








# Shared genetic architecture contributes to risk of major cardiovascular diseases

Received: 17 March 2024

Accepted: 22 July 2025

Published online: 24 September 2025

 Check for updates

Jun Qiao<sup>1,2,3</sup> , Lei Jiang<sup>4</sup> , Liuyang Cai<sup>1,2,3</sup>, Minjing Chang<sup>5</sup>, Can Wang<sup>6</sup>, Rong Zhao<sup>7</sup>, Shan Song<sup>8</sup>, Yuhui Zhao<sup>9</sup>, Miaoran Chen<sup>9</sup>, Shifang Ding<sup>10,11</sup>, Ning Tan<sup>4</sup>, Pengcheng He<sup>4</sup>, Sakthivel Sadayappan<sup>12</sup> , Jinguo Xu<sup>13</sup> , Siim Pauklin<sup>14</sup> , Zhengbing Liu<sup>10,15</sup>  & Yuliang Feng<sup>1,2,3</sup> 

The extensive co-occurrence of cardiovascular diseases (CVDs), as evidenced by epidemiological studies, is supported by positive genetic correlations identified in comprehensive genetic investigations, suggesting a shared genetic basis. However, the precise genetic mechanisms underlying these associations remain elusive. By assessing genetic correlations, genetic overlap, and causal connections, we aim to shed light on common genetic underpinnings among major CVDs. Employing multi-trait analysis, we pursue diverse strategies to unveil shared genetic elements, encompassing SNPs, genes, gene sets, and functional categories with pleiotropic implications. Our study systematically quantifies genetic overlap beyond genome-wide genetic correlations across CVDs, while identifying a putative causal relationship between coronary artery disease (CAD) and heart failure (HF). We then pinpointed 38 genomic loci with pleiotropic influence across CVDs, of which the most influential pleiotropic locus is located at the *LPA* gene. Notably, 12 loci present high evidence of multi-trait colocalization and display congruent directional effects. Examination of genes and gene sets linked to these loci unveiled robust associations with circulatory system development processes. Intriguingly, distinct patterns predominantly driven by atrial fibrillation, coronary artery disease, and venous thromboembolism underscore the significant disparities between clinically defined CVD classifications and underlying shared biological mechanisms, according to functional annotation findings.

Cardiovascular diseases (CVDs), particularly coronary artery disease (CAD) and stroke, are the leading global causes of mortality and disability<sup>1</sup>. In detail, heart failure (HF) is a complex clinical syndrome that often represents the final stage of various heart diseases, commonly resulting from CAD<sup>2</sup>. Atrial fibrillation (AF), on the other hand, is a primary risk factor for both stroke and venous thromboembolism (VTE)<sup>3</sup>. Vascular diseases, such as peripheral artery disease (PAD) and CAD, further increase the risk of both stroke and VTE<sup>4,5</sup>. Additionally, atherosclerosis and hypertension serve as major precursors to many

CVDs, although the precise mechanisms triggering these diseases remain poorly understood<sup>6</sup>. Moreover, these conditions exhibit a complex genetic pattern, which suggests that children of individuals with CVDs face an increased risk of developing a range of CVDs, not merely the specific disease their parents experienced<sup>7,8</sup>. Furthermore, the frequent co-occurrence of multiple CVDs, known as comorbidity, underscores the complex interplay among various CVDs.

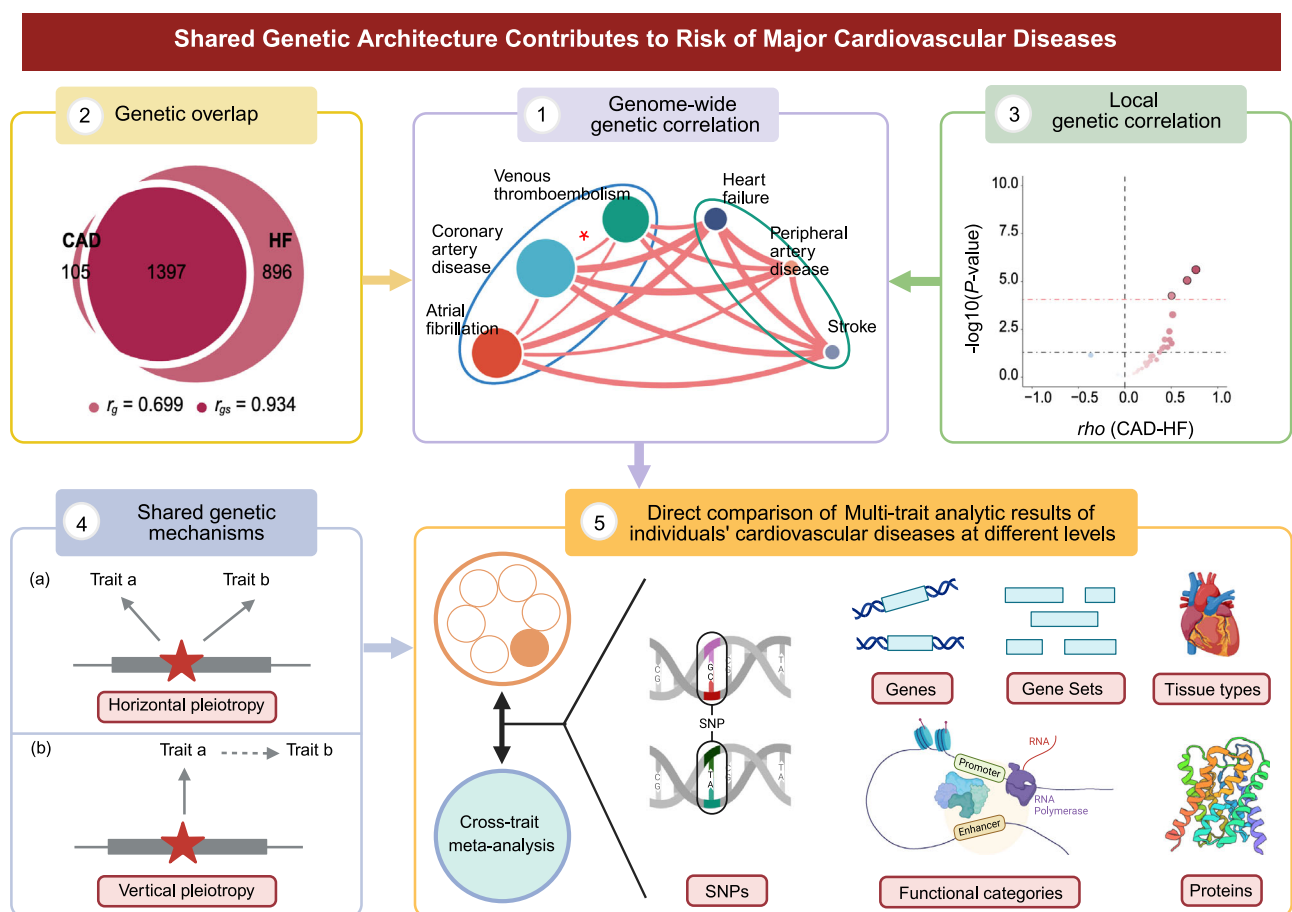
Recent large-scale genomic studies have established the substantial influence of genetic variation on the risk of various CVDs<sup>9</sup>.

A full list of affiliations appears at the end of the paper. ✉ e-mail: [jianglei0731@gmail.com](mailto:jianglei0731@gmail.com); [xujinguo@ahmu.edu.cn](mailto:xujinguo@ahmu.edu.cn); [siim.pauklin@ndorms.ox.ac.uk](mailto:siim.pauklin@ndorms.ox.ac.uk); [njzylzb@126.com](mailto:njzylzb@126.com); [fengyl@sustech.edu.cn](mailto:fengyl@sustech.edu.cn)

Genome-wide association studies (GWASs) have pinpointed numerous genetic risk loci for major CVDs such as AF, CAD, VTE, HF, PAD, and stroke<sup>10–13</sup>. Strikingly, these loci often exert effects across multiple phenotypes, revealing a remarkable interconnectedness among CVDs. While pleiotropic effects are widely acknowledged, their systematic presence and downstream consequences in the context of CVDs have yet to be fully assessed. Furthermore, the common genetic variants discovered to date account for only a limited portion of the heritable risk for these diseases. The ongoing search for new risk loci associated with traits necessitates the continual expansion of sample sizes. Unfortunately, challenges in recruiting large study populations have notably hindered variant discovery for certain traits, such as PAD, which lags behind more common CVDs like CAD and stroke<sup>14</sup>. In response to these challenges, the concept of multi-trait analysis in GWAS has emerged to combine GWAS summary data from diverse cohorts with genetically correlated traits to increase the study sample size<sup>15</sup>. Recent statistical advancements, represented by multi-trait analysis of GWAS (MTAG)<sup>16</sup>, enhance the capacity to detect pleiotropic

variants associated with multiple diseases, thereby unraveling the genetic basis for shared risk across diseases. To date, cross-disease GWAS multi-trait analyses have predominantly focused on psychiatric and autoimmune disorders<sup>17–22</sup>. Given the pervasive comorbidities observed in epidemiological studies and the existing evidence indicating a shared genetic background across multiple CVDs, the need for a comprehensive, large-scale analysis became evident.

In this work, we leverage MTAG to conduct an integrative analysis of the largest available GWAS datasets encompassing six major CVDs in individuals of European ancestry—AF, CAD, VTE, HF, PAD, and stroke—to characterize the underlying biology of the novel risk loci in the context of CVDs. This study addressed four central inquiries regarding the shared genetic basis of these six CVDs (Fig. 1): (i) Can we identify shared genetic architectures amidst the diverse landscape of these clinically distinct CVDs? (ii) Can we determine whether the shared genetic architectures related to major CVDs are driven by causal associations (i.e., vertical pleiotropy)? (iii) Can we detect additional genomic loci for multiple CVDs (i.e., pleiotropic loci), and can we



**Fig. 1 | Schematic representation of analyses performed for all 6 major cardiovascular diseases in the current study.** This figure illustrates the comprehensive analytical approach used for all six major cardiovascular diseases within this study. The initial GWAS findings were obtained from various repositories. Before integrating cardiovascular diseases into a multi-trait analysis, genetic characteristics, such as genome-wide and local genetic correlations, and genetic overlap were individually estimated to identify genetic overlap across the six CVDs beyond genetic correlations. The shared genetic basis was interpreted as genetic variants influencing multiple complex phenotypic traits through vertical and horizontal pleiotropy. Mendelian randomization was used to elucidate the important role of vertical pleiotropy in CVDs. Subsequently, the results of SNP and genomic loci analyses from the multi-trait assessment, along with cross-trait analysis for replication, were compared with the results from individual cardiovascular diseases,

uncovering many novel pleiotropic loci. Various methods were employed to comprehensively characterize the shared genetic mechanisms for each of the six cardiovascular diseases. First, we examined the convergence of SNPs, genomic loci, and mapped genes. Subsequent stages involved biological pathway and functional category analyses, evaluating the enrichment of genetic signals across 9398 distinct gene sets and 49 tissue types for each disease. The enrichment assessments were further extended using LDSC-SEG and GARFIELD, covering 489 and 1005 functional genomic categories, respectively. The diagram was created using BioRender and included with permission for publication (Created in BioRender. Feng, Y. (2025) <https://BioRender.com/mp87c7d>). AF atrial fibrillation, CAD coronary artery disease, VTE venous thromboembolism, HF heart failure, PAD peripheral artery disease.

determine whether some of the loci exhibit opposite allelic effects across CVDs? (iv) Can we identify functional features of the pleiotropic loci that could account for their widespread impact on cardiovascular pathology?

## Results

### Genome-wide genetic overlap beyond genetic correlation across the six CVDs

We obtained GWAS summary statistics for six major CVDs<sup>10–13,23,24</sup>, with sample sizes varying widely (median: 1,098,263, ranging from 511,634 for PAD to 1,500,861 for VTE) (Supplementary Data 1 provides an overview and acronyms; Supplementary Note 1 lists data sources; Fig. 1 illustrates the study flowchart). After harmonizing and filtering of SNPs shared across GWAS summary statistics, we used linkage disequilibrium (LD) score regression (LDSC)<sup>25,26</sup> to estimate SNP-based heritability ( $h^2_{\text{SNP}}$ ) and genome-wide genetic correlation ( $r_g$ ). The statistical power and genetic signal strength of the GWASs varied substantially across the six CVDs, with the median  $h^2_{\text{SNP}}$  being 1.4% (ranging from 0.6% to 3.2%; Fig. 2a and Supplementary Data 2a). Estimated  $r_g$ s ranged from moderate to strong (median = 0.435; ranging from 0.148 to 0.677) and were all significantly positive after Bonferroni correction ( $P = 8.33 \times 10^{-3}$ ; Fig. 2b and Supplementary Data 2b). Substantial  $r_g$ s were observed between CVD pairs regardless of whether the individual traits had low or high  $h^2_{\text{SNP}}$ . For example, HF and stroke, with  $h^2_{\text{SNP}}$  values of 0.6 and 0.8%, respectively, showed a relatively high  $r_g$  of 0.552 (standard error [SE] = 0.051), whereas AF and CAD, despite having higher  $h^2_{\text{SNP}}$  (2.5 and 3.2%, respectively), exhibited a lower  $r_g$  of 0.197 (SE = 0.024).

Although  $r_g$  is a valuable measure to assess the genetic similarity between two phenotypes, it fails to capture all dimensions of genetic overlap. As  $r_g$  is a genome-wide summary measure, it cannot distinguish genetic overlap resulting from a mixture of concordant and discordant effects from the absence of genetic overlap, leading to an estimated  $r_g$  close to 0 in both cases. Therefore, it was imperative to use multiple methods with different modeling assumptions to capture this “missing dimension” of genetic overlap and comprehensively describe the shared genetic architecture of CVDs beyond genetic correlation, including the bivariate causal mixture model (MiXeR)<sup>27</sup>, local analysis of (co)variant association (LAVA)<sup>28</sup>, and pairwise GWAS (GWAS-PW)<sup>29</sup>.

To further characterize the extent of polygenic overlap beyond genetic correlation, we applied MiXeR, a modeling framework that estimates the total number of trait-influencing (i.e., “causal”) variants for each phenotype via univariate analysis, as well as the number of shared and trait-specific variants between pairs of phenotypes via bivariate analysis. First, we used univariate MiXeR analysis to estimate the number of trait-influencing variants for each trait (i.e., polygenicity) and the average magnitude of additive genetic associations among these variants (i.e., discoverability). Among the six CVDs, HF and CAD exhibited higher estimated polygenicity, indicative of more extensive involvement of genetically correlated genomic regions compared to other CVDs (Supplementary Data 3a). In contrast, PAD exhibited higher discoverability but lower polygenicity than other CVDs. As a result, PAD exhibited comparatively fewer genomic loci than other CVDs (Supplementary Data 3a). Notably, the estimated sample size required to achieve 90%  $h^2_{\text{SNP}}$  was over 24 times smaller for PAD than for HF (Supplementary Fig. 1). Overall, there were substantial genetic overlaps among the six CVDs, with the median Dice coefficient (i.e., the proportion of shared variants to the total number of variants) was 0.362 (ranging from 0.207 to 0.732; Fig. 2c, Supplementary Fig. 2, and Supplementary Data 3b), both in the presence of moderate positive  $r_g$  (CAD-HF) and minimal  $r_g$  (AF-PAD). For example, the genetic overlap between CAD and HF was particularly striking, with 1397 (standard deviation [SD] = 254) shared variants, representing 93.3% of the variants influencing CAD and 60.9% of the variants

influencing HF. This mirrored their robust positive genome-wide  $r_g$  of 0.677 (SE = 0.030) and genetic correlation of shared variants ( $r_{gs} = 0.699$ , SE = 0.013). In contrast, fewer shared variants (74, SD = 11) were observed between AF and PAD, representing 14.0% of AF-influencing variants and 18.9% of PAD-influencing variants, consistent with their weak genome-wide  $r_g$  of 0.204 (SE = 0.046). This was associated with weak positive genetic correlations for both shared variants ( $r_{gs} = 0.204$ , SE = 0.046) and genome-wide levels ( $r_g = 0.140$ , SE = 0.015). Extensive genetic overlap was also observed in the presence of weak genetic correlations. For example, CAD and VTE shared a large number of “causal” variants (290, SD = 88), with only 35 unique VTE variants, despite having a weak genome-wide genetic correlation ( $r_g = 0.182$ , SE = 0.026) and genetic correlation of shared variants ( $r_{gs} = 0.220$ , SE = 0.008). This pattern of extensive genetic overlap but weak genetic correlation suggests a mixture of concordant and discordant genetic effects, as indicated by the MiXeR-estimated proportion of shared “causal” variants with concordant effects (0.696, SE = 0.074).

To further investigate the regional architecture of genetic sharing between CVDs, we applied LAVA to estimate pairwise local genetic correlations (local  $r_g$ ) across 2495 1-Mb semi-independent genomic regions. This approach facilitated the detection of localized genetic sharing and mixed-effect directions, even in disease pairs with minimal genome-wide  $r_g$ . Among 896 unique genomic regions that met the significance threshold ( $P < 1 \times 10^{-4}$ ), 495 (55.25%) exhibited univariate genetic signals for multiple CVDs, with 21 regions (2.34%) displaying signals for more than half of the CVDs (Fig. 2c and Supplementary Data 4). In total, 130 unique genomic regions were significant in bivariate analyses ( $P < 0.05$ ) across CVDs, with 18 of these regions shared more than one CVD pair, further providing evidence of mixed-effect directions. This phenomenon was most pronounced between CAD and VTE, where 14 regions were positively correlated and 5 were negatively correlated. Even CAD and PAD, despite their strong positive  $r_g$ , exhibited a mixture of effect directions, with ten positively and 1 negatively correlated regions. However, after Bonferroni correction, only 24 distinct genomic regions remained significant for local  $r_g$ , and just two regions displayed significant correlations for more than two CVDs ( $P = 0.05/\text{number of bivariate tests} = 0.05/577 = 8.67 \times 10^{-5}$ ; Fig. 2c, Supplementary Data 5, and Supplementary Note 2).

To further investigate whether the genomic regions identified by LAVA harbored variants associated with multiple CVDs, we applied GWAS-PW to estimate the posterior probabilities of pleiotropy across the 2495 LAVA-defined regions shared by both traits. Overall, GWAS-PW analysis identified 2112 unique genomic regions with variants influencing pairs of CVDs (i.e.,  $\text{PPA3} > 0.5$  or  $\text{PPA4} > 0.5$ , Supplementary Data 6), suggesting the presence of shared genomic regions and causal variants. Notably, 1238 unique genomic regions demonstrated robust pleiotropic signals across loci shared by more than two trait pairs. Additionally, 15 out of the 24 unique genomic regions were significantly detected by both the bivariate LAVA and GWAS-PW analyses. For example, within the locus (LD block 1398; chr9: 21,674,033–23,091,193), six trait pairs exhibited strong positive local  $r_g$  at the nominal significance level ( $P < 0.05$ ). Four of these six trait pairs (including AF-CAD, CAD-HF, CAD-PAD, and CAD-Stroke) remained significant after Bonferroni correction and were also validated by GWAS-PW. Further analysis of this region using hypothesis prioritization in multi-trait colocalization (HyPrColoc) provided strong colocalization evidence across all CVDs except VTE and PAD, with a posterior probability (PP) greater than 0.7. This locus encompassed a shared causal SNP (rs4977574), located within an intron of cyclin-dependent kinase inhibitor 2B antisense RNA 1 (CDKN2B-AS1) on chromosome 9q21.3. Previous studies suggest a potential association between the CDKN2B-AS1 gene rs4977574 A/G polymorphism and susceptibility to coronary heart disease, which may influence the progression and severity of vascular calcification in vascular smooth muscle cells<sup>30</sup>.





**Fig. 2 | Genome-wide and local genetic correlations, and genetic overlap of the six cardiovascular diseases.** **a** Error-bar plot of the SNP-based heritability ( $h^2_{\text{SNP}}$ ) point estimates for the six cardiovascular diseases, calculated using univariate LDSC. The magnitude and precision of the estimates varied substantially across the CVDs ( $h^2_{\text{SNP}}$  range from 0.0059 to 0.0324; standard error (SE) range from 0.0005 to 0.0033). Error bars represented one standard deviation (SD) ( $1.96 \times \text{SE}$ ). **b** Network visualization of the Bonferroni-corrected significant global  $r_g$  among the cardiovascular diseases, calculated using bivariate LDSC. Connections represented significant  $r_g$ , with the correlation value along connections, thicker lines indicating stronger correlations, and dark gray denoting more significant correlations. The size of the nodes is weighted by the sample size and  $h^2_{\text{SNP}}$  of the given cardiovascular disease (size =  $h^2_{\text{SNP}} \times \sqrt{n}$ ) (AF,  $n = 1,030,836$ ; CAD,  $n = 1,165,690$ ; VTE,  $n = 1,500,861$ ; HF,  $n = 977,323$ ; PAD,  $n = 511,634$ ; Stroke,  $n = 1,308,460$ ). **c** Along the diagonal, univariate MiXeR estimates are provided for each CVD.  $h^2_{\text{SNP}}$  = SNP-based heritability estimate; polygenicity<sub>90</sub> = number (in thousands) of causal variants with the strongest effects required to explain 90% of SNP-based heritability. Mixer-modeled genome-wide genetic overlap and genetic correlations (top-right) and

LAVA local correlations (bottom-left) across the six CVDs are shown. Top-right: MiXeR Venn diagrams showing the number of shared and disorder-specific “causal” variants for each pair of diseases. Genome-wide genetic correlation ( $r_g$ ) and genetic correlation of shared variants ( $r_{gs}$ ) are represented by the color of the disease-specific ( $r_g$ ) and shared regions ( $r_{gs}$ ), respectively, in the Venn diagrams. For AF-HF, CAD-PAD, and HF-PAD, both AICs (best\_vs\_min\_AIC and best\_vs\_max\_AIC) were negative, indicating that the analysis lacked sufficient power to provide precise estimates of genetic overlap. Bottom-left: Volcano plots of LAVA local genetic correlation coefficients ( $\rho$ , y-axis) against  $-\log_{10}$  ( $P$  values) for each pairwise analysis at each locus.  $P$  values of the local  $r_g$  were corrected for the total number of bivariate tests conducted ( $P = 0.05 / \text{the number of bivariate tests} = 0.05 / 1672 = 2.9 \times 10^{-5}$ ). All statistical tests were two-sided. Larger dots with black circles represent loci significantly correlated after the Bonferroni correction. MiXeR-estimated  $r_g$  and  $r_{gs}$ , and LAVA estimated  $\rho$  were represented on the same blue to red color scale. Note that the Volcano plots were plotted at  $P$  values truncated by  $1 \times 10^{-10}$  for better visualization. AF atrial fibrillation, CAD coronary artery disease, VTE venous thromboembolism, HF heart failure, PAD peripheral artery disease.

construction of the latent causal variable (LCV) model<sup>31</sup> allowed us to estimate the partial genetic causality between significantly correlated CVD pairs. We observed convincing evidence for three CVD pairs (Supplementary Data 7), where the absolute value of genetic causality proportion (GCP) exceeded 0.6 and the  $P$  value passed Bonferroni correction ( $P < 0.05 / \text{the number of independent CVD pairs} = 0.05 / 6 = 8.33 \times 10^{-3}$ ; Supplementary Data 7). By considering the sign of the  $r_g$ , it was plausible to infer that CAD could elevate HF risk, while VTE might have an adverse impact on both CAD and stroke. Notably, the AF-HF trait pair (GCP = 0.575,  $P = 1.07 \times 10^{-3}$ ) exhibited trends toward partial genetic causation, suggesting AF might partially influence HF, but fell short of the stringent GCP threshold. To validate the reproducibility of these partial causal associations, the MRlap method was employed, confirming the positive causal associations between CAD-HF and VTE-Stroke (Supplementary Fig. 3 and Supplementary Data 8a). Furthermore, a trait pair (VTE-CAD) also showed a partial positive causality, although it did not withstand multiple testing corrections ( $P < 0.05 / \text{the number of combinations} / \text{the number of MR-tests} = 0.05 / 6 / 2 = 4.17 \times 10^{-3}$ ). However, the AF-HF pair exhibited bias due to reverse causality, which may have skewed the results. In all, we found a high degree of agreement between the MRlap and LCV results, with both methods inferring causality that aligns with existing epidemiological evidence. To investigate potential bias due to sample overlap, we repeated the analyses using GWAS summary statistics of AF and HF based on GWAS excluding the UK Biobank sample (the main source of overlap). We observed small differences in the estimates, but the interpretation remained the same for all results (Supplementary Data 8a).

We extended the univariate results with multivariate MR<sup>32</sup> to assess whether causal effects observed for CAD-HF and VTE-Stroke were confounded by other CVDs using both raw and UKB-excluded summary data. We found that the association between genetic liability to CAD and HF remained statistically significant even after adjusting for other CVDs, a pattern also observed for the VTE-Stroke pair (Supplementary Data 8b). Furthermore, the association between genetic liability to CAD and HF was strengthened in the UKB-excluded summary data, with the odds ratio (OR) rising from 1.33 (CAD-HF) to 2.72 (CAD-HF\_UKB-excluded). In contrast, the association between genetically contributed VTE susceptibility and stroke slightly weakened after adjusting for other CVDs, with the OR changing from 1.09 (VTE-Stroke) to 1.08 (VTE-Stroke\_UKB-excluded). Overall, these findings further support the robust causal relationships between CAD-HF and VTE-Stroke. In conclusion, the MR analyses illuminate that the shared genetic basis for most CVDs is not potentially mediated by vertical pleiotropy, given the limited causal relationships observed across CVDs despite their strong genetic associations. This implies that pleiotropy is more likely horizontal in nature, driven by shared risk factors or upstream biological mechanisms rather than direct causal pathways.

### Shared and novel loci for the six CVDs

Previous studies have demonstrated that cross-trait meta-analysis is a powerful approach for increasing statistical power to identify pleiotropic SNPs, shared genetic architecture, and common biological annotations across distinct psychiatric and immunological diseases<sup>18,22</sup>. Building on this, we conducted cross-trait GWAS meta-analyses using the MTAG method<sup>16</sup> to enhance the identification of novel genetic associations for each CVD by pinpointing pleiotropic SNPs contributing to CVD susceptibility. As a result, we observed an increase in maximum effective sample sizes, with the following increments: from 1,030,836 to 1,076,012 for AF; from 1,165,690 to 1,237,236 for CAD; from 1,500,861 to 1,631,058 for VTE; from 977,323 to 1,219,100 for HF; from 511,634 to 843,568 for PAD; and from 1,308,460 to 1,735,909 for stroke (Table 1 and Fig. 3a–f). Importantly, analysis of heritability Z-scores derived from MTAG results revealed stronger genetic signals compared to the original GWAS outcomes (Table 1). Furthermore, LDSC intercepts were close to one, suggesting that the observed increase in mean  $\chi^2$  statistics was attributable to polygenicity rather than confounding biases such as population stratification. The quantile–quantile (Q-Q) plots and genomic inflation factors ( $\lambda_{\text{GC}}$ ) for both MTAG and original GWAS results are presented in Supplementary Figs. S4, S5. The  $\lambda_{\text{GC}}$  values ranged from 1.096 to 1.421 in the GWAS results and from 1.014 to 1.446 in the MTAG results, suggesting no evidence of systematic inflation. Detailed information on novel genomic loci and mapped genes identified from the MTAG analysis of the six CVDs is provided in Supplementary Notes 4, 5, Supplementary Fig. 5, and Supplementary Data 9–16.

### Exploration of shared biological mechanisms among the six CVDs (i.e., horizontal pleiotropy)

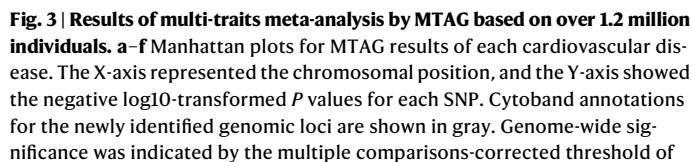
In light of the challenges posed by prior GWAS results in discerning shared biological mechanisms among CVDs, our subsequent efforts involved a comprehensive assessment of genetic overlap within the MTAG results of six CVDs, spanning genomic loci, SNPs, genes, and pathways.

Among the noteworthy observations, a total of 38 pleiotropic loci showed genetic signals across multiple CVDs, of which 9 (22.0%) were shared by more than half of the considered CVDs (Supplementary Data 17a). However, when testing whether these overlapping loci shared causal variants through hypothesis prioritization in multi-trait colocalization (HyPrColoc)<sup>33</sup>, only 12 (31.58%) loci exhibited strong colocalization evidence of sharing causal variants (posterior probability [PP] > 0.70). Out of these, 11 loci were further corroborated by Metasoft analysis<sup>34</sup> (Supplementary Fig. 6). The most prominent pleiotropic locus is located within an intron of the lipoprotein(a) gene (*LPA*) at 6q25.3. This locus harbored the shared causal SNP rs10455872, which colocalized across all CVDs except MTAG\_VTE\_EUR. Notably, rs10455872 has been established as linked to elevated lipoprotein (a)

Table 1 | Statistical summary of GWAS and MTAG results

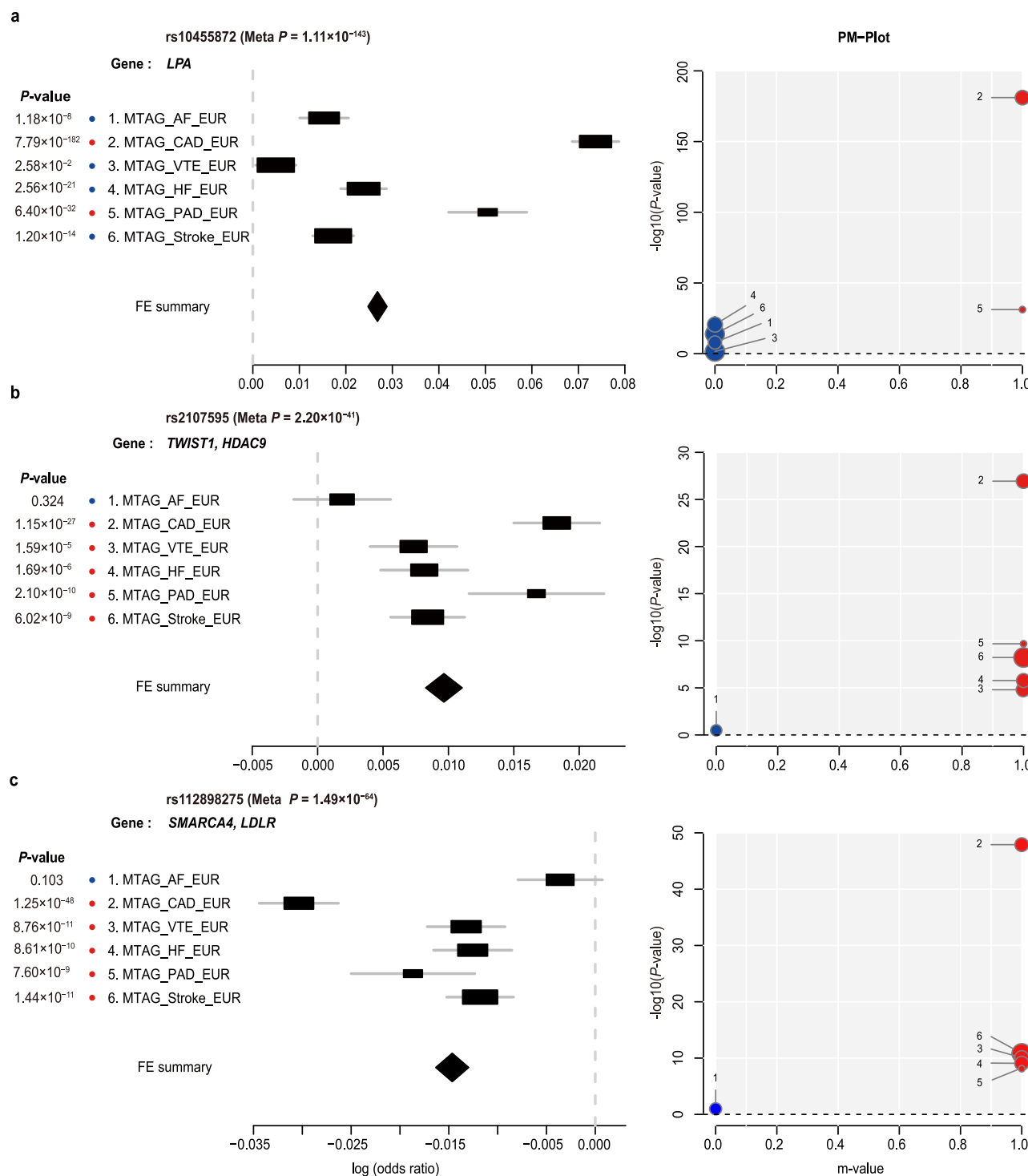
	Atrial fibrillation		Coronary artery disease		Venous thromboembolism		Heart failure		Peripheral artery disease		Stroke	
	GWAS	MTAG	GWAS	MTAG	GWAS	MTAG	GWAS	MTAG	GWAS	MTAG	GWAS	MTAG
mean $\chi^2$	1.5451	1.4888	1.7648	1.8020	1.5698	1.4830	1.1591	1.2062	1.1025	1.1253	1.2280	1.1944
$R^2_{\text{SNP}}$	0.0250	0.0245	0.0324	0.0345	0.0182	0.0183	0.0080	0.0118	0.0094	0.0157	0.0060	0.0084
$\lambda_{\text{GC}}$	1.2932	1.2697	1.4210	1.4460	1.3203	1.2498	1.1270	1.1555	1.0957	1.0135	1.1876	1.1459
Intercept	1.0538	1.0025	1.0218	1.0067	1.0511	0.9501	1.0073	0.9797	1.0157	0.9795	1.0738	0.9750
Ratio	0.0978	0.0052	0.0302	0.0083	0.0898	<0	0.0416	<0	0.1560	<0	0.3245	<0
Sample size	1,030,836	1,076,012	1,165,690	1,237,236	1,500,861	1,631,058	977,323	1,219,100	511,634	843,568	1,308,460	1,735,909
Total SNPs	6,925,287	5,782,223	6,925,287	5,782,223	6,925,287	5,782,223	6,925,287	5,785,906	6,925,287	5,785,906	6,925,287	5,785,906
nSig SNPs	10,784	8401	15,789	12,118	9070	6511	272	449	92	426	941	585
nGenomic locus	117	108	203	171	91	72	11	13	4	37	24	18
nInd Sig SNPs	749	592	1049	862	765	598	21	40	8	60	50	48
nLead SNPs	224	191	362	295	222	176	12	14	5	43	25	20
nGenes mapped by position using FUMA	397	345	662	567	447	375	38	47	21	63	67	66
nGenes mapped by MAGMA	298	241	551	474	401	303	20	31	5	17	43	43
nGenes mapped by eqtl using FUMA	1018	900	1921	1768	1147	1056	70	121	25	151	156	157
nGenes mapped by TWAS	387	322	831	755	598	494	19	32	11	33	43	44

All reported  $P$  values are two-sided.  
mean  $\chi^2$  mean chi-squared value,  $R^2_{\text{SNP}}$  SNP-based heritability,  $\lambda_{\text{GC}}$  Lambda GC of 1 indicates no inflation of the median  $\chi^2$  association statistic, Intercept LD score intercept, a value of 1 indicates that inflation is not due to population stratification biases, Ratio (intercept-1)/(mean chi-squared value-1), the ratio indicates what percentage of the inflation can be attributed to other sources than polygenicity, nSig SNPs number of genome-wide significant SNPs after additional Bonferroni correction for three independent traits ( $P < 5 \times 10^{-9}/3 = 1.67 \times 10^{-9}$ ), nGenomic locus number of the genomic locus, nLead SNPs number of lead SNPs in the genomic locus ( $P < 5 \times 10^{-9}$  and  $r^2 < 0.1$ ), nInd Sig SNPs number of the independent significant SNPs in the genomic locus ( $P < 5 \times 10^{-9}$  and  $r^2 < 0.6$ ), nGenes mapped by position using FUMA Number of genes mapped by position using FUMA, nGenes mapped by MAGMA number of genes mapped by MAGMA, nGenes mapped by eqtl using FUMA Number of genes mapped by eqtl using FUMA, nGenes mapped by TWAS Number of genes mapped by TWAS, FUMA functional mapping and annotation of genetic associations, MAGMA multi-marker analysis of genomic annotation, TWAS transcriptome-wide association study.



$P = 1.67 \times 10^{-8}$  (red dotted line). Colored dots indicated an independent genome-wide significant association with the smallest  $P$  value (Top lead SNP). Only SNPs common to all summary statistics were included. All statistical tests are two-sided. MTAG multi-trait analysis of GWAS, AF atrial fibrillation, CAD: coronary artery disease, VTE venous thromboembolism, HF heart failure, PAD peripheral artery disease.

Multi-trait colocalization analysis using HyPrColoc further highlighted 7 pleiotropic loci that were colocalized to share a causal variant, supporting the significant role of metabolic traits—including anthropometric traits<sup>37</sup>, blood pressure traits<sup>38</sup>, lipidemic traits<sup>39</sup>, glycemic traits<sup>40</sup>, kidney function<sup>41</sup>, liver function<sup>42</sup>, and other traits, such as carotid artery intima-media thickness (cIMT)<sup>43</sup> and C-reactive protein (CRP)<sup>44</sup>—in the pathogenesis of CVDs (Supplementary Data 17b, c). Of these, three loci showed substantial evidence of colocalization among multiple CVDs and more than one metabolic trait. For example, a locus located at 1p36.22 was found to be colocalized between multiple CVDs (MTAG\_AF\_EUR and MTAG\_Stroke\_EUR) with multiple metabolic traits—including blood pressure, lipidemic, and glycemic traits (e.g., glycated hemoglobin [HbA1c])—with the index SNP rs880315 (mapped to *CASZ1* [cancer zinc finger 1]) identified as the likely shared causal variant. Notably, the locus at 16q12.2 (index SNP: rs11075990, mapped to *FTO* [FTO alpha-ketoglutarate dependent dioxygenase]) was found to be colocalized with a broad spectrum of metabolic traits—including diastolic blood pressure (DBP) and pulse pressure (PP) (blood pressure traits), high-density-lipoprotein cholesterol (HDL-C) and triglycerides (TC) (lipidemic traits), HbA1c (glycemic trait), alanine aminotransferase (ALT) (liver function), and CRP—although the PP did not exceed the stringent threshold ( $PP > 0.70$ ).



**Fig. 4 | The three most pleiotropic loci associated with all six cardiovascular diseases.** **a–c** Forest plots with PM-plots show disease-specific effects of the index SNP in each locus, including rs10455872 on 6q25.3, rs2107595 on 7p21.1, and rs112898275 on 19p13.2. For each locus, the disease-specific effects of the causal SNP were illustrated using ForestPMPlot, based on Metasoft analysis. The first panel was a forest plot displaying the disease-specific association  $P$  value, log odds ratios (ORs), and standard errors (SE) for the SNP. Each black square represents the point estimate of the effect size (log OR) for an individual study, with the size of the square proportional to the weight assigned to that study in the meta-analysis. The width of the line in each study represents the confidence interval (CI) for the effect size estimate. The diamond at the bottom indicates the overall effect estimate from the fixed-effects (FE) model, with its width corresponding to the 95% CI of the pooled estimate. The meta-analysis  $P$  value and corresponding summary statistic

were displayed at the top and bottom of the forest plot, respectively. The second panel was a PM-plot where the X-axis represents the  $m$ -value, the posterior probability that the effect exists in each disease, and the Y-axis represents the disease-specific association  $P$  value as  $-\log_{10}(P \text{ value})$ . Diseases were represented by dots, where the size corresponds to the sample size of the individual GWAS (AF,  $n = 1,076,012$ ; CAD,  $n = 1,237,236$ ; VTE,  $n = 1,631,058$ ; HF,  $n = 1,219,100$ ; PAD,  $n = 843,568$ ; Stroke,  $n = 1,735,909$ ). Diseases with estimated  $m$ -values of at least 0.9 are colored red, indicating that the SNP has an effect on the disease, while those with  $m$  values below 0.1 are marked blue, suggesting that the SNP does not have an effect. Diseases with estimated  $m$ -values between 0.1 and 0.9 are colored green, indicating that the SNP may have an uncertain effect on the disease. All statistical tests were two-sided. AF atrial fibrillation, CAD coronary artery disease, VTE venous thromboembolism, HF heart failure, PAD peripheral artery disease.



Besides, two loci (1p34.3 and 7p21.2) were colocalized between multiple CVDs and blood pressure traits, especially systolic blood pressure (SBP). Two other loci exhibited strong evidence of colocalization between multiple CVDs and lipidemic traits (6q25.3) or glycemic traits (18q21.1), respectively (Supplementary Data 17d).

We further tested whether these pleiotropic loci could be colocalized with common risk factors, such as hypertension<sup>45</sup>, type 2 diabetes (T2D)<sup>46</sup>, chronic kidney disease (CKD)<sup>47</sup>, and non-alcoholic fatty liver disease (NAFLD)<sup>48</sup>. For example, the locus (4q21.21) was identified to have strong colocalization signals with MTAG\_AF\_EUR, MTAG\_CAD\_EUR, estimated glomerular filtration rate (eGFR), and CKD, where the index SNP rs1458038 (located near the *FGF5* [fibroblast growth factor 5] gene) was identified as the likely shared causal variant. Strikingly, the locus located at the *FTO* gene also exhibited evidence of colocalization with NAFLD, although the PP did not meet the stringent threshold.

All of these—680 lead SNPs, 924 susceptibility genes identified by multi-marker analysis of genomic annotation (MAGMA)<sup>49</sup> and 1474 tissue-specific genes identified by transcriptome-wide association study (TWAS)<sup>50</sup>—achieved genome-wide significance for any of the six CVDs. Specifically, only 74 (10.88%) lead SNPs ( $P < 1.67 \times 10^{-8}$ ; Supplementary Data 18), 118 (12.77%) susceptibility genes ( $P < 9.45 \times 10^{-7}$ ; Supplementary Data 15), and 156 (10.58%) tissue-specific genes ( $P < 8.71 \times 10^{-7}$ ; Supplementary Data 16) reached genome-wide significance for more than one CVD. Among these, only 17 (2.50%) lead SNPs, 16 (1.73%) susceptibility genes, and 11 (0.75%) tissue-specific genes achieved genome-wide significance in more than half of the six CVDs (i.e., at least four). Together with strong evidence from colocalization analyses, direct SNP- and gene-level comparisons across the six CVDs support the presence of shared genetic signals in overlapping loci.

While the observation of a relatively limited number of overlapping genes and regions among cardiovascular diseases is apparent, the distinct genes associated with these conditions could potentially converge within shared biological pathways or display enrichment in similar tissues or functional categories. To explore this notion, we undertook a comprehensive analysis that included: (i) gene set enrichment based on MAGMA-identified susceptibility genes, (ii) pathway enrichment analysis of tissue-specific genes using Metascape, and (iii) tissue and functional category enrichment using SNP-based heritability from LDSC-SEG<sup>51</sup>. After meticulous correction for 9398 tests across the six CVDs (that is, 7744 and 1654 gene sets for GO BP and Reactome, respectively;  $P = 0.05/9398/3 = 1.77 \times 10^{-6}$ ; Supplementary Data 19), we observed little convergence among sets of CVDs for gene sets. Of these, three gene sets—circulatory system development, circulatory system process, and heart development—were significantly enriched in more than one CVD, and none showed significance in more than two. Subsequently, more lenient enrichment tests using Metascape similarly revealed limited convergence, with circulatory system process being the only pathway enriched across multiple CVDs based on tissue-specific genes (Supplementary Data 20). Similarly, after separately correcting for tissues or functional categories and six CVDs (that is, 49 tissues and 489 functional categories;  $P$ -value for tissues =  $0.05/49/3 = 3.40 \times 10^{-4}$ ;  $P$  value for functional categories =  $0.05/489/3 = 3.41 \times 10^{-5}$ ; Supplementary Figs. 8, 9 and Supplementary Data 21, 22), no test surpassed the significance threshold. Taking a less stringent approach, only correcting separately for tissues or functional categories, there were still no convergence among CVDs for tissues and functional categories. To assess whether certain functional genomic categories (e.g., regulatory elements in particular tissues or cell types) are enriched among multiple CVDs, we applied GARFIELD (GWAS analysis of regulatory and functional information enrichment with LD correction)<sup>52</sup>. After correction for the 1005 tested functional categories across all CVDs ( $P = 0.05/1005/3 = 1.66 \times 10^{-5}$ ; Supplementary Fig. 10 and Supplementary Data 23), we

observed significant enrichment in the exon region and for multiple regulatory features. These included 240 (23.88%) DNaseI hypersensitive sites, 86 (8.56%) chromatin accessibility peaks, 37 (3.68%) histone-modified regions, 19 (1.89%) transcription-factor footprints, 8 (0.796%) chromatin states, 1 transcription-factor binding site (TFBS), and 2 formaldehyde-assisted isolation of regulatory elements (FAIRE) of different tissues or cell types in more than one CVD—excluding MTAG\_HF\_EUR, MTAG\_PAD\_EUR, and MTAG\_Stroke\_EUR.

These findings suggest that shared functional category signals are largely driven by AF, CAD, and VTE, which are the CVDs with the strongest genetic associations, thus limiting their interpretation as general CVD susceptibility signals. Likewise, of the 74 lead SNPs, 118 MAGMA-derived susceptibility genes, and 156 TWAS-identified tissue-specific genes that were genome-wide significant across more than one CVD, most—45 (60.81%), 100 (84.75%), and 113 (72.44%), respectively—were shared among multiple diseases, again excluding MTAG\_HF\_EUR, MTAG\_PAD\_EUR, and MTAG\_Stroke\_EUR. Ultimately, these findings suggest that signals indicating shared biological mechanisms were predominantly driven by AF, CAD, and VTE, rather than reflecting a generalizable genetic basis across all CVDs. This is likely due to their relatively higher heritability and larger GWAS sample sizes, both of which enhance the statistical power to detect shared genetic associations. In contrast, other CVDs, including HF, PAD, and Stroke, may exhibit more heterogeneous etiologies or weaker genetic architectures, which complicates the identification of shared genetic components.

## Discussion

In this work, we systematically investigate the overlap of genomic loci, SNPs, genes, gene sets, tissue types, and functional categories across six major CVDs, leveraging data from over 1.2 million individuals. This comprehensive approach uncovers locally shared genetic mechanisms that would otherwise go undetected and yields four key insights into the shared genetic architecture of CVDs. Specifically, we quantify genetic overlap beyond genome-wide genetic correlations across CVDs, and identify a putative causal relationship between CAD and HF. We further identify 38 genomic loci with pleiotropic effects, the most prominent of which is located at the *LPA* gene. Analysis of associated genes and gene sets reveals strong enrichment in pathways related to circulatory system development. Notably, distinct genetic patterns are primarily driven by AF, CAD, and VTE, suggesting a mismatch between clinical classifications and underlying shared biological mechanisms.

First, our findings systematically quantified the genetic architecture of CVDs beyond genome-wide genetic correlations by integrating multiple methods, including LDSC, MiXeR, LAVA, and GWAS-PW. We observed marked differences in polygenicity across CVDs, accompanied by extensive genetic overlap and relatively few disease-specific variants. These findings were further supported by LAVA local  $r_g$ , which revealed heterogeneous effect directions that were not captured by genome-wide estimates. However, the Dice coefficient may underestimate the true extent of genetic overlap, particularly when one trait harbors significantly fewer causal variants, even if a substantial proportion of them are shared with the other trait. This limitation highlights the need to interpret overlap metrics cautiously, especially in cases with asymmetrical statistical power for variant discovery, such as the HF-PAD pair. In such settings, shared architecture may be obscured by the imbalance in variant discovery power (Supplementary Note 3). Overall, our results suggest that phenotypic specificity in CVDs is primarily shaped not by disease-specific variants, but by the distribution of effect sizes and directions among shared pleiotropic variants.

Second, our MR analyses illuminated that genetic predisposition to CAD is significantly associated with an elevated risk of HF, consistent with findings from previous epidemiological studies. The improved survival rate following acute myocardial infarction, along

with the declining prevalence of hypertension and valvular heart disease as primary causes of incident HF, has made CAD the leading risk factor for HF development<sup>53</sup>. Notably, nearly two-thirds of HF cases are now attributed to underlying CAD, underscoring its central role in the pathogenesis of HF<sup>54</sup>. We also identified a potential risk effect of VTE on stroke. However, current research mainly highlights an increased short-term risk of VTE following ischemic stroke, while evidence directly linking VTE to stroke as a causal risk factor remains limited, necessitating cautious interpretation. In sum, the shared genetic basis between most pairs of CVDs is potentially mediated by horizontal pleiotropy rather than vertical pleiotropy, pointing to a broader genetic framework underlying the risk of cardiovascular pathology. Moreover, most CVDs frequently co-occur with other cardiometabolic disorders and/or factors, suggesting a shared genetic susceptibility<sup>4,5,55</sup>. Our MR analysis of 24 metabolic traits or diseases and six major CVDs further confirmed that most metabolic traits and disorders exert a causal influence on CVD risk (Supplementary Note 6 and Supplementary Data 24). Among these, hypertension emerged as the most robust causal risk factor.

Third, to address limitations in previous cross-disease genetic overlap research, we employed MTAG analysis to enhance power for detecting pleiotropic variants, which allowed us to elucidate the genetic foundation for shared risks across diseases. Remarkably, we identified novel genomic loci associated with various CVDs, significantly increasing the effective sample sizes for these conditions. Our variant-level analyses strongly supported the existence of horizontal pleiotropy, with -10.88% of lead SNPs influencing more than one of the examined CVDs. Intriguingly, multi-trait colocalization analysis from HyPrColoc corroborated that 12 loci displayed particularly extensive pleiotropic profiles across CVDs. Of these, seven loci were further found to be colocalized across multiple CVDs and a variety of metabolic traits (such as blood pressure, lipidemic, and glycemic traits). The most highly pleiotropic locus was located in the *LPA* gene, which encodes apolipoprotein(a), a key component of Lp(a) that was strongly associated with increased CVD risk<sup>35,56,57</sup>. This result was supported by the colocalization between multiple CVDs and lipidemic traits (such as low-density lipoprotein and total cholesterol), which exemplified the complex genetic interplay underlying related diseases and their risk factors. Similarly, the second-most pronounced pleiotropic locus has been demonstrated to regulate the expression of the  *Twist1* gene in vascular smooth muscle cells, which was also supported by the colocalization between multiple CVDs and blood pressure traits, particularly SBP. The typical example was the locus 1p36.22, located in the *CASZ1* gene, which was colocalized with multiple CVDs and common metabolic traits (including blood pressure, lipidemic, and glycemic traits). *CASZ1* encodes a zinc finger transcription factor that can inhibit tumorigenesis. Interestingly, *CASZ1* promotes vascular development by directly regulating the EGFL7/RhoA-mediated pathway, which may contribute to stroke phenotypes<sup>58</sup>. Managing risk factors is also crucial to prevent stroke, because the association of *CASZ1* variant rs880315 with stroke has been proven to be affected by its risk factors<sup>59</sup>. Another interesting example was the locus 16q12.2, located in the intron of the *FTO* gene, which had been reported to be associated with fat mass and obesity in multiple populations<sup>60</sup>. *FTO* is known to encode a nucleic acid methylase dependent on  $\alpha$ -ketoglutarate and iron (Fe II), which is ubiquitous in human tissues<sup>61</sup>. Although this locus showed low colocalization signals between multiple CVDs and most metabolic traits, it was highly colocalized with MTAG\_VTE\_EUR, MTAG\_HF\_EUR, MTAG\_PAD\_EUR, and obesity-related traits (such as body mass index [BMI] and waist-hip ratio [WHR]). The index SNP rs1421085, located -18.9 kb from the previously identified causal variant rs11075990, was identified as a shared causal variant. This result highlights the *FTO* gene as a common genetic connection between metabolic traits (such as obesity) and CVDs.

Fourth, we unearthed extensive evidence linking the cross-disease genetics of CVDs to circulatory system development. Alongside *LPA*, pleiotropic genes like *IL6R* and *SMAD7*, and gene set assessment, further highlighted the connection between pleiotropy and genetic effects on circulatory system development<sup>62–65</sup>. Our functional enrichment analyses using GARFIELD revealed that shared functional category signals were predominantly driven by AF, CAD, and VTE. These results underscore the existence of two distinct groups of CVDs: the first group, which includes AF, CAD, and VTE, is characterized by stronger genetic signals and more prominent shared biological mechanisms. In contrast, the second group, comprising HF, PAD, and stroke, exhibits weaker genetic signals and fewer overlapping biological mechanisms. This observation challenges the classical categorical classification of CVDs. Additionally, to evaluate the potential causal roles of circulating proteins in CVD, we conducted proteome-wide MR and colocalization analyses using whole-proteome data from two major proteomic platforms (i.e., SomaScan and Olink), encompassing 2730 proteins. However, due to challenges such as differences between proteome profiling platforms and unavoidable sample overlap between exposure and outcome datasets, we did not identify any proteins with consistent causal associations across multiple CVDs (Supplementary Note 7, Supplementary Fig. 11, and Supplementary Data 25, 26).

However, our study has several limitations. First, although the dataset encompassed six major CVDs, larger sample sizes and the inclusion of additional disease types could further enhance the statistical power to detect cross-disorder genetic effects. Second, our study predominantly focused on individuals of European ancestry, and more research is needed to validate findings in other populations. To address this limitation, we incorporated GWAS summary statistics from East Asian cohorts for replication, which partially confirmed the consistency of the genetic architecture identified in the European sample (Supplementary Note 8 and Supplementary Data 27–39). Nonetheless, future research adopting a trans-ancestral approach will be crucial for assessing the broader generalizability of these findings. Finally, while GWAS designs capture common variant aspects of genetic architecture, studies examining copy-number variants and rare mutations are needed for a comprehensive understanding.

In conclusion, our study provides new insights into the complex genetic relationships among CVDs, uncovering shared genetic mechanisms, pleiotropy, and the role of circulatory system development. These findings not only advance our understanding of CVD genetics but also highlight potential therapeutic targets and emphasize the need for further research to disentangle the intricate interplay between genetic and environmental factors in cardiovascular health.

## Methods

### Data sources and quality control

We sought the largest GWAS summary statistics of CVDs from publicly available data sources of European ancestry, due to the limited availability of well-powered GWAS for other ancestries, and selected GWAS with sample sizes larger than 50,000 to ensure enough statistical power. Finally, we collected GWAS summary statistics for six major CVDs. Detailed information about these diseases and their publication sources was provided in Supplementary Data 1 and Supplementary Note 1<sup>10,13,15</sup>. Only GWAS summary statistics from European ancestry were used in the main analysis. Pre-analysis quality control steps were performed on the GWAS summary statistics, including (1) aligning to 1000 Genomes Project v3 European reference (hg19 genome build); (2) excluding non-autosomal SNPs; (3) filtering out SNPs without rsID or with duplicated rsIDs; and (4) keeping only biallelic SNPs with minor allele frequency (MAF) >0.01. To assure fair and interpretable comparisons across the six CVDs, we filtered all summary statistics to retain only SNPs available for all six CVDs, resulting in a final set of 6,925,287 SNPs. Additional data processing was carried out according to the

specific requirements of the different methods used in the subsequent analyses.

### Estimating SNP-based heritability and genome-wide genetic correlation using LDSC

We used univariate linkage disequilibrium (LD) score regression (LDSC) to estimate the SNP-based heritability ( $h^2_{\text{SNP}}$ , i.e., the proportion of the phenotypic variance in a trait that could be explained by common genetic variants included in the analysis) of each of the six CVDs<sup>26,66</sup>. All GWAS summary statistics were reformatted to match the precomputed LD scores from the 1000 Genomes Project European reference panel. SNPs were excluded if they did not intersect with the reference panel or were located within the MHC region (chromosome 6: 26–35 Mb). Bivariate LDSC was used to estimate genetic correlation ( $r_g$ , i.e., the proportion of genetic variance shared by two traits divided by the square root of the product of their SNP-based heritability estimates). To account for multiple testing across the six CVDs in our analysis, we used the matrix spectral decomposition (matSpD) software<sup>67</sup> to examine the degree of independence among each trait. We determined that the effective number of independent traits was four, resulting in six independent trait pairs. To correct for multiple testing, the Bonferroni correction was applied to the  $r_g$   $P$  values ( $P = 0.05/6$  combinations =  $0.05/6 = 8.33 \times 10^{-3}$ ). Like univariate LDSC, bivariate LDSC only requires GWAS summary statistics and produces estimates that are unbiased by sample overlap.

### Estimating polygenicity and genetic overlap using MiXeR

We used the univariate causal mixture modeling method (MiXeR) analysis to estimate the number of trait-influencing (i.e., “causal”) variants for each trait (i.e., polygenicity) and the average magnitude of additive genetic associations among these variants (i.e., discoverability)<sup>27</sup>. Following recommended guidelines, we used the 1000 Genomes Project Phase 3 European reference panel and excluded the MHC region (chromosome 6: 26–35 Mb). Bivariate MiXeR models were applied to infer additive genetic effects as a mixture of four bivariate Gaussian components: (a) variants not associated with either trait, (b) variants uniquely associated with the first trait, (c) variants uniquely associated with the second trait, and (d) variants associated with both traits (i.e., shared disease-influencing variants)<sup>68</sup>. Results were presented as Venn diagrams displaying the proportion of shared and unique variants that collectively accounted for 90% of SNP-based heritability in each GWAS. To evaluate genetic overlap, MiXeR was implemented to calculate the Dice coefficient (i.e., the ratio of shared variants to the total number of variants). Model fit was assessed using the Akaike information criterion (AIC), which was based on the maximum likelihood of GWAS z-scores and was visualized by modeled versus observed conditional quantile–quantile (Q-Q) plots.

### Estimating local genetic correlation using LAVA and GWAS-PW

We used univariate local analysis of (co)variant association (LAVA) to test for local SNP-based heritability (local  $h^2_{\text{SNP}}$ ) within each CVD. We used genomic regions defined as autosomal linkage disequilibrium (LD) blocks ( $N = 2495$ ) by ref. 28, which are characterized by minimal LD between blocks, contain at least 2500 variants each, and have an average size of ~1 Mb. The LD structure was estimated using the European sample reference panel from the 1000 Genomes Project Phase 3. The detection of valid and interpretable local genetic correlation (local  $r_g$ ) for bivariate testing requires the presence of sufficient local genetic signal. To this end, we used a threshold of  $P < 1 \times 10^{-4}$  to filter out non-associated loci (i.e., loci devoid of any univariate heritability) that nonetheless might have contributed to the overall  $h^2_{\text{SNP}}$ , to get a more global overview of the shared genetic signal. A total of 577 bivariate tests were conducted, as multiple tests could occur within a single region when different CVDs exhibit univariate signals in the same locus. Significance testing was performed using simulation-

based  $P$  values. To correct for multiple testing, the Bonferroni correction was applied to the local  $r_g$   $P$  values ( $P = 0.05/\text{the number of bivariate tests} = 0.05/577 = 8.67 \times 10^{-5}$ ). Sample overlap across GWAS datasets was accounted for by incorporating pairwise genetic covariance estimates from bivariate LDSC, which were subsequently standardized into a correlation matrix.

We applied the Bayesian pleiotropy association method from the pairwise GWAS framework (GWAS-PW) to detect genomic regions with shared genetic effects among CVDs<sup>29</sup>. As in the LAVA analysis, we used the 1000 Genomes Project Phase 3 European panel as the LD reference and partitioned it into 2495 semi-independent blocks of ~1 Mb each. For each region, GWAS-PW estimates the posterior probabilities (PPAs) for four scenarios: (i) the locus is associated with the first trait (PPA\_1); (ii) with the second trait (PPA\_2); (iii) with both traits via the same causal variant (PPA\_3); and (iv) with both traits via different causal variants (PPA\_4). A posterior probability greater than 0.5 for either PPA\_3 or PPA\_4 was considered evidence of shared genetic signal within the region.

### Mendelian randomization analysis

Genome-wide or local genetic correlations and overlaps may reveal important insights into shared biological mechanisms between two traits; however, these findings should not be interpreted as evidence of causal relationships in either direction. To evaluate potential causal relationships, we performed the latent causal variable (LCV) model to evaluate partial genetic causality between traits<sup>31</sup>. The LCV approach leverages the bivariate effect size distribution of SNPs from two GWAS and their LD scores to estimate the posterior mean genetic causality proportion (GCP), enabling partial genetic causality to be distinguished from  $r_g$ . GCP estimates close to zero for genetically correlated traits suggest that the observed relationship may result from horizontal pleiotropy. In such a case, the traits share genetic pathways but are unlikely to influence each other through vertical pleiotropy within the same causal mechanism. LCV estimates the GCP on a scale ranging from –1 to 1, with values >0 indicating partial genetic causality from the first trait to the second, and values <0 indicating the reverse direction. A strong causal signal was defined as a posterior GCP significantly different from zero (based on a one-sided  $t$ -test) and an absolute GCP greater than 0.6. To correct for multiple testing, the Bonferroni correction was applied to the GCP  $P$  values ( $P = 0.05/6$  combinations =  $0.05/6 = 8.33 \times 10^{-3}$ ). LCV accounts for SNP-based heritability and genetic correlation between traits and is robust to sample overlap. Notably, the LCV model assumes a single latent variable mediating the effect between two traits, with the effect being unidirectional. However, this assumption may be violated in the presence of bidirectional causality or multiple latent mediators, which could confound the results.

Given the limitations of the LCV method, we performed a sensitivity analysis using the MRlap method for CVD pairs that exhibited partial genetic causality<sup>69</sup>. MRlap corrected for biases such as sample overlap, weak instrument strength, and the winner’s curse, thereby offering more reliable causal estimates. This method first estimates observed MR-based effects, followed by corrections based on genetic covariance calculated through LDSC. We compared IVW results before and after correction to assess whether sample overlap significantly influenced the findings. If a significant impact was observed, the corrected  $P$  values were considered the primary results. We repeated the analyses using AF and HF GWAS summary statistics after excluding the UKB sample, which contributed most of the overlap, and compared the results. We also examined associations between the genetic liability of one CVD and another, adjusting for the presence of other CVDs using multivariable Mendelian randomization (MVMR)<sup>32</sup>. MVMR allows us to estimate causal relationships between two traits while controlling for the influence of other related traits. However, since MVMR relies on regression analysis, it does not formally test for mediation effects.



Instead, it provides adjusted causal estimates accounting for potential mediating factors, offering insights into the direct relationships between traits.

### Multi-trait analysis of GWAS using MTAG

Multi-trait analysis of GWAS (MTAG) applies a generalized inverse-variance-weighted meta-analysis approach to multiple correlated traits and aims to detect novel genetic associations for each trait by boosting statistical power by borrowing the correlations among correlated traits<sup>16</sup>. Briefly, MTAG takes summary statistics from single-trait GWAS as input and estimates trait-specific effects for one common set of SNPs, and the resulting *P* value could be interpreted similarly to those from single-trait GWAS. MTAG incorporates bivariate LDSC to account for possibly unknown sample overlap among the GWASs of multiple correlated traits. MTAG relies on the key homogeneity assumption that all SNPs across traits share the same variance-covariance matrix of effect sizes. However, the MTAG estimator remains consistent even when this assumption is violated, when some SNPs influence only a subset of the traits. We calculate the upper bound of the false discovery rate (“maxFDR”) to evaluate the extent of overall inflation due to potential violations of the homogeneity assumption. In the initial MTAG analysis, the “maxFDR” values were  $1.656 \times 10^{-3}$  for AF,  $3.092 \times 10^{-4}$  for CAD,  $2.798 \times 10^{-3}$  for VTE, 0.134 for HF, 0.354 for PAD, and 0.569 for stroke. The “maxFDR” values of HF, PAD, and stroke were greater than 0.05, suggesting potential overall inflation due to violation of the homogeneity assumption. We then repeated the MTAG analysis for HF, PAD, and stroke, obtaining revised ‘maxFDR’ values of 0.0398, 0.0419, and 0.0585, respectively.

To investigate whether violations of the assumptions of equal SNP heritability for each trait and perfect genetic covariance between traits could bias our MTAG results, we performed pairwise cross-trait meta-analysis using cross-phenotype association (CPASSOC) as a sensitivity analysis<sup>70</sup>. CPASSOC assumes the presence of heterogeneous effects across traits and estimates the cross-trait statistic *SHet* and *P*-value through a sample size-weighted, fixed-effect meta-analysis of GWAS summary statistics. We denote the summary statistics from single-trait GWAS as GWAS\_AF, GWAS\_CAD, GWAS\_VTE, GWAS\_HF, GWAS\_PAD, and GWAS\_Stroke, and the summary statistics from MTAG analysis as MTAG\_AF, MTAG\_CAD, MTAG\_VTE, MTAG\_HF, MTAG\_PAD, and MTAG\_Stroke, respectively. MatSpD estimated that the six traits were equivalent to three independent traits based on the MTAG results. Independent SNPs that were genome-wide significant after an additional Bonferroni correction for three independent traits ( $P < 5 \times 10^{-8} / 3 = 1.67 \times 10^{-8}$ ) in the cross-trait meta-analyses using both MTAG and CPASSOC were considered for further verification.

### Genomic loci definition and functional annotation using FUMA

Functional mapping and annotation of genetic associations (FUMA) was an online platform that annotates SNPs with their biological functionality and maps implicated genes<sup>71</sup>. We performed FUMA annotation with default settings, using the 1000 Genomes Project v3 of European samples as a reference panel. Before annotation, FUMA first defined independent significant SNPs with a genome-wide significant *P* value ( $5 \times 10^{-8}$ ) and were independent at  $r^2 < 0.6$  within 1 Mb. Based on LD information, a subset of these independent significant SNPs was labeled as lead SNPs (independent from each other at  $r^2 < 0.1$ ). Genomic loci were identified by merging the LD blocks of lead SNPs that are closely located to each other (less than 250 kb apart). The top lead SNP was defined as the SNP with the lowest *P*-value in a specific region. Biological functionality for lead SNPs, including potential regulatory functions (RegulomeDB score), deleteriousness score (CADD score), effects on gene functions (using ANNOVAR), and mRNA expression levels (using eQTL data), was also estimated by FUMA. In detail, the nearest genes and the functional consequences of each SNP on gene functions were annotated based on ANNOVAR. Combined

annotation dependent depletion (CADD) score indexes the deleteriousness of variants computed based on 67 annotation resources. SNPs with a CADD score higher than 12 were considered to confer deleterious effects. The RegulomeDB provides a categorical score that describes how likely a SNP is to play a regulatory role based on the integration of high-throughput datasets. The RDB score of 1a suggests the strongest evidence, while a score 7 represents the least support for a regulatory potential. eQTL mapping provides significant cis-SNP-gene pairs (up to 1 Mb apart) in CVD-related tissue types from GTEx. SNPs were mapped to genes based on their physical position in the genome and expression quantitative trait locus (eQTL) associations (obtained from ten related tissues in GTEx v8). The SNP locations were defined with reference to the human genome Build 37 (GRCh37/hg19), and only protein-coding genes were included in the analysis. In addition, genome-wide significant SNPs from CPASSOC and original GWAS results were also annotated by FUMA for comparison.

### Testing overlap of genomic risk loci across CVDs

FUMA-annotated genomic risk loci were overlapped based on their chromosomal location to identify overlapping genomic risk loci across multiple CVDs. We further performed Multi-trait colocalization analysis on these overlapping genomic loci using the HyPrColoc (hypothesis prioritization in multi-trait colocalization) method to identify the potential shared causal variants, thus revealing potential biological mechanisms among CVDs<sup>33</sup>. HyPrColoc, an extension of the colocalization method, enables colocalization analysis for multiple traits and uses a deterministic Bayesian divisive clustering algorithm to identify clusters of colocalized traits and candidate causal variants within the same genomic locus, providing the posterior probability (PP) of colocalization for each cluster. A genomic risk locus with a PP greater than 0.7, whose traits were identified by HyPrColoc, are consistent with those obtained by overlapping the genomic risk loci, was declared a colocalized locus. Next, we estimated posterior probabilities (known as the *m* value) for each pleiotropic locus to quantify the best-fit model of cross-disease genotype-phenotype relationships using MetaSoft<sup>34</sup> with the fixed-effects model (FE). An *m* value >0.9 indicated that a particular variant had an effect on a given disease, while an *m* value <0.1 predicted that the SNP had no effect on the disease.

We further examined whether these pleiotropic loci could be colocalized between multiple CVDs and metabolic traits using multi-trait colocalization analysis by HyPrColoc, including anthropometric traits (body mass index [BMI], waist-to-hip ratio [WHR], and waist-to-hip ratio adjusted for BMI [WHRadj])<sup>37</sup>, blood pressure traits (systolic blood pressure [SBP], diastolic blood pressure [DBP], and pulse pressure [PP])<sup>38</sup>, lipidemic traits (Low-density lipoprotein cholesterol [LDL-C], High-density lipoprotein cholesterol [HDL-C], triglycerides [TG], and total cholesterol [TC])<sup>39</sup>, glycemic traits (fasting insulin [FI], fasting glucose [FG], 2-h glucose [2hGlu], and glycated hemoglobin levels [HbA1c])<sup>40</sup>, kidney function (estimated glomerular filtration rate [eGFR])<sup>41</sup>, liver function (alanine aminotransferase [ALT], alkaline phosphatase [ALP], and  $\gamma$ -glutamyl transferase [GGT])<sup>42</sup>, and other traits like carotid artery intima-media thickness [cIMT] and C-reactive protein [CRP])<sup>37,38</sup>. The common risk factors, including hypertension, type 2 diabetes (T2D), chronic kidney disease (CKD), and non-alcoholic fatty liver disease (NAFLD), were also included in the above analysis<sup>33,45,46,48</sup>.

### Gene-level analysis using MAGMA and TWAS

Moving beyond SNP-level studies, we performed gene-based association analysis using multi-marker analysis of genomic annotation (MAGMA), which yields gene-level *P* values by evaluating the joint association of all SNPs within each gene, while accounting for LD between SNPs<sup>49</sup>. SNP positions were defined based on the human genome reference Build 37 (GRCh37/hg19), and only protein-coding genes (*n* = 17,363) were included in the analysis. Gene boundaries were

extended by  $\pm 10$  kb from the gene body, consistent with the default settings of positional mapping in FUMA. We applied Bonferroni correction for multiple testing based on the number of protein-coding genes ( $P = 0.05/\text{the number of protein-coding genes} = 0.05/17,636 = 2.84 \times 10^{-6}$ ), and additionally also indicated which results remained significant after further correction for the number of independent CVDs ( $P = 0.05/\text{the number of protein-coding genes/the number of independent CVDs} = 0.05/17,636/3 = 9.45 \times 10^{-7}$ ).

In addition, we leveraged a transcriptome-wide association study (TWAS) using the functional summary-based Imputation (FUSION) software to further investigate tissue-specific genes across the ten CVD-related tissues<sup>50</sup>. Briefly, we used the FUSION method to prioritize CVD-associated genes by integrating MTAG-based summary statistics with precomputed gene expression weights reference from trait-relevant tissues from the genotype-tissue expression (GTEx) project version 8, while considering LD structures. The cis-genetic components of tissue-specific gene expression were imputed using MTAG-based summary statistics and five predictive models: best linear unbiased prediction (BLUP), Bayesian sparse linear mixed model (BSLMM), least absolute shrinkage and selection operator (LASSO), Elastic Net (ENET), and top single-nucleotide polymorphisms (SNPs). The gene expression weights from the best-performing model were then combined with MTAG results to identify significant associations between gene expression and CVDs. In TWAS, correcting for multiple comparisons using the total number of tested genes is expected to be too stringent and not reflect true biology (i.e., gene expression covariance)<sup>72</sup>. To address this, we estimated the effective number of independent genes per tissue using matSpD software. Briefly, matSpD estimates the effective number of independent variables in a correlation matrix by examining the eigenvalues derived from spectral decomposition. The expression values for genes whose differential expression was predicted by FUSION were extracted from normalized gene expression matrices obtained from GTEx. A total of 19,144 protein-coding genes were tested across the ten GTEx tissues relevant to CVDs, therefore, we applied a Bonferroni-corrected significance threshold ( $P = 0.05/\text{the number of independent genes/the number of independent CVDs} = 0.05/19,144/3 = 8.71 \times 10^{-7}$ ). All gene-mapping analyses were conducted using both GWAS summary statistics and MTAG-based summary statistics across all six CVD traits.

### Pathway-level analysis using MAGMA, LDSC-SEG, and GARFIELD

We conducted gene set enrichment analysis using MAGMA on 9398 gene sets, including canonical pathways and biological process sets from the Molecular Signatures Database (MSigDB, v.2023.1; categories C2: CP REACTOME and C5: GO BP)<sup>73</sup>. Multiple testing correction was applied to adjust for the number of gene sets analyzed ( $P < 0.05/(7,744 + 1,654) = 5.33 \times 10^{-6}$ ), and additionally also indicated which results remained significant after an additional Bonferroni correction for the number of independent CVD traits ( $P < 0.05/(7,744 + 1,654)/3 = 1.78 \times 10^{-6}$ ). All statistical tests were two-sided.

LDSC applied to specifically expressed genes (LDSC-SEG) was used to identify tissue-specific enrichments in gene expression and chromatin modifications. The analysis incorporated multi-tissue gene expression datasets (from GTEx and the Franke lab) and multi-tissue chromatin datasets (from Roadmap Epigenomics and ENCODE)<sup>51</sup>. For tissue-specific gene expression, we used annotations derived from RNA-seq data of human tissues from GTEx, with the Bonferroni-corrected significant threshold set at  $P < 0.05/48/3 = 3.47 \times 10^{-4}$ . For tissue-specific histone marks, we used annotations derived from the Roadmap Epigenomics Project, focusing on narrow peaks for DNase hypersensitivity and chromatin marks, including H3K27ac, H3K4me1, H3K4me3, H3K9ac, and H3K36me3, with the Bonferroni-corrected significant threshold set at  $P < 0.05/489/3 = 3.41 \times 10^{-5}$ .

We also applied GWAS analysis of regulatory and functional information enrichment with LD correction (GARFIELD), a powerful

enrichment tool for comparing different sets of annotation marks across specific cell types or tissues, to correlate GWAS summary statistics with regulatory and functional annotations. This approach identifies features relevant to a phenotype of interest across various GWAS  $P$  value thresholds, while adjusting for LD, MAF, and distance to the transcription start site (TSS)<sup>52</sup>. The annotations included 1005 features derived from ENCODE, GENCODE, and Roadmap Epigenomics projects, including DNase I hypersensitivity hotspots, chromatin accessibility peaks, transcription factor footprints, formaldehyde-assisted isolation of regulatory elements (FAIRE), histone modifications, chromatin segmentation states, genic annotations, and transcription factor binding sites (TFBS), among others, across multiple publicly available cell lines or tissues. Enrichment  $P$  values were empirically determined using a permutation-based procedure that accounts for the structure of associated genomic regions, based on the number of SNPs and their mean LD. We also assessed multiple testing corrections to account for the number of functional categories tested ( $P < 0.05/1005 = 4.98 \times 10^{-5}$ ), and further identified results that remained significant after additional correction for the number of CVDs ( $P < 0.05/1005/3 = 1.66 \times 10^{-5}$ ).

### Reporting summary

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

### Data availability

All analyses in this study were conducted using publicly available datasets. Download links for publicly available GWAS summary statistics used as inputs in this study are provided in Supplementary Data 1. Genome-wide summary statistics for AF, HF, and stroke are available from the GWAS Catalog under accession codes GCST006414, GCST009541, and GCST90104539, respectively. GWAS summary statistics for CAD and PAD are publicly available for download from the Cardiovascular Disease Knowledge Portal (CVDKP) at <https://cvd.hugeamp.org/datasets.html>. Genome-wide summary statistics for VTE were obtained from the deCODE Genetics website: <https://www.decode.com/summarydata/>. The MTAG results for six cardiovascular diseases generated in this study have been deposited in the Zenodo repository under the accession code 10.5281/zenodo.15324321. Source data used for generating the figures are available in the Source Data file with this paper. Source data are provided with this paper.

### Code availability

All software (and version, where applicable) used to conduct the analyses in this paper are freely available online: matSpD (<http://neurogenetics.qimrberghofer.edu.au/matSpD>), LDSC (v1.0.1; <https://github.com/bulik/ldsc>), MiXeR (v1.3; <https://github.com/precimed/mixer>), LAVA (v0.1.0; <https://github.com/josefin-werme/LAVA>), GWAS-PW (v0.21; <https://github.com/joepickrell/gwas-pw>), LCV; MRlap (v0.0.3; <https://github.com/n-mounier/MRlap>), TwoSampleMR (v0.6.8; <https://mrceiu.github.io/TwoSampleMR>), MTAG (v0.9.0 beta; <https://github.com/JonJala/mtag>), CPASSOC (<http://hal.case.edu/~xxz10/zhu-web/>), FUMA (v1.5.4; <http://fuma.ctglab.nl/>), HyPrColoc (v1.0; <https://github.com/jrs95/hyprcoloc>), MetaSoft (<http://genetics.cs.ucla.edu/meta/>), MAGMA (v1.08; <https://ctg.cncr.nl/software/magma>), TWAS (<http://gusevlab.org/projects/fusion/>), GARFIELD (v2; <https://www.ebi.ac.uk/birney-srv/GARFIELD/>), and R (v4.1.3; <https://www.r-project.org/>).

### References

- Roth, G. A. et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **76**, 2982–3021 (2020).
- Roger, V. L. Epidemiology of heart failure: a contemporary perspective. *Circ. Res.* **128**, 1421–1434 (2021).



3. Vinter, N. et al. Temporal trends in lifetime risks of atrial fibrillation and its complications between 2000 and 2022: Danish, nationwide, population based cohort study. *BMJ* **385**, e77209 (2024).
4. Aday, A. W. & Matsushita, K. Epidemiology of peripheral artery disease and polyvascular disease. *Circ. Res.* **128**, 1818–1832 (2021).
5. Glynn, R. J. & Rosner, B. Comparison of risk factors for the competing risks of coronary heart disease, stroke, and venous thromboembolism. *Am. J. Epidemiol.* **162**, 975–982 (2005).
6. Frak, W. et al. Pathophysiology of cardiovascular diseases: new insights into molecular mechanisms of atherosclerosis, arterial hypertension, and coronary artery disease. *Biomedicines*. **10**, 1938 (2022).
7. Hajar, R. Genetics in cardiovascular disease. *Heart Views* **21**, 55–56 (2020).
8. Lloyd-Jones, D. M. et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA* **291**, 2204–2211 (2004).
9. Arking, D. E. & Chakravarti, A. Understanding cardiovascular disease through the lens of genome-wide association studies. *Trends Genet.* **25**, 387–394 (2009).
10. Nielsen, J. B. et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239 (2018).
11. Aragam, K. G. et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).
12. Ghouse, J. et al. Genome-wide meta-analysis identifies 93 risk loci and enables risk prediction equivalent to monogenic forms of venous thromboembolism. *Nat. Genet.* **55**, 399–409 (2023).
13. Shah, S. et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
14. Palotie, A., Widen, E. & Ripatti, S. From genetic discovery to future personalized health research. *N. Biotech.* **30**, 291–295 (2013).
15. Evangelou, E. & Ioannidis, J. P. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **14**, 379–389 (2013).
16. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
17. Genomic Relationships Novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**, 1469–1482 (2019).
18. Romero, C. et al. Exploring the genetic overlap between twelve psychiatric disorders. *Nat. Genet.* **54**, 1795–1802 (2022).
19. Grotzinger, A. D. et al. Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic and molecular genetic levels of analysis. *Nat. Genet.* **54**, 548–559 (2022).
20. Ellinghaus, D. et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat. Genet.* **48**, 510–518 (2016).
21. Li, Y. R. et al. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat. Med.* **21**, 1018–1027 (2015).
22. Demela, P., Pirastu, N. & Soscic, B. Cross-disorder genetic analysis of immune diseases reveals distinct gene associations that converge on common pathways. *Nat. Commun.* **14**, 2743 (2023).
23. van Zuydam, N. R. et al. Genome-wide association study of peripheral artery disease. *Circ. Genom. Precis. Med.* **14**, e2862 (2021).
24. Mishra, A. et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature* **611**, 115–123 (2022).
25. Ding, Y. et al. Polygenic scoring accuracy varies across the genetic ancestry continuum. *Nature* **618**, 774–781 (2023).
26. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
27. Holland, D. et al. Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLoS Genet.* **16**, e1008612 (2020).
28. Werme, J., van der Sluis, S., Posthuma, D. & de Leeuw, C. A. An integrated framework for local genetic correlation analysis. *Nat. Genet.* **54**, 274–282 (2022).
29. Pickrell, J. K. et al. Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717 (2016).
30. Li, Y. Y., Wang, H. & Zhang, Y. Y. CDKN2B-AS1 gene rs4977574 A/G polymorphism and coronary heart disease: A meta-analysis of 40,979 subjects. *J. Cell. Mol. Med.* **25**, 8877–8889 (2021).
31. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* **50**, 1728–1734 (2018).
32. Sanderson, E., Spiller, W. & Bowden, J. Testing and correcting for weak and pleiotropic instruments in two-sample multi-variable Mendelian randomization. *Stat. Med.* **40**, 5434–5452 (2021).
33. Foley, C. N. et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun.* **12**, 764 (2021).
34. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
35. Nordestgaard, B. G. et al. Lipoprotein(a) as a cardiovascular risk factor: current status. *Eur. Heart J.* **31**, 2844–2853 (2010).
36. Nurnberg, S. T. et al. Genomic profiling of human vascular cells identifies TWIST1 as a causal gene for common vascular diseases. *PLoS Genet.* **16**, e1008538 (2020).
37. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700,000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
38. Evangelou, E. et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
39. Graham, S. E. et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).
40. Chen, J. et al. The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* **53**, 840–860 (2021).
41. Stanzick, K. J. et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nat. Commun.* **12**, 4350 (2021).
42. Pazoki, R. et al. Genetic analysis in European ancestry individuals identifies 517 loci associated with liver enzymes. *Nat. Commun.* **12**, 2579 (2021).
43. Yeung, M. W. et al. Twenty-five novel loci for carotid intima-media thickness: a genome-wide association study in >45,000 individuals and meta-analysis of >100,000 individuals. *Arterioscler. Thromb. Vasc. Biol.* **42**, 484–501 (2022).
44. Said, S. et al. Genetic analysis of over half a million people characterises C-reactive protein loci. *Nat. Commun.* **13**, 2198 (2022).
45. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
46. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
47. Wuttke, M. et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat. Genet.* **51**, 957–972 (2019).
48. Ghodsian, N. et al. Electronic health record-based genome-wide meta-analysis provides insights on the genetic architecture of non-alcoholic fatty liver disease. *Cell Rep. Med.* **2**, 100437 (2021).

49. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
50. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
51. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
52. Iotchkova, V. et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.* **51**, 343–353 (2019).
53. Lloyd-Jones, D. M. et al. Lifetime risk for developing congestive heart failure: the Framingham Heart Study. *Circulation* **106**, 3068–3072 (2002).
54. Gheorghade, M. et al. Navigating the crossroads of coronary artery disease and heart failure. *Circulation* **114**, 1202–1213 (2006).
55. Frostegard, J. Immunity, atherosclerosis and cardiovascular disease. *BMC Med.* **11**, 117 (2013).
56. Trinder, M., Uddin, M. M., Finneran, P., Aragam, K. G. & Natarajan, P. Clinical utility of lipoprotein(a) and LPA genetic risk score in risk prediction of incident atherosclerotic cardiovascular disease. *JAMA Cardiol.* **6**, 287–295 (2021).
57. Rider, D. A. et al. Pre-clinical assessment of SLN360, a novel siRNA targeting LPA, developed to address elevated lipoprotein (a) in cardiovascular disease. *Atherosclerosis* **349**, 240–247 (2022).
58. Charpentier, M. S. et al. CASZ1 promotes vascular assembly and morphogenesis through the direct regulation of an EGFL7/RhoA-mediated pathway. *Dev. Cell.* **25**, 132–143 (2013).
59. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
60. Loos, R. J. & Yeo, G. S. The bigger picture of FTO: the first GWAS-identified obesity gene. *Nat. Rev. Endocrinol.* **10**, 51–61 (2014).
61. Zhao, X., Yang, Y., Sun, B. F., Zhao, Y. L. & Yang, Y. G. FTO and obesity: mechanisms of association. *Curr. Diabetes Rep.* **14**, 486 (2014).
62. Bick, A. G. et al. Genetic interleukin 6 signaling deficiency attenuates cardiovascular risk in clonal hematopoiesis. *Circulation* **141**, 124–131 (2020).
63. Georgakis, M. K. et al. Associations of genetically predicted IL-6 signaling with cardiovascular disease risk across population subgroups. *BMC Med.* **20**, 245 (2022).
64. Wei, L. H. et al. Deficiency of Smad7 enhances cardiac remodeling induced by angiotensin II infusion in a mouse model of hypertension. *PLoS ONE* **8**, e70195 (2013).
65. Wei, L. H. et al. Smad7 inhibits angiotensin II-induced hypertensive cardiac remodelling. *Cardiovasc. Res.* **99**, 665–673 (2013).
66. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
67. Nyholt, D. R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74**, 765–769 (2004).
68. Frei, O. et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat. Commun.* **10**, 2417 (2019).
69. Mounier, N. & Kutalik, Z. Bias correction for inverse variance weighting Mendelian randomization. *Genet. Epidemiol.* **47**, 314–331 (2023).
70. Zhu, X. et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* **96**, 21–36 (2015).
71. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
72. Ghaffar, A. & Nyholt, D. R. Integrating eQTL and GWAS data characterises established and identifies novel migraine risk loci. *Hum. Genet.* **142**, 1113–1137 (2023).
73. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

## Acknowledgements

This work was funded by the Natural Science Foundation of China Excellent Young Scientists Fund (Overseas) (Grant No. K241141101), Guangdong Basic and Applied Basic Research Foundation for Distinguished Young Scholars (Grant No. 2024B1515020047), Shenzhen Basic Research General Projects of Shenzhen Science and Technology Innovation Commission (Grant No. JCYJ20230807093514029), National Natural Science Foundation of China (Grant No. 82470452), Natural Science Foundation of Xinjiang Uygur Autonomous Region (Grant No. 2024D01D15) (To Y.F.), National Natural Science Foundation (Grant No. 82170339 and 82270241), NSFC Incubation Project of Guangdong Provincial People's Hospital (Grant No. KY0120220021), Natural Science Foundation of Guangdong Province (Grant No. 2023B1515020082) (To L.J.), Clinical and Translational Research Project of Anhui Province (Grant No. 202427b10020131), Scientific Research Fund project of Anhui Medical University (Grant No. 2023xkj147), Cancer Research UK Career Development Fellowship (Grant No. C59392/A25064) (To S.P.), Shenzhen Science and Technology Program, Shenzhen, China (Grant No. GJHZ20240218111401002), National Science and Technology Major Project (Grant No. 2023ZD0505902) (To Jin-Song Bian), and Center for Computational Science and Engineering at Southern University of Science and Technology. The funder had no role in the design, implementation, analysis, interpretation of the data, approval of the manuscript, or decision to submit the manuscript for publication.

## Author contributions

J.Q., Y.F., Z.L., S.P., J.X., and L.J. conceptualized and supervised the project and wrote the manuscript. J.Q., L.J., L.C., M.C. (Minjing Chang), and C.W. performed the main analyses and wrote the manuscript. J.Q., W.C., R.Z., S.S. (Shan Song), Y.Z., and M.C. (Miaoran Chen) performed the statistical analyses and assisted with the interpretation of the results. S.D., N.T., P.H., and S.S. (Sakthivel Sadayappan) provided expertise in cardiovascular biology. All authors discussed the results and commented on the manuscript. Initials are expanded for authors with identical abbreviations to avoid confusion.

## 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/s41467-025-62419-0>.

**Correspondence** and requests for materials should be addressed to Lei Jiang, Jinguo Xu, Siim Pauklin, Zhengbing Liu or Yuliang Feng.

**Peer review information** *Nature Communications* thanks the anonymous reviewers 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.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

---

<sup>1</sup>Department of Pharmacology, SUSTech Homeostatic Medicine Institute, School of Medicine, Southern University of Science and Technology, Shenzhen, China. <sup>2</sup>Department of Geriatrics, Shenzhen People's Hospital (The First Affiliated Hospital of Southern University of Science and Technology), Shenzhen, China. <sup>3</sup>Joint Laboratory of Guangdong-Hong Kong Universities for Vascular Homeostasis and Diseases, Shenzhen, China. <sup>4</sup>Department of Cardiology, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China. <sup>5</sup>School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen, China. <sup>6</sup>Department of Health Statistics, School of Public Health, Shanxi Medical University, Taiyuan, China. <sup>7</sup>School of Medicine, South China University of Technology, Guangzhou, China. <sup>8</sup>Department of Rheumatology and Clinical Immunology, Peking University First Hospital, Beijing, China. <sup>9</sup>Department of Rheumatology, The Second Hospital of Shanxi Medical University, Shanxi Key Laboratory of Rheumatism Immune Microecology, Shanxi Precision Medical Engineering Research Center for Rheumatology, Taiyuan, China. <sup>10</sup>The First School of Clinical Medicine, Southern Medical University, Guangzhou, China. <sup>11</sup>Department of Cardiology, General Hospital of Central Theater Command, Wuhan Clinical Medicine College of Southern Medical University, Wuhan, China. <sup>12</sup>Heart, Lung and Vascular Institute, Department of Internal Medicine, Division of Cardiovascular Health and Disease, University of Cincinnati, Cincinnati, OH, USA. <sup>13</sup>Department of Cardiovascular Surgery, First Affiliated Hospital of Anhui Medical University, Hefei, China. <sup>14</sup>Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Headington, Oxford, UK. <sup>15</sup>Department of Cardiology, Jinling Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China. ✉e-mail: [jianglei0731@gmail.com](mailto:jianglei0731@gmail.com); [xujinguo@ahmu.edu.cn](mailto:xujinguo@ahmu.edu.cn); [siim.pauklin@ndorms.ox.ac.uk](mailto:siim.pauklin@ndorms.ox.ac.uk); [njzylzb@126.com](mailto:njzylzb@126.com); [fengyl@sustech.edu.cn](mailto:fengyl@sustech.edu.cn)