outcome (mode of inheritance: autosomal dominant or recessive) and which was not significantly enriched in any Mendelian blood pressure SNP set. We did not run this control example for BMI exposure, since we did find enrichment of autosomal dominant disease-assigned SNPs among mental health disease SNPs in the dataset, mostly driven by intellectual development disorders ( $\chi^2=10.3$ ,  $N=71$ ,  $p$ -value=0.001, **Supplementary Table 3**).

### *Random sampling of SNP subsets*

Finally, we decided to empirically determine how often the difference in MR estimates between randomly drawn SNP subsets equals or exceeds the one observed for Mendelian or coloc-partitioned instruments. When simulating subsets for comparison with Mendelian disease-partitioned instruments, we drew random  $n_1$  SNPs without replacement and then separately  $n_2$  SNPs without replacement (where  $n_1$  and  $n_2$  correspond to the number of SNPs in the original SNP subsets) to represent two SNP subsets which can randomly overlap. For coloc, the procedure for drawing SNPs was modified to randomise the association between SNP, tissue and colocalisation probability. In that case, we first randomly permuted all  $PP_{H4}$  values across both tissues and variants, following which we extracted SNPs with  $PP_{H4}$  values above the chosen threshold (0.9 in our analysis, and 0.8 in replication of Leyden et al. (2022), see below) to be used as two random SNP subsets. We then ran equivalent 1 sample and 2 sample MR analyses as for the “true” coloc and Mendelian SNP subsets. The entire procedure was repeated 1,000 times per each exposure-outcome and analysis type combination.

### *Replication of Leyden et al. (2022)*

We re-analysed the Leyden et. al (2022)<sup>11</sup> dataset for select outcomes to compare their MR results using BMI exposures assigned by coloc to the adipose (86 SNPs)

and brain (140 SNPs) tissues with our Mendelian disease-partitioned approach. We also established the extent to which the difference between coloc-partitioned SNP subsets in the MR analyses could emerge by chance, using random instrument re-sampling as described above.

## Results

### *Assignment of variants to pathways - Mendelian*

Using MendelVar, we assessed overrepresentation of Mendelian disease genes (and their symptoms) in strong LD with blood pressure GWAS top loci (**Figure 1**). In particular, we were interested in investigating enrichment of Mendelian phenotypes related to kidney as opposed to vasculature. The two most enriched terms found were “abnormal renal morphology” (DBP: 90 SNPs, p-value=8x10<sup>-5</sup>; SBP: 84 SNPs, p-value=2.8x10<sup>-4</sup>, **Supplementary Table 4-7**) and “abnormal blood vessel morphology” (DBP: 79 SNPs, p-value=1.2x10<sup>-4</sup>; SBP: 83 SNPs, p-value=2x10<sup>-5</sup>) from the Human Phenotype Ontology (HPO)<sup>64</sup>. The instruments in the “renal” and “vessel” subsets showed substantial overlap: 37 for diastolic SNPs and (p-value= 1.2x10<sup>-4</sup>) 42 for systolic SNPs (p-value=2x10<sup>-6</sup>, **Supplementary Figure 1A-B**).

We then used independent ontologies (featured in ConsensusPathDB<sup>58</sup> and ToppGen<sup>57</sup> resources) not related to Mendelian disease to check if they provide orthogonal enrichment evidence for the role of the assigned genes in a given pathway or tissue (**Supplementary Tables Enrichment (STE) and Supplementary Results**). We found that kidney-related terms were significantly enriched in the “renal” gene set (**Supplementary Dataset STE 1-4**), e.g. renal system development (Gene Ontology<sup>65</sup>, DBP q-value: 2.6x10<sup>-9</sup>; SBP q-value: 4.7x10<sup>-6</sup>). However, we also observed a significant overrepresentation of gene sets related to metabolism, hormonal regulation, type 2 diabetes and cancer. Our “vessel” gene set was found to contain a strong overrepresentation of cardiovascular terms across many ontologies (**Supplementary Dataset STE 5-8**): e.g. blood vessel development (Gene Ontology, DBP q-value: 7.7x10<sup>-6</sup>; SBP q-value: 4.8x10<sup>-10</sup>). To a lesser extent, we also found enrichment of kidney-related terms (*renin secretion*, *EPO signalling pathway*) in the “vessel” gene set and cardiovascular-related terms (*blood vessel development*, *heart development*) in the “renal” set which is not unexpected given substantial overlap of “renal” and “vessel” SNPs.

For BMI, we used the Alliance of Genome Resources slim (28 general disease types) version of Disease Ontology<sup>66</sup> where the top most enriched categories were “disease of metabolism” (45 SNPs, p-value=8.8x10<sup>-4</sup>, **Supplementary Tables 8-9**) and “disease of mental health”/“developmental disorder of mental health” (39 SNPs, p-value=4.1x10<sup>-3</sup>). Choosing this ontology allowed us to contrast the contribution of variants related to metabolism and brain function, which have previously shown distinct patterns across disease outcomes<sup>11</sup>. The “metabolic” and “mental health” SNP subsets were largely disjoint, with only 6 shared SNPs (p-value=0.71, **Supplementary Figure 1 C**). The good separation of “mental health” and “metabolic” gene sets was reflected in the top terms enriched in non-Mendelian disease functional ontologies (**Supplementary Dataset STE 9-12**). The most enriched terms in the “mental health” set related to the brain, in particular synaptic signalling, e.g. neuronal system (Reactome<sup>67</sup>, q-value: 1x10<sup>-3</sup>). Reassuringly, the

strongest enrichment of terms in the “metabolic” gene set related to metabolism and type 2 diabetes, e.g. metabolism (Reactome, q-value=  $5 \times 10^{-11}$ ).

#### *Assignment of variants to tissues – coloc*

We then applied the previously proposed<sup>11</sup> approach with colocalisation used to assign lead blood pressure SNPs to tissues as an alternative method to investigate the difference in SNP sets with roles primarily in kidney or vasculature (**Supplementary Tables 10-11**). Use of high minimum posterior probability of colocalisation threshold ( $PP_{H4} > 0.9$ ) for blood pressure top loci resulted in assignment of 117 SNPs to the aorta or coronary (referred to as “artery” collectively) tissue for DBP and 132 for SBP. In kidney, 87 SNPs colocalised in the glomerular or tubulointerstitial tissues (referred to as “nephro” collectively) for DBP and 77 for SBP (**Supplementary Tables 12-13**). Among “artery” SNPs, the majority of loci colocalised in the aorta (108 for SBP, 126 for DBP; **Supplementary Figure 1 D-E**), and the loci shared between the aorta and coronary artery amounted to approximately one-third of all. Among “nephro” SNPs, tubulointerstitial tissue dominated with 65 and 64 SNPs for DBP and SBP (**Supplementary Figure 1 D-E**), but less sharing between the two tissue types was found than in the “artery” subsets – one-quarter and one-fifth was common to both tubulointerstitial and glomerular in the DBP and SBP variant sets, respectively. Comparison of “artery” and “nephro” SNP sets revealed some overlap, with 36 SNPs shared in DBP ( $p\text{-value} < 1 \times 10^{-6}$ ) and 42 SNPs shared in SBP ( $p\text{-value} < 1 \times 10^{-6}$ ).

Coloc-derived assignment showed limited alignment with functional pathways relevant to each tissue (**Supplementary Dataset STE 13-20**), with very few weakly enriched terms found overall: 4 for “artery” SNPs in DBP (hypertrophic cardiomyopathy, ACE inhibitor pathway, metabolism of lipids, mitochondrial electron transport chain; q-value=0.029-0.047), 3 for “nephro” SNPs in DBP (O-glycosylation of TSR domain-containing proteins, aquaporin-mediated transport, transport of small molecules; q-value=0.024-0.047) and 3 cardiomyopathy terms were enriched for among “nephro” SBP genes (q-value=0.017-0.024).

BMI colocalization results were obtained from Leyden et al. (2022). Among those, 140 SNPs were assigned to the brain and 86 to the adipose tissue, with 43 overlapping ( $p\text{-value} < 1 \times 10^{-6}$ ). Enrichment of genes with colocalisation evidence for BMI in adipose and brain tissue sets showed limited overlap with biologically relevant terms, especially for the brain (**Supplementary Dataset STE 21-24**); this replicated analyses previously described in Leyden et al. (2022).

#### *Comparison of Mendelian- and coloc- derived SNP sets*

Finally, comparison of SNP sets derived using Mendelian- and coloc- assignment methods showed that they are largely distinct and could potentially offer orthogonal evidence (**Supplementary Figure 1 F-G**). Altogether, 33 DBP ( $p\text{-value}=0.35$ ) and 34 SBP ( $p\text{-value}=0.41$ ) SNPs were found in both Mendelian and coloc-based SNP subsets, as opposed to 234 and 234 non-shared SNPs, accordingly. Similar observations were made for BMI (**Supplementary Figure 1 H-I**): only 5 out of the 45 “mental health” Mendelian SNPs were shared with coloc groupings as assigned previously by the Leyden et al. (2022) study – 4 with “brain” and 1 with both



“adipose”/“brain” (adipose p-value=0.98, brain p-value=0.77). For “metabolic” Mendelian SNPs higher overlap was found: 17 out of 39 SNPs were shared with coloc subsets (p-value=0.02): 3 SNPs with “adipose”, 5 SNPs with both “adipose”/“brain” and, unexpectedly, 9 SNPs with “brain” colocating SNPs (adipose p-value=0.06, brain p-value=0.006).

#### *Pathway-based Mendelian Randomization analyses for blood pressure*

Having established the Mendelian disease- and coloc-derived SNP sets for blood pressure, we then proceeded to use them as exposures in one sample MR studies carried out in individuals of European ancestry in the UK Biobank. We focussed on selected cardiometabolic outcomes (**Supplementary Table 2**), as they have been firmly established to be causally influenced by blood pressure<sup>22,68-70</sup> and were previously evaluated using the coloc-based approach with respect to BMI by Leyden et al. (2022).

We confirmed that DBP and SBP have a strong causal relationship with coronary heart disease (**Figure 2, Supplementary Table 14**):  $OR_{95\%CI}=1.09-1.11$  and  $OR_{95\%CI}=1.06-1.07$ , respectively. Selecting Mendelian-based SNP subsets (“renal” and “vessel”) and adjusting for shared effect in multivariable MR analysis (**Supplementary Table 15**) containing both subsets indicates that the “vessel” subset is driving the positive effect of DBP and SBP on CHD (DBP:  $OR=1.17$ ,  $OR_{95\%CI}=1.14-1.21$ , p-value= $1.5 \times 10^{-28}$ ; SBP:  $OR=1.13$ ,  $OR_{95\%CI}=1.11-1.15$ , p-value= $2 \times 10^{-43}$ ), relative to “renal” subset (DBP:  $OR=1.04$ ,  $OR_{95\%CI}=1.01-1.07$ , p-value= $6.2 \times 10^{-3}$ ; SBP:  $OR=1.01$ ,  $OR_{95\%CI}=0.99-1.02$ , p-value= $1.9 \times 10^{-2}$ ), with 95% confidence intervals of the two subsets distinct and non-overlapping.

However, comparison with coloc-based subsets (**Supplementary Table 16**) revealed the opposite trend with the kidney “nephro” subset (DBP:  $OR=1.11$ ,  $OR_{95\%CI}=1.08-1.14$ , p-value= $2.8 \times 10^{-16}$ ; SBP:  $OR=1.08$ ,  $OR_{95\%CI}=1.06-1.09$ , p-value= $6.8 \times 10^{-29}$ ) showing a stronger effect size than the “artery” SNP subset in the MVMR analysis (DBP:  $OR=1.07$ ,  $OR_{95\%CI}=1.05-1.10$ , p-value= $6.8 \times 10^{-9}$ ; SBP:  $OR=1.06$ ,  $OR_{95\%CI}=1.04-1.07$ , p-value= $3 \times 10^{-18}$ ). All of the coloc exposures provided evidence of causal effect on CHD for SBP and DBP, unlike Mendelian “renal” SNP subset for SBP which does not seem to causally influence CHD (p-value=0.19).

Limiting exposures to SNPs specific to each subset resulted in similar causal estimates as when using all SNPs. We also show similar results for myocardial infarction (**Supplementary Figure 2**) which is closely genetically correlated to CHD (**Supplementary Figure 3**). Point estimates for MI display greater uncertainty and while the difference between Mendelian disease-partitioned exposures persists, it was not apparent for “artery” and “nephro” coloc-based subsets.

On the other hand, coloc and Mendelian-based SNP subsets showed consistent direction of effect for an important cardiac function outcome, left ventricular stroke volume (SV, **Figure 3**). Analyses including all SNPs show reduction in stroke volume on increase in DBP (beta=-0.015,  $\beta_{95\%CI}=(-0.021,-0.008)$ , p-value= $2.9 \times 10^{-6}$ ), but the opposite for SBP (beta=0.006,  $\beta_{95\%CI}=(0.002,0.01)$ , p-value= $2.7 \times 10^{-3}$ ). There is limited support for “renal” ( $\beta_{95\%CI}=(-0.013,0.034)$ ) and “nephro” ( $\beta_{95\%CI}=-0.011,0.034$ ) diastolic BP exposures being causally associated with SV. We find that

it is the “vessel” ( $\beta = -0.052$ ,  $\beta_{95\%CI} = (-0.076, -0.028)$ ,  $p\text{-value} = 2.8 \times 10^{-5}$ ) and “artery” diastolic BP ( $\beta = -0.03$ ,  $\beta_{95\%CI} = (-0.051, -0.009)$ ,  $p\text{-value} = 5.6 \times 10^{-3}$ ) exposures that drive the negative effect of DBP on stroke volume. Mendelian “vessel” SNPs also have a certain negative effect on systolic BP ( $\beta_{95\%CI} = (-0.034, -0.004)$ ) while the “renal” SNPs contribute to the overall positive effect of SBP ( $\beta_{95\%CI} = (0.0003, 0.0289)$ ), however there is no clear distinction for coloc-partitioned exposures.

Dissecting the effect of various SNP subsets on type 2 diabetes reveals a heterogeneous landscape (**Figure 4**). We first confirmed the increase in the odds of T2D associated with elevated blood pressure (DBP  $OR_{95\%CI} = 1.02\text{-}1.04$ ; SBP  $OR_{95\%CI} = 1.03\text{-}1.04$ ). Next, we found that Mendelian-disease partitioned subsets show comparable effect on the risk of T2D for DBP, but the 95% confidence intervals are wide. Whilst “vessel” SBP exposure produces higher odds of T2D ( $OR = 1.07$ ,  $OR_{95\%CI} = 1.05\text{-}1.09$ ,  $p\text{-value} = 1.5 \times 10^{-12}$ ) than “renal” SBP exposure ( $OR = 1.03$ ,  $OR_{95\%CI} = 1.01\text{-}1.05$ ,  $p\text{-value} = 4.2 \times 10^{-4}$ ), effect size estimates are partially overlapping. In agreement with this pattern, coloc-based “artery” SNPs have a positive causal effect on the risk of T2D for both DBP ( $OR = 1.08$ ,  $OR_{95\%CI} = 1.05\text{-}1.11$ ,  $p\text{-value} = 4.3 \times 10^{-8}$ ) and SBP ( $OR = 1.06$ ,  $OR_{95\%CI} = 1.04\text{-}1.07$ ,  $p\text{-value} = 8.6 \times 10^{-17}$ ), while null effect was found for “nephro” SNPs (DBP  $OR_{95\%CI} = 0.97\text{-}1.03$ ; SBP  $OR_{95\%CI} = 0.99\text{-}1.02$ ).

Repeating the analysis above and analysing each exposure subset independently using one sample MR univariable framework (**Supplementary Table 15-16**) produces similar patterns between SNP subsets, with any differences in effect estimates accentuated further in MVMR analyses.

#### *Sensitivity analyses for blood pressure*

We used summary-level statistics based on independent samples for exposure and outcome to see if two-sample MR returns results in agreement with those obtained in one-sample MR (**Supplementary Table 17-19**). In general, we discovered that two-sample analyses had less power to detect any differences in effect between SNP subsets, but direction of effect was consistent with that observed in one-sample MR. The one exception was a more pronounced causal effect of “nephro” SBP exposure on myocardial infarction ( $OR = 1.04$ ,  $OR_{95\%CI} = 1.02\text{-}1.06$ ,  $p\text{-value} = 2.3 \times 10^{-4}$ ) relative to “artery” ( $OR = 1$ ,  $OR_{95\%CI} = 0.98\text{-}1.02$ ,  $p\text{-value} = 1$ ), which corresponds to similarly dominant role of “nephro” SBP SNPs in one-sample MVMR for the related outcome – CHD.

We found that mean absolute effect sizes tend not to differ significantly between all SNPs and Mendelian-/coloc-based exposure subsets (**Supplementary Table 20**). Among all SNP subsets, mean F-statistics was  $> 60$  suggesting low risk of weak instrument bias.

We then proceeded to evaluate changes in the average heterogeneity ( $Q_{het} = Q / (Q_{df} - 1)$ ) as we expected that it should decrease within tissue- or pathway-based SNP subsets in comparison to all SNPs representing a variety of pleiotropic processes. However, we did not detect a downward trend in average heterogeneity which instead varied in unexpected ways. For example, we found  $Q_{het}$  to be lower among

Mendelian-partitioned and coloc-partitioned exposures relative to all DBP SNPs (**Supplementary Table 21-23**). However, for SBP Mendelian “renal” and coloc “artery” SNP subsets showed higher  $Q_{het}$  than all SBP SNPs, while Mendelian “vessel” and coloc “nephro” SNP subsets were the opposite.

Next, we repeated our analyses using a negative outcome phenotype: age-related macular degeneration<sup>45</sup>. The one-sample and two-sample MR results (**Supplementary Table 15-16, 18-19**) did not indicate strong evidence of causal effect for our coloc- or Mendelian- based SNP subsets, in line with our expectations. We did find one weak non-null result, however, only at a nominal level (**Supplementary Figure 4**) for SNPs specific to the Mendelian “renal” subset in DBP and SBP (DBP OR=1.07, OR<sub>95%CI</sub>=1-1.14, p-value=0.03; SBP OR=1.04, OR<sub>95%CI</sub>=1.00–1.08, p-value=0.03).

Negative control partitioning of exposures by main modes of Mendelian disease inheritance (autosomal “dominant” or “recessive”) showed no association with the Mendelian “renal”-“vessel” subdivision (**Supplementary Figure 5 A, B**) for DBP (p-value=0.13) and SBP (p-value=0.4, **Supplementary Table 3**). We also assumed that this negative control feature should not be associated with biologically meaningful pathway- or tissue- based partitioning of exposure SNPs, which may not hold. We did indeed detect a substantial difference between “dominant” and “recessive” SNPs sets when focussing on CHD as outcome in one sample MR analysis (**Supplementary Figure 6**). For instance, all “dominant” diastolic BP SNPs had a higher effect on CHD (OR=1.12, OR<sub>95%CI</sub>=1.09-1.14, p-value=3.9x10<sup>-23</sup>) than all “recessive” SNPs (OR=1.08, OR<sub>95%CI</sub>=1.05-1.10, p-value=1.8x 10<sup>-11</sup>) (**Supplementary Table 24**). However, when limiting ourselves to exposure SNPs “specific” to each subset, the opposite conclusion was obtained with higher effect size seen for “recessive” SNPs (OR=1.22, OR<sub>95%CI</sub>=1.17-1.26, p-value=3.9x10<sup>-28</sup>) than “dominant” (OR=1.06, OR<sub>95%CI</sub>=1.03-1.09, p-value=3.4x10<sup>-4</sup>). No meaningful difference between the SNP subsets was observed in two-sample MR results (**Supplementary Table 25, 26**). In addition, much less variation across SNP subsets was found for the T2D outcome (**Supplementary Figure 7**), with the exception of one outlier: specific “recessive” SBP SNP subset.

#### *Pathway-based Mendelian Randomization analyses for body mass index*

We found that the Mendelian disease-partitioned BMI SNPs showed the largest difference between “mental health” and “metabolic” SNPs for atrial fibrillation. All BMI SNPs confirmed a moderate effect on AF (OR=1.05, OR<sub>95%CI</sub>=1.04-1.06, p-value=4.6x10<sup>-55</sup>; **Supplementary Table 27**). This effect size was then matched by “mental health” SNPs in one sample multivariable MR analysis (OR=1.05, OR<sub>95%CI</sub>=1.02-1.08, p-value=5.6x10<sup>-4</sup>, **Supplementary Table 28**) but a larger effect size was found for “metabolic” exposures (OR=1.10, OR<sub>95%CI</sub>=1.07-1.13, p-value=3.3x10<sup>-10</sup>). Similar magnitude of differences between “all”, “metabolic” and “mental health” exposures was obtained in 1 sample univariable analyses (**Supplementary Table 28**) and 2 sample MR analyses (**Supplementary Table 29-30**). This is in contrast to coloc-based SNP stratification carried out by Leyden et al. (2022) and replicated here (**Supplementary Table 31-34**), where the “brain” SNP subset was found to increase the risk of AF more strongly (1 sample MVMR

OR=1.04, OR<sub>95%CI</sub>=1.02-1.06, p-value=6.4x10<sup>-6</sup>, **Supplementary Table 32**), than the “adipose” SNP subset (1 sample MVMR OR=1.02, OR<sub>95%CI</sub>=1-1.04, p-value=0.08) but the effect size difference was relatively attenuated.

#### *Sensitivity analyses for body mass index*

Average heterogeneity ( $Q_{het}$ ) of “mental health” and “metabolic” BMI SNPs for AF was close to average heterogeneity for all BMI SNPs (**Supplementary Table 35-36**) – 2.5, 2.05, 2.19, respectively. We also did not uncover evidence for systematically reduced average heterogeneity among SNP subsets across other outcomes.

Results from negative control outcome (AMD) indicated that “mental health” and “metabolic” SNP subsets analysed independently in 1 sample/2 sample univariable MR or in conjunction with each other (1 sample multivariable MR) show no evidence for causal role, as per expectations (**Supplementary Table 28, 30**).

#### *Random sampling of SNP subsets*

Overall, we noticed that one sample MVMR results showed a variety of effect size differences between exposure SNP subsets, ranging from negligible (e.g. diastolic BP Mendelian partitioning with respect to T2D), medium with overlapping confidence intervals (systolic BP Mendelian partitioning with respect to T2D) to totally distinct (systolic BP Mendelian partitioning with respect to CHD). Conclusions from our negative control outcomes and exposures were unclear as we sometimes found substantial differences in effects between Mendelian and coloc exposure subsets which appeared not to satisfy our null expectations. Therefore, we found it difficult to establish whether any differences detected were likely to be both biologically meaningful and not driven by chance assortment. To address this issue, we employed a simulation technique, where we re-ran our analysis pipeline 1000 times for each exposure-outcome pair using 2 subsets of SNP sampled randomly from across all the exposure SNPs. This allowed us to quantify the frequency of the absolute differences in effect sizes such as observed in our MR analysis (or greater) relative to background (in the process obtaining two-tailed p-value with the floor value of 10<sup>-3</sup>), and empirically derive 95% confidence intervals of the difference (**Supplementary Table 37**).

**Figure 5** provides an overview of this sensitivity approach run for all the exposures (BMI, SBP, DBP), SNP subsets (Mendelian, coloc), MR methods (1 sample or 2 sample) and MR models (univariable, multivariable). Across all the MR analyses with blood pressure as exposure, we find strong evidence for the differential effect of Mendelian “renal” and “vessel” exposures on CHD (p-value range: 0.001-0.042) along with MI. In addition, coloc “artery” and “nephro” systolic BP subsets show a strong difference of effect on T2D, CHD and MI (p-value range: 0.007-0.042) but the latter two only in the two-sample MR setting. Significant evidence based on only 1 sample MVMR was found among coloc-based SNP subsets for diastolic BP and T2D, as well as BMI and CHD/MI. Both one-sample and two-sample MVMR analyses support a differential effect of coloc-based “adipose” and “brain” SNPs on BMI (p-value=0.001-0.013).

We also find weaker evidence for a difference between “renal” and “vessel” blood pressure subsets with respect to the stroke volume outcome (1 sample MVMR SBP



p-value= 0.043, DBP p-value=0.025). There was little evidence against the null hypothesis (of chance SNP assortment) for any of the BMI or BP pathway-based results with respect to atrial fibrillation, heart failure, stroke, LVEDV, LVESV or LVEF, with the exception of the following two-sample univariable MR analyses involving Mendelian disease-partitioned instruments: BMI effect on atrial fibrillation (p-value=0.012), systolic and diastolic BP effect on heart failure (p-value < 0.02).

## Discussion

In this investigation, we employed MR to dissect the causal associations between blood pressure or body mass index and cardiometabolic traits, leveraging biological pathway information. To do so, we classified genetic instruments into Mendelian disease categories with MendelVar or assigned SNPs to tissues through colocalization. Our findings underscore the significance of renal and vascular genes in blood pressure-related conditions, revealing distinct sets of SNPs from Mendelian and colocalization methods. Notably, while both Mendelian and colocalisation-based SNP sets implicated diastolic and systolic BP in coronary heart disease, they emphasized different physiological routes – vessel and kidney, respectively. We report similarly contrasting results for the effect of BMI on atrial fibrillation, when comparing estimates using Mendelian- (“metabolic”-led) and coloc- (“brain”-led) derived instrument subsets. Sensitivity analyses emphasized the consistent direction of effects in one-sample and two-sample MR (univariable and multivariable) but showed some variability in magnitude. We also conducted simulations to assess the probability that differences in causal estimates between SNP subsets arose by chance, and these confirm the validity of interpretation of our main findings.

In the main results, we presented examples of representative cardiac traits – CHD and SV, which as shown in Supplementary Figure 2 are highly genetically correlated with MI and LVEDV/LVESV, respectively. As expected, and demonstrated in Figure 5, magnitude of effect size difference between Mendelian disease- and coloc-partitioned SNP subsets for those related exposures is highly congruent. For ventricular function traits, we only focus on one-sample MR as two-sample results can be biased due to sample overlap of UK Biobank individuals with our exposure sample. The interesting outlier effect of increased importance of “metabolic” SNPs relative to “mental health” observed for atrial fibrillation in two-sample univariable MR could be also potentially affected by the sample overlap bias.

A certain limitation to the current Mendelian disease-partitioning approach is the inability to model the subsets jointly using the multivariable approach in a two-sample MR setting. Altogether, we find a comparable number of results (14) with empirically determined significant differences (p-value < 0.05) in one-sample (univariable: 3, multivariable: 11) versus two-sample MR (univariable: 10, multivariable: 4), with univariable and multivariable results mostly agreeing in the direction and magnitude of effect for the same SNP subsets. In some cases, we found reassuring agreement across all MR types and model specifications: Mendelian subsets for diastolic BP versus CHD, coloc subsets for systolic BP and BMI versus T2D. In certain cases, pronounced differences were only observed in a single setting using one model type but not the others, such as: coloc-based partitioning for BMI with respect to CHD (1 sample multivariable), Mendelian



disease-based partitioning of DBP and SBP with respect to heart failure (2 sample univariable). In the first instance, one may expect for some differential effect to become apparent only in the multivariable analysis when adjusting for the shared effect. Nevertheless, lack of effect difference replication in corresponding two-sample MVMR is troubling, as the instrumental variable subsets were found not to suffer from weak instrument bias previously<sup>12</sup>. The second example involving effect difference observed only in two-sample univariable analysis should be treated as less reliable due to poor replication in one-sample setting, especially in multivariable analysis evaluating direct effects.

Therefore, to gain a better understanding of tissue- or pathway-partitioned effects multiple types of MR analyses should be undertaken, whenever possible. For instance, in a recent publication, Leyden et al. (2023)<sup>12</sup> applied the coloc method to study the effect of adipose and brain tissues in BMI on cancer in two-sample setting only, due to availability of one-sample datasets with sufficiently large numbers of cases. While they noted that adipose-derived variants may predominantly drive the association with endometrial cancer, our random sampling control found weak support for a significant difference between adipose and brain in a multivariable analysis (p-value=0.161).

We expected that stratification of SNPs by pathway and tissues should result in reduced heterogeneity compared to using all SNPs, which can capture a wider range of biological processes. However, our coloc- and Mendelian disease-based results did not consistently support this expectation. For some SNP subsets, the heterogeneity was indeed lower, which was in line with the initial hypothesis. However, other SNP subsets showed greater heterogeneity than all SNPs combined, suggesting that those specific pathways or tissues might still be highly pleiotropic or indicate presence of some bias in the method, such as misclassification of SNP functional category.

Our new approach uses enrichment of ontology terms assigned to Mendelian diseases whose causal genes share the same genetic locus as the exposure SNPs, which can sometimes result in ambiguous assignment to appropriate biological pathways. For example, rs12630999 from BMI was allocated to both the “mental health” and “metabolic” set due to location between two neighbouring Mendelian disease genes: *STAG1* and *PCCB*, respectively. In another case, one systolic BP variant (rs3915499) was associated with two different pathways (“renal” and “vessel”) due to the two distinct monogenic diseases caused by disruption of smooth muscle myosin heavy-chain 11 gene<sup>71</sup> whose intron the SNP resides in. Under such a scenario, MendelVar cannot provide a more nuanced prediction regarding the “correct” pathway(s). In this example, “vessel”-only assignment could be more suitable as the associated disease, familial thoracic aortic aneurysm 4 is defined by profound structural abnormalities in the aorta, while a renal symptom (hydronephrosis) is only a marginal feature of megacystis-microcolon-intestinal hypoperistalsis syndrome 2. Nonetheless, we found that in many cases MendelVar unequivocally assigns exposure SNPs to genes causal for disorders with strong links to a single category: “renal” - *CEP164* (nephronophthisis 15), *NR1P1* and *PBX1* (congenital anomalies of kidney and urinary tract syndromes); “vessel” - *EIF2AK4*

(familial pulmonary capillary hemangiomas), *PRDM6* (patent ductus arteriosus), *HTRA1* (cerebral arteriopathy with subcortical infarcts and leukoencephalopathy); “mental health” - *PPP3CA* (developmental and epileptic encephalopathy 91, ACCIID), *KCNMA1* (cerebellar atrophy, developmental delay, and seizures, Liang-Wang syndrome); “metabolic” – *PPARG* (familial lipodystrophy), *SLC2A2* (Fanconi-Bickel syndrome). Correct SNP assignment to the gene in each case was also supported by either presence within its intron or coding region.

Similarly to Darrous et al. (2023)<sup>9</sup>, we conclude that our new method offers complementary, orthogonal SNP stratification to the existing colocalisation approach as evidenced by very weak overlap. Comparison of MR results using both approaches suggests agreement for the dominant role of vasculature-related SNPs in determining the left ventricular stroke volume and the risk of type 2 diabetes, but discordance with regards to coronary heart disease, and influence of different BMI pathways on atrial fibrillation. It is not straightforward to identify the cause of discrepancy. While coloc-based gene selection shows poor enrichment of relevant gene functional categories in general (STE 21-24) unlike Mendelian-based selection, the latter gene categorisation is based on disease symptom ontologies which are a-priori more directly related to gene function than expression pattern. We find that more Mendelian “metabolic” than “mental health” SNPs are shared with coloc “brain” SNPs which could be partially due to weak evidence for involvement of coloc “brain” genes in neurodevelopmental pathways and/or brain-centric expression of some of the key genes regulating metabolism<sup>72,73</sup>. Furthermore, although both Mendelian “metabolic” and coloc “adipose” subsets are enriched for metabolic pathway genes, the “metabolism” term is broad and there could still be stark differences in proportion of metabolic pathways represented in each subset. Disease Ontology used for assignment of genes to Mendelian disease categories for BMI is also quite sparsely annotated as indicated by reduction in the number of pathway-assigned SNPs relative to blood pressure where Human Phenotype Ontology was used; that may well have introduced selection bias.

While this study introduced and evaluated a novel, Mendelian disease-centric approach to dissecting the impact of different biological pathways on complex risk factors in MR, several issues warrant further investigation. Applying the method to other complex traits and physiological pathways could result in more nuanced understanding of shared genetic risk factors in cross-disease analyses. Mendelian disease-partitioning method is reliant on mainly manual curation of disease ontology terms based on descriptions of clinical features<sup>64,66</sup>, and expanding the automation of that time-consuming process could increase power and potentially accuracy of our method. The specificity of the colocalisation method suffers from co-ordinated expression of genes across many tissues which complicates selection of biologically causal tissues over merely tagging but new methods are being developed to address this confounding factor<sup>74,75</sup>. Furthermore, we emphasise that the modest eQTL sample sizes analysed in this study may have reduced the number of instruments we were able to identify with robust colocalization evidence<sup>76</sup>. Next, integrating different features used for exposure partitioning into a single joint model may offer improvement in pathway-based stratification. Here, we used Mendelian and gene

expression information but other orthogonal evidence such as protein-protein interactions<sup>77</sup> and PheWAS<sup>9</sup> could be combined. Such (and other) future new methods would be useful to address discordance in the dominant biological pathways identified for certain disease outcomes between coloc- and Mendelian disease- based instrument partitioning.

## Funding

This work was funded by the UK Medical Research Council (MRC) as part of the MRC Integrative Epidemiology Unit (MC\_UU\_00032/03).

## Competing interests

TRG receives funding from Biogen for unrelated research. TGR is a full-time employee of GlaxoSmithKline outside of this research.

## Author contributions

MKS: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review and editing

TGR: Resources, Writing - review and editing

GML: Resources, Writing - review and editing

TG: Supervision, Methodology, Resources, Writing – review and editing, Funding acquisition

## Data and code availability

GWAS summary statistics availability is provided by relevant publications. We accessed the following summary statistics via OpenGWAS: ebi-a-GCST006414, ebi-a-GCST006906, ebi-a-GCST009541, ieu-a-7, ieu-a-26, ieu-a-798, ieu-b-38, ieu-b-39, ieu-b-40.

## Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 81499.

## Ethics approval

UK Biobank received ethical approval from the Research Ethics Committee (REC reference: 11/NW/0382).

## References

1. Smith, G. D. & Ebrahim, S. Mendelian randomization: prospects, potentials, and limitations. *Int. J. Epidemiol.* **33**, 30–42 (2004).
2. Zheng, J. *et al.* Recent Developments in Mendelian Randomization Studies. *Curr. Epidemiol. Reports* **4**, 330–345 (2017).
3. Kutalik, Z. Commentary on: “The contribution of tissue-specific BMI-associated gene sets to cardiometabolic disease risk: a Mendelian randomization study”. *Int. J. Epidemiol.* **49**, 1257–1258 (2020).

4. Davey Smith, G., Holmes, M. V, Davies, N. M. & Ebrahim, S. Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. *Eur. J. Epidemiol.* **35**, 99–111 (2020).
5. Foley, C. N., Mason, A. M., Kirk, P. D. W. & Burgess, S. MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates. *Bioinformatics* **37**, 531–541 (2021).
6. Liang, X. *et al.* MR-AHC: A fast, efficient and robust method for two-sample summary data Mendelian randomization based on agglomerative hierarchical clustering. *medRxiv* (2023) doi:10.1101/2023.03.18.23287164.
7. Winkler, T. W. *et al.* A joint view on genetic variants for adiposity differentiates subtypes with distinct metabolic implications. *Nat. Commun.* **9**, 1946 (2018).
8. Grant, A. J., Gill, D., Kirk, P. D. W. & Burgess, S. Noise-augmented directional clustering of genetic association data identifies distinct mechanisms underlying obesity. *PLoS Genet.* **18**, e1009975 (2022).
9. Darrous, L., Hemani, G., Smith, G. D. & Kutalik, Z. PheWAS-based clustering of Mendelian Randomisation instruments reveals distinct mechanism-specific causal effects between obesity and educational attainment. *medRxiv* (2023) doi:10.1101/2023.04.06.23288264.
10. Verkouter, I. *et al.* The contribution of tissue-grouped BMI-associated gene sets to cardiometabolic-disease risk: A Mendelian randomization study. *Int. J. Epidemiol.* **49**, 1246–1256 (2020).
11. Leyden, G. M. *et al.* Harnessing tissue-specific genetic variation to dissect putative causal pathways between body mass index and cardiometabolic phenotypes. *Am. J. Hum. Genet.* **109**, 240–252 (2022).
12. Leyden, G. M. *et al.* Disentangling the aetiological pathways between body mass index and site-specific cancer risk using tissue-partitioned Mendelian randomisation. *Br. J. Cancer* **128**, (2023).
13. Hemerich, D. *et al.* Effect of tissue-grouped regulatory variants associated to type 2 diabetes in related secondary outcomes. *Sci. Rep.* **13**, 3579 (2023).
14. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, (2014).
15. Tambets, R., Kolde, A., Kolberg, P., Love, M. I. & Alasoo, K. Extensive co-regulation of neighbouring genes complicates the use of eQTLs in target gene prioritisation. *bioRxiv* (2023) doi:10.1101/2023.09.29.560109.
16. Freund, M. K. *et al.* Phenotype-Specific Enrichment of Mendelian Disorder Genes near GWAS Regions across 62 Complex Traits. *Am. J. Hum. Genet.* **103**, 535–552 (2018).
17. Kim, S. S. *et al.* Improving the informativeness of Mendelian disease-derived pathogenicity scores for common disease. *Nat. Commun.* **2020 111 11**, 1–15 (2020).
18. Burch, K. S. *et al.* Partitioning gene-level contributions to complex-trait

- heritability by allele frequency identifies disease-relevant genes. *Am. J. Hum. Genet.* **109**, 692–709 (2022).
19. Sobczyk, M. K., Gaunt, T. R. & Paternoster, L. MendelVar: gene prioritization at GWAS loci using phenotypic enrichment of Mendelian disease genes. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btaa1096.
  20. Kannel, W. B. Elevated systolic blood pressure as a cardiovascular risk factor. *Am. J. Cardiol.* **85**, 251–255 (2000).
  21. Fuchs, F. D. & Whelton, P. K. High Blood Pressure and Cardiovascular Disease. *Hypertension* **75**, 285–292 (2020).
  22. Malik, R. *et al.* Relationship Between Blood Pressure and Incident Cardiovascular Disease: Linear and Nonlinear Mendelian Randomization Analyses. *Hypertension* **77**, 2004–2013 (2021).
  23. Lindholm, L. H. *et al.* Risk of new-onset diabetes in the Losartan Intervention For Endpoint reduction in hypertension study. *J. Hypertens.* **20**, (2002).
  24. Emdin, C. A., Anderson, S. G., Woodward, M. & Rahimi, K. Usual Blood Pressure and Risk of New-Onset Diabetes: Evidence From 4.1 Million Adults and a Meta-Analysis of Prospective Studies. *J. Am. Coll. Cardiol.* **66**, 1552–1562 (2015).
  25. Aikens, R. C. *et al.* Systolic Blood Pressure and Risk of Type 2 Diabetes: A Mendelian Randomization Study. *Diabetes* **66**, 543–550 (2016).
  26. Warren, H. R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* **49**, 403–415 (2017).
  27. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
  28. Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **51**, 51–62 (2019).
  29. Eales, J. M. *et al.* Uncovering genetic mechanisms of hypertension through multi-omic analysis of the kidney. *Nat. Genet.* **53**, 630–637 (2021).
  30. Tomaszewski, M. *et al.* Kidney omics in hypertension: from statistical associations to biological mechanisms and clinical applications. *Kidney Int.* **102**, 492–505 (2022).
  31. Sheng, X. *et al.* Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. *Nat. Genet.* **53**, 1322–1333 (2021).
  32. Wadei, H. M. & Textor, S. C. The role of the kidney in regulating arterial blood pressure. *Nat. Rev. Nephrol.* **8**, 602–609 (2012).
  33. Ivy, J. R. & Bailey, M. A. Pressure natriuresis and the renal control of arterial blood pressure. *J. Physiol.* **592**, 3955–3967 (2014).
  34. Yu, Z. *et al.* A bidirectional Mendelian randomization study supports causal



- effects of kidney function on blood pressure. *Kidney Int.* **98**, 708–716 (2020).
35. Touyz, R. M. *et al.* Vascular smooth muscle contraction in hypertension. *Cardiovasc. Res.* **114**, 529–539 (2018).
  36. Lüscher, T. F. & Barton, M. Biology of the Endothelium. *Clin. Cardiol.* **20**, II-3-II–10 (1997).
  37. Rees, D. D., Palmer, R. M. & Moncada, S. Role of endothelium-derived nitric oxide in the regulation of blood pressure. *Proc. Natl. Acad. Sci.* **86**, 3375–3378 (1989).
  38. Gibbons, G. H. Endothelial function as a determinant of vascular function and structure: A new therapeutic target. *Am. J. Cardiol.* **79**, 3–8 (1997).
  39. Pober, J. S. & Sessa, W. C. Evolving functions of endothelial cells in inflammation. *Nat. Rev. Immunol.* **7**, 803–815 (2007).
  40. Yengo, L. *et al.* Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
  41. Nielsen, J. B. *et al.* Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).
  42. Shah, S. *et al.* Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, (2020).
  43. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
  44. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
  45. Kuan, V. *et al.* Association of Smoking, Alcohol Consumption, Blood Pressure, Body Mass Index, and Glycemic Risk Factors With Age-Related Macular Degeneration: A Mendelian Randomization Study. *JAMA Ophthalmol.* **139**, 1299–1306 (2021).
  46. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
  47. Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. *bioRxiv* (2020) doi:10.1101/2020.08.10.244293.
  48. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).
  49. Lyon, M. S. *et al.* The variant call format provides efficient and robust storage of GWAS summary statistics. *Genome Biol.* **22**, 32 (2021).
  50. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).

51. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
52. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
53. Lee, P. H., O’Dushlaine, C., Thomas, B. & Purcell, S. M. INRICH: Interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* **28**, 1797–1799 (2012).
54. Han, S. K. *et al.* Mapping genomic regulation of kidney disease and traits through high-resolution and interpretable eQTLs. *Nat. Commun.* **14**, 2229 (2023).
55. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* (80-. ). **369**, 1318–1330 (2020).
56. Wallace, C. Eliciting priors and relaxing the single causal variant assumption in colocalisation analyses. *PLoS Genet.* **16**, 1–20 (2020).
57. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
58. Kamburov, A. & Herwig, R. ConsensusPathDB 2022: molecular interactions update as a resource for network biology. *Nucleic Acids Res.* **50**, D587–D595 (2022).
59. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
60. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
61. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
62. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
63. Bowden, J. & Holmes, M. V. Meta-analysis and Mendelian randomization: A review. *Res. Synth. Methods* **10**, 486–496 (2019).
64. Köhler, S. *et al.* The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* **49**, D1207–D1217 (2021).
65. Carbon, S. *et al.* The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
66. Schriml, L. M. *et al.* The Human Disease Ontology 2022 update. *Nucleic Acids Res.* **50**, D1255–D1261 (2022).
67. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).

68. Vasan, R. S. *et al.* Impact of High-Normal Blood Pressure on the Risk of Cardiovascular Disease. *N. Engl. J. Med.* **345**, 1291–1297 (2001).
69. Aikens, R. C. *et al.* Systolic Blood Pressure and Risk of Type 2 Diabetes: A Mendelian Randomization Study. *Diabetes* **66**, 543–550 (2017).
70. Arvanitis, M. *et al.* Linear and Nonlinear Mendelian Randomization Analyses of the Association Between Diastolic Blood Pressure and Cardiovascular Events. *Circulation* **143**, 895–906 (2021).
71. Amberger, J. S., Bocchini, C. A., Scott, A. F. & Hamosh, A. OMIM.org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* **47**, D1038–D1043 (2019).
72. Ahima, R. S., Qi, Y., Singhal, N. S., Jackson, M. B. & Scherer, P. E. Brain Adipocytokine Action and Metabolic Regulation. *Diabetes* **55**, S145–S154 (2006).
73. Roh, E., Song, D. K. & Kim, M.-S. Emerging role of the brain in the homeostatic regulation of energy and glucose metabolism. *Exp. Mol. Med.* **48**, e216–e216 (2016).
74. Arvanitis, M., Tayeb, K., Strober, B. J. & Battle, A. Redefining tissue specificity of genetic regulation of gene expression in the presence of allelic heterogeneity. *Am. J. Hum. Genet.* **109**, 223–239 (2022).
75. Amariuta, T., Siewert-Rocks, K. & Price, A. L. Modeling tissue co-regulation estimates tissue-specific contributions to disease. *Nat. Genet.* **55**, 1503–1511 (2023).
76. Brotman, S. M. *et al.* Adipose tissue eQTL meta-analysis reveals the contribution of allelic heterogeneity to gene expression regulation and cardiometabolic traits. *bioRxiv* (2023) doi:10.1101/2023.10.26.563798.
77. Kain, J. *et al.* Mendelian randomization and pathway analysis demonstrate shared genetic associations between lupus and coronary artery disease. *Cell reports. Med.* **3**, 100805 (2022).

## Figures legends

**Figure 1.** An overview of this study's workflow for Mendelian disease gene-derived partitioning of genetic instruments and colocalization-derived partitioning of genetic instruments with the aim of investigating pathway-specific effect of blood pressure and body mass index on cardiometabolic traits.

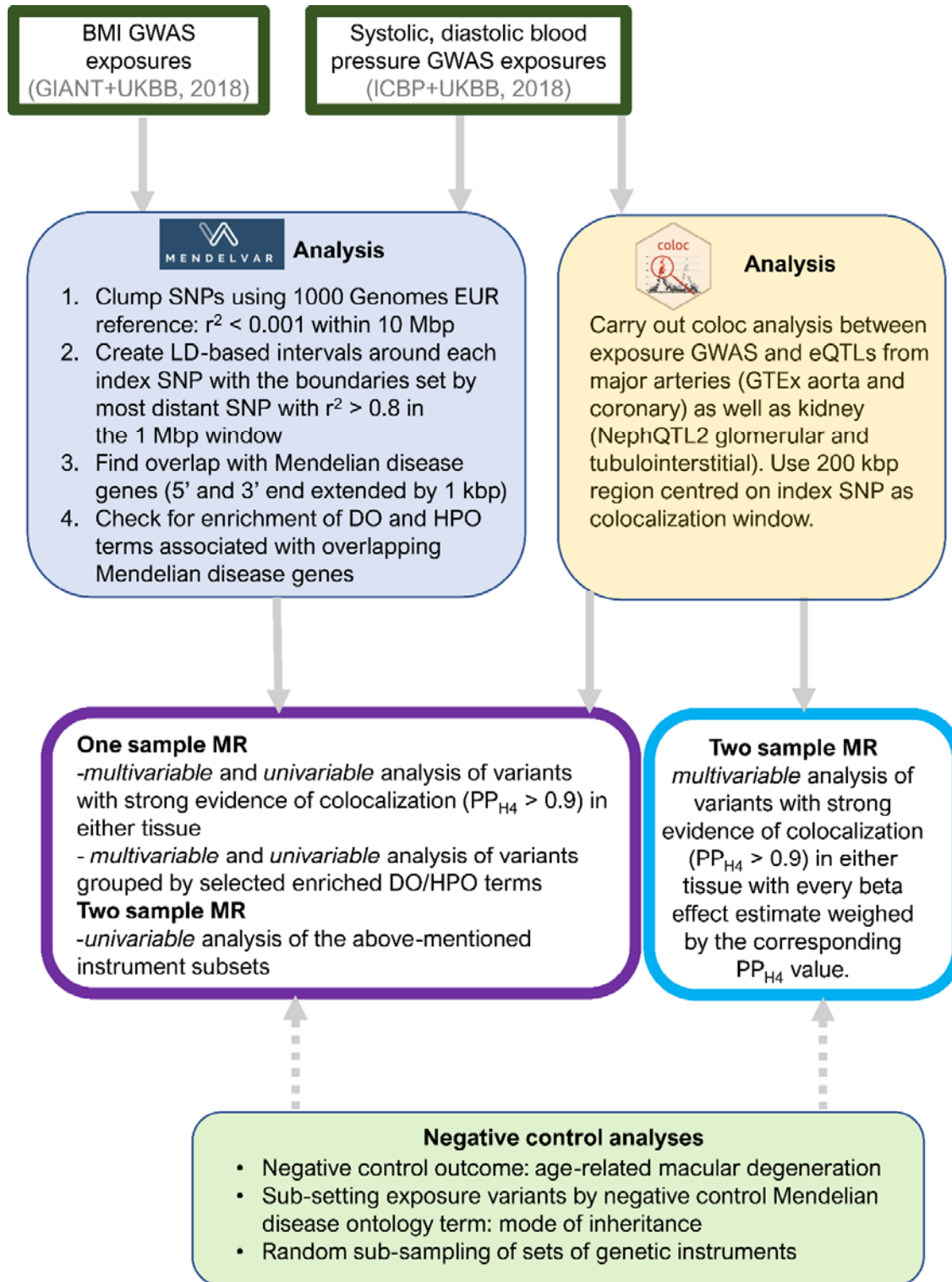
**Figure 2. Coronary heart disease:** one sample Multivariable Mendelian Randomization analysis of the effect of diastolic blood pressure (DBP) and systolic blood pressure (SBP) on CHD using all SNPs, all/specific Mendelian disease partitioned (disease with abnormalities in the *renal* or *blood vessel* system) genetic instruments and all/specific Coloc partitioned (*nephro* – kidney tissues: glomerular and tubulointerstitial, *artery* – aorta and coronary artery tissues) instruments. Effect sizes are scaled to per one SD change in blood pressure.

**Figure 3. Stroke volume:** one sample Multivariable Mendelian Randomization analysis of the effect of diastolic blood pressure (DBP) and systolic blood pressure (SBP) on stroke volume (SV) using all SNPs, all/specific Mendelian disease partitioned (disease with abnormalities in the *renal* or *blood vessel* system) genetic instruments and all/specific Coloc partitioned (*nephro* – kidney tissues: glomerular and tubulointerstitial, *artery* – aorta and coronary artery tissues) instruments. Effect sizes are scaled to per one SD change in blood pressure.

**Figure 4. Type 2 diabetes:** one sample Multivariable Mendelian Randomization analysis of the effect of diastolic blood pressure (DBP) and systolic blood pressure (SBP) on T2D using all SNPs, all/specific Mendelian disease partitioned (disease with abnormalities in the *renal* or *blood vessel* system) genetic instruments and all/specific Coloc partitioned (*nephro* – kidney tissues: glomerular and tubulointerstitial, *artery* – aorta and coronary artery tissues) instruments. Effect sizes are scaled to per one SD change in blood pressure.

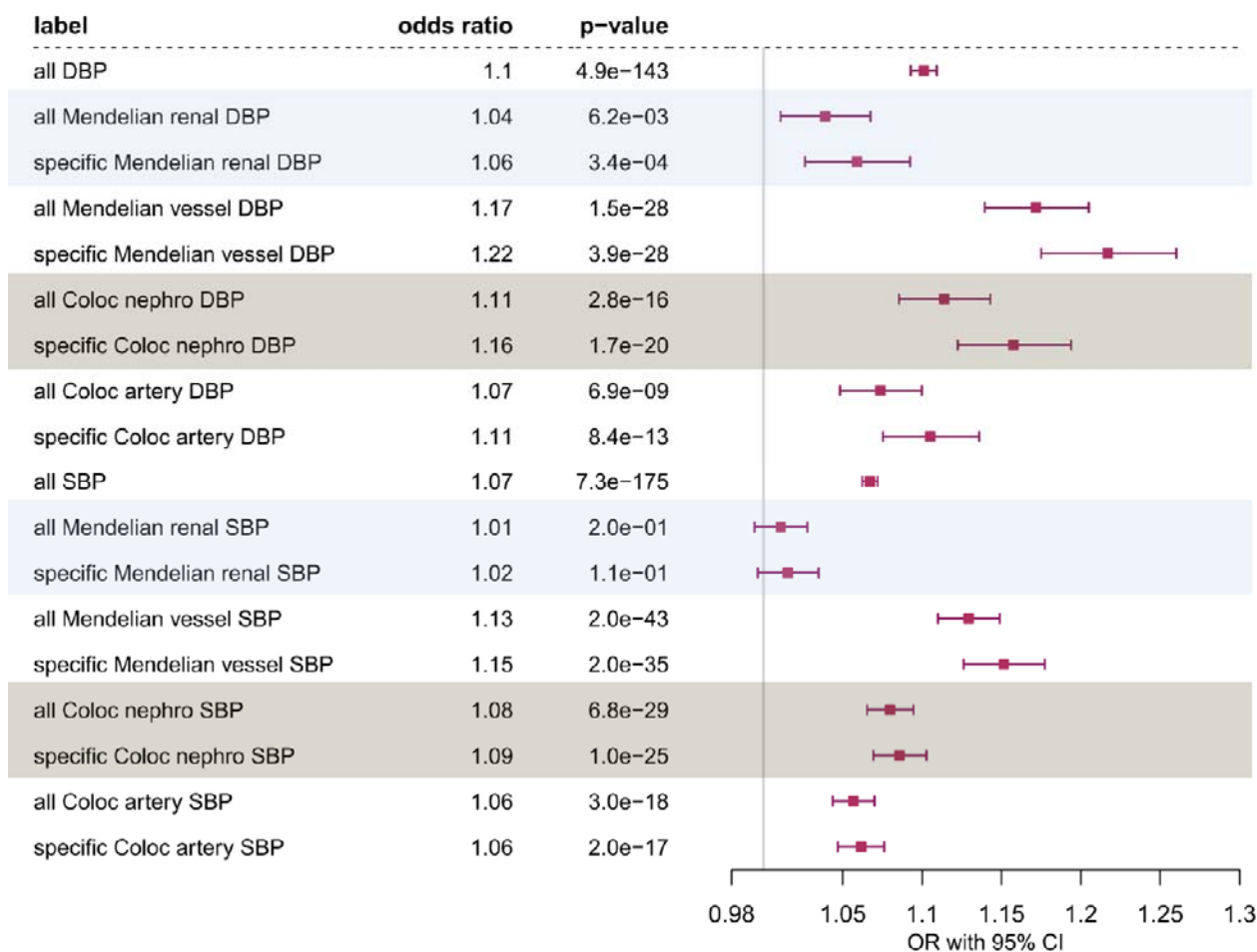
**Figure 5.** Matrix of empirically-derived (1,000 replicates) p-values for distribution of effect size differences between Mendelian disease-partitioned and coloc-partitioned instruments in 1 sample and 2 sample MR setting (Univariable and Multivariable) using body mass index (BMI) and blood pressure – systolic (SBP) and diastolic (DBP) as exposures and cardiometabolic traits as outcomes (AF, CHD, MI, HF, stroke, T2D, LDVEDV, LVEF, LVESV, SV). P-values < 0.05 are highlighted in dark red.

Figure 1

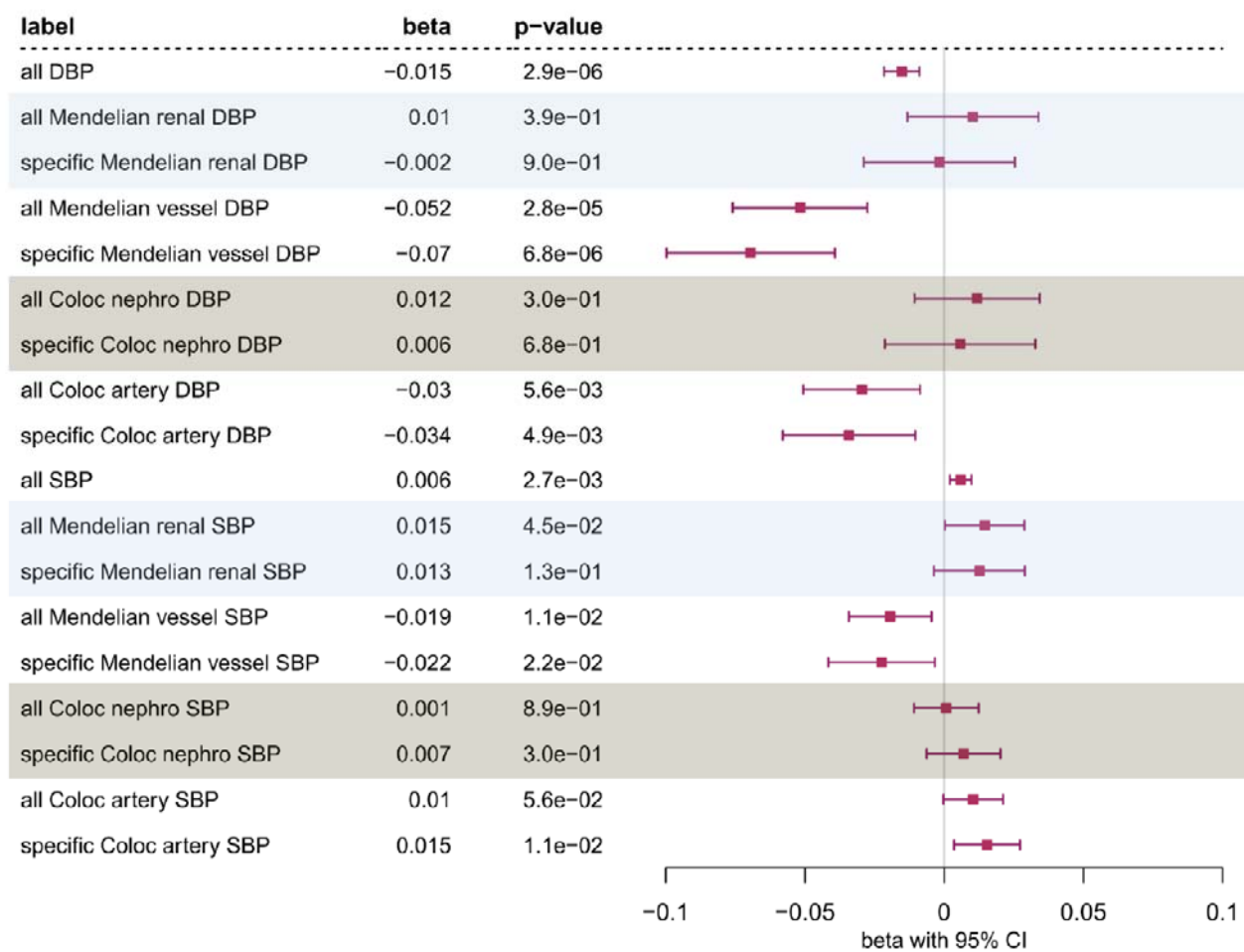




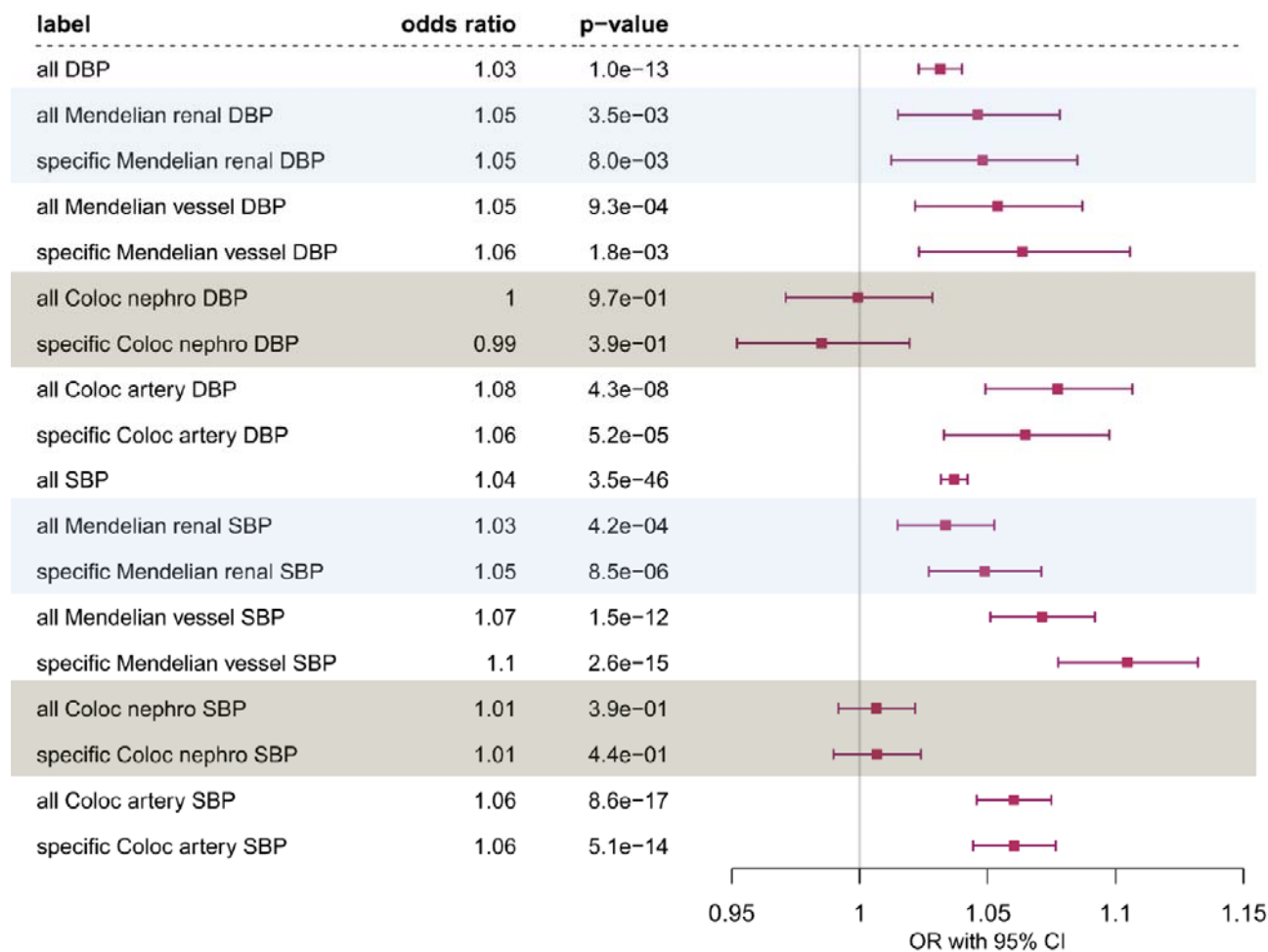
**Figure 2**



**Figure 3**



**Figure 4**



**Figure 5**

exposure	subset	MR	model	CHD	MI	AF	HF	STROKE	T2D	LVEF	LVEDV	LVESV	SV
BMI	Mendelian	1 sample	Multivariable	0.994	0.384	0.157	0.264	0.415	0.502	0.492	0.813	0.756	0.906
BMI	Mendelian	1 sample	Univariable	1.000	0.776	0.213	0.337	0.482	0.544	0.539	0.840	0.784	0.929
BMI	Mendelian	2 sample	Univariable	0.248	0.194	0.012	0.181	0.465	0.150	0.894	0.406	0.622	0.308
BMI_Leyden	Coloc	1 sample	Multivariable	0.022	0.001	0.260	0.162	0.076	0.001	0.431	0.107	0.274	0.133
BMI_Leyden	Coloc	1 sample	Univariable	0.365	0.784	0.729	0.903	0.374	0.112	0.572	0.179	0.346	0.212
BMI_Leyden	Coloc	2 sample	Multivariable	0.409	0.489	0.293	0.092	0.893	0.013	0.656	0.084	0.185	0.161
BMI_Leyden	Coloc	2 sample	Univariable	0.710	0.908	0.787	0.595	0.787	0.279	0.802	0.118	0.207	0.140
DBP	Coloc	1 sample	Multivariable	0.331	0.909	0.715	0.879	0.848	0.025	0.464	0.070	0.098	0.118
DBP	Coloc	1 sample	Univariable	0.389	0.982	0.839	0.990	0.990	0.098	0.604	0.186	0.218	0.248
DBP	Coloc	2 sample	Multivariable	0.311	0.540	0.881	0.934	0.337	0.976	0.475	0.161	0.223	0.102
DBP	Coloc	2 sample	Univariable	0.652	0.825	0.599	0.634	0.112	0.616	0.663	0.337	0.682	0.147
DBP	Mendelian	1 sample	Multivariable	0.002	0.005	0.217	0.270	0.344	0.858	0.074	0.279	0.695	0.025
DBP	Mendelian	1 sample	Univariable	0.032	0.078	0.413	0.467	0.538	0.889	0.234	0.458	0.781	0.151
DBP	Mendelian	2 sample	Univariable	0.037	0.070	0.200	0.011	0.264	0.896	0.903	0.168	0.289	0.149
SBP	Coloc	1 sample	Multivariable	0.287	0.393	0.062	0.667	0.472	0.018	0.060	0.841	0.893	0.500
SBP	Coloc	1 sample	Univariable	0.273	0.397	0.111	0.788	0.612	0.057	0.119	0.854	0.639	0.565
SBP	Coloc	2 sample	Multivariable	0.034	0.020	0.134	0.494	0.061	0.029	0.052	0.557	0.185	0.986
SBP	Coloc	2 sample	Univariable	0.042	0.024	0.152	0.421	0.345	0.007	0.642	0.847	0.970	0.788
SBP	Mendelian	1 sample	Multivariable	0.001	0.001	0.906	0.214	0.465	0.131	0.095	0.130	0.482	0.043
SBP	Mendelian	1 sample	Univariable	0.003	0.020	0.977	0.396	0.632	0.327	0.331	0.326	0.638	0.178
SBP	Mendelian	2 sample	Univariable	0.006	0.014	0.433	0.016	0.041	0.273	0.893	0.423	0.595	0.335