medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Using clustering of genetic variants in Mendelian randomization to interrogate the causal pathways underlying multimorbidity

Xiaoran Liang*[a] Ninon Mounier[a], Nicolas Apfel[b], Sara Khalid[c], Timothy M Frayling[a,d], and Jack Bowden[a], on behalf of the GEMINI Consortium

[a]Genetics of Complex Traits, Exeter Medical School, University of Exeter, Exeter, UK

[b]Department of Economics, University of Southampton, Southampton, UK

[c]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

[d]Faculty of Medicine, Department of Genetic Medicine and Development, CMU, Geneva, Switzerland

November 7, 2023

*correspondence: x.liang2@exeter.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Abstract

Mendelian randomization (MR) is an epidemiological approach that utilizes genetic variants as instrumental variables to estimate the causal effect of a modifiable but likely confounded exposure on a health outcome. This paper investigates an MR scenario in which different subsets of genetic variants identify different causal effects. These variants may aggregate into clusters, and such variant clusters are likely to emerge if they affect the exposure and outcome via distinct biological pathways. In the framework of multi-outcome MR, where a common risk factor causally impacts several disease outcomes simultaneously, these variant clusters can reflect the heterogeneous effects this shared risk factor concurrently exerts on all the diseases under examination. This, in turn, can provide insights into the disease-causing mechanisms underpinning the co-occurrence of multiple long-term conditions, a phenomenon known as multimorbidity. To identify such variant clusters, we adapt the general method of Agglomerative Hierarchical Clustering (AHC) to the summary data MR setting, enabling cluster detection based on the variant-specific causal estimates, using only genome-wide summary statistics. In particular, we tailor the method for multi-outcome MR to aid the elucidation of the potentially multifaceted causal pathways underlying multimorbidity stemming from a shared risk factor. We show in various Monte Carlo simulations that our 'MR-AHC' method detects variant clusters with high accuracy, outperforming the existing multi-dimensional clustering methods. In an application example, we use the method to analyze the causal effects of high body fat percentage on a pair of well-known multimorbid conditions, type 2 diabetes (T2D) and osteoarthritis (OA), discovering distinct variant clusters reflecting heterogeneous causal effects. Pathway analyses of these variant clusters indicate interconnected cellular processes underlying the co-occurrence of T2D and OA; while the protective effect of higher adiposity on T2D could possibly be linked to the enhanced activity of ion channels related to insulin secretion.

Keywords: Mendelian randomization; clustering analysis; multimorbidity; robust MR; hierarchical clustering; heterogeneous causal effects.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Introduction

Mendelian randomization (MR) is a widely used method in epidemiology that leverages genetic variants (usually in the form of single nucleotide polymorphisms, SNPs) as instrumental variables (IV) for estimating the causal effect of a potentially confounded exposure on an outcome [1, 2]. If a genetic variant is sufficiently associated with the exposure, independent of possible confounders of the exposure-outcome relationship, and affects the outcome only through the exposure, then it is a valid instrument for assessing causality [3]. With further parametric assumptions, for example that relationships between all variables are additive and linear, and all variants included as instruments encode a single homogeneous causal effect from the exposure to the outcome, then the causal parameter of interest can be estimated using simple meta-analytic methods based on genome-wide summary statistics [4–6]. In this setting, all the variant-specific causal estimates are expected to target the same, true causal effect, and their 'ratio' estimates (derived as the ratio of the variant-outcome to variant-exposure association), from which the overall meta-analysis is performed, should vary by sampling error alone [7, 8]. Excess 'heterogeneity' amongst the ratio estimates is therefore a sign that one or more of the assumptions has been violated [9].

A major source of excess heterogeneity is undoubtedly *horizontal pleiotropy*, the phenomenon whereby a variant affects multiple traits and therefore is associated with the outcome through pathways other than via the exposure [9, 10]. This has been extensively studied with improved methods for pleiotropy detection [8, 11] and robust estimation [5, 12, 13]. Violation of the causal effect homogeneity assumption, has, by contrast, been far less researched, despite this being a plausible feature of many analyses. For example, it is suspected that general adiposity, which is often proxied by a single trait like body mass index (BMI), exerts a heterogeneous causal effect on type 2 diabetes (T2D) depending on the location of the adipose tissue in the body (e.g. if it is peripheral or visceral) [14]. In this case, variants associated with different physiological aspects of the exposure may

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

target distinct causal effects.

In the presence of excess heterogeneity from both sources, the genetic variants can be grouped into distinct clusters, such that all variants in each cluster indicate the same effect. Several studies have explored variant clusters in the MR framework. It is well recognized that it is impossible to discern whether each cluster embodies genuine causal mechanisms between the exposure and outcome, or is formed due to pleiotropic pathways, without further domain knowledge or modelling assumptions. Therefore, overdispersion caused by both sources can be summarized under an umbrella term such as "clustered heterogeneity", as proposed by Foley et al. [15], or "mechanistic heterogeneity" by Iong et al. [16].

In this paper, we propose a method to identify variant clusters under mechanistic heterogeneity, building upon the Agglomerative Hierarchical Clustering (AHC) method developed by Apfel and Liang [17] in the field of econometrics for IV selection. We adapt the method to the summary-data MR setting, hence referring to it as "MR-AHC", to group variants based on their ratio estimates using genome-wide summary statistics. More notably, we have tailored the method to the multi-outcome MR setting, in which a shared exposure causally impacts several disease outcomes simultaneously. This extension is specially crafted for investigating the causal mechanisms underpinning multimorbidity, which refers to the co-existence of two or more long-term conditions in one individual [18].

A substantial and growing proportion of the adult population is affected by multimorbidity, and it has been recognized as a global priority for health research [19, 20]. It is therefore important to comprehend the underlying disease-causing pathways. Numerous studies have identified common risk factors associated with a broad range of conditions. New methods have also been introduced for identifying these factors [21]. However, such shared risk factors are often complex traits and may exert heterogeneous influences on diseases through multifaceted mechanisms. For example, obesity, one of the most

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

well-established risk factors contributing to various forms of multimorbidity [22, 23], is recognized to impact diseases through a variety of distinct pathways [14, 24].

Given the multitude of potential causal pathways stemming from a common risk factor, particularly in the case of complex traits like obesity, to enable effective clinical prevention and intervention, it is necessary to elucidate the mechanisms through which this common risk factor induces the co-occurrence of the conditions. A starting point can be identifying the variant clusters associated with diverse causal effects within a multi-outcome MR framework. To illustrate this, consider the hypothetical example depicted in Figure 1. Here, variants linked to the exposure X are divided into three groups (G_1 to G_3), as they influence the two disease outcomes Y_1 and Y_2 through three different aspects of the exposure X (denoted by X_1 to X_3) that might not be easy to measure directly. Among the three groups, G_1 is associated with an increasing effect on Y_1 but a protective effect against Y_2, and G_3 indicates an increasing effect on Y_2 but no effect on Y_1. Only the group G_2 corresponds to pathways through which the shared risk factor increases the risk of both diseases. Therefore, identifying variant clusters can unveil the potentially heterogeneous causal effects, and subsequently shed light on the mechanisms linking the common risk factor to the co-occurrence of the conditions.

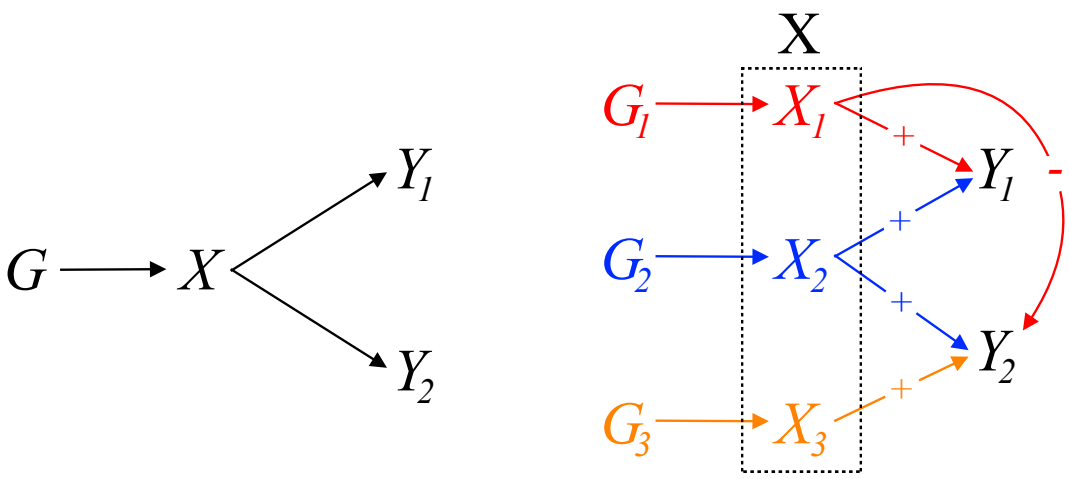


Figure 1: Left: Multi-outcome MR involving two disease outcomes and a common risk factor; right: clusters formed by the variants associated with the common exposure, which reflect heterogeneous causal pathways.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Several clustering approaches have been proposed within the MR framework to group genetic variants based on their causal estimates, such as MR-Clust [15] and MR-PATH [16]. However, these methods are primarily tailored to settings involving a single exposure and a single outcome, making them less suitable for handling the complexities of multimorbidity, as they lack multi-dimensional clustering options. On the other hand, methods such as NAvMix [25] do allow for multi-dimensional clustering of genetic variants, but are not inherently rooted in the MR framework, since they group variants based on their direct variant-trait associations, rather than causal estimates. This may limit their utility for causal inference. The mclust method [26] does permit multi-dimensional clustering using causal estimates, but we show that the method's accuracy can be sub-optimal. In contrast, our MR-AHC method allows for multi-dimensional causal clustering based on MR estimates, whilst achieving a high clustering accuracy, which we have demonstrated in extensive Monte Carlo simulations.

We apply MR-AHC to investigate the causal effects of body fat percentage (BFP), as a shared risk factor, on a pair of multimorbid conditions, T2D and osteoarthritis (OA). Our analysis identifies four variant clusters indicating heterogeneous effects on both conditions. To provide insights into the underlying causal pathways, we conducted comprehensive gene-set analyses on the clusters, combining evidence from both canonical pathway analyses and gene-set Phenome-wide association analyses (PheWAS). While the clustering results cannot directly label a cluster as signifying genuine mechanisms or pleiotropic pathways, we show how post-clustering analyses may enable this distinction. Our findings on the cluster associated with increasing risks of both conditions indicate shared pathways underpinning the co-occurrence of T2D and OA through interconnected cellular processes related to gene expression transcription and cellular responses to stimuli. We provide further evidence using cluster-specific MR for the unifying pathway from obesity to the T2D-OA multimorbidity through elevated oxidative stress. Another cluster exhibits a protective effect against T2D, with integrated canonical pathway and PheWAS evidence supporting a possible mechanism involving enhanced activity of the ion chan-

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

nels related to insulin secretion. This might be linked to elevated levels of high-density lipoprotein cholesterol (HDL-C) associated with smaller waist-to-hip ratios.

Results

The MR-AHC method for clustering genetic variants in summary-data Mendelian randomization

We assume the following summary statistics for J genetic variants involved in an MR investigation: the variant-exposure association estimate $\widehat{\gamma}_j$, and the variant-outcome association estimate $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{j}}$, where $j = 1, ..., J$. In an MR setting with a common exposure and multiple P outcomes, $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{j}}$ is the vector of the P variant-outcome associations. We maintain the assumption that the variant-exposure associations are measured from a sample independent from all the variant-outcome association samples, but overlap between the outcome samples is allowed. We also assume that all the genetic variants are themselves mutually uncorrelated (i.e. not in linkage disequilibrium). The causal estimates of the exposure on the P outcomes using only variant j as instrument (i.e. the ratio estimates) can then be obtained as the P-dimensional vector $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} = \widehat{\boldsymbol{\Gamma}}_{\boldsymbol{j}}/\widehat{\gamma}_j$. Let $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}$ be an estimate of the covariance matrix of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$.

We propose MR-AHC, a two-step procedure with $\left\{\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}\right\}_{j=1}^{J}$ as inputs, to group genetic variants indicating the same causal effects, or in other words, having similar observed ratio estimates $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$, into the same cluster. We illustrated the method with a simple hypothetical example, shown in Figure 2. For ease of illustration, we consider the case with a single outcome, but the same procedure applies generally with multiple outcomes. The first step of the method, the merging step, is illustrated in the left panel of Figure 2. It shows a situation with six variants that form three clusters (one of them comprised of a single variant). The dotted lines at β_1 and β_2 are the true heterogeneous causal effects from the exposure to the outcome, and the circles above the real line denote the variant-specific ratio estimates. The differences in the size of the circles reflect the fact that summary data

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

estimates exhibit varying degrees of uncertainty. In the explanation below, we refer to these estimates and their corresponding variants by the numbers 1 to 6, from left to right.

In the initialization step of the merging process (Step 0 in the illustration), each variant-specific estimate has its own cluster. Next, we merge the two estimates which are closest in terms of their weighted squared Euclidean distance, i.e. those estimated with Variant 3 and 4 (the two red circles). These two estimates are merged into one cluster and we now have five clusters left. We re-calculate the pairwise distances with the five clusters and merge the closest two into a new cluster. We continue with this procedure until step 5 where all variants are in a single cluster.

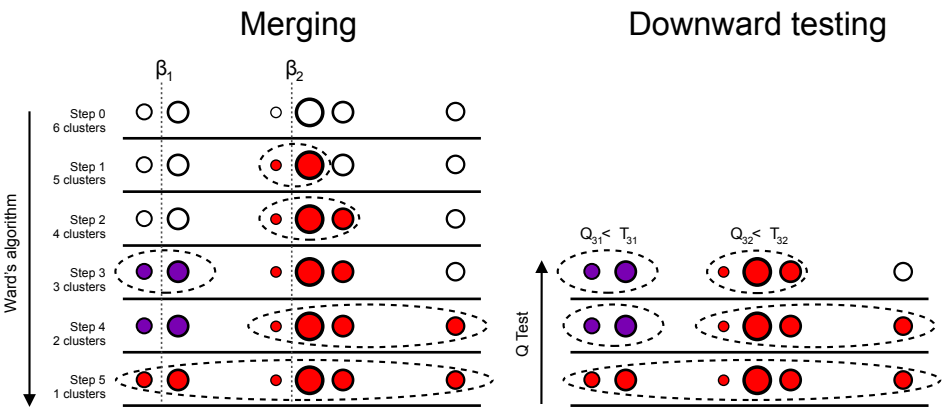


Figure 2: Illustration of the MR-AHC method for a hypothetical example adapted from Apfel and Liang [17]. Left panel: Ward's algorithm defines a clustering path. Right panel: A downward testing procedure is applied until the step-specific heterogeneity statistics can not be rejected at a specified threshold.

By the end of the merging step, we have generated a clustering path. Along each step of the path, the number of clusters, denoted by $\mathcal{K}$, varies from $\mathcal{K} = 1$ to $\mathcal{K} = J$ by increments of 1. Next, in the second step of MR-AHC, we re-trace the clustering path to select the optimal value of $\mathcal{K}$ using a downward testing procedure, operates as follows: starting from the largest cluster containing all variants, apply Cochran's Q test [27] to examine the degree of heterogeneity of all the ratio estimates by calculating the test statistic and comparing it with a pre-specified significance threshold. If the null hypothesis of "no excess heterogeneity" gets rejected, then move to the next level of the clustering path and apply the Q test to all the sub-clusters on that level. We repeat this process until reach-

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

ing a level where no sub-cluster heterogeneity statistic rejects at the given significance threshold. In our illustrative example, we would expect the downward testing procedure to re-trace from step 5 to step 3 of the clustering path, thus determining three groups formed by Variants 1-2, Variants 3-5, and Variant 6 alone.

In the original AHC algorithm proposed by Apfel and Liang [17], the inputs are essentially just the ratio estimates $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$, hence the clustering objects are treated as non-random fixed data points. MR-AHC adjusts the algorithm to take into account the uncertainty of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$ by incorporating the covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}$ into the weighted squared Euclidean distance in the merging process. We show in Appendix A that this distance between two clusters is essentially the Wald statistic for testing the null hypothesis that "the two clusters indicate the same causal effect". Therefore, merging two clusters with the smallest distance can be interpreted as merging two clusters with the highest similarity in their cluster-specific causal effects.

In terms of the covariance matrix estimate $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}$, we show in Appendix A that if all the outcome samples are non-overlapping and/or the phenotypic correlations between the outcome traits are zero, then all the ratio estimates of a given variant are uncorrelated. In this case, all the covariance terms are zero and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}$ is just a diagonal matrix with the non-zero entries being the variances of the ratio estimates, which can be easily estimated from the GWAS summary statistics. If the covariances are non-zero, we show that $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}$ can be estimated via linkage disequilibrium (LD) score regression [28] and seamlessly incorporated into the analysis.

In the downward testing procedure, following the recommendation in Belloni et al. [29], we define the threshold p-value for the Q test as $\zeta = 0.1/\log(n)$ where n is the sample size. We prove in Appendix B that this threshold p-value results in a consistent clustering procedure. That is, as n increases, the probability of correctly identifying all true members of each cluster tends to 1. If the exposure and outcome samples are of different

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

sizes, we recommend using the sample size of the smallest outcome sample. For binary outcomes, an effective sample size can be approximated with the number of cases and controls, see Han and Eskin [30].

MR-AHC does not require pre-specification of the number of clusters. It can also easily identify a "null cluster" and a "junk cluster", following the terminology of Foley et al. [15], which refer to, respectively, the cluster identifying a zero causal effect, and the cluster containing variants not assigned to any detected clusters. Specifically, we conduct a post-clustering Wald test on each cluster-specific causal estimate for the null hypothesis of a zero causal effect using $\zeta = 0.1/\log(n)$ as the threshold significance p-value. For the junk cluster, we simply classify all variants that do not fit into any other clusters as junk variants. To further improve the clustering accuracy of MR-AHC, we also extend the basic algorithm illustrated above to an outlier-robust version, to correct for the outliers in the ratio estimates, see the method section for details.

Simulation results

We conduct Monte Carlo simulations to evaluate the performance of the MR-AHC method in detecting variant clusters and estimating the causal effects in various settings that mimic the multimorbidity scenarios we are interested in, which involve a shared exposure causally affecting multiple outcome conditions. We consider 12 simulation designs, where the number of outcomes is either $P = 2$ or $P = 3$, the number of substantive variant clusters is either $K = 1$ or $K = 4$, and the sample correlation between the outcomes is either $\rho = 0$, $\rho = 0.2$ or $\rho = 0.7$ (see the method section for a detailed definition of ρ). In all designs, we have $J = 100$ SNPs with 10 designated as true 'junk' variants.

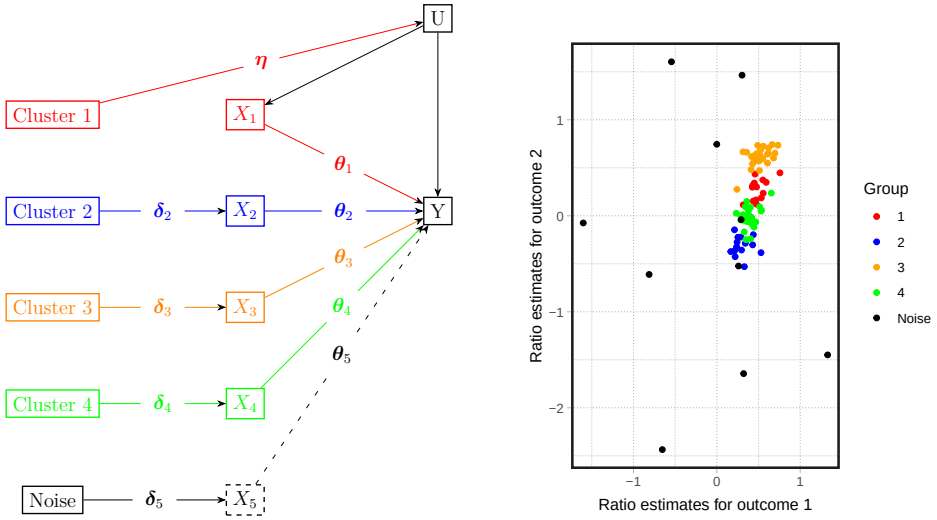
The two classes of scenarios stratified by the number of variant clusters are illustrated in Figure 3. The directed acyclic graph (DAG) in Panel (a) illustrates the data generation process when there are four substantive clusters and one noise cluster. Multiple outcomes (two or three) are represented by the single notation Y. Variant clusters are formed due

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

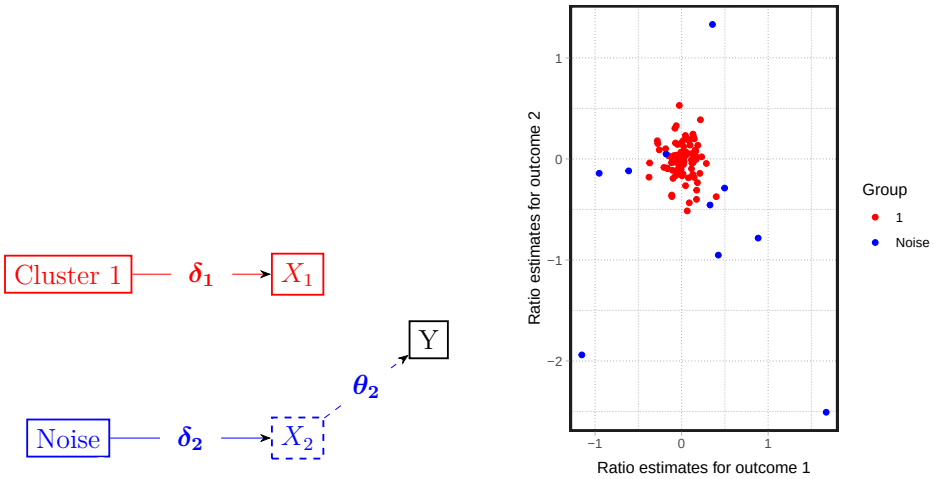
to differential sub-components of the exposure, denoted by X_1 to X_5. Variant Cluster 1 and X_1 represent a correlated pleiotropy pathway, and Clusters 2 to 4 correspond to genuine heterogeneous causal mechanisms from the exposure to the outcomes. The scatter plot on the right of Panel (a) is based on a representative simulated dataset of the two-outcome case. We also examine the performance of the method when there is actually no mechanistic heterogeneity, i.e. there is only one real cluster and one noise cluster. The design is shown by a DAG and representative dataset in Figure 3 Panel (b). The arrow from X_1 to Y is absent, meaning that the only substantive cluster is also a null cluster and there is no causal effect between the exposure and the outcomes. See the method section for a detailed design specification.

We compare MR-AHC with two other multi-dimensional clustering approaches. The first is the mclust algorithm [26], a general clustering method that accommodates the utilization of ratio estimates as inputs. We select this method as it is one of the most widely used clustering method based on Gaussian mixture models [26]. We implement mclust in two ways: the basic setting without a noise component, and the setting incorporating a Poisson noise component. An initial value of the proportion of noise variants is required, and we set this value favourably as 10% which is the ground truth. The second method, NAvMix, proposed by Grant et al. [25], groups genetic variants based on the variant-trait associations instead of the ratio estimates. We choose this method for comparison as it is also motivated by elucidating the biological mechanisms that can possibly be revealed by the patterns of the genetic variants associated with various traits. We employed two sets of input data for NAvMix: the variant-trait associations, as initially proposed; and the ratio estimates as in MR-AHC. Similar to mclust, we set the initial proportion of noise variants for NAvMix at 10%. For these two methods, cluster membership is assigned based on the highest probability. As a general clustering method for mechanistic heterogeneity, MR-AHC also works in the one-outcome case. We compare MR-AHC with MR-Clust [15], a popular method for conducting one-dimensional clustering based on ratio estimates, see Table S4 in Appendix D.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .



(a) Simulation designs with 4 substantive variant clusters.



(b) Simulation designs with 1 substantive variant cluster.

Figure 3: Directed acyclic graphs (DAG) of the data generation process and scatter plots of representative simulated data (with two outcomes) of the simulation designs with different numbers of variant clusters.

We report the following statistics from the simulations for each approach: the number of substantive clusters detected by the methods ("*#clusters*"); the Rand index which measures the similarity between the true clustering structure and the detected clusters for the substantives clusters ("*Rand index*"); the number of variants classified into the junk cluster ("*#junk variants*"); the number of true noise variants classified by the methods as junk ("*correct junk*"); the mean absolute error ("*MAE*") and the mean squared error ("*MSE*"). For simulation designs with one substantive cluster indicating zero causal effects, we

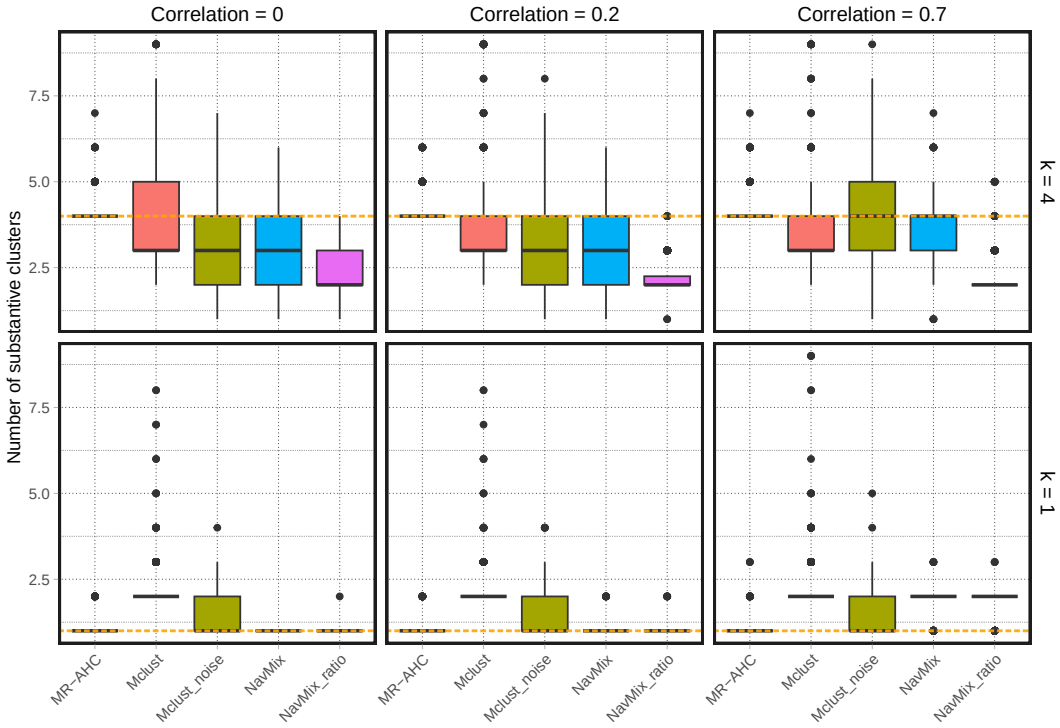
medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

additionally report the frequency of correctly identifying the null cluster ("*Freq.null*"). Definitions of the statistics can be found in the method section.

For scenarios involving non-zero outcome sample correlations, since the GWAS results typically lack a direct estimate for the correlation, we initially apply all methods treating the correlation ρ as zero. The two-outcome simulation results are presented in Figure 4 and Table 1. Across all settings, MR-AHC consistently demonstrates high accuracy in identifying the number of clusters, aligning closely with the ground truth in both mean and median assessments. In the boxplot of MR-AHC (Panel (a) of Figure 4), all the quantiles are concentrated around the median, showing that the method consistently divides variants into the correct number of clusters with little fluctuation. By comparison, both settings of mclust tend to underestimate the number of clusters when there are four true clusters and overestimate it when only one substantive cluster exists. The NAvMix method, employing two different sets of input, also exhibits a tendency to underestimate the number of clusters when $K = 4$. While it successfully identifies one cluster when $K = 1$ with low outcome correlations, it overestimates the cluster number when the outcome correlation is high ($\rho = 0.7$). In line with the cluster number results, MR-AHC performs very well in terms of grouping the variants correctly, as measured by the Rand index. It consistently achieves Rand indices close to 1, significantly outperforming all other approaches in all settings. One potential drawback of MR-AHC is its tendency to assign slightly more noise variants to the "junk" cluster than the true count, but the number of true noise variants selected as junk of MR-AHC is only marginally lower than that of mclust with the noise component, outperforming all other approaches. Regarding estimation bias, both MR-AHC and NAvMix with ratio estimates input exhibit comparable MAE and MSE in general, both of which are smaller than those of other methods in most of the settings. For scenarios where $K = 1$, MR-AHC accurately identifies the null cluster with frequencies close to 1. The two variations of NAvMix also exhibit high accuracy in this aspect, although this accuracy diminishes for NAvMix with ratio estimates when the outcome correlation is high.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Panel (a) – The number of detected substantive clusters.



Panel (b) – Rand index for variants in the substantive clusters.

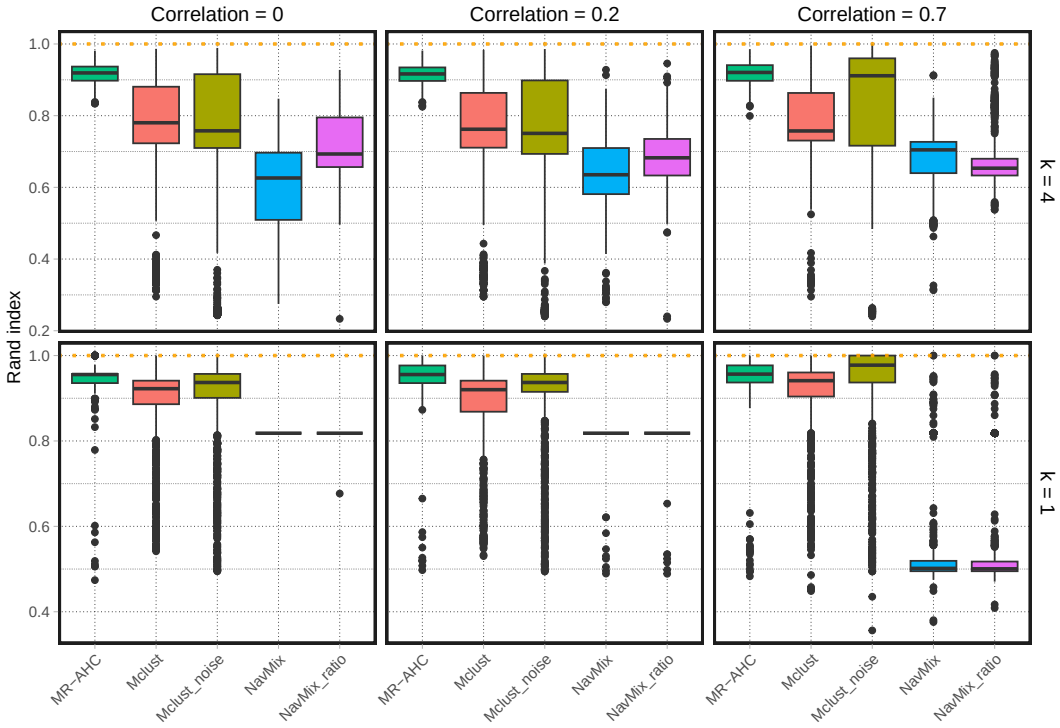


Figure 4: Two outcomes – boxplots for the number of detected substantive clusters and Rand index with different cluster numbers ($K = 4$ or $K = 1$) and outcome correlations ($\rho = 0$, $\rho = 0.2$ and $\rho = 0.7$). All methods are conducted treating the outcome correlations as 0. The dotted horizontal lines represent the true values. "*mclust noise*" stands for the mclust algorithm with a noise component, and "*NAvMix ratio*" for the NAvMix method with ratio estimates as input. Results are based on 1000 replications.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

Table 1: Simulation results for designs with two outcomes. All methods are conducted treating the outcome correlations as 0. "*mclust noise*" stands for the mclust algorithm with a noise component, and "*NAvMix ratio*" for the NAvMix method with ratio estimates as input. Statistics are calculated as the mean over 1000 replications.

	MR-AHC	mclust	mclust noise	NAvMix	NAvMix ratio
Two outcomes, K = 4, correlation = 0					
# clusters	4.196	3.716	2.964	3.018	2.307
Rand index	0.917	0.757	0.737	0.615	0.710
# junk variants	10.966	0.000	7.726	6.517	4.080
# correct junk	6.975	0.000	7.134	1.446	3.951
MAE	0.088	0.120	0.123	0.162	0.106
MSE	0.030	0.041	0.041	0.052	0.035
Two outcomes, K = 4, correlation = 0.2					
# clusters	4.207	3.665	2.933	3.130	2.254
Rand index	0.915	0.744	0.725	0.630	0.694
# junk variants	10.916	0.000	7.835	6.930	3.999
# correct junk	7.031	0.000	7.234	1.493	3.906
MAE	0.088	0.118	0.122	0.153	0.101
MSE	0.030	0.039	0.040	0.048	0.032
Two outcomes, K = 4, correlation = 0.7					
# clusters	4.229	3.652	3.754	3.649	2.202
Rand index	0.918	0.784	0.835	0.684	0.677
# junk variants	10.993	0.000	8.044	5.347	3.925
# correct junk	7.144	0.000	7.433	1.302	3.844
MAE	0.089	0.091	0.088	0.119	0.087
MSE	0.029	0.027	0.027	0.035	0.024
Two outcomes, K = 1, correlation = 0					
# clusters	1.026	2.211	1.312	1.000	1.001
Rand index	0.947	0.887	0.895	0.818	0.818
# junk variants	10.194	0.000	8.211	0.000	0.000
# correct junk	7.703	0.000	7.073	0.000	0.000
MAE	0.012	0.020	0.015	0.017	0.017
MSE	0.000	0.004	0.001	0.000	0.000
Freq.null	0.955	0.726	0.767	0.998	0.997
Two outcomes, K = 1, correlation = 0.2					
# clusters	1.019	2.226	1.297	1.009	1.007
Rand index	0.949	0.881	0.893	0.816	0.816
# junk variants	10.275	0.000	8.438	0.058	0.060
# correct junk	7.833	0.000	7.199	0.009	0.008
MAE	0.012	0.021	0.015	0.017	0.018
MSE	0.000	0.004	0.001	0.000	0.001
Freq.null	0.965	0.724	0.791	0.996	0.990
Two outcomes, K = 1, correlation = 0.7					
# clusters	1.025	2.221	1.296	1.883	1.881
Rand index	0.948	0.910	0.924	0.546	0.545
# junk variants	9.685	0.000	9.458	29.057	27.936
# correct junk	7.995	0.000	8.323	5.099	5.020
MAE	0.013	0.019	0.014	0.026	0.113
MSE	0.000	0.003	0.001	0.001	0.015
Freq.null	0.954	0.760	0.785	0.928	0.102

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

When $K = 4$ with non-zero outcome correlations, MR-AHC tends to identify more clusters than the ground truth. This feature can be rectified by incorporating accurate outcome correlation information, see the results generated by applying the method with the true correlation parameter (Table S2 in Appendix D). We show in Appendix A that the outcome correlation depends on both the extent of sample overlap between the outcome samples, and the phenotypic correlation between the outcome traits. Hence, high outcome correlations are uncommon in practice. To achieve, for instance, a correlation of $\rho = 0.7$, one would need perfect sample overlap and a phenotypic correlation of 0.7 between the two outcome traits. Even in this extreme scenario, implementing MR-AHC while assuming a zero correlation performs reasonably well. The simulation results for scenarios with three outcomes are presented in Appendix D, Table S1 and S3. Once again, MR-AHC exhibits good performance, producing clustering results that closely align with the ground truth and generally surpassing the performance of all other approaches.

Estimating the causal effects of higher adiposity on type 2 diabetes and osteoarthritis

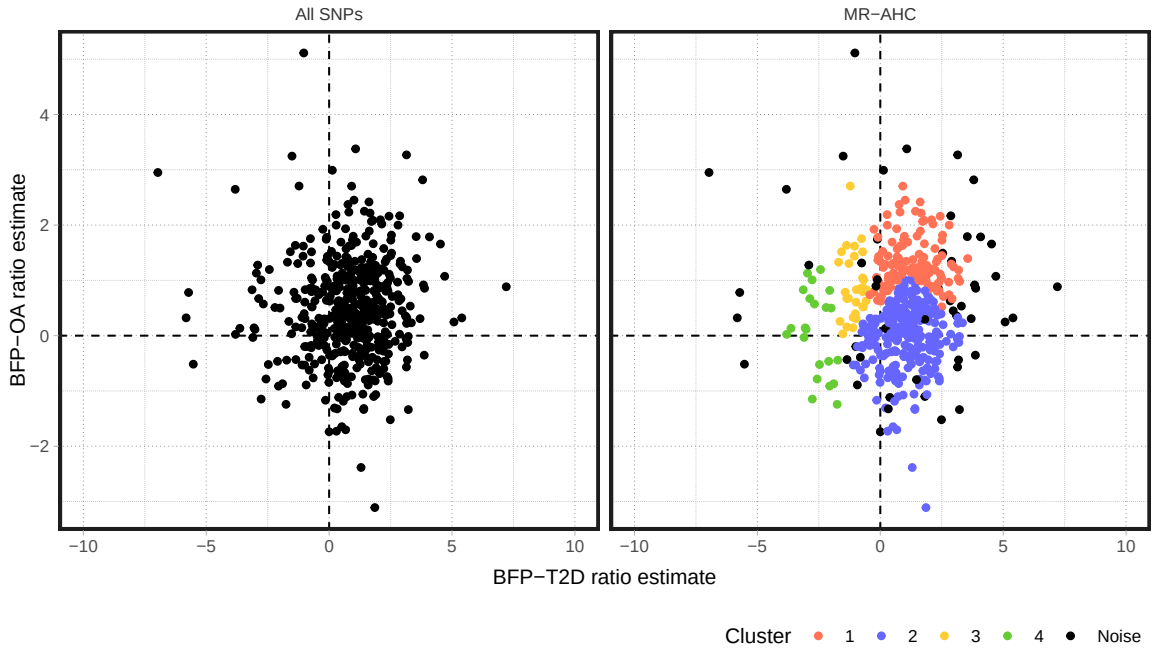
We apply the MR-AHC method to investigate the causal relationship between body fat percentage (BFP), as a measure of adiposity, and a pair of multimorbid conditions, T2D and OA. We use a three-sample summary-data MR design with 487 SNPs associated with BFP as instruments, accounting for the causal effects of the common risk factor BFP on both of the conditions simultaneously. For comparison, we also perform variant clustering using the mclust algorithm and the NAvMix method.

The clustering results of MR-AHC are presented in Figure 5, Panel (a). It detects 4 substantive clusters indicating heterogeneous causal effects. The cluster-specific estimation results, obtained with the inverse-variance weighted (IVW) approach [5], are depicted in Figure 5, Panel (b). Among the 4 clusters, Cluster 1 with 124 SNPs is the only cluster associated with increasing risk for both conditions; Cluster 2 with 258 SNPs indicates an increasing risk for T2D but a null effect for OA; both Cluster 3 (32 SNPs) and Cluster 4

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

(22 SNPs) are associated with a protective effect against T2D, and for OA, a causative effect and a null effect, respectively. See Appendix E for detailed estimation results.

Panel (a) – (left) The scatter plot of the 487 SNPs associated with BFP; On the x-axis are the ratio estimates for T2D, and y-axis for OA. Each point represents a specific SNP. (right) the clustering results of MR-AHC.



Panel (b) – The cluster-specific IVW estimates and 95% confidence intervals in odds ratio for each cluster detected by MR-AHC.

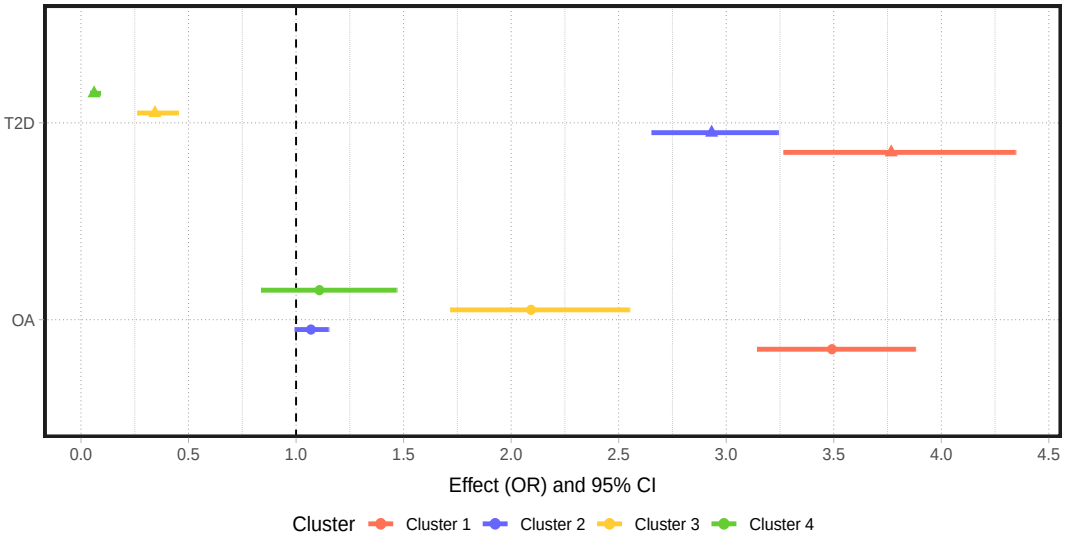


Figure 5: MR-AHC clustering and estimation results of the 487 SNPs associated with BFP based on their ratio estimates on T2D and OA.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

These results align with the conclusions drawn from previous research. For example, Martin et al. [24] examined the causal effects of higher adiposity on a variety of conditions including T2D and OA. Their findings suggest that adiposity exerts heterogeneous effects on the risk of T2D: in general, higher adiposity increases the risk of T2D, but there is a metabolically "favourable" component of adiposity that reduces the risk of the condition. For OA, all adiposity measures, including the metabolically favourable adiposity, consistently identify an increasing risk. This suggests a non-metabolic weight-bearing effect as a likely cause. Given this, it is reasonable to partition the variants into distinct clusters along both outcome dimensions: on the T2D-estimate dimension, clustering occurs due to the indication of opposing effects by different variants; on the OA-estimate dimension, clustering is also likely to occur, as we may expect an adverse effect if the variants are associated with fat located around the articulations in a load-bearing way, but no effect elsewhere.

The clustering results generated with mclust and NAvMix are presented in Figure 6. Both methods fail to segregate the variants along the OA-estimate dimension, as all clusters indicate increasing effects, hence might have underestimated the number of clusters, which also appears as an over-arching feature of the methods in the simulations. Even for the T2D-estimate clustering, their results may be dubious: mclust assigns SNPs in nonadjacent regions with largely opposing estimates into the same cluster (Cluster 2 in blue); NAvMix either labels a large number of SNPs as 'junk' if setting a non-zero initial noise proportion, or does not identify any noise at all with a zero initial proportion. More importantly, for clusters generated by these two methods, variants tend to display substantial within-cluster heterogeneity in their ratio estimates, which can be a significant concern for causal inference.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

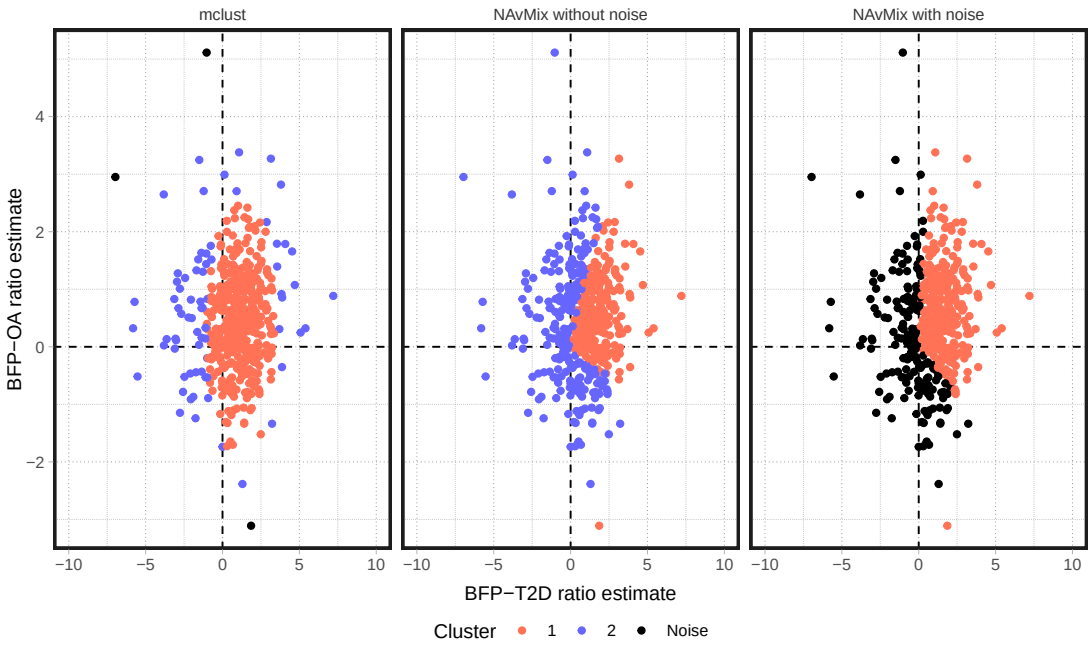


Figure 6: From left to right: the clustering results of the mclust algorithm with an initial noise proportion 5%; the clustering results of the NAvMix method with an initial noise proportion 0; the clustering results of the NAvMix method with an initial noise proportion 5%.

Biological insights into the variant clusters

To gain insights into the biological mechanisms linking obesity to the T2D-OA multimorbidity from the variant clusters detected by MR-AHC, we use an approach similar to the one taken by previous works such as Grant et al. [25] and Wang et al. [31]. For each of the clusters identified by MR-AHC, we first map the SNPs in the cluster to genes, then perform gene set enrichment analysis with the mapped genes. Both steps are conducted using the Functional Mapping and Annotation Platform (FUMA) [32]. SNPs are mapped to genes using a three-way mapping strategy (positional, eQTL and chromatin interactions mapping). The gene set enrichment analysis is to test if the mapped genes are over-represented in a given pre-defined gene set which corresponds to a canonical biological pathway or is associated with a phenotype reported from the GWAS catalog. We refer to the latter as the gene-set Phenome-wide association analysis (PheWAS), or just "PheWAS" for short. We integrate both lines of evidence from the pathway and PheWAS analyses that can complement or validate each other, to infer the possible biological mechanisms underlying each cluster. See Supplementary material S2 for a summary of

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

the enrichment analyses results.

First, it is likely that Cluster 2 (containing 258 SNPs, associated with increasing risk of T2D) is highly pleiotropic. Based on the PheWAS analysis, this cluster is enriched with a large number of phenotypes, double that for Cluster 1 which has the second most (112 versus 56). These phenotypes fall into a wide range of categories, displaying no clear pattern. The majority of the canonical pathways enriched for this cluster are related to intermediate filament, which might not have a strong direct link with the causal relationship under examination.

Cluster 1 (containing 124 SNPs, indicating increasing risks of both conditions) holds particular significance as it aligns with our primary objective of exploring the multimorbidity of T2D and OA through obesity. The majority of the canonical pathways uniquely enriched for Cluster 1 can be classified into two categories of cellular processes that are closely interconnected: gene expression transcription and cellular responses to stimuli. A significant example in the first category is DNA methylation, while in the second category, one of the most significantly enriched pathways is associated with oxidative stress. For some of the pathways, we can delve deeper into the investigation using readily available GWAS data. As an example, we further inspect the possible unifying pathway from obesity to the T2D-OA multimorbidity via oxidative stress.

Oxidative stress (OS) is the imbalance between the production of reactive oxygen species and the counteracting antioxidant defenses in the direction that favors the former, which may lead to tissue injury [33]. Clinical research has established that obesity can induce systemic OS through various metabolic pathways [34, 35]. Moreover, OS is evidenced to exert direct effects on the development of T2D via mechanisms such as decreasing insulin secretion from pancreatic β cells [36, 37]. It also plays a role in the progression of OA by promoting cartilage degradation [38]. Herein, we examine the role of OS by performing cluster-specific MR: we first analyze how BFP predicted by SNPs in cluster

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

1 is associated with a variety of OS biomarkers. Then for comparison, we conduct the same analysis on Cluster 4, serving as a counterpart to Cluster 1 due to its relatively benign nature for both conditions, manifesting a protective effect against T2D and a null effect on OA.

We select 11 OS biomarkers from diverse categories. First, as endogenous antioxidants are highly responsive to OS [39], we use 4 enzyme antioxidants (GST, CAT, SOD, GPX) as OS injury biomarkers, which have been utilized in previous MR studies [40, 41]. One of the mechanisms through which obesity induces systemic OS is chronic inflammation [34, 35]. We thus incorporate three traits known to mediate the pathway from inflammation to OS (CRP, IL-6, TNF-α) [35] as another set of OS biomarkers. Biochemical research has shown that the production of some cytokines, including IL-1β, IL-12 and IL-8, are enhanced under elevated OS levels [42, 43]. Therefore, we also include these three cytokines in the analysis. Finally, we incorporate GDF-15, which is a biomarker for both inflammation and OS [44]. See the method section for the full form of the abbreviations of the biomarkers.

We estimate the effect of BFP on each of the biomarkers by two-sample MR using SNPs in Cluster 1 and Cluster 4 as instruments separately. The estimates and standard errors are calculated by the IVW approach. Sensitivity checks by MR-PRESSO [45] and the robust adjusted profile score (MR-RAPS) method [13] can be found in Appendix E. Results in Z-scores are presented in Figure 7. For 8 out of 11 of these OS markers, Cluster 1 is associated with increasing effects, while Cluster 4 is associated with declining effects. For CAT and CRP, Cluster 1 and Cluster 4 have effects in the same direction, but Cluster 1 is either associated with a larger increasing effect (CRP), or a smaller decreasing effect (CAT). The only exception is SOD, on which the increasing effect of Cluster 1 is smaller than that of Cluster 4.

Overall, we can see a clear heterogeneity pattern between Cluster 1 and Cluster 4 in

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

their cluster-specific effects on the OS biomarkers, which supports that Cluster 1 is associated with an elevated level of oxidative stress, while it may be the opposite for Cluster 4. These results align with the existing findings regarding adiposity and oxidative stress: higher adiposity is in general associated with elevated oxidative stress, but fat patterns featured with a smaller waist-to-hip ratio (WHR) may be related to less oxidative damage [35, 46]. This correlation between WHR and OS is observed in Cluster 4, as we will show later that this cluster is associated with a decreasing WHR.

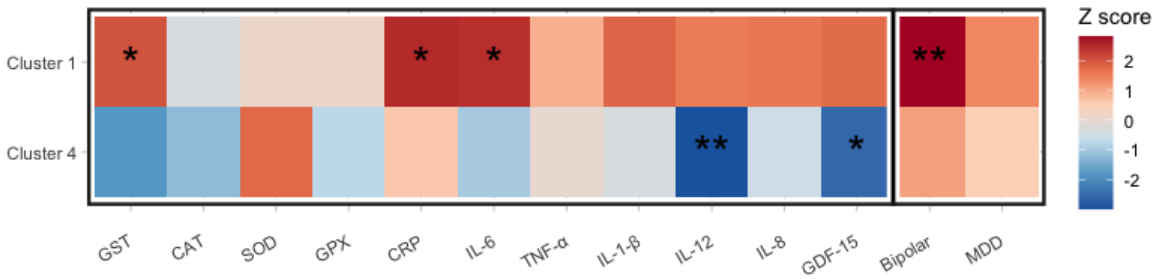


Figure 7: Results of the two-sample MR estimating the effects of BFP on the 11 oxidative stress biomarkers and 2 psychological disorders using variants in Cluster 1 and Cluster 4 as instruments respectively. Estimates are given by the IVW approach, presented in the form of Z-scores (the ratio of the estimate and the standard error). "*" represents significance at the p-value 0.05; "**" for the first 11 traits represents significance at 0.05/11, for the last two traits at 0.05/2.

Complementary evidence that may be related to the shared pathway via oxidative stress can be found in the PheWAS results for Cluster 1. A notable PheWAS pattern associated with this cluster is that it is enriched with quite a few psychological disorders. Clinical research has shown that OS is implicated in the development of such disorders, including bipolar disorder and depression [47], which are both significantly enriched for Cluster 1. We estimate the effects of BFP predicted by variants in respectively Cluster 1 and Cluster 4 on bipolar and major depressive disorder (MDD) using two-sample MR. Results are presented in the last two columns in Figure 7. Cluster 1 is associated with increasing risks of both conditions with a significant effect on bipolar. The effects of Cluster 4, on the other hand, are both insignificant and smaller than those of Cluster 1. These results may suggest a possible direction for exploring the multimorbidity between obesity-related metabolic conditions and psychological disorders.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

It is important to note that there is very likely to be intricate interactions between the pathways involved in the underlying mechanism from obesity to the T2D-OA multimorbidity. For example, another canonical pathway uniquely enriched for Cluster 1 is related to programmed cell death, or apoptosis. It has been well-documented that excess OS plays a role in the activation of apoptosis [48], and pancreatic β-cell and chondrocyte loss due to apoptosis are implicated in the development of T2D and OA respectively [49, 50]. Furthermore, quite a few gene expression transcription pathways enriched for Cluster 1 are related to epigenetic processes. Emerging evidence supports the involvement of OS in epigenetic regulation of gene expression such as inducing DNA methylation changes [51, 52]. Thus, additional research is warranted to further unravel the exact causal roles of these pathways.

Both Cluster 3 and Cluster 4 exhibit a protective effect against T2D. The most noteworthy PheWAS pattern for these two clusters is that they are both enriched with phenotypes related to fat distribution. This is particularly pronounced for Cluster 4, with 17 out of 42 enriched phenotypes associated with fat patterns including the WHR-related traits. Also, Cluster 4 has a clear pattern regarding its enriched biological pathways: 13 out of 16 of the pathways are related to ion channel activities. Ion channels are membrane proteins acting as gated pathways for the passage of ions across the cell membranes [53].

To integrate the evidence from the WHR-enriched PheWAS pattern and the ion-channel-enriched pathway pattern into a potential explanation of the protective mechanism against T2D, one possible link may be that Cluster 4 is also enriched with several HDL-C related phenotypes. Existing studies have found a negative relationship between WHR and HDL-C [54, 55], i.e. smaller WHR may be associated with higher levels of HDL-C. Moreover, the connection between HDL-C levels, ion channel activities, and T2D development might be explained by the primary role of HDL-C in cholesterol clearance [56]. On one hand, ion channels, such as the β-cell voltage-gated calcium channels, are crucial for insulin

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

secretion [57]. On the other hand, the activity of such channels can be suppressed by excess membrane cholesterol [29]. Thus, the depletion of cholesterol facilitated by HDL-C might positively impact the activity of the ion channels related to insulin secretion. This link is evidenced by previous experimental research on mice, which shows that reduced HDL-C levels are correlated with impaired glucose-induced insulin secretion [58]. This is because the increased rigidity of the β-cell membrane due to cholesterol-enrichment reduces the stimulation of ion channels essential for secreting insulin [59, 60].

To examine the possible protective mechanism against T2D stated above, we conduct two-sample MR to examine the effects indicated by Cluster 4 on WHR (adjusted for BMI), HDL-C and total cholesterol levels. The results, shown in Figure 8, are in line with the hypothesized mechanism: this cluster is associated smaller WHRs, higher levels of HDL-C, lower levels of total cholesterol, and consequently decreasing risk of coronary artery disease (CAD).

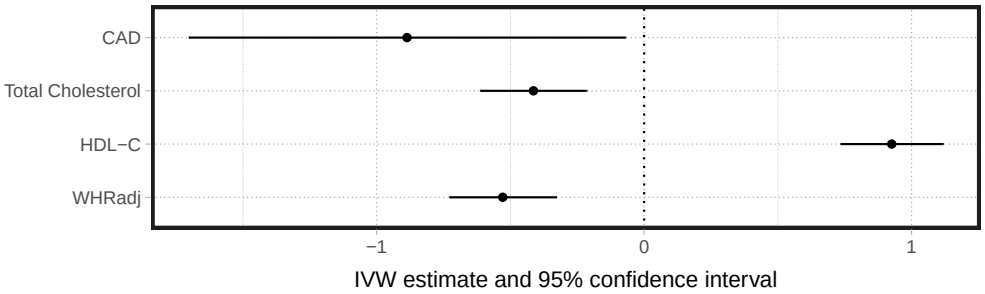


Figure 8: Two-sample MR results estimating the effects of cluster 4-predicted BFP on the waist-to-hip ratio adjusted for BMI, HDL-C, total cholesterol and coronary artery disease. Estimates are given by the IVW approach, presented in the form of the 95% confidence intervals.

Discussion

In this paper we adapt the general method of agglomerative hierarchical clustering to the summary-data Mendelian randomization setup. MR-AHC is a useful tool for interrogating a set of genetic variants to see if they collectively identify a single causal effect, or if it is more plausible that a number of subgroups identify distinct effects driven by different biological mechanisms. The method is of particular interest when the potentially

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

heterogeneous physiological components of the exposure are not known beforehand, or are difficult/expensive to measure. Of special interest is its utility as a multi-dimensional clustering method in the multi-outcome MR setting, where we can elucidate the shared causal pathways that underlie the co-occurrence of a range of conditions through a common risk factor.

In an effort to investigate the intricate mechanisms underpinning disease causation, a number of approaches has been utilized to categorize genetic variants associated with a specific phenotype, based on their GWAS associations with a range of traits linked to that target phenotype, such as the Bayesian nonnegative matrix factorization clustering method [61] and the NAvMix method [25]. MR-AHC is motivated similarly by the purpose of exploring the diverse disease-causing pathways reflected by distinct variant clusters. However, it is distinctly tailored to a different scenario and employs a different clustering strategy. Its primary application is rooted in the domain of causal inference, specifically within the framework of MR. It groups genetic variants based on their causal estimates, which integrates both their associations with the target phenotype (in this context, a common exposure) and their associations with the related traits (herein, downstream outcomes). By the comparison with NAvMix through Monte Carlo simulations and the real-world application, we have shown that MR-AHC has certain advantages over the association-based approaches in MR settings, namely an enhanced capacity to identify the patterns of the genetic variants that may mirror distinct causal mechanisms between specific traits. Furthermore, while the variant clusters discovered through association-based methods may have broad biological implications encompassing a wide range of traits, those detected by MR-AHC are precisely focused on elucidating a specific causal relationship. Consequently, pathway information derived from each cluster identified by MR-AHC offers a higher degree of relevance and specificity for the causal relationship under examination.

MR-AHC possesses the features that it does not require pre-specifying the number of

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

clusters, and that alongside detecting meaningful clusters it can also identify and label null and junk clusters without an initial specification on the proportion of 'noise'. While for hierarchical clustering algorithms, it can be difficult to choose the 'optimal' dissimilarity metric, linkage and number of clusters on the dendrogram to yield reliable clustering results, studies in the field of model selection [17, 62] provide the theoretical basis for MR-AHC to ensure highly accurate results. We have adapted the original AHC method in Apfel and Liang [17] to accommodate the varying degrees of uncertainty exhibited in summary-data estimates due to allele frequency differences across SNPs. Moreover, our method is capable of handling outliers in the variant-specific estimates with our outlier removal procedure. It should be noted that all the aforementioned methods assign variants to clusters in a probabilistic (i.e. 'soft') way, while MR-AHC do the clustering in a deterministic (i.e. 'hard') manner. Although we view this as a strength, some may view its lack of stochasticity as a disadvantage. For this reason we plan to develop a framework to quantify the sensitivity of MR-AHC clustering results to small changes in the data and thresholding rule used.

We showed in simulations that, in situations of sample overlap in the outcome data, incorporating the correct correlation information can improve the performance of the method. Nevertheless, it is in general not a significant concern if the correlation estimates are set to zero. Our method is currently focused on the problem of estimating a causal relationship between the shared exposure and the downstream outcomes without accounting for the direct causality between the outcomes. In our application example, various existing evidence supports the absence of direct causality between T2D and OA [63, 64]. However, we show in Appendix C that even if direct causality exists, our method is still applicable, as the clustering of the variants associated with the common exposure are generally robust to the outcome causality. The challenge then shifts to estimating the direct causal effect of the exposure on a particular outcome while considering other outcome traits as an additional risk factor, or accepting that the original estimates represent total causal effects via the outcome in question. Given this, another potential future

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

extension of our work is to extend the method to the multi-exposure framework, with the additional flexibility to consider genetic sub-structure within each exposure.

Materials and Methods

Model setup

We start from the individual-level model underlying the variant-exposure and variant-outcome summary associations. We assume a general linear IV model for the MR framework allowing for multiple outcomes with a shared exposure, which accounts for pleiotropy and heterogeneous causal mechanisms. We also assume that the causal effects from the exposure to the outcomes via different pathways are additive. Therefore, without loss of generality, we model the causality heterogeneity using additive sub-components of the exposure. Let the common exposure X be denoted by $X = X_1 + ... + X_k$ where X_k is the k-th sub-component in X and $k = 1, ..., K$. Let the p-th observed outcome be denoted by the scalar Y_p where $p = 1, ..., P$ and $P \geq 2$ in a multi-outcome MR. The vector $\mathbf{G} = (G_1, ..., G_J)'$ is used to denote the J genetic variants used as instruments. We then have the following linear structural model:

$$U = \sum_{j=1}^{J} \eta_j G_j + \epsilon_U, \tag{1}$$

$$X_k = \sum_{j=1}^{J} \delta_{kj} G_j + q_{xk} U + \epsilon_{Xk}, \tag{2}$$

$$X = \sum_{k=1}^{K} X_k, \tag{3}$$

$$Y_p = \sum_{k=1}^{K} \theta_{kp} X_k + \sum_{j=1}^{J} \psi_{jp} G_j + q_{yp} U + \epsilon_{Yp}, \tag{4}$$

where

- U represents the uncontrolled confounding between Y_p and any sub-component of X, with the strength of confounding determined by parameters q_{xk} and q_{yp};

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

- ϵ_U, ϵ_{Xk} and ϵ_{Yp} are error terms affecting U, X_k and Y respectively, and we assume $E(\epsilon_U G_j) = E(\epsilon_{Xk} G_j) = 0$ for all $j = 1, ...J$;

- $\theta_{1p}, ..., \theta_{Kp}$ represent the heterogeneous causal effects of X on Y_p.

We maintain the assumption that all the variants are independent with each other, and therefore inspect the relationship between each individual variant G_j and X as well as between G_j and Y_p. First, from (1) and (2), we obtain the reduced-form relationship between X_k and G_j as

$$X_k = \gamma_{kj} G_j + \xi_{Xkj},$$

where the total effect of G_j on X_k is $\gamma_{kj} = \delta_{kj} + q_{xk}\eta_j$. The error term ξ_{Xkj} is defined implicitly, but from the the previous assumptions $E(\epsilon_U G_j) = E(\epsilon_{Xk} G_j) = 0$ and that G_j is independent with all other variants, we have $E(\xi_{Xkj} G_j) = 0$ as well. It follows from (3) that the overall relationship between G_j and the exposure X is

$$X = \gamma_j G_j + \xi_{Xj} \tag{5}$$

where

$$\gamma_j = \sum_{k=1}^{K} \gamma_{kj} = \sum_{k=1}^{K} \delta_{kj} + \eta_j \sum_{k=1}^{K} q_{xk}, \tag{6}$$

and the error term $\xi_{Xj} = \sum_{k=1}^{K} \xi_{Xkj}$ with $E(\xi_{Xj} G_j) = 0$. We assume that the relevance condition for the instruments is satisfied at the scale of the overall exposure X so that $\gamma_j \neq 0$ for $j = 1, ..., J$.

Now we inspect the reduced-relationship between G_j and Y_p. It follows from (4) that the pleiotropic effect of G_j on Y_p can be derived as $\alpha_{jp} = \psi_{jp} + q_{yp}\eta_j$. Additionally by plugging (1) and (2) into (4), the overall reduced-form between Y_p and G_j is

$$Y_p = \Gamma_{jp} G_j + \xi_{Ypj}, \tag{7}$$

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

where

$$\Gamma_{jp} = \sum_{k=1}^{K} \theta_{kp}\gamma_{kj} + \alpha_{jp} = \sum_{k=1}^{K} \theta_{kp}\delta_{kj} + \eta_j(\sum_{k=1}^{K} \theta_{kp}q_{xk} + q_{yp}) + \psi_{jp}. \tag{8}$$

The correlation between the implicitly-defined error term ξ_{Ypj} and G_j depends on the correlation between ϵ_{Yp} from Equation 4 and G_j. If $E(\epsilon_{Yp}G_j) = 0$, then ξ_{Ypj} and G_j are also uncorrelated. In this case, for the G_j-Y_p association estimate (denoted by $\widehat{\Gamma}_{jp}$) generated from a GWAS by regressing Y_p on G_j in a given sample, we have

$$\widehat{\Gamma}_{jp} \xrightarrow{p} \Gamma_{jp}$$

as the sample size $n \to \infty$. Similarly, for the G_j-X association estimated in a GWAS (denoted by $\widehat{\gamma}_j$) in a sample independent from the G_j-Y_p sample, we have

$$\widehat{\gamma}_j \xrightarrow{p} \gamma_j$$

as $n \to \infty$. Then for the variant-specific causal estimate of G_j, defined as $\widehat{\beta}_{jp} = \widehat{\Gamma}_{jp}/\widehat{\gamma}_j$, we have

$$\widehat{\beta}_{jp} \xrightarrow{p} \beta_{jp} \quad \text{and} \quad \beta_{jp} = \frac{\Gamma_{jp}}{\gamma_j} = \frac{\sum_{k=1}^{K} \theta_{kp}\gamma_{kj} + \alpha_{jp}}{\sum_{k=1}^{K} \gamma_{kj}}.$$

In words, as the sample size n goes to infinity, each $\widehat{\beta}_{jp}$ converges to their variant-specific causal estimand β_{jp}, which is the causal effect from X to Y_p identified using G_j as instrument. In the simple case where G_j only instruments one sub-component X_k, the variant-specific causal estimand then becomes:

$$\beta_{jp} = \frac{\theta_{kp}\gamma_{kj} + \alpha_{jp}}{\gamma_{kj}} = \theta_{kp} + \frac{\alpha_{jp}}{\gamma_{kj}}. \tag{9}$$

Equation 9 reflects the possible sources of mechanistic heterogeneity among the variant-specific estimates: heterogeneous causal effect from the exposure to the outcome, and pleiotropic effects. We aim to group the genetic variants into distinct clusters such that within each cluster, all variants identify the same causal effect. More generally, for the

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

multi-outcome MR with $P \geq 2$ outcomes, with a given variant G_j, we combine all $\widehat{\beta}_{jp}$ and β_{jp} for $p = 1, ..., P$ into the vectors $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}} = \left(\widehat{\beta}_{j1}, ..., \widehat{\beta}_{jP}\right)'$ and $\boldsymbol{\beta}_{\boldsymbol{j}} = (\beta_{j1}, ..., \beta_{jP})'$ respectively. We propose the MR-AHC method, elaborated in the subsequent section, to divide the genetic variants into distinct clusters based on the similarity of their ratio estimates $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$, so that variants with the same estimand $\boldsymbol{\beta}_{\boldsymbol{j}}$ are in the same cluster.

Thus far, we have inspected the case where there is no residual correlation between G_j and ϵ_{Yp} in Equation 4, i.e. $E(\epsilon_{Yp} G_j) = 0$. In a multi-outcome MR model, this relationship can be violated if there is direct causality between the outcome variables. For example, consider two outcomes Y_p and Y_q, if Y_q causally affects Y_p directly, then it will enter Equation 4 as part of ϵ_{Yp}, hence the error term may be correlated with G_j. We show in Appendix C that the clustering results of the variants are in general not affected by the additional direct causality between the outcomes, but the causal effects identified by each cluster are the total effects including the outcomes causality, instead of the direct effects from the exposure to the outcomes. In this paper, we mainly focus on the case without the direct outcome causality.

The MR-AHC algorithm

We make the following normality assumption on the summary statistics described in the previous section: $\sqrt{n}\left(\begin{pmatrix} \widehat{\Gamma}_{jp} \\ \widehat{\gamma}_j \end{pmatrix} - \begin{pmatrix} \Gamma_{jp} \\ \gamma_j \end{pmatrix}\right) \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{Yjp} & 0 \\ 0 & \sigma^2_{Xj} \end{pmatrix}\right)$, for $j = 1, ..., J$ and $p = 1, ..., P$. It follows that $\sqrt{n}\left(\widehat{\beta}_{jp} - \beta_{jp}\right) \xrightarrow{d} N\left(0, \sigma^2_{jp}\right)$. The estimates of the standard errors of $\widehat{\gamma}_j$ and $\widehat{\Gamma}_{jp}$, denoted by $se\left(\widehat{\gamma}_j\right)$ and $se\left(\widehat{\Gamma}_{jp}\right)$, are generally given by the corresponding GWAS, hence taken as known. The standard error of $\widehat{\beta}_{jp} = \widehat{\Gamma}_{jp}/\widehat{\gamma}_j$ can then be obtained using the Delta method as $\widehat{v}_{jp} = \sqrt{\frac{se\left(\widehat{\Gamma}_{jp}\right)^2 + \widehat{\beta}^2_{jp} se\left(\widehat{\gamma}_j\right)^2}{\widehat{\gamma}^2_j}}$. But it is typically approximated by $\widehat{v}_{jp} = se\left(\widehat{\Gamma}_{jp}\right)/|\widehat{\gamma}_j|$ since $se\left(\widehat{\gamma}_j\right)$ is deemed negligible when we only use variants that pass a genome-wide significance threshold as instruments [11]. When there are multiple outcomes, let $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}$ be the estimate of the covariance matrix of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}$. The diagonal entries of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}$ are just the variances of $\left(\widehat{\beta}_{j1}, ..., \widehat{\beta}_{jP}\right)'$, i.e. $\left(\widehat{v}^2_{j1}, ..., \widehat{v}^2_{jP}\right)'$.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

The off-diagonal entries of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}$ are the pairwise covariances of $\left(\widehat{\beta}_{j1}, ..., \widehat{\beta}_{jP}\right)'$. We show in Appendix A that the covariances are zero if the outcome samples are non-overlapping and/or the phenotypic correlations between the outcome traits are zero, otherwise, they can be estimated using LD score regression [28].

With the summary statistics $\left\{\widehat{\boldsymbol{\beta}}_{\boldsymbol{j}}, \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{j}}\right\}_{j=1}^{J}$ as inputs, we apply MR-AHC to discover the variant clusters using a two-step procedure: Step 1 (agglomerative hierarchical clustering) generates a decision path from $\mathcal{K} = J$ to $\mathcal{K} = 1$ clusters; Step 2 (downward testing) re-traces the path from $\mathcal{K} = 1$ to $\mathcal{K} = J$ until the optimal cluster choice $\mathcal{K}_{opt}$ is chosen. The first step is summarized as follows:

Step 1. *Ward's algorithm [65]*

1. ***Initialization:*** *Each variant-specific estimate is viewed as a cluster on its own. Hence, initially, the total number of clusters is $\mathcal{K} = J$.*

2. ***Merging:*** *The two clusters that are closest as measured by their weighted squared Euclidean distance are merged into a new cluster. Without loss of generality, assume this is satisfied by cluster $\mathcal{S}_k$ and $\mathcal{S}_l$. $\widehat{\boldsymbol{\beta}}^{IVW}_{\boldsymbol{\mathcal{S}_k}}$ is defined as the inverse-variance weighted mean of all the variant-specific estimates in $\mathcal{S}_k$, as follows:*

$$\widehat{\boldsymbol{\beta}}^{IVW}_{\boldsymbol{\mathcal{S}_k}} = \left(\widehat{\beta}^{IVW}_{\mathcal{S}_k,1}, ..., \widehat{\beta}^{IVW}_{\mathcal{S}_k,P}\right)'$$

where

$$\widehat{\beta}^{IVW}_{\mathcal{S}_k,p} = \frac{\sum_{j\in S_k} \widehat{\beta}_{jp} w_{jp}}{\sum_{j\in S_k} w_{jp}} \tag{10}$$

with $w_{jp} = 1/v^2_{jp}$ for $p = 1, ..., P$. $\widehat{\boldsymbol{\beta}}^{IVW}_{\boldsymbol{\mathcal{S}_l}}$ for Cluster $\mathcal{S}_l$ can be defined similarly. Then the weighted squared Euclidean distance between $\mathcal{S}_k$ and $\mathcal{S}_l$ is defined as

$$\mathcal{D}_{k,l} = \left(\widehat{\boldsymbol{\beta}}^{IVW}_{\boldsymbol{\mathcal{S}_k}} - \widehat{\boldsymbol{\beta}}^{IVW}_{\boldsymbol{\mathcal{S}_l}}\right)' \widehat{\boldsymbol{\Omega}}^{-1}_{k,l} \left(\widehat{\boldsymbol{\beta}}^{IVW}_{\boldsymbol{\mathcal{S}_k}} - \widehat{\boldsymbol{\beta}}^{IVW}_{\boldsymbol{\mathcal{S}_l}}\right). \tag{11}$$

The $P \times P$ matrix $\widehat{\boldsymbol{\Omega}}_{k,l}$ is defined as follows: let $W_{kp} = \sum_{j\in S_k} w_{jp}$, and $W_{lp} = \sum_{j\in S_l} w_{jp}$ for $p = 1, ..., P$. Consider the entry at the i-th column and the r-th row

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

of $\widehat{\mathbf{\Omega}}_{k,l}$ with $i, r \in \{1, ..., P\}$, denoted by cov_{ir}. We have

$$cov_{ir} = \frac{\rho_{ir}}{W_{ki}W_{kr}} \sum_{j \in S_k} \frac{\widehat{\gamma}_j^2}{se\left(\widehat{\Gamma}_{ji}\right) se\left(\widehat{\Gamma}_{jr}\right)} + \frac{\rho_{ir}}{W_{li}W_{lr}} \sum_{j \in S_l} \frac{\widehat{\gamma}_j^2}{se\left(\widehat{\Gamma}_{ji}\right) se\left(\widehat{\Gamma}_{jr}\right)},$$

where ρ_{ir} is the correlation between $\widehat{\Gamma}_{ji}$ and $\widehat{\Gamma}_{jr}$, which is assumed to be constant across $j = 1, ..., J$. See Appendix A for details.

3. ***Iteration:*** *The merging step is repeated until all the variant-specific estimates are in one cluster of size J.*

After generating the clustering path using Step 1, we are left with a $\mathcal{K} = 1$ super-cluster containing all variants. We then re-trace the pathway to select the optimal value of $\mathcal{K}$ using a downward testing procedure originally proposed by Andrews [62], operating as follows:

Step 2. *Downward testing procedure.*

Firstly, define Q_{fg} to be the Cochran's Q statistic [27] associated with the g-th cluster at level f of the clustering path, denoted by $\mathcal{S}_{fg}$. Also define T_{fg} to be the $(1-\zeta)$ significance threshold of a χ^2 distribution on $P \times (|\mathcal{S}_{fg}| - 1)$ degrees of freedom with $\zeta = 0.1/log(n)$, n is the sample size, and $|\mathcal{S}_{fg}|$ is the number of variants in $\mathcal{S}_{fg}$. Q_{fg} is defined as follows: for the p-th outcome, let $\widehat{\boldsymbol{\beta}}^{\boldsymbol{p}}$ be the vector of length $|\mathcal{S}_{fg}|$ with the j-th entry being $\widehat{\beta}_{jp}$ where $j \in \mathcal{S}_{fg}$. Combine all $\widehat{\boldsymbol{\beta}}^{\boldsymbol{p}}$ into a vector of length $P \times |\mathcal{S}_{fg}|$ for all $p = 1, ..., P$, denoted by $\mathcal{B}_{fg}$. Let $\widehat{\boldsymbol{\beta}}^{\boldsymbol{p}}_{\boldsymbol{IVW}}$ be the IVW mean of all the estimates in $\widehat{\boldsymbol{\beta}}^{\boldsymbol{p}}$ as defined in (10), and $\boldsymbol{\iota}$ be a vector of 1 of length $|\mathcal{S}_{fg}|$. Then combine all the $|\mathcal{S}_{fg}|$-length vector $\boldsymbol{\iota}\widehat{\boldsymbol{\beta}}^{\boldsymbol{p}}_{\boldsymbol{IVW}}$ into a vector of length $P \times |\mathcal{S}_{fg}|$ for all $p = 1, ..., P$, denoted by $\mathcal{B}^{IVW}_{fg}$. Then Q_{fg} is

$$Q_{fg} = \left(\mathcal{B}_{fg} - \mathcal{B}^{IVW}_{fg}\right)' \widehat{\mathbf{\Phi}}^{-1}_{fg} \left(\mathcal{B}_{fg} - \mathcal{B}^{IVW}_{fg}\right), \tag{12}$$

where $\widehat{\mathbf{\Phi}}_{fg}$ is a matrix that can be partitioned into $P \times P$ blocks. The block on the i-th column and r-th row, denoted by $\widehat{\boldsymbol{\phi}}_{\boldsymbol{ir}}$, is a $|\mathcal{S}_{fg}| \times |\mathcal{S}_{fg}|$ dimension diagonal matrix. The j-the diagonal entry equals to $\frac{\rho_{ir} se\left(\widehat{\Gamma}_{ji}\right) se\left(\widehat{\Gamma}_{jr}\right)}{\gamma_j^2}$ for $j \in \mathcal{S}_{fg}$.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

1. *Starting from the cluster that contains all the variants, calculate the global Q statistic, Q_{11}, on all the ratio estimates;*

2. *If $Q_{11} < T_{11}$, then stop and assume that all the variants form a single cluster. If $Q_{11} \geq T_{11}$, then revert to the variant clusters on the next level of the path, where the number of clusters is $\mathcal{K}$=2;*

3. *Calculate Q statistics for the two sub-clusters separately, Q_{21} and Q_{22};*

4. *If both $Q_{21} < T_{21}$ and $Q_{22} < T_{22}$, then stop. Otherwise, continue to the next level where $\mathcal{K}$=3;*

5. *Repeat steps 3-4 until a $\mathcal{K} \in (1, .., J)$ is arrived at for which no sub-cluster heterogeneity statistic rejects at its given threshold.*

In implementing the MR-AHC method, in addition to the baseline procedure summarized in Step 1 and 2, we propose an extension of the method to handle outliers in the ratio estimates: after we run Step 1 and 2 and obtain the clustering results, within each detected cluster, calculate each individual variant's contribution to the overall Q statistic. The individual Q statistic, calculated using (12) with only estimates of that variant, approximately follows a χ^2_P distribution [11], and variants with large individual Q (here defined as the p-value of the individual Q below 5%) are viewed as outliers. We remove the outliers from each detected cluster, and re-run Step 1 and 2 with all the remaining variants. All the outliers are then assigned to the junk cluster.

Monte Carlo simulations

For all the simulation designs, we simulate two/three-sample summary data based on the data generating process defined in Model (1)-(4) with all sample sizes equal to $N = 60,000$. Here we fix $\psi_j = 0$. We assume that G_j, X and Y_p are normalized with $Var(G_j) = Var(X) = Var(Y_p) = 1$ and $\mathbb{E}[G_j] = \mathbb{E}[X] = \mathbb{E}[Y_p] = 0$. We also assume that the covariances between the variants are 0, $Cov(G_j, G_i) = 0$ for $i \neq j$. According to Equation

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

(6) and (8), the variant-exposure and variant-outcome summary statistics are generated in the following way:

$$\widehat{\beta}_{Xj} = \sum_{k=1}^{K} \delta_{kj} + \eta_j \sum_{k=1}^{K} q_{xk} + N(0, 1/N),$$
$$\widehat{\beta}_{Yjp} = \sum_{k=1}^{K} \theta_{kp}\delta_{kj} + \eta_j(\sum_{k=1}^{K} \theta_{kp}q_{xk} + q_{yp}) + e_{Ypj},$$

for $j = 1, ..., J$ with $J = 100$ and $p = 1, 2$ or $p = 1, 2, 3$. The normally distributed random variables $N(0, 1/N)$ add the random component to $\widehat{\beta}_{Xj}$ that mimics the asymptotically normal distribution of the statistics obtained from GWAS with standardized data. The random error of $\widehat{\beta}_{Yjp}$, denoted by e_{Ypj}, is generated from a multivariate normal distribution for the multiple outcomes. All the variance terms of this multivariate normal distribution are set to $1/N$. When there are $P = 2$ outcomes, the covariance equals ρ/N with $\rho = 0$, 0.2, and 0.7 for the zero, low, and high outcome correlation settings respectively. When there are $P = 3$ outcomes, the pair-wise outcome correlation equals to $\rho_{ij} = \rho^{|i-j|}$ where $i, j \in \{1, 2, 3\}$ and $i \neq j$ with $\rho = 0$, 0.2 and 0.7 for three different settings. Then the standard errors of $\widehat{\beta}_{Xj}$ and $\widehat{\beta}_{Yjp}$ are given by

$$se(\widehat{\beta}_{Xj}) = se(\widehat{\beta}_{Yjp}) = \sqrt{1/N}.$$

To simulate the summary statistics, we need to set the values of the following parameters: θ_{kp} (the causal effect of the sub-component X_k on Y_p), δ_{kj} (the effect of variant G_j on X_k), η_j (the effect of variant G_j on the uncontrolled confounder U), and q_{xk}, q_{yp} (the effect of U on X_k and Y_p respectively). Let the variation in U explained by all the variants be h^2_U, and the variation in X_k directly explained by all the variants (not through U) be h^2_k. Then given the values of h^2_U and h^2_k, parameters θ_{kp}, q_{xk} and q_{yp} are set as constants under the following restrictions to make $Var(X) = Var(Y_p) = 1$ feasible:

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

$$\sum_{k=1}^{K} h_k^2 + h_U^2 \sum_{k=1}^{K} q_{xk}^2 < 1, \qquad \sum_{k=1}^{K} \theta_{kp}^2 h_k^2 + h_U^2 \sum_{k=1}^{K} q_{xk}^2 \theta_{kp}^2 + q_y^2 h_U^2 < 1.$$

When there are $K = 4$ substantive clusters (with 15, 15, 30, 30 variants respectively) and one noise cluster (with 10 variants), let $h_U^2 = 0.05$ and $h_k^2 = (0, 0.05, 0.1, 0.1, 0.005)$ for $k = 1, ..., 5$ with the last entry for the junk cluster. Set $q_{xk} = 1$ for the first cluster, which corresponds to the correlated pleiotropy pathway, and $q_{xk} = 0$ for all the other clusters. When there are two outcomes, let $q_{y1} = 0.4$ and $q_{y2} = 0.1$. The causal effect parameters are set as $\theta_{k1} = (0.1, 0.3, 0.5, 0.4)$ for the first outcome, and $\theta_{k2} = (0.2, -0.3, 0.6, 0)$ for the second outcome, with $k = 1, ..., 4$. Causal effects of the 10 noise variants are generated from $N(0, 1)$. When there are three outcomes, additionally set $q_{y3} = 0.2$ and $\theta_{k3} = (-0.2, 0.3, 0, 0.3)$.

When there is $K = 1$ substantive cluster (with 90 variants) and one noise cluster (with 10 variants), set $h_U^2 = q_{xk} = q_{yp} = 0$ and $h_k^2 = (0.1, 0.005)$. The causal effects of X on all the outcomes are set to zero. In all simulation designs, δ_{kj} and η_j are generated from the uniform distribution $U[0.1, 0.3]$, and are randomly assigned to be positive or negative, then re-scaled as $\delta_{kj}\sqrt{h_k^2}/\sqrt{\sum_{j=1}^{J} \delta_{kj}^2}$ and $\eta_j\sqrt{h_U^2}/\sqrt{\sum_{j=1}^{J} \eta_j^2}$ to make sure that variations in U and X_k explained by the variants equal to h_u^2 and h_k^2 respectively.

The Rand index [66] is a quantity which measures the similarity between two clustering outcomes with values between 0 and 1. It is given by $R = (a + b)/\binom{p}{2}$. Here, a denotes the number of pairs of objects that are classified as belonging to the same cluster in both clustering outcomes and b is the number of pairs of objects that are classified in different clusters by both clustering outcomes. R values close to 1 indicate good agreement and values close to 0 indicate poor agreement between two clustering outcomes. Here the Rand index is calculated with variants assigned to the substantive clusters by the methods. The mean absolute error (MAE) and mean squared error (MSE) are calculated as follows:

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

$$MAE_p = \frac{1}{J'}\sum_{j=1}^{J'}|\beta_{jk} - \widehat{\beta}_{jk}|, \quad MSE_p = \frac{1}{J'}\sum_{j=1}^{J'}(\beta_{jk} - \widehat{\beta}_{jk})^2,$$

where J' is the number of variants that are not assigned to the junk cluster by the methods, and do not belong to the junk cluster by the ground truth. β_{jk} is the true causal effect associated with the cluster which G_j truly belongs to, and $\widehat{\beta}_{jk}$ is the causal estimate associated with the cluster which G_j is assigned to by the methods. The subscript p denotes the p-th outcome, and the overall MAE and MSE are calculated as means over all the outcomes.

MR-AHC is performed with the outlier-robust variation as described previously. To avoid spurious clusters, we only report the detected clusters containing more than 4 variants. Small clusters with less than 4 variants are subsumed into the junk cluster. The inputs for NAvMix are standardized before being supplied to the algorithm, as recommended in Grant et al. [25]. The cluster-specific causal estimates are obtained using the IVW approach. One exception is that if there is overdispersion within a detected cluster, as indicated by a non-zero I^2, then the cluster-specific estimate and its standard error are calculated using MR-RAPS [13] to account for the within-cluster overdispersion.

Clustering analysis on the BFP associated genetic variants based on the causal estimates of T2D and OA

We use SNP-BFP summary data from a GWAS study based on UK Biobank individuals from Martin et al. [24], including 696 SNPs at genome-wide significance ($p < 5 \times 10^{-8}$). The T2D GWAS statistics are from Mahajan et al. [67], which combine 31 published GWAS studies but exclude the UK Biobank individuals. The SNP-OA summary statistics are from a FinnGen GWAS (code: M13_ARTHROSIS_INCLAVO) [68]. Only SNPs present in all three datasets are used for analyses (487 in total). SNPs are orientated across all three datasets in the direction of increasing the exposure. The T2D and OA

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

samples are non-overlapping, therefore for each SNP, the covariance between the SNP-T2D estimate and SNP-OA estimate is treated as zero.

In implementing the MR-AHC method, we use the effective sample size [30] of the T2D GWAS sample ($n = 193,440$) to calculate the threshold p-value $0.1/\log(n)$ in the binary outcome setting for Cochran's Q test and the post-selection Wald test in detecting the null clusters. Clustering results of MR-AHC are obtained using an iterated outlier removal procedure: this performs the outlier removal and re-fitting indefinitely until the individual p-values of the Q statistics for all SNPs are above 5%. The cluster-specific causal estimates and standard errors are calculated with the IVW approach. For clusters with overdispersion indicated by a non-zero I^2, the estimates are obtained using MR-RAPS to account for the within-cluster overdispersion. We set the initial proportion of noise SNPs as 5% for both mclust and NAvMix.

Post-clustering analysis

We map SNPs in each cluster to genes using the SNP2GENE function in FUMA based on positional mapping (with deleterious coding SNPs) [69], eQTL mapping, and chromatin interaction mapping. This three-way mapping strategy is used in the applied examples in the original paper introducing FUMA [32]. The uploaded SNPs are also set to be the pre-defined lead SNPs. All default settings are applied, with the exception that we set the reference panel population as "UKB release2b 10k European". For eQTL mapping, following the practice in Grant et al. [25], we select tissue types from the following data sources: eQTL catalogue, PsychENCODE, van der Wijst et al. scRNA eQTLs, DICE, eQTLGen, Blood eQTLs, MuTHER, xQTLServer, CommInMind Consortium, BRAINEAC and GTEx v8 [70–79]. For chromatin interaction mapping, we select all available Hi-C datasets. The gene-set enrichment analysis is conducted using the GENE2FUNC function in FUMA. For the mapped genes corresponding to each cluster, we perform the hypergeometric test to check if the mapped genes are over-represented in a pre-defined gene set. Multi-testing correction with the Benjamini-Hochberg proce-

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

dure is applied, with the adjusted p-value ≤ 0.05 as threshold [32]. The pre-defined gene sets for canonical pathways and gene ontology processes are obtained from MsigDB and WikiPathways [80, 81]. Gene sets for phenotypes are from GWAS catalog [82].

To test how the variant clusters are associated with oxidative stress, we create a list of 11 OS biomarkers from various categories, including: glutathione transferase (GST), catalase(CAT), superoxide dismutase(SOD), glutathione peroxidase (GPX), C-reactive protein (CRP), Interleukin 6 (IL-6), Tumor necrosis factor alpha (TNF-α), Interleukin 1 beta (IL-1β), Interleukin 12 (IL-12), Interleukin 8 (IL-8) and Growth/differentiation factor-15 (GDF-15). The GWAS summary statistics for the four antioxidants (GST, CAT, SOD, GPX) and CRP are obtained from the GWAS of Sun et al. [83]; for GDF-15, the GWAS of Gudjonsson et al. [84]; for the rest five cytokines, the GWAS of Ahola-Olli et al. [85] and Kalaoja et al. [86]. Summary statistics for bipolar disorder and major depression disorder are taken from two GWAS studies conducted by the Psychiatric Genomics Consortium [87, 88]. For MR analyses associated with Cluster 4, the WHR (adjust for BMI) statistics are obtained from a GWAS conducted by the GIANT Consortium [89]; for the HDL-C and total cholesterol data, the GWAS from the Global Lipids Genetics Consortium [90]; for CAD, the GWAS from the CARDIoGRAMplusC4D Consortium [91].

References

[1] George Davey Smith and Shah Ebrahim. "Mendelian randomization: prospects, potentials, and limitations". In: *International journal of epidemiology* 33.1 (2004), pp. 30–42.

[2] Debbie A Lawlor et al. "Mendelian randomization: using genes as instruments for making causal inferences in epidemiology". In: *Statistics in medicine* 27.8 (2008), pp. 1133–1163.

[3] Vanessa Didelez and Nuala Sheehan. "Mendelian randomization as an instrumental variable approach to causal inference". In: *Statistical methods in medical research* 16.4 (2007), pp. 309–330.

[4] Tom M Palmer et al. "Using multiple genetic variants as instrumental variables for modifiable risk factors". In: *Statistical methods in medical research* 21.3 (2012), pp. 223–242.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

[5] Jack Bowden and Michael V. Holmes. "Meta-analysis and Mendelian randomization: A review". In: *Research Synthesis Methods* 10.4 (2019), pp. 486–496. DOI: https://doi.org/10.1002/jrsm.1346. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jrsm.1346.

[6] Patrick MA Sleiman and Struan FA Grant. "Mendelian randomization in the era of genomewide association studies". In: *Clinical chemistry* 56.5 (2010), pp. 723–728.

[7] Miguel A Hernán and James M Robins. "Instruments for causal inference: an epidemiologist's dream?" In: *Epidemiology* (2006), pp. 360–372.

[8] Fabiola Del Greco M et al. "Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome". In: *Statistics in medicine* 34.21 (2015), pp. 2926–2940.

[9] Stephen Burgess et al. "Guidelines for performing Mendelian randomization investigations: update for summer 2023". In: *Wellcome open research* 4 (2019).

[10] Gibran Hemani, Jack Bowden, and George Davey Smith. "Evaluating the potential role of pleiotropy in Mendelian randomization studies". In: *Human Molecular Genetics* 27.R2 (May 2018), R195–R208. ISSN: 0964-6906. DOI: 10.1093/hmg/ddy163. URL: https://doi.org/10.1093/hmg/ddy163.

[11] Jack Bowden et al. "Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption". In: *International Journal of Epidemiology* 48.3 (Dec. 2018), pp. 728–742. ISSN: 0300-5771. DOI: 10.1093/ije/dyy258. URL: https://doi.org/10.1093/ije/dyy258.

[12] Jack Bowden et al. "A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization". In: *Statistics in medicine* 36.11 (2017), pp. 1783–1802.

[13] Qingyuan Zhao et al. "Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score". In: *The Annals of Statistics* 48.3 (2020), pp. 1742–1769.

[14] Ruth JF Loos and Tuomas O Kilpeläinen. "Genes that make you fat, but keep you healthy". In: *Journal of internal medicine* 284.5 (2018), pp. 450–463.

[15] Christopher N Foley et al. "MR-Clust: clustering of genetic variants in Mendelian randomization with similar causal estimates". In: *Bioinformatics* 37.4 (2021), pp. 531–541.

[16] Daniel Iong, Qingyuan Zhao, and Yang Chen. "A latent mixture model for heterogeneous causal mechanisms in Mendelian randomization". In: *arXiv preprint arXiv:2007.06476* (2020).

[17] Nicolas Apfel and Xiaoran Liang. "Agglomerative Hierarchical Clustering for Selecting Valid Instrumental Variables". In: *arXiv preprint arXiv:2101.05774* (2021).

[18] Alessandra Marengoni et al. "Aging with multimorbidity: a systematic review of the literature". In: *Ageing research reviews* 10.4 (2011), pp. 430–439.

[19] Academy of medical sciences (Royaume uni). *Multimorbidity: a priority for global health research*. Academy of medical sciences, 2018.

[20] Jane AH Masoli, Luke C Pilling, and Timothy M Frayling. "Genomics and multimorbidity". In: *Age and Ageing* 51.12 (2022), afac285.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

[21] Verena Zuber et al. "Multi-response Mendelian randomization: Identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes". In: *The American Journal of Human Genetics* 110.7 (2023), pp. 1177–1199.

[22] Calypse B Agborsangaya et al. "Multimorbidity prevalence in the general population: the role of obesity in chronic disease clustering". In: *BMC Public Health* 13.1 (2013), pp. 1–6.

[23] M Kivimaki et al. "Body-mass index and risk of obesity-related complex multimorbidity: an observational multicohort study. Lancet Diabetes Endocrinol. 2022; 10: 253–63". In: *Article PubMed PubMed Central* (), p. 253.

[24] Susan Martin et al. "Disease consequences of higher adiposity uncoupled from its adverse metabolic effects using Mendelian randomisation". In: *Elife* 11 (2022), e72452.

[25] Andrew J Grant et al. "Noise-augmented directional clustering of genetic association data identifies distinct mechanisms underlying obesity". In: *PLoS Genetics* 18.1 (2022), e1009975.

[26] Luca Scrucca et al. "mclust 5: clustering, classification and density estimation using Gaussian finite mixture models". In: *The R journal* 8.1 (2016), p. 289.

[27] William G Cochran. "The comparison of percentages in matched samples". In: *Biometrika* 37.3/4 (1950), pp. 256–266.

[28] Brendan Bulik-Sullivan et al. "An atlas of genetic correlations across human diseases and traits". In: *Nature genetics* 47.11 (2015), pp. 1236–1241.

[29] A. Belloni et al. "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain". In: *Econometrica* 80.6 (2012), pp. 2369–2429. DOI: 10.3982/ecta9626.

[30] Buhm Han and Eleazar Eskin. "Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies". In: *The American Journal of Human Genetics* 88.5 (2011), pp. 586–598.

[31] Jingshu Wang et al. "Causal inference for heritable phenotypic risk factors using heterogeneous genetic instruments". In: *PLoS genetics* 17.6 (2021), e1009575.

[32] Kyoko Watanabe et al. "Functional mapping and annotation of genetic associations with FUMA". In: *Nature communications* 8.1 (2017), p. 1826.

[33] D John Betteridge. "What is oxidative stress?" In: *Metabolism* 49.2 (2000), pp. 3–8.

[34] Prasenjit Manna and Sushil K Jain. "Obesity, oxidative stress, adipose tissue dysfunction, and the associated health risks: causes and therapeutic strategies". In: *Metabolic syndrome and related disorders* 13.10 (2015), pp. 423–444.

[35] Heather K Vincent and Ann G Taylor. "Biomarkers and potential mechanisms of obesity-induced oxidant stress in humans". In: *International journal of obesity* 30.3 (2006), pp. 400–418.

[36] Taka-aki Matsuoka et al. "Glycation-dependent, reactive oxygen species-mediated suppression of the insulin gene promoter activity in HIT cells." In: *The Journal of clinical investigation* 99.1 (1997), pp. 144–150.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

[37] Shigetada Furukawa et al. "Increased oxidative stress in obesity and its impact on metabolic syndrome". In: *The Journal of clinical investigation* 114.12 (2017), pp. 1752–1761.

[38] Panagiotis Lepetsos and Athanasios G Papavassiliou. "ROS/oxidative stress signaling in osteoarthritis". In: *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1862.4 (2016), pp. 576–591.

[39] Karol Tejchman, Katarzyna Kotfis, and Jerzy Sieńko. "Biomarkers and mechanisms of oxidative stress—last 20 Years of Research with an emphasis on kidney damage and renal transplantation". In: *International journal of molecular sciences* 22.15 (2021), p. 8010.

[40] Zhe Lu et al. "Oxidative stress and psychiatric disorders: evidence from the bidirectional mendelian randomization study". In: *Antioxidants* 11.7 (2022), p. 1386.

[41] Pu Yifu. "Evidence for causal effects of polycystic ovary syndrome on oxidative stress: a two-sample mendelian randomisation study". In: *BMC Medical Genomics* 16.1 (2023), p. 141.

[42] JD Crapo. "Oxidative stress as an initiator of cytokine release and cell damage". In: *European Respiratory Journal* 22.44 suppl (2003), 4s–6s.

[43] K Ito et al. "Oxidative stress reduces histone deacetylase 2 activity and enhances IL-8 gene expression: role of tyrosine nitration". In: *Biochemical and biophysical research communications* 315.1 (2004), pp. 240–245.

[44] Lars Wallentin et al. "Growth differentiation factor 15, a marker of oxidative stress and inflammation, for risk assessment in patients with atrial fibrillation: insights from the Apixaban for Reduction in Stroke and Other Thromboembolic Events in Atrial Fibrillation (ARISTOTLE) trial". In: *Circulation* 130.21 (2014), pp. 1847–1858.

[45] Marie Verbanck et al. "Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases". In: *Nature genetics* 50.5 (2018), pp. 693–698.

[46] Giovanni Davı et al. "Platelet activation in obese women: role of inflammation and oxidant stress". In: *Jama* 288.16 (2002), pp. 2008–2014.

[47] Samina Salim. "Oxidative stress and psychological disorders". In: *Current neuropharmacology* 12.2 (2014), pp. 140–147.

[48] Maureen Redza-Dutordoir and Diana A Averill-Bates. "Activation of apoptosis signalling pathways by reactive oxygen species". In: *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1863.12 (2016), pp. 2977–2992.

[49] James D Johnson and Dan S Luciani. "Mechanisms of pancreatic β-cell apoptosis in diabetes and its therapies". In: *The Islets of Langerhans* (2010), pp. 447–462.

[50] Marcello Del Carlo Jr and Richard F Loeser. "Cell death in osteoarthritis". In: *Current rheumatology reports* 10.1 (2008), pp. 37–42.

[51] Rodrigo Franco et al. "Oxidative stress, DNA methylation and carcinogenesis". In: *Cancer letters* 266.1 (2008), pp. 6–11.

[52] Yingmei Niu et al. "Oxidative stress alters global histone modification and DNA methylation". In: *Free Radical Biology and Medicine* 82 (2015), pp. 22–28.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

[53] Frances M Ashcroft. *Ion channels and disease*. Academic press, 1999.

[54] Richard E Ostlund Jr et al. "The ratio of waist-to-hip circumference, plasma insulin level, and glucose intolerance as independent predictors of the HDL2 cholesterol level in older adults". In: *New England Journal of Medicine* 322.4 (1990), pp. 229–234.

[55] Rena R Wing et al. "Waist to hip ratio in middle-aged women. Associations with behavioral and psychosocial factors and with changes in cardiovascular risk factors." In: *Arteriosclerosis and thrombosis: a journal of vascular biology* 11.5 (1991), pp. 1250–1257.

[56] Gerd Schmitz and Margot Grandl. "The molecular mechanisms of HDL and associated vesicular trafficking mechanisms to mediate cellular lipid homeostasis". In: *Arteriosclerosis, thrombosis, and vascular biology* 29.11 (2009), pp. 1718–1722.

[57] Shao-Nian Yang and Per-Olof Berggren. "The role of voltage-gated calcium channels in pancreatic β-cell physiology and pathophysiology". In: *Endocrine reviews* 27.6 (2006), pp. 621–676.

[58] Eva Xepapadaki et al. "Impact of apolipoprotein A1-or lecithin: cholesterol acyltransferase-deficiency on white adipose tissue metabolic activity and glucose homeostasis in mice". In: *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1865.6 (2019), pp. 1351–1360.

[59] Marie M Gleason, MS Medow, and Thomas N Tulenko. "Excess membrane cholesterol alters calcium movements, cytosolic calcium levels, and membrane fluidity in arterial smooth muscle cells." In: *Circulation Research* 69.1 (1991), pp. 216–227.

[60] Eva Xepapadaki et al. "HDL and type 2 diabetes: the chicken or the egg?" In: *Diabetologia* 64 (2021), pp. 1917–1926.

[61] Miriam S Udler et al. "Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis". In: *PLoS medicine* 15.9 (2018), e1002654.

[62] Donald W. K. Andrews. "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation". In: *Econometrica* 67.3 (1999), pp. 543–563. DOI: `10.1111/1468-0262.00036`.

[63] Andrew Khor et al. "Diabetes mellitus is not a risk factor for osteoarthritis". In: *RMD open* 6.1 (2020), e001030.

[64] Ana Luiza Arruda et al. "Genetic underpinning of the comorbidity between type 2 diabetes and osteoarthritis". In: *The American Journal of Human Genetics* 110.8 (2023), pp. 1304–1318.

[65] Joe H Ward Jr. "Hierarchical grouping to optimize an objective function". In: *Journal of the American statistical association* 58.301 (1963), pp. 236–244.

[66] William M Rand. "Objective criteria for the evaluation of clustering methods". In: *Journal of the American Statistical association* 66.336 (1971), pp. 846–850.

[67] Anubha Mahajan et al. "Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps". In: *Nature genetics* 50.11 (2018), pp. 1505–1513.

[68] FinnGen. *FinnGen Documentation of R8 release*. 2022. URL: `https://finngen.gitbook.io/documentation/`.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

[69] Martin Kircher et al. "A general framework for estimating the relative pathogenicity of human genetic variants". In: *Nature genetics* 46.3 (2014), pp. 310–315.

[70] Daifeng Wang et al. "Comprehensive functional genomic resource and integrative model for the human brain". In: *Science* 362.6420 (2018), eaat8464.

[71] Monique GP Van Der Wijst et al. "Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs". In: *Nature genetics* 50.4 (2018), pp. 493–497.

[72] Benjamin J Schmiedel et al. "Impact of genetic polymorphisms on human immune cell gene expression". In: *Cell* 175.6 (2018), pp. 1701–1715.

[73] Urmo Võsa et al. "Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis". In: *BioRxiv* (2018), p. 447367.

[74] Elin Grundberg et al. "Mapping cis-and trans-regulatory effects across multiple tissues in twins". In: *Nature genetics* 44.10 (2012), pp. 1084–1089.

[75] Bernard Ng et al. "An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome". In: *Nature neuroscience* 20.10 (2017), pp. 1418–1426.

[76] Menachem Fromer et al. "Gene expression elucidates functional impact of polygenic risk for schizophrenia". In: *Nature neuroscience* 19.11 (2016), pp. 1442–1453.

[77] Adaikalavan Ramasamy et al. "Genetic variability in the regulation of gene expression in ten regions of the human brain". In: *Nature neuroscience* 17.10 (2014), pp. 1418–1428.

[78] GTEx Consortium et al. "The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans". In: *Science* 348.6235 (2015), pp. 648–660.

[79] GTEx Consortium Lead analysts: Aguet François 1 Brown Andrew A. 2 3 4 Castel Stephane E. 5 6 Davis Joe R. 7 8 He Yuan 9 Jo Brian 10 Mohammadi Pejman 5 6 Park YoSon 11 Parsana Princy 12 Segrè Ayellet V. 1 Strober Benjamin J. 9 Zappala Zachary 7 8 et al. "Genetic effects on gene expression across human tissues". In: *Nature* 550.7675 (2017), pp. 204–213.

[80] Arthur Liberzon et al. "Molecular signatures database (MSigDB) 3.0". In: *Bioinformatics* 27.12 (2011), pp. 1739–1740.

[81] Martina Kutmon et al. "WikiPathways: capturing the full diversity of pathway knowledge". In: *Nucleic acids research* 44.D1 (2016), pp. D488–D494.

[82] Jacqueline MacArthur et al. "The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog)". In: *Nucleic acids research* 45.D1 (2017), pp. D896–D901.

[83] Benjamin B Sun et al. "Genomic atlas of the human plasma proteome". In: *Nature* 558.7708 (2018), pp. 73–79.

[84] Alexander Gudjonsson et al. "A genome-wide association study of serum proteins reveals shared loci with common diseases". In: *Nature communications* 13.1 (2022), p. 480.

[85] Ari V Ahola-Olli et al. "Genome-wide association study identifies 27 loci influencing concentrations of circulating cytokines and growth factors". In: *The American Journal of Human Genetics* 100.1 (2017), pp. 40–50.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.18.23287164; this version posted November 7, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .

[86] Marita Kalaoja et al. "The role of inflammatory cytokines as intermediates in the pathway from increased adiposity to disease". In: *Obesity* 29.2 (2021), pp. 428–437.

[87] Niamh Mullins et al. "Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology". In: *Nature genetics* 53.6 (2021), pp. 817–829.

[88] Naomi R Wray, Patrick F Sullivan, et al. "Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. bioRxiv". In: *URL https://www. biorxiv. org/content/early/2017/07/24/167577* (2017).

[89] Dmitry Shungin et al. "New genetic loci link adipose and insulin biology to body fat distribution". In: *Nature* 518.7538 (2015), pp. 187–196.

[90] "Discovery and refinement of loci associated with lipid levels". In: *Nature genetics* 45.11 (2013), pp. 1274–1283.

[91] "A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease". In: *Nature genetics* 47.10 (2015), pp. 1121–1130.

[92] Frank Windmeijer et al. "The confidence interval method for selecting valid instrumental variables". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 83.4 (2021), pp. 752–776.

Funding

This work was supported by the Strategic Priority Fund "Tackling multimorbidity at scale" programme (grant number MC/MR/WO14548/1) delivered by the Medical Research Council and the National Institute for Health and Care Research in partnership with the Economic and Social Research Council and in collaboration with the Engineering and Physical Sciences Research Council. Nicolas Apfel is supported by the ESRC grant EST013567/1.