Article

MR-EILLS: an invariance-based Mendelian randomization method integrating multiple heterogeneous GWAS summary datasets

Lei Hou [1,2], Hao Chen [1,3,7] ✉ & Xiao-Hua Zhou [4,5,6,7] ✉

Diverse genetic structures can lead to heterogeneity among GWAS summary datasets from distinct populations. This makes it more difficult to infer causal effects of exposures on the outcome when multiple GWAS summary datasets are integrated. Here, we propose a Mendelian randomization method called MR-EILLS, which leverages environment invariant linear least squares to establish whether there is a causal relationship that is invariant in all heterogeneous populations. The MR-EILLS model works in both univariate and multivariate scenarios and allows for invalid instrumental variables that violate the exchangeability and exclusion restriction assumptions. In addition, MR-EILLS shows the unbiased causal effect estimations of one or multiple exposures on the outcome, whether there are valid or invalid instrumental variables. Compared to traditional Mendelian randomization and meta methods, MR-EILLS yields the highest estimation accuracy, the most stable type I error rates, and the highest statistical power. Finally, we apply MR-EILLS to explore the independent causal relationships between 11 blood cells and 20 disease-related outcomes, using GWAS summary statistics from five ancestries (African, East Asian, South Asian, Hispanic/Latino and European). The results cover most of the expected causal links that have biological interpretations as well as additional links supported by previous observational studies.

In recent years, with the increasing number of genome-wide association study (GWAS) investigations, there has been a notable increase in the public availability and utilization of GWAS summary data by researchers[1,2]. This inclusive dataset encompasses information from diverse populations and ethnic backgrounds[3–6], a development that researchers find valuable, thus making it a current focal point of research interest. Owing to a range of influences, including geographical landscapes and varied lifestyles, genetic structures exhibit significant diversity among distinct populations[7,8], also called

population stratification, potentially leading to heterogeneity in GWAS summary data across different ethnic groups, such as those of European, Asian, and American descent.

Mendelian randomization (MR)[2,9] is a methodology that relies on the utilization of publicly available GWAS summary data for causal inference. It uses genetic variants as instrumental variables (IVs) to infer the causal effect of one or multiple exposures on an outcome, that is, univariable or multivariable MR[10,11], respectively. A valid IV must satisfy the following three assumptions: (A1) Relevance (IV is strongly

[1]Healthcare Big Data Research Institute, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, P. R. China. [2]Department of Medical Dataology, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, P. R. China. [3]Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, P. R. China. [4]Beijing International Center for Mathematical Research, Peking University, Beijing, P. R. China. [5]Department of Biostatistics, Peking University, Beijing, P. R. China. [6]Chongqing Big Data Research Institute, Peking University, Chongqing, P. R. China. [7]These authors contributed equally: Hao Chen, Xiao-Hua Zhou. ✉e-mail: hao.chen@sdu.edu.cn; azhou@bicmr.pku.edu.cn

associated with at least one of the exposures), (A2) Exchangeability (IV is independent of confounders between exposures and outcomes), and (A3) Exclusion restrictions (IV affects the outcome only through exposures)[9]. These assumptions are also mentioned in the Methods section. When we consider heterogeneous populations, one valid IV in a population may be an invalid IV in another population due to various genetic structures. For example, G_1 is a valid IV in population I; it may be correlated with the confounder U between exposure and outcome in population II, whereas U is not the confounder in population I. In this case, G_1 violates the exchangeability in population II. In addition, G_1 may be correlated (linkage disequilibrium (LD))[12] with another SNP G_2, which directly affects the outcome in population II, but G_1 is independent of G_2 in population I. In this case, G_1 violates the exclusion restriction in population II because the LD references in different populations are different. In addition, the effects of some traits, such as body mass index, educational attainment or depression, on various disease outcomes are mediated or modified by social and environmental factors, which lead to inconsistent causal relationships in different societies or population groups. Therefore, the heterogeneity among populations includes genetic and nongenetic differences, such as social or environmental factors, also creating distinct challenges for IV validity across groups. This complexity increases the difficulty of deducing a purely causal relationship by integrating multiple heterogeneous GWAS summary datasets.

Therefore, the aim of this paper was to explore the pure causal effect, in which "pure" means that we focus on the causal effect that is not affected by social and environmental factors, i.e., the pure causal effect is invariant across heterogeneous populations. One straightforward way to infer pure causal relationships using MR is to first conduct MR analysis separately using valid IVs in different populations, obtain causal effect estimations in each population, and then combine all estimations by meta-analysis[13,14]. Even if there are invalid IVs in the first step, many MR methods[15–18] have been proposed to eliminate the influence of invalid IVs on causal effect estimation. However, the accuracy of meta-analysis results depends on the robustness of different MR methods, and these MR methods require different assumptions[15–18], which may be difficult to satisfy or cannot be tested. This may induce inconsistent causal effect estimation in different populations and make inferring pure causal relationships difficult (see Application section). Another idea is to first conduct GWAS meta-analysis for heterogeneous populations, and then select valid IVs to infer causal relationships via MR. The difficulty with this strategy is that only a short number of independent SNPs (no LD) can be selected because the LD reference panels in different populations are different[8,19]. These two strategies are both two-step processes, and result in doubled statistical errors, which results in a lower accuracy of causal effect estimation. In addition, meta-analysis is a statistical technique used to combine and analyze results from multiple studies[20]; if one result is inaccurate, the meta-analysis results are also incorrect. Meta-analysis is not a causal method in itself and does not necessarily provide causal evidence that holds true in every population included in the analysis. Therefore, we proposed a one-step method that integrates all the information, not only the MR results for each population, and provides causal evidence that holds true (also called invariant effects) in each population.

In this paper, we introduce an MR method called MR-EILLS, which utilizes the environment invariant linear least squares (EILLS)[21] to integrate multiple heterogeneous GWAS summary datasets and then infer pure causal relationships. The MR-EILLS model works in both univariate and multivariate scenarios and allows for invalid IVs that violate exchangeability and exclusion restriction assumptions. In addition, MR-EILLS shows the unbiased causal effect estimation of one or multiple exposures on the outcome, whether there are valid or invalid IVs. Compared with traditional MR and meta methods, MR-EILLS yields the highest estimation accuracy, the most stable type I

error rates, and greater statistical power. Finally, MR-EILLS was applied to explore the independent causal relationships between 11 blood cells and 20 disease-related outcomes, using GWAS summary statistics from five ancestries: African, East Asian, South Asian, Hispanic/Latino, and European.
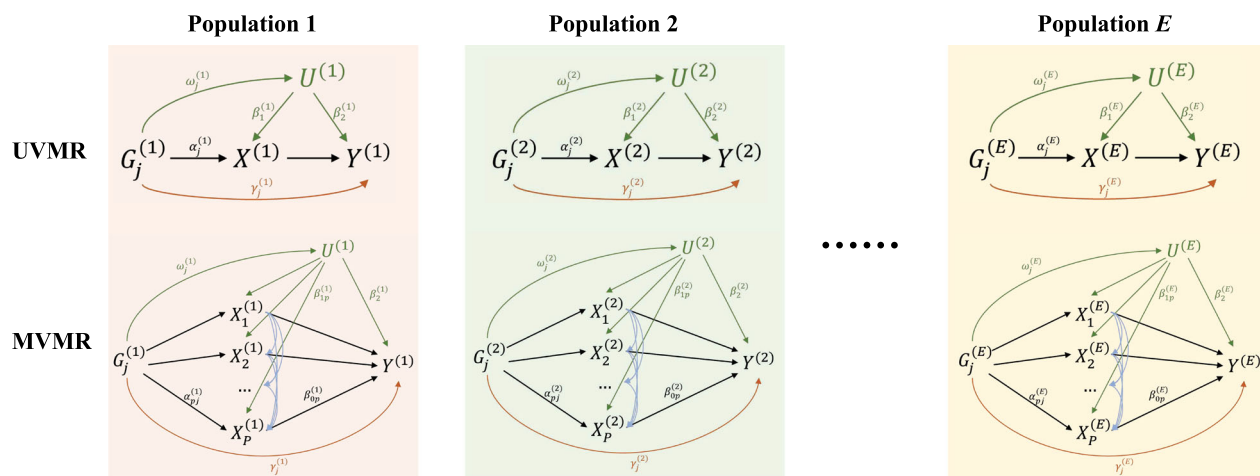
Results
Method overview

The MR-EILLS model integrates the GWAS summary statistics from multiple heterogeneous populations, and identifies the causal exposures that have invariant effects on the outcome. The GWAS summary statistics for E heterogeneous populations include the $G_j - X$ association $\hat{\theta}_{p,j}^{(e)}$ and its standard error $\sigma_{G_j X_p}^{(e)}$, as well as the $G_j - Y$ association $\hat{\Gamma}_{y,j}^{(e)}$ and its standard error $\sigma_{y,j}^{(e)}$, where the subscript p represents the p-th exposure ($p \in \{1, 2, ..., P\}$), j represents the j-th IV ($j \in \{1, 2, ..., J\}$) and the superscript e represents the e-th population ($e \in \mathcal{E}$, $\mathcal{E} = \{1, 2, ..., E\}$). We assume that the causal effects of causal exposures ($\{X_p\}, p \in P^*$) on Y are invariant in different populations, that is $\beta_{0p}^{(1)} = \beta_{0p}^{(2)} = ... = \beta_{0p}^{(E)} = \beta_{0p}^*$ for $p \in P^*$, where P^* is the set of causal exposures, whereas the genetic associations between SNPs and exposures/outcome/confounders may be different, and confounders between exposures and the outcome are also different. The aim of the MR-EILLS model (Fig. 1) is to explore the pure causal effect of causal exposures on the outcome by minimizing the following objective function

$$Q(\beta_{0p}^*; \hat{\theta}_{p,j}^{(e)}, \hat{\Gamma}_{y,j}^{(e)}, \sigma_{y,j}^{(e)2})$$
$$= \sum_{e \in \mathcal{E}} w^{(e)} \mathbb{E}_{j \in S^*}[|w_j^{(e)} \hat{\varepsilon}_j^{(e)}|^2] + \gamma \sum_{p \in P} \sum_{e \in \mathcal{E}} w^{(e)} |\mathbb{E}_{j \in S^*}[\hat{\theta}_{p,j}^{(e)} \cdot w_j^{(e)} \hat{\varepsilon}_j^{(e)}]|^2 \tag{1}$$

where $\mathbb{E}$ is the expectation, $w_j^{(e)}$ is the weight of IV G_j on the causal effect estimation in the e-th population, and $w^{(e)}$ is the weight of e-th population on the pure causal effect estimation. The first part of the objective function (1) is the empirical L_2 loss, which is the multiple-population version of ordinary least squares in one population, and $\hat{\varepsilon}_j^{(e)} = \hat{\Gamma}_{y,j}^{(e)} - \sum_p \hat{\theta}_{p,j}^{(e)} \beta_{0p}^{(e)}$ denotes the pleiotropic effect. Motivating simulation (Fig. 1a and Supplementary Fig. 1a) demonstrated that as the pleiotropic effect increased, the absolute value of $\hat{\varepsilon}_j^{(e)}$ increased. The pleiotropic effect includes correlated and uncorrelated pleiotropy. When only the A3 assumption (exclusion restriction) is violated, we say that an uncorrelated pleiotropic effect exists; when the A2 assumption (exchangeability) is violated, a correlated pleiotropic effect emerges. The second part of the objective function (1) is the empirical focused linear invariance regularizer, which discourages the selection of exposures with strong correlations between $\theta_{p,j}^{(e)}$ and $\varepsilon_j^{(e)}$ in some populations because these correlations represent correlated pleiotropy, which would distort causal effect estimation. The results of the motivating simulation (Fig. 1b and Supplementary Fig. 1b) demonstrated that as the correlated pleiotropic effect increased, the correlation between $\hat{\varepsilon}_j^{(e)}$ and $\hat{\theta}_{p,j}^{(e)}$ became stronger. $\gamma > 0$ is the hyperparameter. In addition, we added the following restriction

$$S^* = \{j : \sum_{e \in \mathcal{E}} |\varepsilon_j^{(e)}| + \sum_{p \in P} \sum_{e \in \mathcal{E}} |\hat{\theta}_{p,j}^{(e)} \varepsilon_j^{(e)}| < \lambda\} \tag{2}$$

to select valid IVs, that satisfy the three assumptions and have no pleiotropy. The first part of Eq. (2) represents the total pleiotropic effect for the $j - th$ IV, and the second part of Eq. (2) represents the correlated pleiotropic effect, that is, the correlation between $\theta_{p,j}^{(e)}$ and $\varepsilon_j^{(e)}$ for the $j - th$ IV. $\lambda > 0$ is the hyperparameter controlling the strictness of filtering IVs. When there are invalid IVs, the ridge plot of $\sum_{e \in \mathcal{E}} |\varepsilon_j^{(e)}| + \sum_{p \in P} \sum_{e \in \mathcal{E}} |\hat{\theta}_{p,j}^{(e)} \varepsilon_j^{(e)}|$ has at least two peaks (Fig. 1c), whereas the ridge plot has only one peak when there is no invalid IV. The
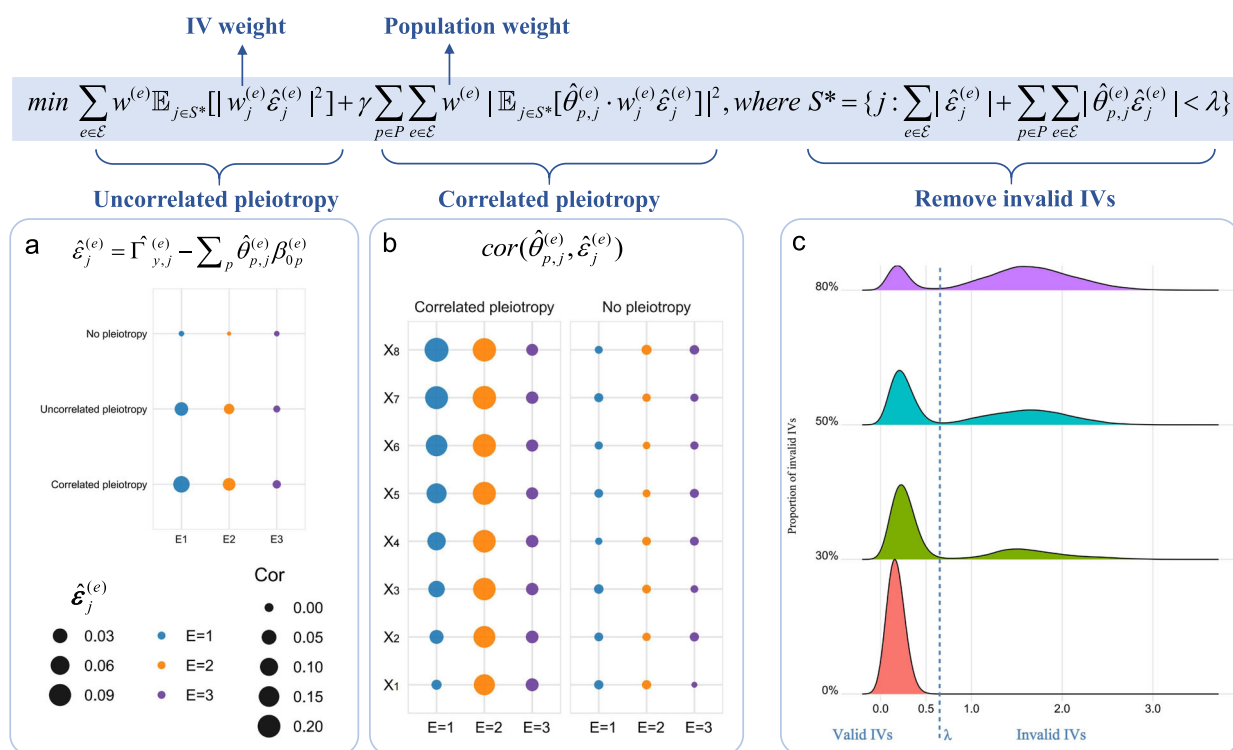
Fig. 1 | MR-EILLS model. The MR-EILLS model aims to infer the causal relationships between one or multiple exposures and one outcome, integrating multiple GWAS summary datasets from heterogeneous populations. There are different pleiotropic effects and IV strengths for the same IVs in heterogeneous populations. The objective function of the MR-EILLS model considers both correlated and uncorrelated pleiotropy and removes invalid IVs. Panels **a**–**c** show the results of the motivating simulation. Panel **a** shows the point plot for the absolute value of $\hat{\varepsilon}_j^{(e)}$ in different populations, where a larger point indicates a larger value of $|\hat{\varepsilon}_j^{(e)}|$. As the pleiotropic effect increases, $|\hat{\varepsilon}_j^{(e)}|$ increases; thus, the first part of the MR-EILLS model minimizes the pleiotropic effect between different populations. Panel **b** shows the correlation between $\hat{\varepsilon}_j^{(e)}$ and $\hat{\theta}_{p,j}^{(e)}$, which represents the correlated pleiotropic effect or the violation of the InSIDE assumption. As the correlated pleiotropic effect increases, this correlation becomes stronger. This corresponds to the second part of MR-EILLS, the empirical focused linear invariance regularizer, which discourages the selection of exposures with strong correlations between $\theta_{p,j}^{(e)}$ and $\varepsilon_j^{(e)}$ in some populations because this correlation would distort the causal effect estimation. Panel **c** shows the ridge plot of $\sum_{e \in E}|\varepsilon_j^{(e)}| + \sum_{p \in P}\sum_{e \in E}|\hat{\theta}_{p,j}^{(e)}\varepsilon_j^{(e)}|$ when there are different proportions of invalid IVs. When there are invalid IVs, the ridge plot has two peaks, whereas the ridge plot has only one peak when there is no invalid IV. The corresponding abscission value at the lowest point between the two peaks is the optimal λ. The third part of the MR-EILLS model removes the invalid IVs by λ.
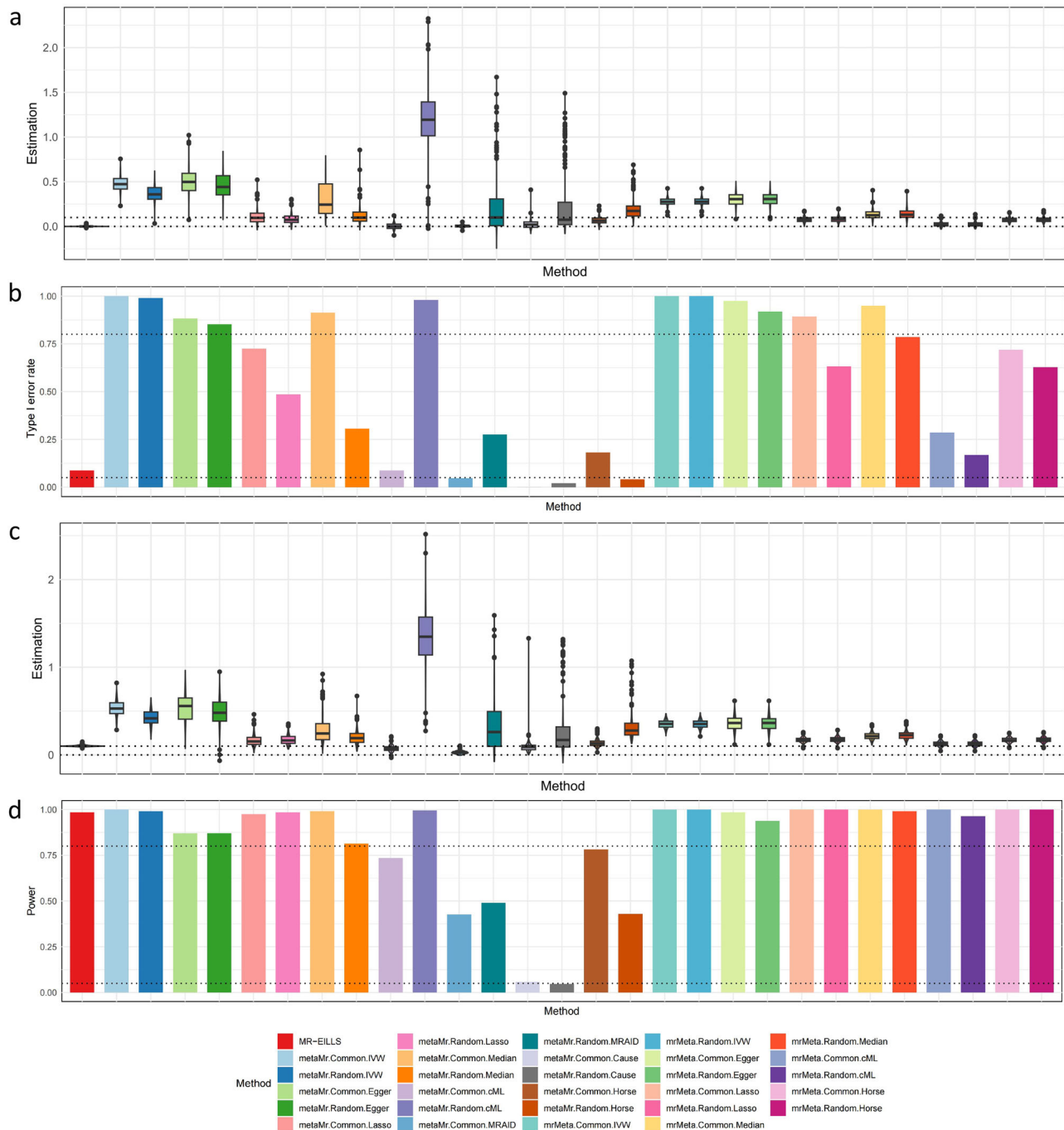
Fig. 2 | Simulation results when *P* = 1 (UVMR). a, **b** Results of the causal effect estimation and type I error rate when the causal effect is zero. **c**, **d** Results of causal effect estimation and statistical power when the causal effect is 0.1. The number of IVs is 100 and the proportion of invalid IVs is 30%. The number of populations is *E* = 3. 200 repeated datasets were generated in all simulations. Data in boxplots are presented as median values and interquartile range.

corresponding abscission value at the lowest point between the two peaks is the optimal λ. Thus, Eq. (2) removes the invalid IVs with pleiotropic effects larger than λ. Details for motivating the simulation are shown in Supplementary Note 1.

Simulation
We generated GWAS summary statistics for heterogeneous populations with varying edge effects, IV strengths, and pleiotropy values under both the UVMR and MVMR scenarios. We then compared MR-EILLS with nine published methods: IVW, MR-Egger, MR-Lasso, MR-Median, MR-cML, MRAID, Cause, MR-BMA, and MR-Horse. For these MR methods, we implemented two analytical strategies: (1) metaMR,

which first performs meta-analysis of GWAS summary statistics across all datasets for each variable, followed by MR analysis; and (2) mrMeta, which first conducts MR analysis separately for each dataset, followed by meta-analysis of the MR results. Both random effects and fixed effects meta-analysis approaches were employed.

In the UVMR scenario (a), with both correlated and uncorrelated pleiotropy (30% invalid IVs), MR-EILLS, Cause, and MR-cML with metaMR demonstrated unbiased causal effect estimation, whereas other methods exhibited bias (Fig. 2 and Supplementary Data 1, 2). Compared with Cause and MR-cML with metaMR, MR-EILLS showed superior performance with higher accuracy, more stable type I error rates when the causal effect was null, and greater statistical power for nonnull effects.
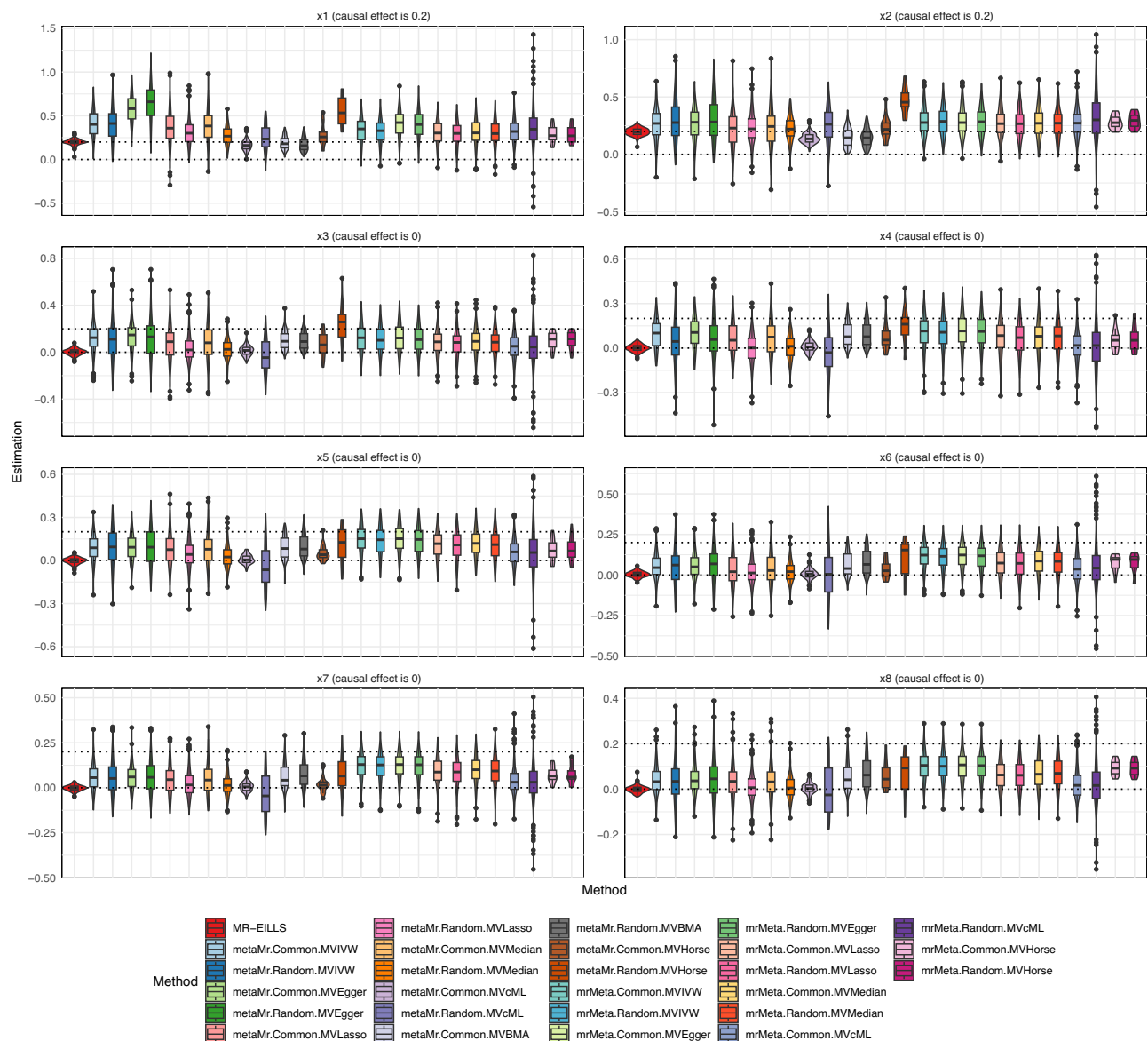
Fig. 3 | Simulation results of causal effect estimation when there is correlated and uncorrelated pleiotropy (MVMR). The number of IVs is 100, and the proportion of invalid IVs is 30%. The number of populations is $E = 3$. Among a total of eight exposures, two (X_1 and X_2) are causal exposures (with a causal effect of 0.2), and the other six ($X_3, ..., X_8$) are spurious exposures (with a causal effect of 0). 200 repeated datasets were generated in all simulations. Data in boxplots are presented as median values and interquartile range.

When 80% of the IVs were invalid, all the MR methods, including MR-cML and Cause, produced biased estimates, whereas MR-EILLS maintained unbiased estimation (Supplementary Figs. 8, 9,18, 19). Notably, MR-EILLS displayed robust performance across all evaluation metrics: it maintained appropriate type I error rates for null effects and achieved >90% power with 300 IVs for nonnull effects (Supplementary Figs. 10, 11 and Supplementary Figs. 20, 21). In scenario (b), MR-IVW and MR-Egger showed significant bias, whereas the other methods performed better with smaller bias. Additionally, all methods yielded more accurate estimates under balanced pleiotropy conditions than under unbalanced pleiotropy conditions (Supplementary Figs. 22–29). For scenario (c), all methods performed similarly well, exhibiting three key characteristics: unbiased effect estimation, well-controlled type I error rates for null effects, and high power (>80%) for detecting nonnull effects (Supplementary Figs. 30–33). The complete simulation results are presented in Supplementary Figs. 2–33 (Supplementary Note 2–5).

For MVMR analyses with eight exposures and 30% of IVs exhibiting either correlated or uncorrelated pleiotropy (case (a)), Fig. 3

(Supplementary Data 3) demonstrates that MR-EILLS provides unbiased causal effect estimates for all exposures, whereas other methods show varying degrees of bias, including slight biases observed with MR-cML using metaMR for certain exposures. Notably, MR-EILLS achieves the highest estimation accuracy among all methods. Figure 4 (Supplementary Data 3) presents type I error rates under null effects and statistical power under nonnull effects, revealing that MR-EILLS maintains the highest statistical power for nonnull effects while showing the most stable type I error control, albeit with marginally lower than nominal rates (0.05) for some exposures—a phenomenon that diminishes with increasing population sizes (Supplementary Figs. 52, 53). Similar performance patterns are observed when $P = 3$ (Supplementary Figs. 46, 47). Under more extreme conditions with 80% invalid IVs, all MR methods produce biased estimates except MR-EILLS, which maintains unbiased estimation (Supplementary Figs. 48, 49). Furthermore, MR-EILLS demonstrates robust type I error control and achieves >90% statistical power with 300 IVs for nonnull effects

Fig. 4 | Simulation results of the type I error rate for spurious exposures and statistical power for causal exposures when there is correlated and uncorrelated pleiotropy (MVMR). The number of IVs is 100, and the proportion of invalid IVs is 30%. The number of populations is $E = 3$. Among total of eight exposures, two (X_1 and X_2) are causal exposures (with a causal effect of 0.2), and the other six ($X_3, \dots, X_8$) are spurious exposures (with a causal effect of 0). This figure displays the type I error rates for X_1 and X_2, the statistical power for $X_3, \dots, X_8$.

(Supplementary Figs. 50, 51). For case (b), the results mirror those from UVMR analyses regardless of pleiotropy balance status (Supplementary Figs. 54–57). In the absence of pleiotropy (case (c)), all methods performed comparably well, yielding unbiased estimates, appropriate type I error rates, and high statistical power (>90%) (Supplementary Figs. 58, 59). Comprehensive simulation results are provided in Supplementary Figs. 38–59 (Supplementary Notes 6–9). When evaluating $P = 15$, Fig. 5 displays the mean F1 score, recall, and precision across methods, with MR-EILLS consistently outperforming all alternatives on these metrics.

Our analyses further revealed significant heterogeneity in causal effect estimates across different populations. Supplementary Figs. 34, 60 summarizes the variation in I^2 across all the simulations. To illustrate this heterogeneity, we randomly selected one simulation, and the detailed forest plots of causal effect estimates for each MR method and dataset are presented in Supplementary Figs. 36, 37, 61–63. These plots demonstrate substantial inconsistencies in effect estimates among populations.

Our simulation studies were conducted under weak IV conditions, with conditional F-statistics in simulations mirroring those observed in the applied analysis (Supplementary Data 4). Compared with alternative MR methods, our proposed approach exhibits minimal variance inflation while preserving superior statistical power (Supplementary Figs. 44, 45). These findings proved robust across different genetic correlation scenarios, maintaining consistent performance when accounting for nonzero genetic correlations between populations (Supplementary Figs. 4–7, 14–17, 40–43). In terms of computational efficiency, MR-EILLS maintains high performance regardless of the number of exposures. Conversely, methods such as MR-cML and MR-BMA—particularly MR-Horse—exhibit progressively increasing computational requirements, demanding significantly greater memory allocation and processing time as the number of exposures increases (Table 1).

Application

We explored the causal relationships between a total of 11 blood cells (five red blood cells: hemoglobin (HGB) concentrations, hematocrit
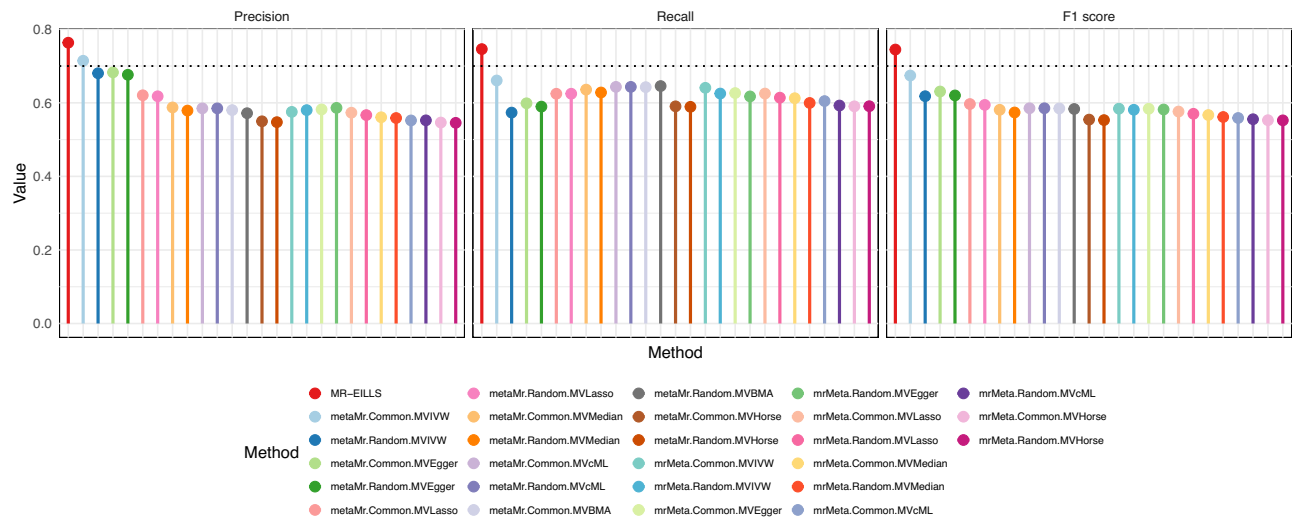
Fig. 5 | Simulation results of the F1 score, precision and recall when *P* = 15 (MVMR). The number of IVs is 100 and the proportion of invalid IVs is 30%. The number of populations is *E* = 3.

Table 1 | Computing time for different methods in the simulation

UVMR							
mrMeta							
IVW	Egger	Lasso	Median	cML	Horse		
0.053 secs	0.045 s	0.186 s	1.060 s	0.384 s	1.588 min		
metaMR							
IVW	Egger	Lasso	Median	cML	MRAID	Cause	Horse
3.478 secs	3.221 s	3.475 s	3.970 s	38.968 s	3.801 s	36.068 s	1.109 min
MR-EILLS							
0.208 secs							
MVMR							
mrMeta							
IVW	Egger	Lasso	Median	cML	Horse		
0.158 secs	0.156 s	0.272 s	18.271 s	4.132 s	126.325 min		
metaMR							
MVIVW	MVEgger	MVLasso	MVMedian	MVcML	MVBMA	MVHorse	
14.931 secs	14.950 s	14.845 s	26.557 s	4.612 min	3.877 min	91.088 min	
MR-EILLS							
1.089 s							

(HCT), mean corpuscular hemoglobin (MCH) concentrations, mean corpuscular volume (MCV), and the mean corpuscular hemoglobin concentration (MCHC); five white blood cells: white blood cell (WBC) counts, neutrophil (Neutro) counts, monocyte (Mono) counts, basophil (Baso) counts, and eosinophil (Eosin) counts; one platelets: platelet (PLT) counts) and 20 disease-related outcomes (asthma, body fat percentage, body mass index, waist circumference, weight, fasting blood glucose, HbA1c measurement, total cholesterol, type 2 diabetes, cardioembolic stroke, ischemic stroke, large artery stroke, stroke, heel bone density, pneumonia, schizophrenia, forced expiratory volume (FEV), forced vital capacity (FVC), the FEV/FVC ratio, and peak expiratory flow (PEF)) using GWAS summary statistics from five ancestries: African, East Asian, South Asian, Hispanics Latinos, and European.

First, we conducted traditional MR analysis in five ancestries, and performed heterogeneous analysis for each MR method. The results are shown in Fig. 6 (Supplementary Data 5). We found that there were large heterogeneities ($I^2 > 0.75$) for blood cells among five ancestries. We subsequently conducted multivariable MR-EILLS analysis to explore the independent causal effects of 11 blood cells on 20 disease-related outcomes. We plot ridge plots for each outcome in five ancestries, and the results are shown in Supplementary Figs. 64–66. We used the ridge plot to set the λ for MR-EILLS (Supplementary Data 6).

In Fig. 7 (Supplementary Data 7), we found that higher counts of some white blood cells, red blood cells or platelets were independently associated with decreased lung function. For FEV, higher WBC, Neutro counts and HGB concentrations causally induced a lower FEV (WBC: beta = −0.14, 95%CI: [−0.24, −0.04]; Neutro: beta = −0.17, 95%CI: [−0.24, −0.04]; HGB: beta = −0.29, 95%CI: [−0.54, −0.03]). The counts of Neutro and HCT were negatively associated with the level of FVC (Neutro: beta = −0.09, 95%CI: [−0.18, −0.01]; HCT: beta = −0.06, 95%CI: [−0.13, −0.002]). In addition, increases in the PLT and Neutro counts were associated with a decreased FEV/FVC ratio (PLT: beta = −0.26, 95%CI: [−0.49, −0.02]; Neutro: beta = −0.16, 95%CI: [−0.30, −0.02]). Higher MCH concentrations might result in a lower PEF level (beta = −0.08, 95%CI: [−0.16, −0.004]). James et al. reported that an increased WBC count is associated with lower levels of lung function and
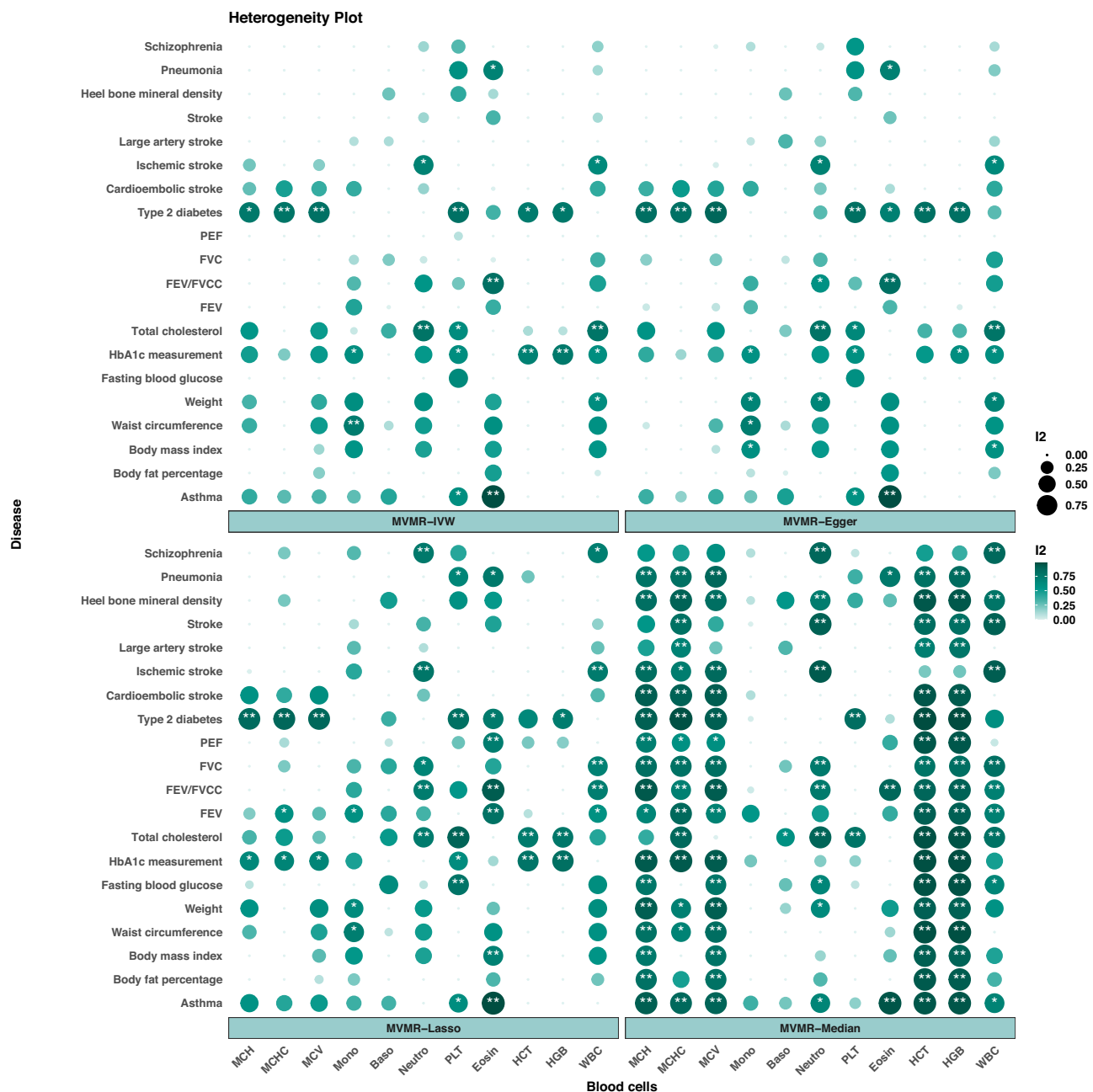
Fig. 6 | Heterogeneity plot of the application results. The heterogeneity of MR estimations in each population using MVMR-IVW, MVMR-Egger, MVMR-Lasso, and MVMR-Median, whereas MVMR-cML and MVMR-Horse cannot output MR estimations owing to the substantial memory consumption and computational time.

provided biological explanations[22]. A 15-year longitudinal study demonstrated that higher blood neutrophil concentrations were associated with accelerated FEV decline[23]. The inverse relationships of FEV and FVC with red blood cell counts were also supported by observational studies[24,25]. A prospective longitudinal analysis revealed that a higher baseline neutrophil count predicted a lower serially obtained FVC[26]. Moreover, a retrospective study revealed a strong correlation between the PLT count and the FEV/FVC ratio[27].

Our study provides compelling evidence for causal relationships between cytokine profiles and stroke pathogenesis. Elevated PLT counts were independently correlated with increased overall stroke risk (OR = 1.93, 95%CI: [1.37, 2.71]), particularly for the cardioembolic stroke subtype. Increased Mono levels were positively associated with stroke susceptibility (OR = 1.39, 95%CI: [1.24, 1.57]), whereas elevated MCHC specifically amplified cardioembolic stroke risk (OR = 1.25, 95%

CI: [1.03, 1.53]). MCH levels exhibited a marginal yet statistically significant association with ischaemic stroke incidence (OR = 1.06, 95%CI: [1.01, 1.12]). Notably, reduced WBC counts conferred protection against overall stroke risk (OR = 0.63, 95%CI: [0.47, 0.84]), with enhanced preventive efficacy observed in the large artery stroke subtype (OR = 0.55, 95%CI: [0.40, 0.75]). Protective associations were further identified across multiple hematological parameters: Eosin (OR = 0.38, 95%CI: [0.34, 0.44]), Neutro (cardioembolic: OR = 0.74, 95%CI: [0.58, 0.95]; large artery: OR = 0.54, 95%CI: [0.39, 0.76]), Baso (OR = 0.39, 95%CI: [0.31, 0.49]), HGB (OR = 0.64, 95%CI: [0.57, 0.72]), and HCT (OR = 0.69, 95%CI: [0.63, 0.76]).

These findings align with established mechanisms of inflammation-mediated vascular pathology. Mansfield et al.[28] demonstrated that cytokine-driven endothelial dysfunction occurs through the IL-6/TNF-α pathway, providing mechanistic support for our
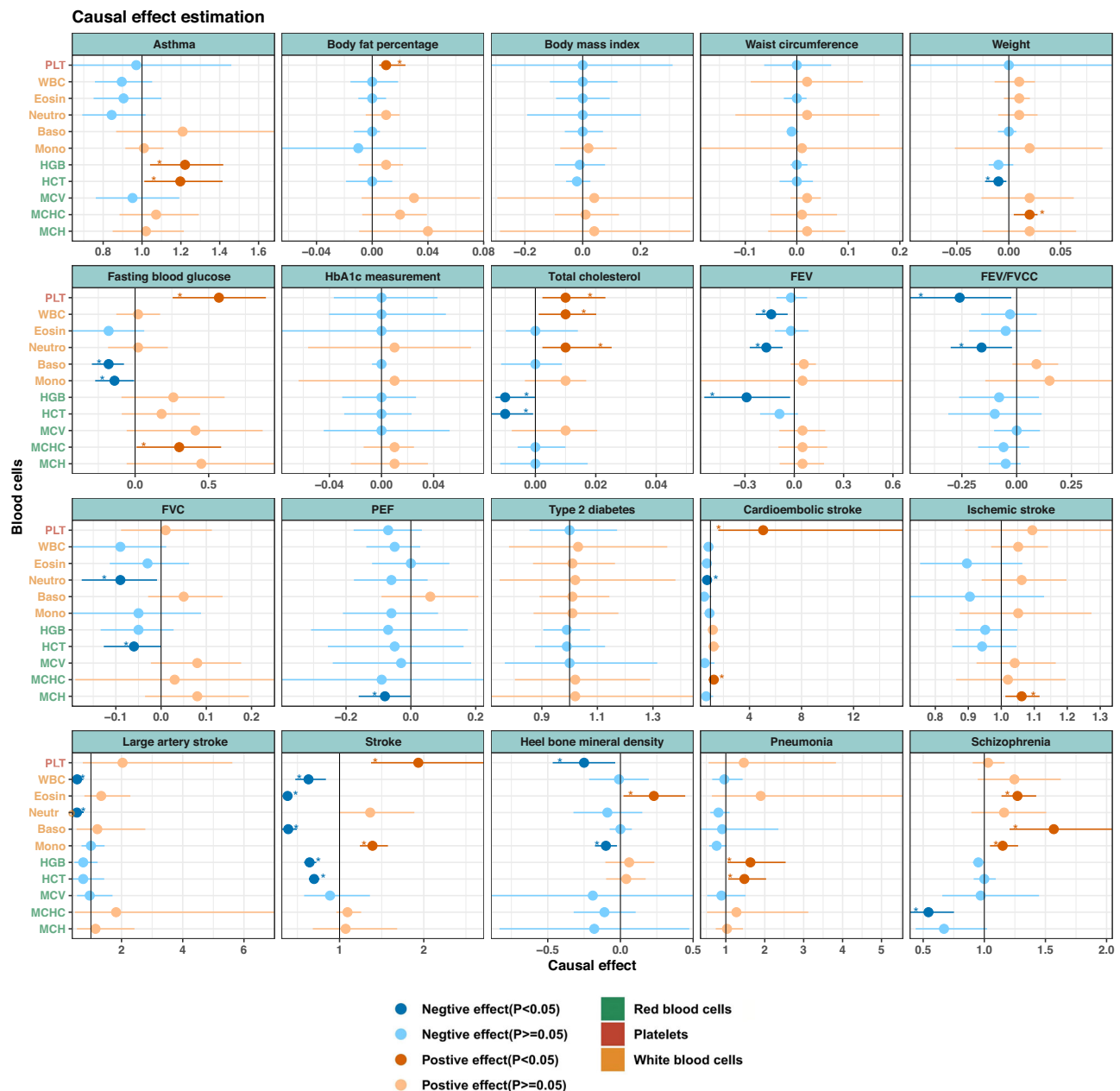
Causal effect estimation

Fig. 7 | Forest plot of the application results. Causal effect estimations of 11 blood cells on disease-related outcomes via the MVMR-EILLS method. Data are presented as mean values and 95% confidence intervals. All the *P* value and confidence interval is estimated via the bootstrap method, and it is two-sided.

observed mono-stroke associations. The neuroprotective neutrophil subset (N2) identified by ref. 29 offers a plausible explanation for the reduced stroke risk at lower Neutro counts through inflammatory modulation. Rodhe et al.[30] further substantiated cytokine-mediated vascular injury through IL-8/TNF-α correlations in geriatric populations, whereas ref. 31 delineated cytokine polarization patterns affecting vascular homeostasis, which is consistent with our PLT and HGB findings. Tyagi et al.[32] established chemokine-specific inflammatory cascades (e.g., CXCL10), reinforcing our conclusions that Eosin and Baso count reductions attenuate stroke risk.

Our analysis demonstrated that elevated PLT counts independently increased body fat percentage (beta = 0.01, 95%CI: [0.005, 0.02]), although no significant causal associations were detected between blood cell indices and body mass index or waist circumference. Higher MCHC values were positively related to weight (beta = 0.02, 95%CI: [0.004, 0.02]), whereas lower HCT levels were causally linked to increased weight (beta = −0.01, 95%CI:

[−0.002, −0.02]). Our findings align with studies linking hematological indices to metabolic outcomes. Zhang et al.[33] demonstrated that elevated HCT in obese individuals with dyslipidaemia correlates with worsened lipid profiles (higher TC, TG, LDL-C; lower HDL-C), supporting our observation that lower HCT levels are causally associated with increased weight. Similarly, ref. 34 established HCT as a key determinant of blood viscosity, which impairs insulin sensitivity and promotes adiposity—mechanisms consistent with our positive association between MCHC and body mass. These studies reinforce the role of red blood cell parameters in metabolic regulation, although causal relationships remain nuanced compared with BMI/waist circumference.

Elevated PLT (beta = 0.57, 95%CI: [0.26, 0.89]) and MCHC (beta = 0.3, 95%CI: [0.01, 0.58]) values were independently positively associated with fasting blood glucose. Conversely, Baso (beta = −0.18, 95%CI: [−0.29, −0.08]) and Mono (beta = −0.14, 95%CI: [−0.17, −0.007]) counts demonstrated inverse correlations with glucose levels. For lipid

metabolism, the PLT (beta = 0.001, 95%CI: [0.002, 0.023]), WBC (beta = 0.01, 95%CI: [0.001, 0.02]), and Neutro (beta = 0.01, 95%CI: [0.002, 0.02]) counts exhibited positive dose-response relationships with total cholesterol. In contrast, HGB (HGB, beta = −0.01, 95%CI: [−0.013, −0.0001]) and HCT (beta = −0.01, 95%CI: [−0.014, −0.0007]) levels displayed protective associations against cholesterol elevation. Our findings align with the established literature linking hematological indices to metabolic outcomes. Elevated PLT and MCHC levels predict fasting blood glucose, which is consistent with studies demonstrating the role of platelet activation in insulin resistance[35] and the impact of erythrocyte rigidity on endothelial function[36]. Lower Baso and Mono counts are correlated with increased glucose levels, supporting their anti-inflammatory roles in glucose metabolism[37,38]. For lipid metabolism, the PLT, WBC, and Neutro counts are positively associated with total cholesterol, mirroring leukocyte-driven atherosclerosis mechanisms[39] and haemorheological contributions to dyslipidaemia[40]. Conversely, lower HGB and HCT levels have protective effects against hypercholesterolemia, which aligns with the role of iron metabolism in lipid regulation[41].

Elevated HGB and HCT independently elevated asthma risk (HGB: OR = 1.22, 95%CI: [1.04, 1.41]; HCT: OR = 1.19, 95%CI: [1.01, 1.41]). Higher Eosin counts were positively associated with heel bone density (beta = 0.23, 95%CI: [0.02, 0.45]), whereas reduced PLT and mono counts were inversely correlated (PLT: beta = −0.25, 95%CI: [−0.47, −0.04]; Mono: beta = −0.1, 95%CI: [−0.17, −0.02]). HGB and HCT elevations were linked to pneumonia risk (HGB: OR = 1.63, 95%CI: [1.04, 2.54]; HCT: OR = 1.48, 95%CI: [1.06, 2.03]). Schizophrenia risk exhibited dose-dependent relationships with Eosin (OR = 1.27, 95%CI: [1.14, 1.42]), Baso (OR = 1.56, 95%CI: [1.21, 2.05]), and Mono (OR = 1.15, 95%CI: [1.04, 1.28]) counts, in contrast with the protective effect of a lower MCHC (OR = 0.54, 95%CI: [0.39, 0.75]). The independent elevation of asthma risk by HGB and HCT aligns with studies showing that iron deficiency anemia (proxied by low HGB/HCT) impairs lung function and increases airway hyperresponsiveness[42]. For bone health, the positive correlation between Eosin counts and heel bone density supports the role of eosinophils in osteoblast activation, as demonstrated by the fact that eosinophil-derived growth factors promote bone formation[43]. Conversely, lower PLT and Mono counts may reflect anti-inflammatory environments that reduce asthma exacerbations[44]. The links between HGB/HCT elevations and pneumonia risk correspond with evidence that iron overload (high HGB/HCT) promotes bacterial adherence and pulmonary inflammation[45]. With respect to psychiatric outcomes, the dose-dependent relationships of Eosin, Baso, and Mono counts with schizophrenia risk align with immune dysregulation theories, where proinflammatory monocytes and basophils exacerbate neuroinflammation[46]. Finally, the protective effect of low MCHCs against schizophrenia may reflect iron homeostasis disruption in schizophrenic patients, as elevated MCHCs are correlated with oxidative stress and dopamine dysregulation[47]. The details of the results are shown in Supplementary Data 5–9 and Supplementary Note 10.

Discussion

In this paper, we proposed an MR method MR-EILLS, which works in both univariable and multivariable frameworks, and it outputs the pure causal effect estimation of multiple heterogeneous populations using only GWAS summary statistics. The results of the simulation revealed the superior performance of MR-EILLS and its application in exploring causal relationships from 11 blood cells to 20 disease-related outcomes, which covered most of the expected causal links.

MR-EILLS integrates GWAS summary datasets from heterogeneous populations, and for each population, GWAS summary datasets for exposure and outcome can be from either the same individuals or different but heterogeneous individuals. This assumption is the same as that in traditional two-sample MR analysis, which requires two

homogeneous but nonoverlapping samples. MR-EILLS assumes that the GWAS summary datasets for each population are from homogeneous but nonoverlapping samples. In the application, we assume that the individuals in each ancestry are homogeneous, and that the genetic diversity in different ancestries leads to heterogeneity among ancestries (different IV strengths and pleiotropy). The heterogeneity among populations includes genetic and nongenetic differences, also creating distinct challenges for IV validity across groups. Genetic architecture differences include allele frequency variations, population-specific linkage disequilibrium (LD) patterns, divergent selection pressures and unique mutation histories, etc. In addition, nongenetic modifiers—such as socioeconomic status, environmental exposures, cultural practices, and disparities in healthcare access—also play a significant role in shaping population heterogeneity. These contextual differences contribute to inconsistent causal relationships across different populations, highlighting the complex interplay between genetic predispositions and external determinants of health.

In our context, we assume that given the causal exposures, the conditional expectation of the outcome is invariant, that is, the causal effects of causal exposures on the outcome are invariant across different populations. The joint distribution of the candidate exposures and outcome may vary across different populations due to population heterogeneity. Therefore, the conditional expectation invariance structure assumption of the original EILLS method is satisfied. Another two identification conditions for identifying pure causal effect are: (1) it is necessary for identifying pre causal effect to select heterogeneous populations with large differences in social and environment factors including genetic difference; (2) a minimal identification condition related to the heterogeneity of the populations: for a exposures' set, if $\hat{\theta}_{p,j}^{(e)}$ and $\hat{\varepsilon}_j^{(e)}$ is not independent in at least of one population, then there must be two populations with different causal effects. However, both of these identification conditions are untestable in practice. Therefore, when applying this method, practitioners must rely on prior knowledge to satisfy the identification assumptions as much as possible.

MR-EILLS is specifically designed for estimating causal effects in biological pathways where variables and pathway effects remain independent of social/environmental factors—either unaffected by or unmodified through these contextual influences. Taking the causal effect of hemoglobin concentration on stroke risk as an example, MR-EILLS tend to estimation the causal effect on the biological pathway like Hemoglobin level → Impaired oxygen transport → Elevated blood viscosity → Thrombosis → Stroke risk, or Hemoglobin level → Anemia development → Compensatory cardiac mechanisms → Left ventricular hypertrophy → Stroke risk, etc., but not hemoglobin level → oxygen transport efficiency → fatigue symptoms → reduced physical activity → elevated stroke risk. MR-EILLS methodology excludes such socio-environmentally mediated pathways when estimating invariant causal effects, as the mediator ("exercise") inherently exhibits environmental variability. MR-EILLS aims to eliminate the influence of socioeconomic factors on the causal invariant effect from exposures to the outcome. The methodology does not require study populations to be globally representative, as it remains applicable to specific observational study cohorts. This is justified because socioeconomic confounders or effect modifiers inherently exist within any defined study population in observational research. Similar to MR, MR-EILLS requires no interaction assumptions for causal exposures and environmental variables. If there is an interaction between exposure and environmental variables, then this exposure tends to be considered a spurious exposure. If there is an interaction between the environmental variable and the mediator variable in the mediation pathway from exposure to outcome, then the pathway where this mediator variable is located is not included in the invariant causal effect. The additional explanation for the pure causal effect is shown in Supplementary Note 11.

MR-EILLS allows different IVs to be set in different populations. However, the strategy for metaMR, that is, first conducting GWAS

meta-analysis and then performing MR analysis, requires the SNPs that are independent (i.e., no LD detected) in all populations, which reduces the number of IVs substantially; however, GWAS meta-analysis helps researchers identify more significant SNPs (i.e., $P < 5 \times 10^{-8}$). In addition, only a few MR methods allow the SNP set to have high LD. MR-EILLS solves this tricky issue and only requires that the IV set in each population is independent without LD.

The MR-EILLS model has two hyperparameters, which require researchers to set appropriate values to estimate the causal effects of exposures on the outcome. For γ, we recommend $\gamma > 0.4$ in UVMR and $\gamma > 0$ in MVMR. The larger γ is, the stronger the role of the empirical focused linear invariance regularizer is. For λ, we suggest that researchers construct ridge plots to find the optimal value. In Model (2), we keep the SNP for which the pleiotropic effect in all populations is lower than λ. When the scales of different populations are different, the Model (2) can be modified to the following Model (2-1)

$$S^* = \{ j : |\varepsilon_j^{(e)}| + \sum_{p \in P} |\hat{\theta}_{p,j}^{(e)} \varepsilon_j^{(e)}| < \lambda_e \, for \, any \, e \} \quad (2-1)$$

Researchers can set different λ_e values for different populations. For example, in our applications, we set different λ_e values for five ancestries and five ridge plots are shown for each outcome. MR-EILLS works only if there are at least $J \geq P$ valid IVs in the IV set; this assumption is less strict than the plurality assumption[17], which requires the valid IVs from the largest group of IVs sharing the same causal parameter value.

There are several limitations for MR-EILLS. The first is that MR-EILLS does not yet work in the high-dimensional case. One future key research direction is to extend MR-EILLS to high-dimensional exposure scenarios, especially for high-dimensional omics biomarkers; for this purpose, correlated IVs are also important issues to be solved. Another point is that inappropriate settings of hyperparameters may lead to incorrect inference of causal relationships between exposures and outcomes. It is important to choose the appropriate hyperparameters, especially for λ. The value of λ determines whether the invalid IVs are removed, and if λ is too large, the causal effect estimation would be distorted. If λ is too small, the number of remaining IVs is small; thus, in the future, it is necessary to extend MR-EILLS to correlated IV scenarios. Our empirical application of MVMR to analyse 11 potentially correlated blood cell traits presents notable methodological challenges. The high degree of intertrait correlation, combined with the difficulty of identifying truly independent IVs for each exposure, renders this analysis particularly vulnerable to weak instrument bias. While acknowledging these fundamental limitations inherent in analyzing highly correlated traits, our proposed methodology demonstrates improved robustness by substantially reducing variance inflation compared to conventional approaches. This enhancement enables more reliable inference despite the analytical challenges posed by trait correlations.

In conclusion, we propose the MR method MR-EILLS, which integrates multiple heterogeneous GWAS summary datasets to infer the pure causal relationships between exposures and outcomes. This study has important guiding significance for the discovery of new disease-related factors. We look forward to offering constructive suggestions for disease diagnosis and applying our method beyond the scope considered here.

Methods
Ethics

Overview of MR methods
For one population, assume P exposures $X_p, p \in \{1, ..., P\}$ and one outcome Y. The J independent IVs $G_j, j \in \{1, ..., J\}$ satisfy the following three assumptions:

A1. Relevance, G_j is associated with at least one of the P exposures;

A2. Exchangeability, G_j is not associated with the confounder between P exposures and the outcome;

A3. Exclusion restriction, G_j affects the outcome only through exposures.

The MR model based on the individual data is as follows:

$$U = \sum_j \omega_j G_j + \varepsilon_{X_U}$$
$$X_p = \sum_j \alpha_{pj} G_j + \sum_{X_k \in pa(X_p)} \beta_{X_k X_p} X_k + \beta_{1p} U + \varepsilon_{X_p} \quad (3)$$
$$Y = \sum_j \gamma_j G_j + \sum_p \beta_{0p} X_p + \beta_2 U + \varepsilon_Y$$

where $\varepsilon_{X_U}, \varepsilon_{X_p}, \varepsilon_Y \sim N(0,1)$. γ_j represents the uncorrelated pleiotropic effect and ω_j represents the correlated pleiotropy. $X_k \in pa(X_p)$ is the father node of X_p, which is the direct cause of X_p, and where $\beta_{X_k X_p}^{(e)}$ represents the effect of X_k on X_p. β_{0p} denotes the causal effect of X_p on Y. We call the exposures with $\beta_{0p} \neq 0$ causal exposures, which we want to discover, whereas the exposures with $\beta_{0p} = 0$ are the spurious exposures, which are not the true cause of the outcome. We define the set of causal exposures as $\{X_p\}, p \in P^* \subseteq \{1, ..., P\}$. When $P = 1$, the above model is called UVMR, whereas when $P > 1$, it is called MVMR. To simplify the expression, our model below uniformly uses P exposures, both of which are applicable to UVMR and MVMR.

The GWAS summary statistics include the $G_j - X_p$ association $\hat{\theta}_{p,j}$ and its variance $\sigma_{p,j}^2$, as well as the $G_j - Y$ association $\hat{\Gamma}_{y,j}$ and its variance $\sigma_{y,j}^2$. Based on Model (3), we have

$$\theta_{p,j} = \alpha_{p,j} + \omega_j \beta_{1p} + \sum_{X_k \in pa(X_p)} \theta_{k,j} \beta_{X_k X_p}$$
$$\Gamma_{y,j} = \omega_j \beta_2 + \gamma_j + \sum_p \theta_{p,j} \beta_{0p} \quad (4)$$

When G_j is a valid IV (no pleiotropy), that is, $\gamma_j = \omega_j = 0$, then $\varepsilon_j = \Gamma_{y,j} - \sum_p \theta_{p,j} \beta_{0p}$ is zero and is dependent on $\theta_{p,j}$. For $j \in \{1, ..., J\}$, we can identify $\beta_{0p}(p \in \{1, ..., P\})$ by the system of linear equations $\Gamma_{y,j} = \sum_p \theta_{p,j} \beta_{0p}$ if and only if $J \geq P$. The causal effects of exposures on the outcome β_{0p} can be estimated by a weighted version of ordinary least squares (OLS), that is, the IVW regression

$$\hat{\Gamma}_{y,j} = \sum_p \hat{\theta}_{p,j} \beta_{0p} + \zeta_j, \zeta_j \sim N(0, \sigma_{y,j}^2) \quad (5)$$

which sets the intercept equal to zero. This model minimizes the empirical L_2 loss objective function

$$\begin{aligned} Q(\beta_{0p}; \hat{\theta}_{p,j}, \hat{\Gamma}_{y,j}, \sigma_{y,j}^2) \\ = E[|w_j \varepsilon_j|^2] \\ = E[|w_j(\hat{\Gamma}_{y,j} - \sum_p \hat{\theta}_{p,j} \beta_{0p})|^2] \end{aligned} \quad (6)$$

where w_j represents the weight of IV G_j in the causal effect estimation. If G_j has uncorrelated pleiotropy ($\gamma_j \neq 0$), that is, if G_j is causally associated with Y not through any X_p, then $\varepsilon_j = \gamma_j$ is no longer equal to zero, and it represents the uncorrelated pleiotropic effect. MR-Egger regression[18] is proposed to solve this problem by allowing the intercept term $\bar{\gamma}$ in Model (5), and the intercept represents the mean pleiotropic effect. The causal effect can be estimated by minimizing

$$Q(\beta_{0p}; \hat{\theta}_{p,j}, \hat{\Gamma}_{y,j}, \sigma_{y,j}^2) = E[|w_j(\hat{\Gamma}_{y,j} - \sum_p \hat{\theta}_{p,j} \beta_{0p} - \bar{\gamma})|^2].$$

MR-Egger regression requires the InSIDE assumption, which means that the pleiotropic effect is independent of $\theta_{p,j}$. If G_j has correlated pleiotropy ($\omega_j \neq 0$), that is, if G_j is causally associated with the

unmeasured confounding U between X_p and Y, then the pleiotropic effect $\varepsilon_j = \omega_j\beta_2 + \gamma_j$ is not independent of $\theta_{p,j}$ because of the common term ω_j. This is a violation of the InSIDE assumption. Equations (5–6) and MR-Egger require that ε_j is independent of $\theta_{p,j}$ because the correlation between the intercept term and the independent variables would distort the causal effect estimation. The results of the motivating simulation for correlated and uncorrelated pleiotropy are shown in Supplementary Fig. 1.

The MR-Lasso method applies lasso-type penalization to the direct effects of IVs on the outcome. The causal estimate is described as a post-lasso estimate and is obtained via the IVW method, which uses only those IVs that are identified as valid by the lasso procedure. The objective function is as follows:

$$Q(\beta_{0p}; \hat{\theta}_{p,j}, \hat{\Gamma}_{y,j}, \sigma^2_{y,j}) = E[|w_j(\hat{\Gamma}_{y,j} - \sum_p \hat{\theta}_{p,j}\beta_{0p} - \gamma_j)|^2 + \lambda_L|\gamma_j|]$$

If γ_j shrinks to zero, the IV is treated as a valid IV. The MR-median estimator is defined as the 50th percentile of either an unweighted or IVW empirical density function of the Wald ratio, and it is consistent even when up to 50% of the information comes from an invalid IV. The aforementioned methods are classified into a category of MR methods, namely, the Wald ratio-based MR method.

The MR-cML and MRAID methods are representative of the likelihood-based MR method. MR-cML infers causal relationships using constrained maximum likelihood. It is robust to both uncorrelated and correlated pleiotropic effects, under the plurality assumption. MRAID models an initial set of candidate IVs that are in potentially high linkage disequilibrium with each other and automatically selects among them the suitable IVs for causal inference, under the joint likelihood framework. MRAID also explicitly models both uncorrelated and correlated horizontal pleiotropy.

The MR-Horse and cause methods produce unbiased causal effect estimates under the framework of Bayesian inference, while avoiding inflated false positive rates, and can account for both correlated and uncorrelated pleiotropy. The disadvantage of these two methods is that when the number of exposures is large, they require a significant amount of memory and computation time.

MR-BMA is an MVMR method used to select likely causal risk factors from high-throughput experiments. It uses Bayesian model averaging and computes the posterior probability for all possible combinations of risk factors, finally estimating the model-averaged causal estimate (MACE) by weighting and summing the causal effect estimation of each model. To some extent, MR-BMA avoids pleiotropy by considering as many risk factors as possible.

These methods are all based on a single dataset, and only information from one dataset can be used. Therefore, we proposed a method that fully exploits the information from multiple existing datasets to identify causally invariant factors, comprehensively improves the estimation precision and enhances the statistical power.

MR-EILLS model: MR integrating multiple heterogeneous populations

When there are E heterogeneous populations, GWAS summary statistics include $\hat{\theta}^{(e)}_{p,j}$, $\sigma^{(e)2}_{G_jX_p}$, $\hat{\Gamma}^{(e)}_{y,j}$ and $\sigma^{(e)2}_{y,j}$ for $e \in \mathcal{E}$. We define $\varepsilon^{(e)}_j = \Gamma^{(e)}_{y,j} - \sum_p \theta^{(e)}_{p,j}\beta^{(e)}_{0p}$ and $\hat{\varepsilon}^{(e)}_j = \hat{\Gamma}^{(e)}_{y,j} - \sum_p \hat{\theta}^{(e)}_{p,j}\beta^{(e)}_{0p}$ in the version of multiple populations. We use the superscript (e) to denote the e-th population. We assume that the pleiotropic effect, IV strength and relationships among exposures are different in heterogeneous populations, whereas the causal effects of causal exposures on Y are invariant, that is $\beta^{(1)}_{0p} = \beta^{(2)}_{0p} = \ldots = \beta^{(E)}_{0p} = \beta^*_{0p}$ for $p \in P^*$; this assumption is called the structure assumption[21]. These assumptions are rational because the IVs satisfying A1–A3 control only the unmeasured confounders between X_p and Y, whereas other unmeasured confounders between IVs and

exposures, or between IVs and outcomes, or between exposures are not controlled, and these unmeasured confounders are also the reason for heterogeneity between populations.

Note that a valid IV in one population may be an invalid IV in the other heterogeneous populations. On the other hand, an IV may be associated with exposure in all heterogeneous populations, while it may have different uncorrelated or correlated pleiotropy in different populations. This leads to inconsistent independence relationships between $\theta^{(e)}_{p,j}$ and $\varepsilon^{(e)}_j$ across different populations and inconsistent causal effect estimation of exposures on the outcome in different heterogeneous populations. Therefore, we leveraged the EILLS[21], the multiple heterogeneous population version of OLS, to construct the MR-EILLS model. The MR-EILLS model integrates GWAS summary statistics from multiple heterogeneous populations and finds causal exposures that have invariant effects on the outcome in heterogeneous populations. The MR-EILLS model aims to minimize the following objective function

$$Q(\beta^*_{0p}; \hat{\theta}^{(e)}_{p,j}, \hat{\Gamma}^{(e)}_{y,j}, \sigma^{(e)2}_{y,j})$$
$$= \sum_{e\in\mathcal{E}} w^{(e)} E_{j\in S^*}[|w^{(e)}_j \hat{\varepsilon}^{(e)}_j|^2] + \gamma \sum_{e\in\mathcal{E}} w^{(e)} \sum_{p\in P} |E_{j\in S^*}[\hat{\theta}^{(e)}_{p,j} \cdot w^{(e)}_j \hat{\varepsilon}^{(e)}_j]|^2 \quad (7)$$

where

$$w^{(e)}_j = \frac{\sigma^{(e)-2}_{y,j}}{\sum_{j\in S^*} \sigma^{(e)-2}_{y,j}} \text{ and } w^{(e)} = \frac{N_e}{N} \quad (8)$$

$w^{(e)}_j$ is the weight of IV G_j on the causal effect estimation in the e-th population, and $w^{(e)}$ is the weight of the eth population on the final causal effect estimation. The first part of the objective function (1) is the empirical L_2 loss, which is the multiple-population version of the objective function (6) in one population. The second part of the objective function (1) is the empirical focused linear invariance regularizer, which discourages the selection of exposures with strong correlations between $\theta^{(e)}_{p,j}$ and $\varepsilon^{(e)}_j$ in some populations because this will distort the causal effect estimation. $\gamma > 0$ is the hyperparameter. In addition, we added the following restriction

$$S^* = \{j : \sum_{e\in\mathcal{E}} |\varepsilon^{(e)}_j| + \sum_{p\in P}\sum_{e\in\mathcal{E}} |\hat{\theta}^{(e)}_{p,j}\varepsilon^{(e)}_j| < \lambda\} \quad (9)$$

to select the valid IVs. The first part of Eq. (2) represents the uncorrelated pleiotropic effect for the $j - th$ IV, and the second part of Eq. (2) represents the correlated pleiotropic effect for the $j - th$ IV. $\lambda > 0$ is the hyperparameter controlling the strictness of filtering IVs. Equation (2) removes the invalid IVs with pleiotropic effects above λ.

The causal effects β^*_{0p} can be identified under the assumption[21] that there are at least P valid IVs in the IV set, that is $J \geq P$. We use a limited-memory modification of the BFGS quasi-Newton method[48] to find the optimal solution β^*_{0p} of the objective function (1) under the restriction of Eq. (2). The confidence interval is estimated via bootstrap method.

Simulation

We generate the GWAS summary statistics of E heterogeneous populations via the following process:

$$\theta^{(e)}_{p,j} = \alpha^{(e)}_{p,j} + \omega^{(e)}_j \beta^{(e)}_{1p} + \sum_{X_k \in pa(X_p)} \theta^{(e)}_{k,j}\beta^{(e)}_{X_kX_p} + \xi^{(e)}_{p,j}$$

$$\Gamma^{(e)}_{y,j} = \omega^{(e)}_j \beta^{(e)}_2 + \gamma^{(e)}_j + \sum_p \theta^{(e)}_{p,j}\beta^{(e)}_{0p} + \xi^{(e)}_{y,j}$$

where $X_k \in pa(X_p)$ is the father node of X_p, which is the direct cause of X_p, and where $\beta^{(e)}_{X_kX_p}$ represents the effect of X_k on X_p. In total P

exposures, the causal exposures are the top 30% of all exposures (e.g., $P = 8$, $floor(P \times 30\%) = 2$, then the top two (X_1 and X_2) are the causal exposures). The effect of causal exposure on Y ($\beta_{0p}^{(e)}$, $p \in P^*$) is 0.2 for MVMR ($P > 1$) and 0.1 for UVMR ($P = 1$), and the effect of other spurious exposures on Y ($\beta_{0p}^{(e)}$, $p \notin P^*$) is 0.$\beta_{X_k X_p}^{(e)} \sim U(-1, 1)$ for the effect of edge $X_k \to X_p$. The structure between the exposures is randomly generated, and the parameter $\beta_{X_k X_p}^{(e)}$ represents the effect from X_k to X_p. We set IV strength $\alpha_{p,j}^{(e)} \sim N(0, 0.2)$ for the eth population and X_p; $\xi_{p,j}^{(e)} \sim N(0, \sigma_{p,j}^{(e)2})$ for eth population and X_p, $\sigma_{p,j}^{(e)2} \sim U(0.01, 0.05)$ for X_p; $\xi_{y,j}^{(e)} \sim N(0, \sigma_y^{(e)2})$ for $E = e$, $\sigma_{y,j}^{(e)2} \sim U(0.05, 0.1)$ and different variances represent different sample sizes; $\beta_{1p}^{(e)} \sim U(0.5, 0.8)$ for X_p; and $\beta_2^{(e)} \sim U(0.5, 0.8)$. We consider three scenarios:

(a) uncorrelated and correlated pleiotropy effects, $\gamma_j^{(e)} \sim U(0, 0.5)$ and $\omega_j^{(e)} \sim U(0, 0.5)$;

(b) uncorrelated pleiotropy effect (balanced pleiotropy: $\gamma_j^{(e)} \sim U(-0.5, 0.5)$, unbalanced pleiotropy: $\gamma_j^{(e)} \sim U(0, 0.5)$);

(c) No pleiotropy.

The parameters of edge effects, IV strength and pleiotropy were randomly selected from a uniform distribution; thus, they are different in different datasets and represent heterogeneous datasets. We varied the number of populations to be $E = 3$ or 8; the number of IVs was 100 or 300; and the number of exposures was $P = 1, 3, 8$, or 15, including the cases of univariable and multivariable MR. In addition, we considered the situation in which genetic correlations among different populations are nonzero, and we generated the IV strength $\alpha_{p,j}^{(e)}$ in different populations using multiple normal distributions, with correlations in the covariance matrix of 0.2 or 0.6. Finally, as the number of exposures increases, weak IVs are more likely to emerge; thus, we observed the performance of our method in the presence of weak IVs, by setting the IV strength ranging from 0.01 to 0.05, and the corresponding conditional F-statistics are shown in Supplementary Data 4.

We conducted 200 repeated simulations to evaluate the performance of MR-EILLS. We also compare nine methods, including IVW, MR-Egger, MR-Lasso, MR-Median, MR-cML, MR-BMA, MARID, Cause and MR-Horse. For $P = 1$, we compare eight methods in the UVMR version except MR-BMA; for $P = 3, 8$, and 15, we compare seven methods in the MVMR version except MARID and Cause, because these two methods have only the UVMR version. For these MR methods, we considered two strategies: (1) metaMR, first meta all the GWAS summary statistics of E datasets for each variable and then conduct the MR analysis; and (2) mrMeta, first conduct the MR analysis (excluding MR-BMA, MARID and Cause because they cannot output the standard errors) in E datasets separately and then meta all the MR results. Meta methods include random effects and fixed effects meta-analyses.

We evaluated the performance of all methods via box-violin plots for causal effect estimation, histograms for type I errors when the causal effect is zero and statistical power when the causal effect is not zero. In addition, we calculated the I^2 statistics in each simulation to evaluate the heterogeneity of causal effect estimation among different datasets for each MR method. We constructed the violin plot of the I^2 statistics for the estimations of each variable, randomly selected three simulations to demonstrate the quartiles of estimation, and then constructed the forest plot of the estimations for each method and each variable. For $P = 15$, we calculated the means of the F1 score, recall and precision for each method. Recall (i.e., power, or sensitivity) measures how many relationships a method can recover from the true causal relationships, whereas precision (i.e., 1-FDR) measures how many correct relationships are recovered in the inferred relationships. The F1 score is a combined index of recall and precision. We also summarize the computing time for different methods to assess the computational efficiency.

Setting of hyperparameters γ and λ

We recommend that practitioners determine the value of λ by constructing a ridge plot. The abscissa is the value of $\sum_{e \in \mathcal{E}} |\varepsilon_j^{(e)}| + \sum_{p \in P} \sum_{e \in E} |\hat{\theta}_{p,j}^{(e)} \varepsilon_j^{(e)}|$ for each IV in Eq. (8). We constructed the ridge plot in the simulations in Supplementary Figs. 67–71. These plots demonstrate that when there is no pleiotropy, the figure has only one peak, and λ only takes the value of abscission after the first peak. When there is pleiotropy, the figure has two peaks, and the corresponding abscission value at the lowest point between the two peaks is the optimal λ. We provide the function of the ridge plot in the R package MR-EILLS.

In addition, we evaluated the root mean square error (RMSE) of causal effect estimation using a grid search: γ ranges from 0.1 to 200, and λ ranges from 0.1 to 1. The results are shown in Supplementary Figs. 72–80. We present the ranges of hyperparameters when the RMSE <0.1 in Supplementary Data 10. For UVMR, we recommend $\gamma > 0.4$. When $\gamma > 0.4$, the RMSE is less than 0.1, especially for the case of correlated and uncorrelated pleiotropy, whereas in other cases, the RMSE is less than 0.05. For MVMR, $\gamma > 0$ is recommended. Compared with all valid IVs, invalid IVs increased the RMSE of causal effect estimation, regardless of whether correlated or uncorrelated pleiotropy was used. Therefore, γ is loosely valued, especially when $P > 1$. The larger γ is, the stronger the role of the empirical focused linear invariance regularizer. Details are shown in Supplementary Note 12.

In our simulations, we employed a grid search approach to identify the optimal lambda value that minimizes the RMSE. In addition, we set $\gamma = 0.5$ for MVMR and $\gamma = 3$ for UVMR.

Application

We explored the causal effect of 11 blood cells on 20 disease-related outcomes using GWAS summary statistics from five ancestries: African, East Asian, South Asian, Hispanic/Latino, and European. GWAS summary statistics for blood cells were extracted from ref. [49], who conducted transethnic and ancestry-specific GWASs in 746,667 individuals from five populations (15,171, 151,807, 8189, 9368, and 563,947 individuals for five ancestries, respectively). GWAS summary statistics for 20 disease-related outcomes[50] were extracted from the MR-Base and GWAS Catalog platforms. The details are shown in Supplementary Data 11.

First, we selected IVs for MR analysis. We separately selected SNPs with $P < 5 \times 10^{-8}$ and clumped the LD with $r^2 > 0.01$ in each population (Supplementary Data 8). In addition, we calculated the conditional F-statistics for IVs to assess the IV strength for each exposure (Supplementary Data 9). Then, we extracted the summary statistics for the IVs and conducted MR-EILLS and MR analysis for each population. We also calculated the I^2 statistic to evaluate the heterogeneity of causal effect estimation among different populations for each MR method. For MR-EILLS, we constructed the ridge plot in each population, and set $\gamma = 0.5$. The settings of λ are shown in Supplementary Data 6. The confidence interval is estimated via the bootstrap method, and details are shown in Supplementary Note 13.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The GWAS summary statistics for 20 disease-related outcomes and 11 blood cell traits are publicly available through MR-Base and the GWAS Catalog. All dataset accession IDs are comprehensively listed in Supplementary Data 11. Data points in Figs. 2–4 are shown in Supplementary Data 1–3. Conditional F-statistics in simulation are shown in Supplementary Data 4. Results of heterogeneity analysis in the application are shown in Supplementary Data 5. Settings of parameters in

the application are shown in Supplementary Data 6. Results of the MR-EILLS analysis in the application are shown in Supplementary Data 7. IVs and their GWAS summary statistics in the application are shown in Supplementary Data 8. Conditional F-statistics in the application are shown in Supplementary Data 9. Optimal ranges of hyperparameters when RMSE <0.1 in the simulation are shown in the Supplementary Data 10.

Code availability

R package MR-EILLS are available in the GitHub repository at https://github.com/hhoulei/MREILLS under the MIT license (https://doi.org/10.5281/zenodo.15951617). All the codes for simulation are available in the GitHub repository at https://github.com/hhoulei/MREILLS_Simul under the MIT license (https://doi.org/10.5281/zenodo.15951779). All the analyses in our article were implemented in R software (version 4.3.2). The R packages used in our analysis include *TwoSampleMR*, *MendelianRandomization*, *meta*, *cause*, *R2jags*, *MARID* and *ggplot2*.

References

1. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–48 (2016).
2. Emdin, C. A., Khera, A. V. & Kathiresan, S. Mendelian randomization. *JAMA* **318**, 5521925–5521926 (2017).
3. Bycroft, C. et al. The UK Biobank resource with deep phenotype and genomic data. *Nature* **562**, 203–209 (2018).
4. Kubo, M. & Editors, G. uest BioBank Japan project: epidemiological study. *J. Epidemiol.* **27**, S1 (2017).
5. Zhao, H. et al. Proteome-wide Mendelian randomization in global biobank meta-analysis reveals multi-ancestry drug targets for common diseases. *Cell Genom.* **2** (2022).
6. Feng, Y. A. et al. Taiwan Biobank: a rich biomedical research database of the Taiwanese population. *Cell Genom.* **2**, 100197 (2022).
7. Lewis, A. C. F. et al. Getting genetic ancestry right for science and society. *Science* **376**, 250–252 (2022).
8. Petrovski, S. & Goldstein, D. B. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biol.* **17**, 1–3 (2016).
9. Sanderson, E. et al. Mendelian randomization. *Nat. Rev. Methods Prim.* **2**, 6 (2022).
10. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int. J. Epidemiol.* **48**, 713–727 (2019).
11. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).
12. Slatkin, M. Linkage disequilibrium-understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* **9**, 477–485 (2008).
13. Fang, A., Zhao, Y., Yang, P., Zhang, X. & Giovannucci, E. L. Vitamin D and human health: evidence from Mendelian randomization studies. *Eur. J. Epidemiol.* **39**, 467–490 (2024).
14. Bowden, J. & Holmes, M. V. Meta-analysis and Mendelian randomization: a review. *Res. Synth. methods* **10**, 486–496 (2019).
15. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
16. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
17. Burgess, S., Zuber, V., Gkatzionis, A. & Foley, C. N. Modal-based estimation via heterogeneity - penalized weighting: model averaging for consistent and efficient estimation in Mendelian

18. randomization when a plurality of candidate instruments are valid. *Int. J. Epidemiol.* **47**, 1242–1254 (2018).
18. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
19. Ellegren, H. & Galtier, N. Determinants of genetic diversity. *Nat. Rev. Genet.* **17**, 422–433 (2016).
20. Stroup, D. F. et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of observational studies in epidemiology (MOOSE) group. *JAMA* **283**, 2008–2012 (2000).
21. Fan, J., Fang, C., Gu, Y. & Zhang, T. Environment invariant linear least squares. *Ann. Stat.* **52**, 2268–2292 (2024).
22. James, A. L. et al. Associations between white blood cell count, lung function, respiratory illness and mortality: the Busselton Health Study. *Eur. Respir. J.* **13**, 1115–1119 (1999).
23. Zeig-Owens, R. et al. Blood leukocyte concentrations, FEV1 decline, and airflow limitation. A 15-year longitudinal study of World Trade Center-exposed firefighters. *Ann. Am. Thorac. Soc.* **15**, 173–183 (2018).
24. Grant, B. J. et al. Relation between lung function and RBC distribution width in a population-based study. *Chest* **124**, 494–500 (2003).
25. Huang, Y. et al. Relationship of red cell index with the severity of chronic obstructive pulmonary disease. *Int. J. Chronic Obstr. Pulm. Dis.* **16**, 825–834 (2021).
26. Wareing, N. et al. Blood neutrophil count and neutrophil-to-lymphocyte ratio for prediction of disease progression and mortality in two independent systemic sclerosis cohorts. *Arthritis Care Res.* **75**, 648–656 (2023).
27. Ulasli, S. S., Ozyurek, B. A., Yilmaz, E. B. & Ulubay, G. Mean platelet volume as an inflammatory marker in acute exacerbation of chronic obstructive pulmonary disease. *Polskie Archiwum Med. Wewnetrznej* **122**, 284–290 (2012).
28. Mansfield, K. J., Chen, Z., Moore, K. H. & Grundy, L. Urinary tract infection in overactive bladder: an update on pathophysiological mechanisms. *Front. Physiol.* **13**, 886782 (2022).
29. Sas, A. R. et al. A new neutrophil subset promotes CNS neuron survival and axon regeneration. *Nat. Immunol.* **21**, 1496–1505 (2020).
30. Rodhe, N., Löfgren, S., Strindhall, J., Matussek, A. & Mölstad, S. Cytokines in urine in elderly subjects with acute cystitis and asymptomatic bacteriuria. *Scand. J. Prim. Health Care* **27**, 74–79 (2009).
31. Ghoniem, G. et al. Differential profile analysis of urinary cytokines in patients with overactive bladder. *Int. Urogynecol. J.* **22**, 953–961 (2011).
32. Tyagi, P. et al. Elevated CXC chemokines in urine noninvasively discriminate OAB from UTI. *Am. J. Physiol. Ren. Physiol.* **311**, F548–F554 (2016).
33. Nebeck, K. et al. Hematological parameters and metabolic syndrome: findings from an occupational cohort in Ethiopia. *Diab. Metab. Syndr.* **6**, 22–27 (2012).
34. Baskurt, O. K. & Meiselman, H. J. Blood rheology and hemodynamics. *Semin. Thromb. Hemost.* **29**, 435–450 (2003).
35. Bhatta, S., Singh, S., Gautam, S. & Osti, B. P. Mean platelet volume and platelet count in patients with type 2 diabetes mellitus and impaired fasting glucose. *J. Nepal Health Res. Counc.* **16**, 392–395 (2019).
36. Mahmood, A. et al. Association of red blood cell and platelet parameters with metabolic syndrome: a systematic review and meta-analysis of 170,000 patients. *Horm. Metab. Res.* **56**, 517–525 (2024).
37. Li, R., Li, L., Liu, B., Luo, D. & Xiao, S. Associations of levels of peripheral blood leukocyte and subtypes with type 2 diabetes: a

longitudinal study of Chinese government employees. *Front. Endocrinol.* **14**, 1094022 (2023).

38. Khan, I. M. et al. Postprandial monocyte activation in individuals with metabolic syndrome. *J. Clin. Endocrinol. Metab.* **101**, 4195–4204 (2016).

39. Zernecke, A. et al. Meta-analysis of leukocyte diversity in atherosclerotic mouse aortas. *Circ. Res.* **127**, 402–426 (2020).

40. AlShareef, A. A. et al. Association of hematological parameters and diabetic neuropathy: a retrospective study. *Diab. Metab. Syndr. Obes.* **17**, 1191–1200 (2024).

41. Yao, C. A., Yen, T. Y., Hsu, S. H. & Su, T. C. Glycative stress, glycated hemoglobin, and atherogenic dyslipidemia in patients with hyperlipidemia. *Cells* **12**, 640 (2023).

42. Li, M., Chen, Z., Yang, X. & Li, W. Causal relationship between iron deficiency anemia and asthma: a Mendelian randomization study. *Front. Pediatrics* **12**, 1362156 (2024).

43. Andreev, D. & Porschitz, P. Emerging roles of eosinophils in bone. *Curr. Osteoporos. Rep.* **23**, 17 (2025).

44. Nacaroglu, H. T. et al. Can mean platelet volume be used as a biomarker for asthma?. *Postepy Dermatol. Alergol.* **33**, 182–187 (2016).

45. Kumar, R., Wallace, W. A., Ramirez, A., Benson, H. & Abelmann, W. H. Hemodynamic effects of pneumonia. II. Expansion of plasma volume. *J. Clin. Investig.* **49**, 799–805 (1970).

46. Mazza, M. G. et al. Monocyte count in schizophrenia and related disorders: a systematic review and meta-analysis. *Acta Neuropsychiatr.* **32**, 229–236 (2020).

47. Cheng, B. et al. Mendelian randomization study of the relationship between blood and urine biomarkers and schizophrenia in the UK Biobank cohort. *Commun. Med.* **4**, 40 (2024).

48. Eisen, M., Mokhtari, A. & Ribeiro, A. Decentralized quasi-Newton methods. *IEEE Trans. Signal Process.* **65**, 2613–2628 (2017).

49. Chen, M. H. et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198–1213.e14 (2020).

50. Shrine, N. et al. Multi-ancestry genome-wide association analyses improve resolution of genes and pathways influencing lung function and chronic obstructive pulmonary disease risk. *Nat. Genet.* **55**, 410–422 (2023).

Acknowledgements

Author contributions

L.H. and X.-H.Z. conceived the study. L.H. contributed to the theoretical derivation with assistance from X.-H.Z. L.H. and H.C. contributed to the data simulation and application. L.H. and X.-H.Z. wrote the manuscript with input from all the other authors. L.H. and H.C. revised the manuscript. All the authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-62823-6.

Correspondence and requests for materials should be addressed to Hao Chen or Xiao-Hua Zhou.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.