1 **Disentangling shared genetic etiologies for kidney function and cardiovascular**

2 **diseases**

3 Jun Qiao[1#], Kaixin Yao[2,3#], Yujuan Yuan[4,5#], Xichen Yang[2,3], Le Zhou[2,3], Yinqi Long[2,3],

4 Miaoran Chen[2,3], Wenjia Xie[2,3], Yixuan Yang[2,3], Yangpo Cao[1], Siim Pauklin[6*], Jinguo

5 Xu[7*], Yining Yang[4,5*], Yuliang Feng[1*]

6 [1] Department of Pharmacology, Joint Laboratory of Guangdong-Hong Kong

7 Universities for Vascular Homeostasis and Diseases, School of Medicine, Southern

8 University of Science and Technology, Shenzhen, China.

9 [2] Department of Nephrology, Shanxi Kidney Disease Institute, Second Hospital of

10 Shanxi Medical University, Taiyuan, China.

11 [3] Kidney Research Center of Shanxi Medical University, Taiyuan, China.

12 [4] Department of Cardiology, People's Hospital of Xinjiang Uygur Autonomous Region,

13 Urumqi, China.

14 [5] Xinjiang Key Laboratory of Cardiovascular Homeostasis and Regeneration Research,

15 Urumqi, China

16 [6] Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and

17 Musculoskeletal Sciences, University of Oxford, Headington, Oxford, UK.

18 [7] Department of Cardiovascular Surgery, First Affiliated Hospital of Anhui Medical

19 University, Hefei, China.

20 [#]These authors contributed equally to this work.

21

22 *Correspondence to:

23   Siim Pauklin PhD

24   Group leader and CRUK Career Development Fellow

25   Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and

26   Musculoskeletal Sciences, University of Oxford, Headington, Oxford OX3 7LD, UK.

27   Tel: +44 (0)-1865226492

28   Email: siim.pauklin@ndorms.ox.ac.uk

29

30   Jinguo Xu, MD, PhD

31   Associate Professor

32   Department of Cardiovascular Surgery, First Affiliated Hospital of Anhui Medical

33   University, Hefei, Anhui, 230022, China

34   Tel: +86 (551)-65908735

35   Email: xujinguo@ahmu.edu.cn

36

37   Yining Yang MD, PhD

38   Professor

39   Department of Cardiology, People's Hospital of Xinjiang Uygur Autonomous Region,

40   Urumqi, Xinjiang Uygur Autonomous Region, 830000, China.

41   Xinjiang Key Laboratory of Cardiovascular Homeostasis and Regeneration Research,

42   Urumqi, Xinjiang Uygur Autonomous Region, 830000, China.

43   Tel:+86 (0991)-960200

44   Email: yangyn5126@xjrmyy.com

45

46   Yuliang Feng MD, PhD

47   Associate Professor

48   Department of Pharmacology, Joint Laboratory of Guangdong-Hong Kong

49   Universities for Vascular Homeostasis and Diseases, School of Medicine, Southern

50   University of Science and Technology, Shenzhen, Guangdong, 518055, China.

51   Tel:+86 (755)-88012564

52   Email: fengyl@sustech.edu.cn

53

54 **Abstract**

55 Cardiovascular diseases (CVDs) are the leading cause of death worldwide, with

56 chronic kidney disease (CKD) identified as a significant risk factor. CKD is primarily

57 monitored through the estimated glomerular filtration rate (eGFR), calculated using the

58 CKD-EPI equation. Although epidemiological and clinical studies have consistently

59 demonstrated strong associations between eGFR and CVDs, the genetic underpinnings

60 of this relationship remain elusive. Recent genome-wide association studies (GWAS)

61 have highlighted the polygenic nature of these conditions and identified several risk

62 loci correlating with their cross-phenotypes. Nonetheless, the extent and pattern of their

63 pleiotropic effects have yet to be fully elucidated. We analyzed the most

64 comprehensive GWAS summary statistics, involving around 7.5 million individuals, to

65 investigate the shared genetic architectures and the underlying mechanisms between

66 eGFR and CVDs, focusing on single nucleotide polymorphisms (SNPs), genes,

67 biological pathways, and proteins exhibiting pleiotropic effects. Our study identified

68 508 distinct genomic locations associated with pleiotropic effects across multiple traits,

69 involving 379 unique genes, notably *L3MBTL3* (6q23.1), *MMP24* (20q11.22), and

70 *ABO* (9q34.2). Additionally, pathways such as stem cell population maintenance and

71 the glutathione metabolism pathway were pivotal in mediating the relationships

72 between these traits. From the perspective of vertical pleiotropy, our findings suggest a

73 causal relationship between eGFR and conditions such as atrial fibrillation and venous

74 thromboembolism. These insights significantly enhance our understanding of the

75   genetic links between eGFR and CVDs, potentially guiding the development of novel

76   therapeutic strategies and improving the clinical management of these conditions.

77

78   **Keywords**

79   Estimated glomerular filtration rate, Cardiovascular disease, Shared genetic

80   architectures, GWAS

81

Introduction

Cardiovascular diseases (CVDs) are the leading cause of mortality and disability globally, accounting for one-third of all deaths.[1] The American Heart Association anticipates that by 2035, the economic impact of CVDs will escalate to approximately $1.1 trillion. Among the diverse risk factors, chronic kidney disease (CKD) is particularly significant, as it exacerbates heart disease.[2] CKD leads to an increased production of reactive oxygen species (ROS) through oxidative stress, resulting in endothelial dysfunction—a critical precursor to atherosclerosis and an elevated risk of cardiovascular events.[3] Affecting about 9.1% of the global population, as the World Health Organization reported in 2023, CKD is diagnosed and monitored by measuring the estimated glomerular filtration rate (eGFR), utilizing the CKD-EPI equation. An eGFR below 60 mL/min/1.73 m² signifies notable kidney impairment.[4,5] Research by Go et al. has demonstrated a clear, graded association between declining eGFR and increased cardiovascular events.[6,7] Additionally, studies focusing on older adults reveal that the risk of CVDs increases from 15% for those with an eGFR of 90 mL/min/1.73 m² to 40% for individuals at 30 mL/min/1.73 m² over three years.[8] The Framingham Heart Study supports these findings, showing that individuals with mildly reduced eGFR levels (60 to 79 mL/min per 1.73 m²) face a higher risk of CVDs than those with higher eGFR.[9] These studies underscore the significance of eGFR as a robust biomarker for assessing cardiovascular risk, highlighting its crucial role in clinical practice.

Genome-wide association studies (GWAS) have identified hundreds of variants contributing to the risk of eGFR and CVDs, some of which are shared risk loci. For example, Graham and colleagues pinpointed 147 loci associated with eGFR, seven demonstrating significant co-localization with cardiovascular traits, highlighting a genetic link between eGFR and various CVD phenotypes.[10] Genetic pleiotropy, where a single genetic variant influences multiple traits, is prevalent in complex human diseases, particularly in loci associated with CVDs. This pleiotropy can be vertical,

110   where a variant's effect on one trait impacts another, or horizontal, where a variant

111   independently affects several traits. Mendelian randomization (MR) analyses utilizing

112   vertical pleiotropy have estimated causal relationships between eGFR and CVDs. For

113   example, Kelly et al. established a causal link between declining eGFR and increased

114   stroke risk,[11] while Geurts et al. identified a bidirectional causal relationship between

115   eGFR and atrial fibrillation (AF).[12,13] However, recent studies suggest that eGFR does

116   not significantly affect other CVDs, such as AF, coronary artery disease (CAD), stroke,

117   and heart failure (HF). These inconsistencies in MR results, possibly due to residual

118   biases and unmeasured confounders, necessitate further investigation. Furthermore, the

119   study by Gong et al. highlights the role of horizontal pleiotropy in explaining the

120   common genetic architecture of human phenotypes, within the context of recent

121   advances in genomic statistical tools.[14] Despite these advances, the extent of horizontal

122   pleiotropy between eGFR and CVDs still needs to be explored, and the common

123   genetic mechanisms must be fully elucidated. Therefore, there is a critical need for

124   systematic research to investigate the shared genetic architectures between eGFR and

125   CVDs, to identify shared polygenic risk variants, and to explore the involvement of

126   specific molecular biological pathways.

127   In this study, we conducted an extensive analysis leveraged the latest GWAS summary

128   statistics for European ancestry to explore the genetic associations between eGFR and

129   six major CVDs, including AF, CAD, Venous Thromboembolism (VTE), HF,

130   Peripheral Artery Disease (PAD), and Stroke. Firstly, the shared genetic basis is

131   determined by quantifying the overall and local genetic correlations and exploring the

132   overall genetic overlap. On this common genetic basis, in-depth research was

133   conducted on the potential genetic mechanisms underlying these traits. Regarding

134   horizontal pleiotropy, first, pleiotropy variation at the single nucleotide polymorphism

135   (SNP) level is detected, and then co-localization analysis is performed to locate

136   candidate causal variations accurately. Subsequently, position mapping and expression

137   quantitative trait loci (eQTL) mapping were used at the gene level to identify candidate

138   multi-effect genes. In addition, we conducted enrichment analysis on biological

139 pathways and elucidated the biological functions of genes associated with multiple

140 effector sites. We also investigated the potential association between plasma protein

141 levels and disease susceptibility and evaluated its potential as a therapeutic target.

142 Finally, MR analysis was conducted at the level of vertical pleiotropy, using whole

143 genome SNP data to estimate bidirectional causal effects while considering

144 confounding factors and sample overlap.In conclusion, our comprehensive analysis of

145 pleiotropy revealed the intricate genetic interactions between eGFR and CVDs. This

146 work provides a promising path for a deeper understanding of their genetic associations,

147 laying a solid foundation for future studies and driving new insights and targets in

148 prevention, early diagnosis, and personalized treatment.

149

150 **Methods**

151 **Study Design**

152 Figure 1 presents the workflow for this study.

153

154 **Data selection and quality control**

155 GWAS summary statistics for eGFR were sourced from the most extensive publicly

156 accessible meta-analysis to date, encompassing two primary datasets (n = 1,004,040): (i)

157 from the CKDGen consortium and (ii) from the UK Biobank.[15] Similarly, we selected

158 six major CVDs from large meta-analyses for greater statistical power and clinical

159 relevance. We extracted GWAS summary statistics for AF from a genome-wide

160 meta-analysis that included 60,620 cases and 970,216 controls of European ancestry

161 across six studies: The Nord-Trøndelag Health Study (HUNT), deCODE, the Michigan

162 Genomics Initiative (MGI), DiscovEHR, UK Biobank, and the AFGen Consortium.[16]

163 GWAS summary statistics for CAD were from a meta-analysis encompassed 181,522

164 cases and 984,168 controls of European ancestry, amalgamating data from nine studies

165 not previously included along with data from the UK Biobank and

166 CARDIoGRAMplusC4D consortium.[17] GWAS summary statistics for VTE were

167   obtained from a meta-analysis of seven cohorts, including the Copenhagen Hospital

168   Biobank Cardiovascular Disease Cohort (CHB-CVDC), Danish Blood Donor Study

169   (DBDS), Intermountain Healthcare, deCODE, UK Biobank, Million Veteran Program

170   (MVP), and FinnGen, totaling 81,190 cases and 1,419,671 controls.[18] GWAS summary

171   statistics for HF included data from 47,309 cases and 930,014 controls across 26 studies

172   from the Heart Failure Molecular Epidemiology for Therapeutic Targets (HERMES)

173   Consortium.[19] GWAS summary statistics for PAD involved 12,086 cases and 499,548

174   controls from a genome-wide meta-analysis comprising 11 independent studies.[20]

175   GWAS summary statistics for Stroke were sourced from a meta-analysis by the

176   GIGASTROKE consortium, which included 73,652 patients and 1,234,808 control

177   individuals.[21] All summary statistics are based on individuals of European ancestry due

178   to the limited availability of well-powered GWAS for other ancestries. Further details

179   about the traits are available in eTable 1 in Supplement 1.

180

181   Before further analysis, we implemented rigorous quality control measures on these

182   GWAS summary statistics by aligning them with the 1000 Genomes Project Phase 3

183   European population based on the hg19 genome build, removing SNPs lacking rsIDs or

184   with duplicate rsIDs, restricting the analysis to autosomal chromosomes, and retaining

185   only SNPs with a minor allele frequency (MAF) greater than 0.01. A total of 6,907,393

186   SNPs were included in the final analysis.

187

188   **Genetic correlation between eGFR and CVDs**

189   In order to investigate the possible shared genetic basis of eGFR and CVDs, we

190   estimated SNP-based heritability (h^2_{SNP}) and evaluated genome-wide genetic

191   correlation (r_g) by means of cross-trait linkage disequilibrium (LD) score regression

192   (LDSC) method.[22] LDSC estimates the genetic contribution to complex diseases and

193   traits by quantifying the LD between each SNP and its neighbors. Initially, we

194   performed univariate LDSC to estimate h^2_{SNP}, which reflects the proportion of

195   phenotypic variation in a trait explained by shared genetic variants. LD scores were

196   calculated for each SNP using genotypes of common SNPs within a 10 Mb window

197   sourced from the 1000 Genomes Project Phase 3 European population. SNPs in the

198   major histocompatibility complex (MHC) region (chromosome 6: 25-35 Mb) were

199   excluded due to their intricate LD structure. We then conducted bivariate LDSC to

200   assess the r_g between eGFR and CVDs, using regression of LD scores against the

201   product of z statistics from the respective diseases and traits. The r_g ranges from -1,

202   representing a complete negative correlation, to +1, representing a complete positive

203   correlation, with values closer to these extremes indicating stronger correlations. The

204   intercept from LDSC may indicate a possible sample overlap among the GWAS data

205   sets; however, it is possible to estimate the r_g in an unbiased manner, even with

206   overlapping samples. After evaluating all pairwise r_g, we applied a Bonferroni

207   correction for multiple comparisons, setting the significance threshold at $P < 8.33 \times 10^{-3}$

208   (0.05 / 6).

209

210   In order to evaluate the biological basis of the shared genetic predisposition to eGFR

211   and CVDs, we performed stratified LDSC applied to specifically expressed genes

212   (LDSC-SEG) to identify possible enrichment of related tissue and cell types. This

213   method integrates multi-tissue gene expression data from sources like the

214   Genotype-Tissue Expression (GTEx) and Franke Lab, as well as multi-tissue chromatin

215   information from the Roadmap Epigenomics and ENCODE datasets. The GTEx

216   project provided statistics representing key tissue types and their specific expression

217   across 49 tissues using baseline models and complete genomes. Additionally, we

218   incorporated tissue-specific histone marker annotations from the Roadmap

219   Epigenomics project. This included narrow peaks for DNase Hypersensitivity,

220   H3K27ac, H3K4me1, H3K4me3, H3K9ac, and H3K36me3 chromatid. For identified

221   relevant tissues or cell types, we used the false discovery rate (FDR) method to correct

222   for each dataset, with a significance threshold of FDR < 0.05.

223

224   Genetic overlap between eGFR and CVDs

225   LDSC primarily measures the average genetic correlation of effect sizes for all SNPs

226   across the genome. This approach can potentially obscure the shared genetic

227   architectures involving a mixture of concordant and discordant effect directions and

228   may fail to capture specific overlapping loci and relevant genes. To address this

229   limitation and quantify the polygenic overlap between eGFR and CVDs, we employed

230   the causal mixture modeling approach (MiXeR).[23] Firstly, univariate MiXeR analyses

231   were performed to estimate polygenicity (i.e., the count of variants explaining 90% of

232   h^2_{SNP}) and discoverability (the proportion of phenotypic variance attributable to causal

233   variant effect sizes).[23] We used data from the 1000 Genomes Project Phase 3 European

234   population for our analyses and excluded SNPs within the structurally complex MHC

235   region (CHR 6: 26-35 Mb). Subsequently, bivariate MiXeR analyses were conducted to

236   ascertain the total count of trait-specific and shared causal SNPs visually represented in

237   Venn diagrams. This delineation covers four components: unique causal variants for

238   trait 1, unique for trait 2, shared causal variants, and shared non-causal variants. MiXeR

239   calculates Dice coefficient scores (ranging from 0 to 1) to gauge polygenic overlap and

240   estimates the proportion of SNPs with concordant effects within the shared genetic

241   component. Additionally, MiXeR computes the r_g and the correlation of effect sizes

242   within the shared genetic component ($r_g s$). To evaluate model fit, indicative of MiXeR's

243   predictive accuracy against actual GWAS data, we constructed conditional

244   quantile-quantile (Q-Q) plots, used the Akaike Information Criterion (AIC), and

245   generated log-likelihood plots. A model-based Q-Q plot that closely aligns with the

246   actual Q-Q plot indicates strong predictive capability. A positive AIC value suggests

247   the model adequately distinguishes itself from comparative models, indicating a

248   reliable fit. Conversely, a negative AIC implies that the MiXeR model does not

249   significantly differentiate from scenarios of maximum or minimum genetic overlap,

250   rendering the genetic overlap estimate potentially unreliable.

251

252   **Local genetic correlation between eGFR and CVDs**

253   To estimate local genetic correlations (local-$r_g s$) between eGFR and CVDs within

254   specific genomic regions and identify loci with mixed effect directions, we utilized the

255   Local Analysis of [co]Variant Annotation (LAVA).[24] LAVA captures detailed patterns

256   of shared genetic loci across individuals, revealing mixed effect directions obscured by

257   genome-wide r_g estimates due to opposing effects across genomic regions. Unlike the

258   genome-wide correlation analysis using LDSC, which considers the entire genome,

259   LAVA divides the genome into 2,495 local LD blocks and estimates r_g within each

260   block. Differing from MiXeR's method of estimating the proportion of shared 'causal'

261   variants with concordant effects, LAVA identifies mixed effect directions by capturing

262   significantly correlated genetic loci. Our study began with a univariate LAVA analysis

263   to estimate local heritability for each phenotype, considering loci with p-values <

264   1×10^{-4} as significant genetic signals. This approach identified 425 loci for subsequent

265   bivariate tests to estimate pairwise bivariate local-r_gs across the genome. We adjusted

266   the p-values for local-r_gs based on the number of bivariate tests, setting a

267   Bonferroni-corrected significance threshold at $P < 1.18\times10^{-4}$ (0.05/425).

268

269   Additionally, we employed Hypothesis Prioritisation for multi-trait Colocalization

270   (HyPrColoc) on genomic regions with shared risk loci across multiple phenotypes to

271   elucidate potential biological mechanisms linking traits. HyPrColoc, a Bayesian

272   approach, identifies clusters of co-localization and candidate cause variants in the same

273   genomic locus. It estimates the posterior probability (PP) of colocalization of multiple

274   traits within a single causal variant, considering PP > 0.7 as significant evidence of

275   colocalization. This method enhances our understanding of how genetic variants

276   contribute to multiple traits, offering insights into complex trait interactions.

277

278   **Mendelian randomization analysis between eGFR and CVDs**

279   The genetic correlation and overlap analysis explored the shared genetic foundation

280   between eGFR and CVDs; however, whether their relationship is mediated by vertical

281   or horizontal pleiotropy remains unclear. Latent Heritable Confounder Mendelian

282   Randomization (LHC-MR)[25] provides valuable insights into the causal relationship

283   between eGFR and CVDs by utilizing vertical pleiotropy. This approach makes full use

284   of all genome-wide variations, improves statistical capacity, and corrects for sample

285   overlap, rather than just significant loci throughout the genome. Critically, LHC-MR is

286   able to distinguish between SNPs according to their co-association with a set of traits,

287   and to differentiate heritable confounding that results in r_g. This capability allows

288   LHC-MR to provide concurrent unbiased estimates of bidirectional causal and

289   confounder effects. The LHC-MR framework accommodates multiple pathways

290   through which SNPs can affect the traits, including allowing for null effects, thereby

291   enabling precise causal effect estimations. Causal estimates from LHC-MR are

292   presented as odds ratios (ORs) with corresponding 95% confidence intervals (CIs).

293   Causality is considered unidirectional if P is lower than the Bonferroni-corrected

294   threshold ($0.05 / 6 / 2 = 4.17 \times 10^{-3}$) and the P-value of the effect in the opposite

295   direction is greater than 0.05. Bidirectional causality is considered if $P < 4.17 \times 10^{-3}$ in

296   both directions. Additionally, we employed several methods as sensitivity analyses to

297   validate the results, including the inverse variance weighted (IVW) method, the

298   weighted median, MR-Egger, simple mode, and weighted mode. These analyses help

299   confirm the robustness of the causal inferences drawn from our study.

300

301   **SNP-level analysis**

302   **Identification of pleiotropic loci between eGFR and CVDs**

303   In order to explore the effect of horizontal pleiotropy between eGFR and CVDs, we

304   used Pleiotropic Analysis under the Composite Null Hypothesis (PLACO) to clarify the

305   shared genetic mechanisms underlying these conditions.[26] PLACO identifies

306   pleiotropic loci between two traits by testing the compound null hypothesis that a locus

307   is related to either none or only one of them. This hypothesis is further subdivided into

308   specific sub-scenarios: (1) M00, indicating no association with any trait; (2) M10 and

309   M01, each indicating an association with only one of the traits; and (3) M11,

310   representing a pleiotropic association with both traits. This method calculates the

311   PLACO statistics by multiplying the Z-scores of the two traits ($Z_{trait1} \times Z_{trait2}$) while

312   excluding SNPs with squared Z-values above 100 to mitigate spurious signals of

313   pleiotropy. SNPs are considered to exhibit significant genome-wide pleiotropy if their

314   P_{PLACO} is less than 5×10^{-8}.

315

316   **Genomic loci definition and functional analysis**

317   To identify pleiotropic SNPs associated with eGFR and CVDs, we utilized the

318   Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA)

319   platform. FUMA integrates data from a variety of biological databases to improve

320   functional annotation, gene prioritization, and interactive visualization of GWAS

321   outcomes.[27] We calculated LD scores using the 1000 Genomes Phase III European

322   population as a reference. Firstly, independently significant SNPs achieving

323   genome-wide significance ($P < 5.0 \times 10^{-8}$ and $r^2 < 0.6$) were identified. Lead SNPs

324   were then determined based on their independence ($r^2 < 0.1$). Genomic risk sites were

325   delineated by merging lead SNPs within 500 kb of each other, allowing these sites to

326   contain multiple lead SNPs. The top lead SNP at each site, defined by the lowest P

327   value, was selected. Additionally, the directional impact of these sites was assessed by

328   comparing Z-scores between eGFR and CVD, providing insights into their potential

329   roles in these conditions. A GWAS site was considered novel if none of its lead,

330   independent, or candidate SNPs overlapped with SNPs previously reported in GWAS

331   meta-analyses. Using LocusZoom, we created a regional association plot with gene

332   tracks, allowing us to examine the details of the relationship between each locus. To

333   predict the functional outcomes of Top SNPs, we matched SNPs to databases

334   containing established functional annotations, including the Annotate Variation

335   (ANNOVAR) category, Combined Annotation-Dependent Depletion (CADD) score,

336   RegulomeDB (RDB) score, and chromatin state.[28-30] The CADD score is used to

337   predict the deleteriousness of SNP effects, incorporating 63 functional annotations. A

338   threshold of 12.37 is generally recognized as indicative of deleterious variants, and thus,

339   we filtered SNPs with a CADD score greater than 12.37 for location mapping. In order

340   to prioritize eQTL genes and explore their potential regulatory role, we used

341   RegulomeDB, which provides functional interpretation of SNPs based on curated

342   references. The RegulomeDB scoring system, ranging from 1a (most regulatory

343   potential) to 7 (least), helps evaluate the regulatory likelihood of SNPs. Additionally,

344   chromatin states were used to delineate the regulatory landscape of genomic regions,

345   with 15 classes of states predicted using 5 chromatin markers across 127 epigenomes

346   via the ChromHMM tool.[31] Two methods were employed to map SNPs to genes: (i)

347   positional mapping, which assigns SNPs to genes based on their physical proximity

348   (within a 10kb window) to known protein-coding genes in the human reference

349   assembly, and (ii) eQTL mapping, which links SNPs to genes based on significant

350   eQTL associations, where allelic variations in an SNP correlate with variations in gene

351   expression levels, as identified using the GTEx database. This approach improves our

352   understanding of the genetic structure by linking SNPs to potential functional

353   outcomes.

354

355   **Colocalization analysis**

356   For FUMA-annotated pleiotropic loci, we performed a Bayesian colocalization

357   analysis to identify potential common causal variations between trait pairs. COLOC

358   employs a Bayesian framework to calculate the PP of five distinct hypotheses

359   concerning shared causal variation within a genomic region.[32] These hypotheses assess

360   whether one or both traits share causal variations at a specific locus. Specifically, the

361   hypotheses are as follows: i) PPH0: Neither trait has causal variation at the locus; ii)

362   PPH1: Only the first trait has causal variation at the locus; iii) PPH2: Only the second

363   trait has causal variation at the locus; iv) PPH3: Each trait has a different causal

364   variation within the locus; v) PPH4: Both traits share a causal variation within the

365   locus.[33] The SNP exhibiting the highest PPH4 within the locus is identified as the

366   candidate causal variant. Loci are considered colocalized if the PPH4 exceeds 0.7,

367   indicating a strong likelihood that the locus harbors shared causal variations for the

368   traits under study.

369

370

371   **Gene-level analysis using MAGMA, eMAGMA and TWAS**

372   To identify candidate pleiotropic genes further, we conducted Multi-marker Analysis of

373   GenoMic Annotation (MAGMA) based on results from PLACO and individual GWAS.

374   MAGMA uses the SNP average model within the multi-regression framework to derive

375   *P* values, and to estimate the association of the gene, and to adjust for factors such as

376   the size of the gene, the number of SNP for each gene, and the LD between markers.[34]

377   Our analysis referenced the 1000 Genomes Project Phase 3 European population and

378   used the Genome Reference Consortium Human Build 37 (hg19) for SNP locations and

379   gene annotations. We analyzed 17,636 protein-coding genes on the MAGMA software

380   website, focusing on genes containing at least ten SNPs to ensure computational

381   stability. Gene based testing involves extending 10kb upstream and downstream of the

382   gene transcription start and end sites. Because of their complicated LD patterns, we

383   ruled out MHC regions (chr6: 25 - 35 Mb). Strict Bonferroni correction was used for

384   adjustment for multiple trials, and in the MAGMA analysis based on the PLACO

385   results, a significant threshold of $P < 4.23 \times 10^{-7}$ (0.05 / 17636 / 6) was established.

386

387   The inherent limitations of MAGMA, which assigns SNPs based solely on proximity to

388   genes without accounting for functional associations such as gene regulation, can

389   hinder its effectiveness in elucidating the underlying mechanisms of genetic variants.

390   To overcome this and delve deeper into the functional implications of genetic variants

391   associated with eGFR and CVDs, we employed the EQTL-informed MAGMA

392   (E-MAGMA).[35] Operating within the same statistical framework as MAGMA,

393   E-MAGMA uses a multi-principal component linear regression model to leverage

394   tissue-specific eQTL data from multiple sources, aiming to identify potential causal

395   genes for phenotypic traits. Our study utilized eQTL data from 47 tissues, enriched for

396   differentially expressed genes in tissues as provided by the Genotype-Tissue

397   Expression Project version 8 (GTEx v8), and referenced the 1000 Genomes Project

398   Phase 3 European population. To mitigate potential confounding effects from the broad

399   range of tissues, we focused on 11 specific tissues that were identified by LDSC-SEG

400   analysis, including arterial, adipose, cardiac, whole blood, liver, EBV-transformed

401   lymphocytes, and kidney tissues. We excluded kidney tissue from our analysis due to

402   the absence of reference in E-MAGMA. We applied Bonferroni-adjusted *p*-value

403   thresholds to accommodate multiple testing, calculated based on the number of genes

404   analyzed in E-MAGMA and the number of statistically significant trait pairs tested. For

405   example, the threshold for subcutaneous fat was defined as $P < 8.67 \times 10^{-7}$ (0.05 / 9,603

406   / 6).

407

408 We conducted the Transcriptome-wide Association Study (TWAS) using single-trait

409 GWAS results to explore the tissue specific gene expression correlation between eGFR

410 and CVDs. TWAS combines eQTL and summary association statistics from large scale

411 GWAS to identify genes that are associated with complex traits.[36] For gene expression

412 prediction, we utilized the FUSION tool with default settings, employing various

413 methods, including best linear unbiased prediction (BLUP), Bayesian sparse linear

414 mixed model (BSLMM), least absolute shrinkage and selection operator (LASSO),

415 elastic net (ENET), and Top SNP, which were used to estimate the cis-genetic

416 component of tissue-specific gene expression. We selected the most effective predictive

417 model to determine gene expression weights and used GWAS summary statistics to

418 conduct the TWAS. In addition, we used the Bonferroni correction to adjust for

419 multiple comparisons between the different tissue types analyzed.

420

421 **Pathway-level analysis using MAGMA and Metascape**

422 We conducted a MAGMA gene set analysis to elucidate the biological functions of

423 genes exhibiting pleiotropic effects on eGFR and CVDs. This analysis employs a

424 competitive gene set framework to assess whether specific gene sets are more strongly

425 associated with a particular phenotype than the broader genomic background. Our

426 approach integrates gene definitions and their respective signals using MAGMA's

427 gene-based multi-marker method. The analyzed gene set is derived from Gene

428 Ontology (GO) and Reactome pathways, as listed in the Molecular Feature Database

429 (MSigDB v7.5). The significance threshold was set to $P < 0.05 / (7744 + 1654) / 6$,

430 adjusted for the number of GO and Reactome pathways, as well as the number of trait

431 pairs analysed. To elucidate the biological processes and signaling pathways associated

432 with eGFR and CVDs, we conducted Metascape analysis on the overlapping genes

433 identified by MAGMA and EMAGMA. Metascape provides comprehensive annotation

434 of gene and protein, enrichment analysis, and protein-protein interaction networks,

435 which can help to better understand the functions of genes.[37] We used Metascape to

436 perform GO annotation and Reactome enrichment analysis, which utilizes the

437   hypergeometric test to identify significant ontology terms. The GO resource offers a

438   framework and concepts to describe gene product functions across all organisms, while

439   the Reactome knowledgebase details cellular processes, such as signaling, transport,

440   DNA replication, and metabolism, as ordered networks of molecular transformations.

441   Pathways with a p-value<0.01 were considered significant.

442

Proteome-wide Mendelian Randomization analysis using SMR

444   To investigate potential associations between plasma protein levels and disease

445   susceptibility, we employed Summary data-based Mendelian Randomization (SMR).

446   This analysis integrated plasma protein quantitative trait locus (pQTL) summary

447   statistics from the UK Biobank Pharmaceutical Proteomics Project (UKB-PPP) with

448   GWAS summary statistics for various diseases. The UKB-PPP data, derived from the

449   Olink proteomics platform, includes genetic associations for 2,940 plasma proteins

450   across a cohort of 34,557 Europeans. In this context, cis-pQTLs are defined as SNPs

451   located in the 1Mb window surrounding the transcription start site (TSS) of each

452   protein. For SMR analysis, we only took into account the cis-pQTLs related to plasma

453   protein levels at the genome-wide significance threshold ($P < 5 \times 10^{-8}$). SMR, a method

454   that utilizes summary-level data, assesses potential causal relationships between

455   exposures (e.g., plasma protein levels) and outcomes (e.g., traits or diseases). In order

456   to distinguish between pleiotropy and linkage in the cis-pQTL region, we tested

457   multiple SNPs using the heterogeneity of instrument-dependent (HEIDI) method.

458   Results indicating pleiotropy (HEIDI test $P < 0.01$) were excluded from further analysis.

459   In recognition of the limitations of single-SNP analyses, we also performed a

460   sensitivity analysis using a multi-SNP-SMR test, with a P value of < 0.05 being

461   significant. A Bonferroni correction was applied to the number of unique proteins to be

462   analyzed, and a significance threshold was set at $P < 3.65 \times 10^{-6}$ (0.05 / 1958 / 7).

463   Additionally, we performed COLOC analysis to determine if the same causal variation

464   is responsible for the association of protein levels with disease phenotype. The PP.H4 >

465   0.7 for shared causal variants indicates significant colocalization of GWAS and pQTL.

466

Result

Genome-wide genetic correlation between eGFR and CVDs

After implementing stringent quality control, we utilized LDSC to assess h^2_{SNP} and genome-wide r_g between eGFR and CVDs. We used univariate LDSC to estimate h^2_{SNP} for eGFR and each CVD. The SNP heritability estimate was highest for eGFR (h^2_{SNP} = 0.0708, SE = 0.0036). Among these CVDs, heritability estimates ranged from 0.0060 for Stroke, the lowest (SE = 0.0005), to 0.0324 for CAD, the highest (SE = 0.0019) (Supplementary Fig. 1a and Supplementary Table 2a).We then conducted bivariate LDSC analysis to explore the r_g between eGFR and CVDs. This revealed significant negative r_g for two trait pairs: eGFR and Stroke (r_g = -0.0985, SE = 0.0309) and eGFR and VTE (r_g = -0.0783, SE = 0.0238), both surpassing the Bonferroni-corrected significance threshold ($P < 8.33 \times 10^{-3}$) (Supplementary Fig. 1b and Supplementary Table 2b). However, no significant r_g was found between eGFR and AF, CAD, HF, or PAD.

Polygenic overlap and local genetic correlation between eGFR and CVDs

Nevertheless, previously described r_g may underestimate the genetic overlap between eGFR and CVDs. The fact that there is no significant r_g does not necessarily mean that there is no common genetic component among these traits. Indeed, the r_g measure does not distinguish between mixtures of concordant and discordant genetic effects and a true absence of genetic overlap. To address this limitation and more comprehensively elucidate the shared genetic underpinnings of complex polygenic features, we employed advanced analytical tools, MiXeR and LAVA, specifically designed to provide more detailed insights into genetic architectures that traditional analyses might obscure.

We conducted MiXeR analyses to quantify the polygenic genetic overlap between eGFR and CVDs, accounting for mixed effect directions. Initially, univariate MiXeR analyses estimated that approximately 2,044 variants (SD = 111) influence eGFR. Among the CVDs, HF had the highest polygenicity, with 2,286 'causal' variants

497 accounting for 90% of HF h^2_{SNP} (SD = 215), followed by CAD with 1,431 variants (SD

498 = 311), and stroke with 1,039 variants (SD = 120) (Supplementary Table 3a).

499 Subsequently, bivariate analyses revealed significant polygenic overlap between eGFR

500 and these CVDs, with Dice coefficients ranging from 0.024 to 0.457 (Fig. 2a,

501 Supplementary Fig. 2, Supplementary Table 3b). Notably, eGFR displayed a moderate

502 genetic overlap with diseases like CAD (Dice = 0.457), HF (Dice = 0.39), and stroke

503 (Dice = 0.35). Among these, the eGFR-HF pair possessed the highest number of shared

504 'causal' variants (841, SD=151), accounting for substantial proportions of variants

505 affecting each condition (41.1% for eGFR and 36.8% for HF), suggesting highly

506 similar genetic architectures between these two traits. However, the genetic correlation

507 (4.78×10^{-4}, SD=9.83E-3) and the correlation coefficients of genetic risk (9.62×10^{-4},

508 SD=0.024) between eGFR and HF were notably weak. This extensive genetic overlap,

509 coupled with weak r_g, highlights the existence of mixed effects, as demonstrated by the

510 proportion of shared 'causal' variants with consistent effects (0.50, SD = 0.007). A

511 similar relationship was observed between eGFR-CAD and eGFR-Stroke, with the r_g

512 and r_gs (r_g = -0.098; r_gs = -0.269) of eGFR-Stroke consistent with their

513 LDSC-estimated significant negative genetic correlations. Compared to diseases with

514 lower polygenicity, such as AF, VTE, and PAD, eGFR is characterized as a highly

515 polygenic trait, leading to significant disparities in the number of shared and unique

516 'causal' variants. For instance, the eGFR-VTE pair demonstrated mild genetic overlap

517 (Dice = 0.177) with 211 shared variants (SD = 83), accounting for 10.3% of eGFR's and

518 66.2% of VTE's heritability, and exhibited the highest genetic risk correlation (r_g =

519 -0.115, SD = 0.007). Conversely, the eGFR-PAD pair had the fewest shared variants (28,

520 SD = 15), but the shared genetic components showed the strongest correlation (r_gs =

521 0.567, SD =0.351), indicating that while many eGFR-associated variants do not

522 influence PAD, those that do have highly similar effects (Supplementary Fig. 2).

523

524 LAVA, which calculates local-r_gs, was employed to explore further the relationships

525 between eGFR and CVDs, revealing mixed effect directions. First, univariate analyses

526 were conducted to identify heritable regions, filtering out loci devoid of univariate

527   heritability (Supplementary Table 4), leading to 425 bivariate tests. While no

528   significant genome-wide r_g emerged between eGFR and diseases such as AF, CAD, HF,

529   or PAD, LAVA identified specific regions with significant local-r_gs ($P < 0.05$, Fig.2b

530   and Supplementary Table 5). This was particularly notable between eGFR and CAD,

531   where 27 positively and 28 negatively correlated regions were found, suggesting that

532   the overall non-significant genome-wide correlation could result from

533   counterbalancing local effects. Similarly, mixed effects were observed with other

534   diseases: eGFR and AF exhibited 17 positive and 18 negative regions, HF showed 5

535   positive and 4 negative regions, and PAD had an equal number of positive and negative

536   regions (4 each). Contrary to these, eGFR and VTE displayed more negatively

537   correlated regions (12 positive / 23 negative), aligning with the negative genome-wide

538   correlations estimated by LDSC. However, the findings for eGFR and Stroke, which

539   presented more positively than negatively correlated regions, did not consistently align

540   with the negative r_g estimated by LDSC, though 5 negatively correlated regions were

541   still identified. After applying the Bonferroni correction, 36 genetic regions showed

542   significant local-r_gs. Notably, three genetic regions were linked to multiple trait pairs:

543   LD block 96 on chromosome 1 was associated with both eGFR-CAD and eGFR-VTE;

544   LD blocks 673 on chromosome 4 and 1057 on chromosome 6 exhibited correlations

545   between eGFR-AF and eGFR-CAD. Further analysis using HyPrColoc identified

546   shared causal variants, with SNP rs12740374 in LD block 96 showing strong

547   co-localization evidence of influencing eGFR-VTE and CAD with a PP greater than

548   0.7.

549

550   In summary, we systematically quantified the genetic overlap between eGFR and six

551   major CVDs using MiexR and LAVA, going beyond simple genome-wide r_g. These

552   analyses highlighted significant shared genetic bases between eGFR and the CVDs,

553   which were not apparent from traditional genome-wide studies alone. Building on these

554   findings, we performed further pleiotropy analysis to delve deeper into the molecular

555   mechanisms underlying these trait pairs.

556

Causal effects identified between eGFR and CVDs

Based on the above-mentioned common genetic basis between eGFR and CVDs, we implemented multiple pleiotropy methods to reveal the potential mechanisms of vertical or horizontal pleiotropy. Among them, in terms of vertical pleiotropy, we used LHC-MR technology to study causal relationships while controlling bias and confounding factors. Our analysis found that AF had a significant reverse causal effect on eGFR (OR: 0.806; 95% CI = 0.720 to 0.903), which suggests that the presence of AF may reduce eGFR. In addition, we observed another pair of significant negative causal relationships, indicating that an increase in eGFR would reduce the risk of VTE (OR = 0.931; 95% CI = 0.899 to 0.964) (Fig. 3, Supplementary Table 6a). These findings were confirmed by the results of the IVW method. In summary, our MR study provides strong evidence for the causal relationship between AF and eGFR and between eGFR and VTE (Supplementary Fig. 5, Supplementary Table 6b). However, it must be recognized that vertical pleiotropy alone cannot fully explain the common genetic basis between eGFR and CVDs.

Pleiotropic genomic loci identified between eGFR and CVDs

Given the extensive genome-wide genetic overlap, we further tested pleiotropic SNPs between eGFR and CVDs by PLACO, and a total of 20,073 SNPs ($P < 5 \times 10^{-8}$) were identified as significant pleiotropic variants. FUMA aggregated these SNPs into 508 loci, involving 226 unique chromosomal regions, of which 32 loci appeared in over 50% of trait pairs (Fig. 4, Supplementary Fig 6, Supplementary Table 7). Notably, loci such as 6q25.3 (SLC22A1), 8p23.1 (BLK), 9q34.2 (ABO), and 6p21.1 (*CRIP3*) showed overlap across all six trait pairs, while 12q24.12 (*ALDH2*) appeared in five. These broad pleiotropic effects are exemplified by the common variants at 6p21.1, which are associated with both polycystic kidney disease and stroke, and by 12q24.12, a locus significant for CAD and stroke. Interestingly, 24 pleiotropic loci had not been reported in previous eGFR and CVD studies, which reported 81 and 364 loci, respectively. Targeted analysis within the 508 pleiotropic loci revealed that 139 top SNPs (27.4%) indicated an increased risk for these traits, while 118 (23.2%) indicated a reduced risk.

587   The remaining 251 top SNPs (49.4%) showed discordant associations, highlighting

588   potential contrasting biological mechanisms. Functional annotation using FUMA

589   identified that 330 SNPs (65.0%) were intronic variants, 127 (25.0%) were intergenic,

590   and 21 (4.1%) were exonic, including 14 mRNA and 7 non-coding RNA exon variants.

591   Moreover, 40 SNPs in these risk loci, including rs34312154 with the highest CADD

592   score of 31, may be potentially harmful, affecting genes like *HOXB1* to *HOXB8*.

593   Additionally, 58 SNPs identified by RegulomeDB as potentially affecting transcription

594   factor binding underscore the regulatory significance of these findings, with rs4918003

595   showing the most potent evidence of regulatory potential (RDB: 1b).

596

597   Further analysis of co-localization at 508 identified loci revealed that 74 loci (14.6%)

598   exhibit high PP (PPH4 > 0.7), indicating strong evidence of shared genetic signals (Fig.

599   4, Supplementary Fig. 7, Supplementary Table 7). Of these, 71 top SNPs were

600   identified as candidate common pathogenic variants. Notably, the 1p13.3 locus showed

601   significant pleiotropy, influencing the trait pairs of eGFR-CAD, eGFR-VTE, and

602   eGFR-HF, with PP.H4 values ranging from 0.956 to 0.998. Subsequent HyPrColoc

603   analysis identified rs660240 in the CELSR2 gene at this locus as exhibiting strong

604   co-localization for both eGFR and HF and eGFR and VTE, with PP exceeding 0.7.

605

606   **Identification of pleiotropic genes between eGFR and CVDs**

607   We used MAGMA analysis to convert the SNP level signals of identified eGFR-CVDs

608   into gene level signals, focusing on potential pleiotropy within or overlapping 508

609   pleiotropic loci. This analysis identified 802 significant pleiotropic genes, with 579

610   being unique. Notably, 379 genes were implicated in more than one trait pair, providing

611   robust statistical evidence for shared genetic inheritance across diverse traits (Fig. 5,

612   Supplementary Table 11). Specific genes such as *ATXN2* (located at 12q24.12)

613   significantly impacted five trait pairs. Additionally, genes including *ABO* (9q34.2),

614   *BRAP* (12q24.12), *SH2B3* (12q24.12), *L3MBTL3* (6q23.1), *MMP24* (19q13.32),

615   *MRPS21* (1q21.2), and *PRPF3* (1q21.2) were significant across four trait pairs,

616   underscoring their pivotal roles in these genetic networks. Remarkably, the 12q24.12

617   locus, home to *ATXN2*, *BRAP*, and *SH2B3*, is critical in assessing genetic risks for

618   hypertension[38] and CKD, with *ATXN2* linked to myocardial infarction, renal function,

619   and hypertension.[39] Among the genes identified, 118 were not previously explored in

620   eGFR studies, and 410 were new to CVD research, highlighting significant novel

621   findings. Positional mapping in FUMA confirmed 97.88% of these risk genes identified

622   by MAGMA, further solidifying the evidence for a shared genetic framework across

623   these trait pairs (Supplementary Table 9).

624

625   We employed the LDSC-SEG method to perform tissue-specific genetic analysis,

626   identifying distinct tissue associations for each trait. The analysis revealed significant

627   gene expression enrichment for eGFR in the renal cortex and liver tissues. Similarly,

628   AF showed considerable enrichment in the atrial appendage and cardiac left ventricular

629   tissues, while CAD was significantly enriched in the aorta, coronary artery, and tibial

630   artery tissues, with all exceeding an FDR threshold of 0.05. Subsequent LDSC-SEG

631   chromatin analysis corroborated these findings, enhancing the resolution by detecting

632   significant tissue chromatin signals associated with eGFR, AF, and CAD traits

633   (Supplementary Fig. 3, Supplementary Table 14).

634

635   We conducted E-MAGMA analysis on ten related tissues using tissue-specific eQTL

636   information, which identified 1,276 pleiotropic genes that remained significant after

637   Bonferroni correction, showing strong enrichment in at least one of the analyzed tissues

638   (Supplementary Table 15). Particularly noteworthy, 26 genes were prevalent in over 50%

639   of trait pairs, with *KAT5* and *PROCR* appearing in five trait pairs each and *BLK* and

640   *SLC22A1* in four. *KAT5*, or lysine acetyltransferase 5, serves a dual role; it regulates the

641   level of acetylation of target proteins, affecting the expression of inflammation related

642   genes and the intensity of inflammatory responses. This regulatory function is crucial,

643   as abnormal expression or dysfunction of *KAT5* can elevate cardiovascular disease risks

644   by affecting the inflammatory pathway.[40] Additionally, *KAT5* acts as a DNA repair

645   factor, vital for preventing ischemic acute kidney injury (AKI), primarily by regulating

646   glomerular filtration.[41] Therefore, the normal function and activity of *KAT5* are critical

647   in preventing or alleviating AKI and indirectly maintaining or enhancing eGFR levels.

648   We also performed the TWAS to assess the tissue specificity of pleiotropic genes for

649   each trait, leading to the discovery of 354 new genes associated with eGFR and 142

650   new genes related to various CVDs (Supplementary Table 16). Furthering our

651   investigation, we utilized FUMA for eQTL mapping, identifying 2,049 genes, of which

652   71.40% were confirmed by our analysis to be associated with the traits under study

653   (Supplementary Table 9).

654

655   MAGMA and E-MAGMA analyses have jointly identified 424 pleiotropic genes,

656   including 314 unique ones, shedding light on the common mechanisms underlying

657   various trait pairs (Supplementary Table 11). Notably, 76 of these genes recur across

658   multiple trait pairs, with genes such as *MMP24*, *L3MBTL3*, and *ABO* being prominent

659   in half or more of the analyzed traits. Specific genes like *ABO* (9q34.13), *PROCR*

660   (20q11.22), *BLK* (8p23.1), *ALDH2* (12q21.31), and *SLC22A1* (6q25.3) each featured in

661   three different trait pairs, highlighting their significant roles in multiple pathologies.

662   For instance, *L3MBTL3* has been implicated in renal dysfunction and cardiovascular

663   pathogenesis, illustrating the interconnectedness of renal and cardiovascular systems.

664   *MMP24*'s dysregulated expression, associated with renal tubular atrophy and diabetic

665   nephropathy (DN), may contribute to structural and functional kidney impairments that

666   exacerbate cardiovascular risk. In addition, the *ABO* gene, which is associated with

667   changes in von Willebrand factor (VWF) plasma levels, plays a key role in regulating

668   platelet function and atherosclerosis, thus affecting the severity of CAD. While the

669   precise mechanisms connecting ABO genes to eGFR remain elusive, their association

670   with conditions like type 2 diabetes (T2D) and hypertension, which are main risk

671   factors for CKD, likely impacts eGFR.[42]

672

673   **Shared biological mechanisms between eGFR and CVDs**

674   We employed the MAGMA algorithm to assess pathway enrichment among the

675 identified genes, setting a stringent Bonferroni-corrected threshold ($P < 8.87 \times 10^{-7}$).

676 This analysis revealed 78 significantly enriched pathways, comprising 73 Gene

677 Ontology Biological Process (GO BP) terms and 5 Reactome gene set pathways,

678 indicative of key biological processes (Supplementary Table 17a). Notably, pathways

679 such as the positive regulation of macromolecule biosynthetic processes, transcription

680 by RNA polymerase II, and RNA metabolic processes were enriched in half or more of

681 the trait pairs, highlighting their crucial roles in the regulation of core biological

682 processes that contribute to the development of the studied traits.

683

684 Further analysis of pathway enrichment for overlapping genes, identified in multiple

685 trait pairs from MAGMA and E-MAGMA analyses, was conducted using the

686 Metascape website (Supplementary Table 17b). This analysis pinpointed 16

687 significantly enriched pathways predominantly involved in regulating biosynthetic and

688 metabolic processes, lipid metabolism, and receptor signal transduction. The pathways

689 related to glutathione breakdown and stem cell maintenance were critical and had

690 profound implications for eGFR and CVDs. The antioxidant and anti-inflammatory

691 properties of glutathione are essential for cardiovascular protection. However, its

692 degradation may elevate the risk of CVDs through mechanisms linked to renal

693 insufficiency. Moreover, the protective roles of endothelial progenitor cells (EPC) and

694 stem cell factor (SCF) within the kidney are critical for maintaining eGFR and CVDs

695 health, implying that diminished activity in these pathways could adversely affect renal

696 and cardiovascular systems.

697

698 **Proteome-wide Mendelian Randomization analysis for eGFR and CVDs**

699 We conducted a SMR analysis utilizing GWAS summary data for eGFR, six major

700 CVDs, and plasma protein pQTLs. This analysis identified 78 risk proteins, 32 linked

701 to eGFR and 46 to CVDs, validated through HEIDI evaluation ($P > 0.01$), multiple

702 SNP-SMR sensitivity analysis ($P < 0.05$), and Bonferroni corrections ($P_{SMR} < 0.05$ /

703 (1958×7)). Subsequently, we investigated potential pleiotropic effects within specific

704 eGFR-CVD pairs, identifying three signature pairs that contain pleiotropic proteins

705   (Supplementary Table 18). Notably, CELSR2 was common between eGFR-CAD and

706   eGFR-VTE, while FGF5 linked eGFR-AF and eGFR-CAD. Additionally, a

707   colocalization analysis of these 78 proteins sought to identify common pathogenic

708   variants associated with specific eGFR or CVDs signatures. This analysis showed

709   strong evidence of colocalization for 49 proteins (41 unique), with 17 proteins

710   associated with eGFR and 32 with CVDs. Proteins such as CELSR2 and FGF5,

711   emphasized in the SMR analysis, displayed strong evidence of colocalization across

712   three identified signature pairs, underscoring a shared pathogenic pathway across these

713   conditions.

714

715   **Discussion**

716   In this study, a comprehensive genome-wide pleiotropy analysis was carried out to

717   investigate the genetic correlation and overlap between eGFR and CVDs, uncovering a

718   complex shared genetic foundation. Our detailed investigations identified key

719   pleiotropic loci and genes, including *L3MBTL3*, *MMP24*, and *ABO*, and illuminated

720   crucial biological pathways such as stem cell population maintenance and glutathione

721   metabolism. Additionally, using MR analysis, we indicated a causal relationship

722   between eGFR and VTE, as well as between AF and eGFR. Collectively, the results

723   provide a better understanding of the shared genetic etiology and biological

724   mechanisms that underlie the relationship between eGFR and CVDs, opening new

725   avenues for targeted therapeutic strategies and further research.

726

727   Our analysis probed the shared genetic architectures of eGFR and six major CVDs

728   using complementary genetic tools—LDSC, MiXeR, and LAVA—to reveal extensive

729   shared genetic bases across all trait pairs. LDSC highlighted significant negative global

730   r_g between eGFR and VTE and Stroke, contrasting with previous studies that reported

731   no significant associations. This underscores the advantage of our approach and the

732   benefits of a larger sample size. Despite the absence of significant global correlations

733   for certain traits, MiXeR identified substantial overlap for CAD, HF, and AF with

734   eGFR, indicating mixed effect directions even when global correlations are not evident.

735   LAVA provided further insights into local-r_gs, revealing mixed effect directions not

736   apparent from genome-wide estimates, such as those observed between eGFR and

737   CAD, which exhibited both positively and negatively correlated loci. This suggests a

738   more nuanced genetic relationship than previously recognized. Moreover, employing

739   bivariate MiXeR revealed extensive genetic overlap between eGFR and CAD,

740   supported by LAVA's local-r_gs, highlighting concordant and discordant effect

741   directions. Our comprehensive analysis confirms the shared genetic foundations

742   between eGFR and various CVDs, indicating multiple common genetic variants

743   influencing critical biological pathways.

744

745   Our findings elucidate the shared genetic bases between eGFR and CVDs, attributable

746   to both horizontal pleiotropy and direct causal mechanisms (vertical pleiotropy). Our

747   MR analysis effectively separated potential causal effects from other genetic influences.

748   Notably, the LHC-MR results identified a reverse causal relationship of eGFR on VTE,

749   a finding supported by Yuan et al.[43] This causal relationship is likely driven by impaired

750   kidney function, which increases coagulation factors and decreases endogenous

751   anticoagulants, collectively promoting thrombus formation. Additionally, our analysis

752   indicates a potential causal influence of AF on eGFR, aligning with previous studies

753   that suggest AF can lead to renal impairments via mechanisms such as

754   thromboembolism and reduced renal microvascular blood flow.[12,44] However, our

755   analysis did not establish causal relationships for other trait pairs, suggesting that while

756   genetic overlaps between eGFR and CVDs are partially driven by causality, the

757   complexity of these relationships varies across different conditions.

758

759   Our cross-trait meta-analysis has identified many genome-wide significant pleiotropic

760   loci associated with eGFR and CVDs. Notably, loci such as 8p23.1 (*BLK*) and 6q25.3

761   (*SLC22A1*) demonstrated significance across six trait pairs, while others like 12q24.12

762   (*ALDH2*) and 20q11.22 (*PROCR*) were significant in over 50% of the trait pairs.

763   Specifically, the BLK gene at 8p23.1 enhances insulin synthesis and secretion, linking

764   it to CKD symptoms and complications in nephropathy and diabetes by increasing the

765   production of pro-inflammatory cytokines. BLK is also identified as a susceptibility

766   gene for Kawasaki disease,[45] a systemic vasculitis that can impact coronary arteries and

767   increase CVDs risks.[46] Furthermore, the locus *SLC22A1* at 6q25.3 is associated with

768   DN, a severe complication of diabetes characterized by persistent albuminuria,

769   declining eGFR, and elevated arterial blood pressure, which elevates the risk for CVDs

770   and cerebrovascular diseases.[47] The role of this locus in hypertension underscores its

771   potential as a mediator of underlying disease risks. The consistency of effect alleles in

772   nearly half of the significant SNPs for both eGFR and CVD supports a positive

773   association between these conditions and provides insights into their complex r_g.

774

775   Our analysis has identified several key pleiotropic genes in over 50% of the trait pairs,

776   including *L3MBTL3* (6q23.1), *MMP24* (20q11.22), and *ABO* (9q34.2). For example,

777   *L3MBTL3* significantly influences the risk of coronary heart disease (CHD) by

778   modulating adipogenesis, the process of fat cell differentiation. Dysfunctional adipose

779   tissue can lead to ectopic fat accumulation, which promotes insulin resistance and other

780   metabolic disorders, escalating the risk for obesity-related CHD. These metabolic

781   disturbances further amplify the risk of ischemic stroke by intensifying common risk

782   factors (such as hypertension and diabetes). Although *L3MBTL3*'s expression in the

783   kidney is documented, the specific mechanisms by which it affects renal function and

784   structure remain to be fully explored. Matrix metalloproteinase-24 (*MMP24*) plays a

785   key role in extracellular matrix degradation, and is linked with renal tubular atrophy

786   and the severity of alcohol-induced renal injury and fibrosis. Dysregulated *MMP24*

787   expression detrimentally impacts both renal and cardiovascular health. In DN, *MMP24*

788   is implicated in abnormal extracellular matrix accumulation, leading to significant

789   damage to the glomeruli and renal tubules and indirectly impacting eGFR.[48] Given the

790   close association of DN with T2D, a prominent risk factor for CVDs, *MMP24*'s role in

791   DN could indirectly heighten the risk of CVDs.[49] The *ABO* gene, vital for determining

792   blood group phenotypes, also affects crucial hemostatic factors such as VWF and factor

793   VIII (FVIII), which are essential for platelet adhesion at injury sites. Individuals with

794   non-O blood types generally exhibit higher VWF levels and are more susceptible to

795   VTE.[50] In addition, researchers have discovered a new connection between *ABO* and

796   primary glomerulonephritis, especially IgA nephropathy, where certain blood groups

797   are linked to an increased risk of renal function decline.[51] These findings highlight the

798   complex interaction of genetics with the pathophysiology of renal and CVDs, which

799   provides a strong basis for further study of the mechanisms and risks associated with

800   genetic diseases.

801

802   Our study also examines the role of shared genetic determinants in two critical

803   biological pathways, including stem cell population maintenance and glutathione

804   catabolism, and their implications for kidney function and cardiovascular health. In the

805   stem cell maintenance pathway, *MMP24*[52], *PSRC1*[53], and *CHMP1A*[54] play pivotal roles

806   by regulating the stem cell cycle, ensuring cellular quiescence, and preventing

807   premature differentiation, respectively. These molecular actions are crucial for

808   promoting tubular cell growth and enhancing renal function, mainly through the

809   secretion of regenerative factors such as bone morphogenetic protein-7 and hepatocyte

810   growth factor.[55] The effectiveness of stem cell therapy in heart repair, particularly in

811   those with elevated levels of inflammation, further underscores the importance of

812   robust stem cells for optimal cardiovascular and renal outcomes. In the glutathione

813   pathway, *DPEP1*[56], *ORMDL3*, *MRPS21*, and *CPS1*[57,58] are crucial for maintaining

814   cellular redox balance and defending against oxidative stress, and they are vital for

815   stroke prevention and kidney function maintenance.[59,60] Variations in kidney function,

816   reflected by changes in eGFR, can significantly affect glutathione turnover and its

817   protective roles, directly linking renal and cardiovascular health.[61] Our findings suggest

818   that targeted modulation of these key genes could significantly enhance kidney and

819   heart repair and regeneration, thus improving eGFR and decreasing the risk of CVDs.

820

821   Our study has some limitations. Firstly, the variation in sample sizes across disorders,

822   from 511,634 cases for PAD to 1,500,861 for VTE, restricted our ability to detect

823   pleiotropic effects in less-represented diseases. Future studies with larger, more

824   balanced samples will likely improve our capacity to discern these effects more

825 precisely. Additionally, including subjects with concurrent eGFR abnormalities and

826 CVDs may have introduced bias in our assessment of genetic overlap between these

827 conditions. Our analysis focused on shared genetic variants, excluding rare mutations

828 that could further elucidate the complex genetic relationships involved. Moreover, the

829 concentration on individuals of European ancestry limits the generalizability of our

830 findings to other populations. It is crucial to conduct cross-ancestry follow-up studies to

831 determine whether these results hold across diverse ethnic groups. Overcoming these

832 limitations will require enhancing the diversity of GWAS populations and conducting

833 more extensive experimental validations and cohort analyses.

834

835 **Conclusion**

836 Together, this study has significantly improved our understanding of the genetic basis

837 for the role of eGFR in susceptibility to major CVDs. By identifying crucial genetic

838 components, including key pleiotropic loci and genes such as *L3MBTL3*, *MMP24*, and

839 *ABO*, we elucidate potential shared mechanisms that may influence critical biological

840 pathways associated with CVDs, notably stem cell population maintenance and

841 glutathione metabolism. These results provide a better understanding of the genetic

842 basis of CVDs and provide a basis for the development of more targeted therapeutic

843 and preventive strategies that could potentially transform cardiovascular care.

844

845 **Disclosure**

846 All authors declare no competing interests.

847

848 **Data Statement**

849 The study used only openly available GWAS summary statistics on estimated

850 glomerular filtration rate and six major cardiovascular diseases that have originally

851 been conducted using human data. GWAS summary statistics on eGFR are available at

852   https://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/

853   genetische-epidemiologie/gwas-summary-statistics/index.html. GWAS summary

854   statistics on AF, HF, and Stroke are available at the GWAS Catalog (GCST90104539,

855   GCST009541, and GCST90104539). GWAS summary statistics on CAD and PAD are

856   publicly available for download at the Cardiovascular Disease Knowledge Portal

857   (CVDKP) website: https://cvd.hugeamp.org/datasets.html. GWAS summary statistics

858   on VTE are obtained from the deCODE genetics website: https://www.decode.com/

859   summarydata/. Blood-based cis-pQTL from UKB-PPP are obtained from

860   https://www.synapse.org/Synapse:syn51365303.

861

874   manuscript for publication.

875

876   **Author contributions**

877   J.Q., Y. F., Y.Y., J.X., and S.P. conceptualized and supervised this project and wrote the

878   manuscript. J.Q., K.Y., and Y.Y. performed the main analyses and wrote the manuscript.

879   J.Q., X.Y., Y.Y., L.Z., and Y.L. performed the statistical analysis and assisted with

880   interpreting the results. M.C., W.X., and Y.Y. provided expertise in cardiovascular

881   biology and GWAS summary statistics. All authors provided intellectual content and

882   approved the final version of the manuscript.

883

884 **Reference**

885 1.   Roth, G. A. et al. Global, Regional, and National Burden of Cardiovascular
886   Diseases for 10 Causes, 1990 to 2015. Journal of the American College of
887   Cardiology 70, 1-25, doi:10.1016/j.jacc.2017.04.052 (2017).

888 2.   Bays, H. E. et al. Ten things to know about ten cardiovascular disease risk
889   factors. American journal of preventive cardiology 5, 100149,
890   doi:10.1016/j.ajpc.2021.100149 (2021).

891 3.   Popolo, A., Autore, G., Pinto, A. & Marzocco, S. Oxidative stress in patients
892   with cardiovascular disease and chronic renal failure. Free radical research 47,
893   346-356, doi:10.3109/10715762.2013.779373 (2013).

894 4.   Ammirati, A. L. Chronic Kidney Disease. Revista da Associacao Medica
895   Brasileira (1992) 66Suppl 1, s03-s09, doi:10.1590/1806-9282.66.S1.3 (2020).

896 5.   Eckardt, K. U. et al. Evolving importance of kidney disease: from subspecialty
897   to global health burden. Lancet (London, England) 382, 158-169,
898   doi:10.1016/s0140-6736(13)60439-0 (2013).

899 6.   Visseren, F. L. J. et al. 2021 ESC Guidelines on cardiovascular disease
900   prevention in clinical practice. European journal of preventive cardiology 29,
901   5-115, doi:10.1093/eurjpc/zwab154 (2022).

902 7.   Lees, J. S. et al. Glomerular filtration rate by differing measures, albuminuria
903   and prediction of cardiovascular disease, mortality and end-stage kidney
904   disease. Nat Med 25, 1753-1760, doi:10.1038/s41591-019-0627-8 (2019).

905 8.   Manjunath, G. et al. Level of kidney function as a risk factor for cardiovascular
906   outcomes in the elderly. Kidney international 63, 1121-1129,
907   doi:10.1046/j.1523-1755.2003.00838.x (2003).

908 9.   Ataklte, F. et al. Association of Mildly Reduced Kidney Function With
909   Cardiovascular Disease: The Framingham Heart Study. Journal of the American
910   Heart Association 10, e020301, doi:10.1161/jaha.120.020301 (2021).

911 10. Graham, S. E. et al. Sex-specific and pleiotropic effects underlying kidney
912   function identified from GWAS meta-analysis. Nature communications 10,
913   1847, doi:10.1038/s41467-019-09861-z (2019).

11.   Kelly, D. M. et al. Interplay Between Chronic Kidney Disease, Hypertension, and Stroke: Insights From a Multivariable Mendelian Randomization Analysis. Neurology 101, e1960-e1969, doi:10.1212/wnl.0000000000207852 (2023).

12.   Geurts, S. et al. Disentangling the association between kidney function and atrial fibrillation: a bidirectional Mendelian randomization study. International journal of cardiology 355, 15-22, doi:10.1016/j.ijcard.2022.03.004 (2022).

13.   Park, S. et al. Atrial fibrillation and kidney function: a bidirectional Mendelian randomization study. Eur Heart J 42, 2816-2823, doi:10.1093/eurheartj/ehab291 (2021).

14.   Gong, W. et al. Role of the Gut-Brain Axis in the Shared Genetic Etiology Between Gastrointestinal Tract Diseases and Psychiatric Disorders: A Genome-Wide Pleiotropic Analysis. JAMA psychiatry 80, 360-370, doi:10.1001/jamapsychiatry.2022.4974 (2023).

15.   Stanzick, K. J. et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. Nature communications 12, 4350, doi:10.1038/s41467-021-24491-0 (2021).

16.   Nielsen, J. B. et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nature genetics 50, 1234-1239, doi:10.1038/s41588-018-0171-3 (2018).

17.   Aragam, K. G. et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. Nature genetics 54, 1803-1815, doi:10.1038/s41588-022-01233-6 (2022).

18.   Ghouse, J. et al. Genome-wide meta-analysis identifies 93 risk loci and enables risk prediction equivalent to monogenic forms of venous thromboembolism. Nature genetics 55, 399-409, doi:10.1038/s41588-022-01286-7 (2023).

19.   Shah, S. et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. Nature communications 11, 163, doi:10.1038/s41467-019-13690-5 (2020).

20.   van Zuydam, N. R. et al. Genome-Wide Association Study of Peripheral Artery Disease. Circulation. Genomic and precision medicine 14, e002862,

944   doi:10.1161/circgen.119.002862 (2021).

945 21. Mishra, A. et al. Stroke genetics informs drug discovery and risk prediction

946   across ancestries. Nature 611, 115-123, doi:10.1038/s41586-022-05165-3

947   (2022).

948 22. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding

949   from polygenicity in genome-wide association studies. Nature genetics 47,

950   291-295, doi:10.1038/ng.3211 (2015).

951 23. Hindley, G. et al. Charting the Landscape of Genetic Overlap Between Mental

952   Disorders and Related Traits Beyond Genetic Correlation. The American

953   journal of psychiatry 179, 833-843, doi:10.1176/appi.ajp.21101051 (2022).

954 24. Werme, J., van der Sluis, S., Posthuma, D. & de Leeuw, C. A. An integrated

955   framework for local genetic correlation analysis. Nature genetics 54, 274-282,

956   doi:10.1038/s41588-022-01017-y (2022).

957 25. Darrous, L., Mounier, N. & Kutalik, Z. Simultaneous estimation of

958   bi-directional causal effects and heritable confounding from GWAS summary

959   statistics. Nature communications 12, 7274, doi:10.1038/s41467-021-26970-w

960   (2021).

961 26. Ray, D. & Chatterjee, N. A powerful method for pleiotropic analysis under

962   composite null hypothesis identifies novel shared loci between Type 2 Diabetes

963   and Prostate Cancer. PLoS genetics 16, e1009218,

964   doi:10.1371/journal.pgen.1009218 (2020).

965 27. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional

966   mapping and annotation of genetic associations with FUMA. Nature

967   communications 8, 1826, doi:10.1038/s41467-017-01261-5 (2017).

968 28. Kircher, M. et al. A general framework for estimating the relative pathogenicity

969   of human genetic variants. Nature genetics 46, 310-315, doi:10.1038/ng.2892

970   (2014).

971 29. Boyle, A. P. et al. Annotation of functional variation in personal genomes using

972   RegulomeDB. Genome research 22, 1790-1797, doi:10.1101/gr.137323.112

973   (2012).

974   30.   Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes.
975   Nature 518, 317-330, doi:10.1038/nature14248 (2015).

976   31.   Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies
977   predicts complex trait gene targets. Nature genetics 48, 481-487,
978   doi:10.1038/ng.3538 (2016).

979   32.   Wallace, C. A more accurate method for colocalisation analysis allowing for
980   multiple causal variants. PLoS genetics 17, e1009440,
981   doi:10.1371/journal.pgen.1009440 (2021).

982   33.   van der Linden, R. J. et al. Genetic overlap between Alzheimer's disease and
983   blood lipid levels. Neurobiology of aging 108, 189-195,
984   doi:10.1016/j.neurobiolaging.2021.06.019 (2021).

985   34.   de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA:
986   generalized gene-set analysis of GWAS data. PLoS computational biology 11,
987   e1004219, doi:10.1371/journal.pcbi.1004219 (2015).

988   35.   Gerring, Z. F., Mina-Vargas, A., Gamazon, E. R. & Derks, E. M. E-MAGMA:
989   an eQTL-informed method to identify risk genes using genome-wide
990   association study summary statistics. Bioinformatics (Oxford, England) 37,
991   2245-2249, doi:10.1093/bioinformatics/btab115 (2021).

992   36.   Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide
993   association studies. Nature genetics 48, 245-252, doi:10.1038/ng.3506 (2016).

994   37.   Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis
995   of systems-level datasets. Nature communications 10, 1523,
996   doi:10.1038/s41467-019-09234-6 (2019).

997   38.   Levy, D. et al. Genome-wide association study of blood pressure and
998   hypertension. Nature genetics 41, 677-687, doi:10.1038/ng.384 (2009).

999   39.   Yamada, Y. et al. Identification of 13 novel susceptibility loci for early-onset
1000   myocardial infarction, hypertension, or chronic kidney disease. International
1001   journal of molecular medicine 42, 2415-2436, doi:10.3892/ijmm.2018.3852
1002   (2018).

1003 40.   Liu, C. et al. STUB1 is acetylated by KAT5 and alleviates myocardial

1004   ischemia-reperfusion injury through LATS2-YAP-β-catenin axis.
1005   Communications biology 7, 396, doi:10.1038/s42003-024-06086-9 (2024).

1006 41.   Hishikawa, A. et al. DNA repair factor KAT5 prevents ischemic acute kidney
1007   injury through glomerular filtration regulation. iScience 24, 103436,
1008   doi:10.1016/j.isci.2021.103436 (2021).

1009 42.   Cano, E. A. et al. Association between ABO Blood Groups and Type 2 Diabetes
1010   Mellitus: A Meta-Analysis. Curr Diabetes Rev 19, e270422204139,
1011   doi:10.2174/1573399818666220427124448 (2023).

1012 43.   Yuan, S., Bruzelius, M. & Larsson, S. C. Causal effect of renal function on
1013   venous thromboembolism: a two-sample Mendelian randomization
1014   investigation. Journal of thrombosis and thrombolysis 53, 43-50,
1015   doi:10.1007/s11239-021-02494-4 (2022).

1016 44.   Watanabe, H. et al. Close bidirectional relationship between chronic kidney
1017   disease and atrial fibrillation: the Niigata preventive medicine study. American
1018   heart journal 158, 629-636, doi:10.1016/j.ahj.2009.06.031 (2009).

1019 45.   Kim, J. J. et al. Identification of B-cell-related HSPG2 and CDSN as
1020   susceptibility loci for Kawasaki disease. Human immunology 84, 567-570,
1021   doi:10.1016/j.humimm.2023.07.001 (2023).

1022 46.   Newburger, J. W. et al. Diagnosis, treatment, and long-term management of
1023   Kawasaki disease: a statement for health professionals from the Committee on
1024   Rheumatic Fever, Endocarditis and Kawasaki Disease, Council on
1025   Cardiovascular Disease in the Young, American Heart Association. Circulation
1026   110, 2747-2771, doi:10.1161/01.Cir.0000145143.19711.78 (2004).

1027 47.   Sallinen, R. et al. Association of the SLC22A1, SLC22A2, and SLC22A3 genes
1028   encoding organic cation transporters with diabetic nephropathy and
1029   hypertension. Annals of medicine 42, 296-304,
1030   doi:10.3109/07853891003777109 (2010).

1031 48.   Garcia-Fernandez, N. et al. Matrix Metalloproteinases in Diabetic Kidney
1032   Disease. Journal of clinical medicine 9, doi:10.3390/jcm9020472 (2020).

1033 49.   Chen, S., Chen, L. & Jiang, H. Prognosis and risk factors of chronic kidney

1034   disease progression in patients with diabetic kidney disease and non-diabetic

1035   kidney disease: a prospective cohort CKD-ROUTE study. Renal failure 44,

1036   1309-1318, doi:10.1080/0886022x.2022.2106872 (2022).

1037 50. Zhang, H., Mooney, C. J. & Reilly, M. P. ABO Blood Groups and

1038   Cardiovascular Diseases. Int J Vasc Med 2012, 641917,

1039   doi:10.1155/2012/641917 (2012).

1040 51. Wang, M. et al. Associations of ABO blood type and galactose-deficient

1041   immunoglobulin A1 with adverse outcomes in patients with IgA nephropathy.

1042   Nephrology, dialysis, transplantation : official publication of the European

1043   Dialysis and Transplant Association - European Renal Association 36, 288-294,

1044   doi:10.1093/ndt/gfz171 (2021).

1045 52. Porlan, E. et al. MT5-MMP regulates adult neural stem cell functional

1046   quiescence through the cleavage of N-cadherin. Nature cell biology 16, 629-638,

1047   doi:10.1038/ncb2993 (2014).

1048 53. Wilson, N. K. et al. Combined Single-Cell Functional and Gene Expression

1049   Analysis Resolves Heterogeneity within Stem Cell Populations. Cell stem cell

1050   16, 712-724, doi:10.1016/j.stem.2015.04.004 (2015).

1051 54. Coulter, M. E. et al. The ESCRT-III Protein CHMP1A Mediates Secretion of

1052   Sonic Hedgehog on a Distinctive Subtype of Extracellular Vesicles. Cell reports

1053   24, 973-986.e978, doi:10.1016/j.celrep.2018.06.100 (2018).

1054 55. Hopkins, C., Li, J., Rae, F. & Little, M. H. Stem cell options for kidney disease.

1055   J Pathol 217, 265-281, doi:10.1002/path.2477 (2009).

1056 56. Strauer, B. E. & Steinhoff, G. 10 years of intracoronary and intramyocardial

1057   bone marrow stem cell therapy of the heart: from the methodological origin to

1058   clinical practice. Journal of the American College of Cardiology 58, 1095-1104,

1059   doi:10.1016/j.jacc.2011.06.016 (2011).

1060 57. Lau, A. et al. Dipeptidase-1 governs renal inflammation during ischemia

1061   reperfusion injury. Science advances 8, eabm0142,

1062   doi:10.1126/sciadv.abm0142 (2022).

1063 58. Wu, J. et al. Regulation of the urea cycle by CPS1 O-GlcNAcylation in response

1064   to dietary restriction and aging. Journal of molecular cell biology 14,

1065   doi:10.1093/jmcb/mjac016 (2022).

1066 59.   Matuz-Mares, D., Riveros-Rosas, H., Vilchis-Landeros, M. M. & Vá

1067   zquez-Meza, H. Glutathione Participation in the Prevention of Cardiovascular

1068   Diseases. Antioxidants (Basel, Switzerland) 10, doi:10.3390/antiox10081220

1069   (2021).

1070 60.   Bajic, V. P. et al. Glutathione "Redox Homeostasis" and Its Relation to

1071   Cardiovascular Disease. Oxidative medicine and cellular longevity 2019,

1072   5028181, doi:10.1155/2019/5028181 (2019).

1073 61.   Hinchman, C. A. & Ballatori, N. Glutathione-degrading capacities of liver and

1074   kidney in different species. Biochemical pharmacology 40, 1131-1135,

1075   doi:10.1016/0006-2952(90)90503-d (1990).

1076
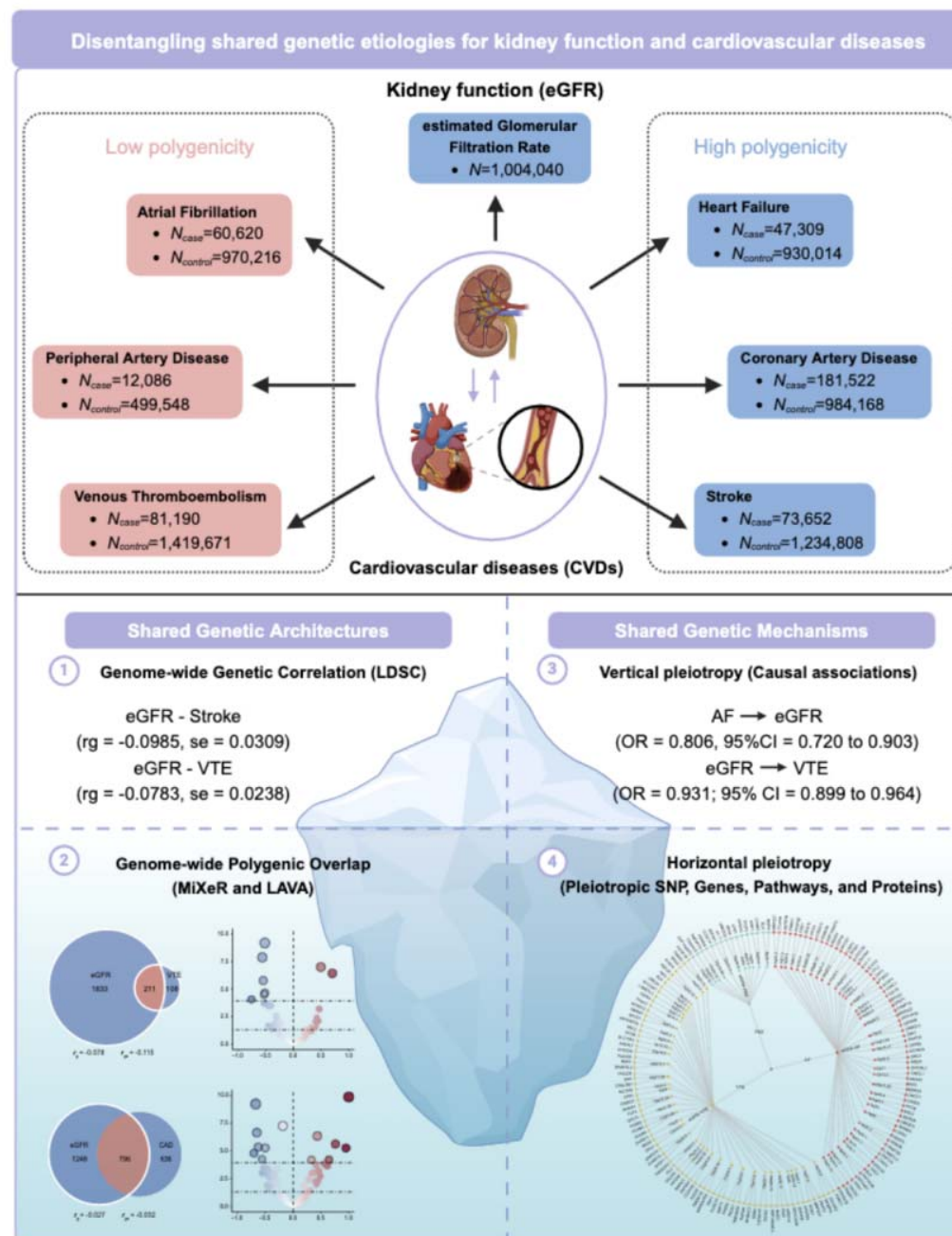
1077

Figure 1: Schematic diagram of the analysis of genetic associations between glomerular filtration rate and six major cardiovascular diseases in this study.

We analyzed the most comprehensive GWAS summary statistics to investigate the shared genetic architecture and potential mechanisms between eGFR and CVDs. First, the shared genetic basis was determined by quantifying global and local genetic correlations and exploring the global genetic overlap. Then, their two genetic pleiotropy were analyzed to investigate the shared genetic

1084   architecture, which can be vertical pleiotropy, that is, the effect of the variant on one trait affects

1085   another trait or horizontal pleiotropy, that is, the variant affects multiple traits independently. We

1086   first used Mendelian randomization on vertical pleiotropy to illustrate their causal relationship.

1087   Then, various statistical genetic methods were used on horizontal pleiotropy to sequentially apply

1088   pleiotropy analysis at the SNP, gene level, biological pathway, and protein target level to investigate

1089   the common genetic mechanism. Through this comprehensive genetic pleiotropy analysis, our

1090   understanding of the genetic link between eGFR and CVDs has been dramatically enhanced, which

1091   may guide the development of new treatment strategies and improve the clinical management of
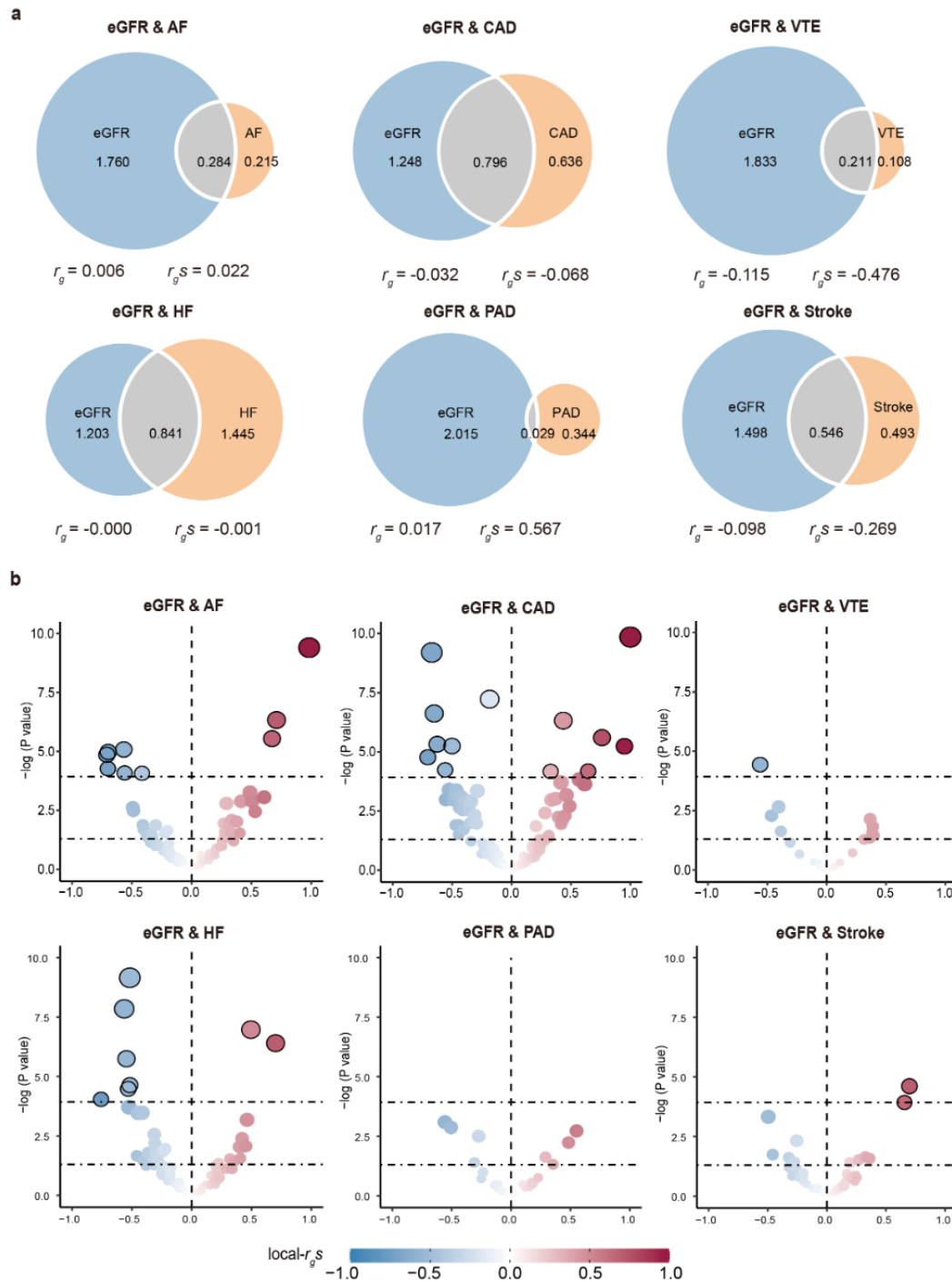
1092   these diseases.

Figure 2: Genetic overlap between estimated glomerular filtration rate and six major cardiovascular diseases exceeds genome-wide genetic correlation.

Genetic overlap and local genetic correlation between eGFR and six major CVDs investigated by MiXeR and LAVA. a: MiXeR Venn diagrams showing common and unique effect trait variants

1098   showing polygenic overlap (grey) between eGFR (blue) and CVDs (orange). Numbers in the Venn

1099   diagrams indicate the estimated number of common and unique effect trait variants (in thousands)

1100   that explain 90% of the SNP heritability and standard error in each phenotype. Circle size indicates

1101   the degree of polygenicity for each trait, with larger circles indicating stronger polygenicity. We also

1102   provide genome-wide genetic correlation (r_g) and genetic correlation of shared variants (r_gs). b:

1103   LAVA volcano plots showing local genetic correlation coefficients (local-r_gs, y-axis) for eGFR and

1104   CVDs with -log10 p-values for each trait pair of analyses for each locus. Loci above the horizontal

1105   line are significant at $P < 0.05$ (negative correlation for blue dots, positive correlation for red dots).

1106   Larger points with black circles indicate loci significantly associated after Bonferroni correction (P

1107 $< 0.05 / 425$). LAVA-estimated local-r_gs is shown on the blue-red scale. eGFR, estimated glomerular

1108   filtration rate; AF, atrial fibrillation; CAD, coronary artery disease; VTE, venous thromboembolism;

1109   HF, heart failure; PAD, peripheral arterial disease.
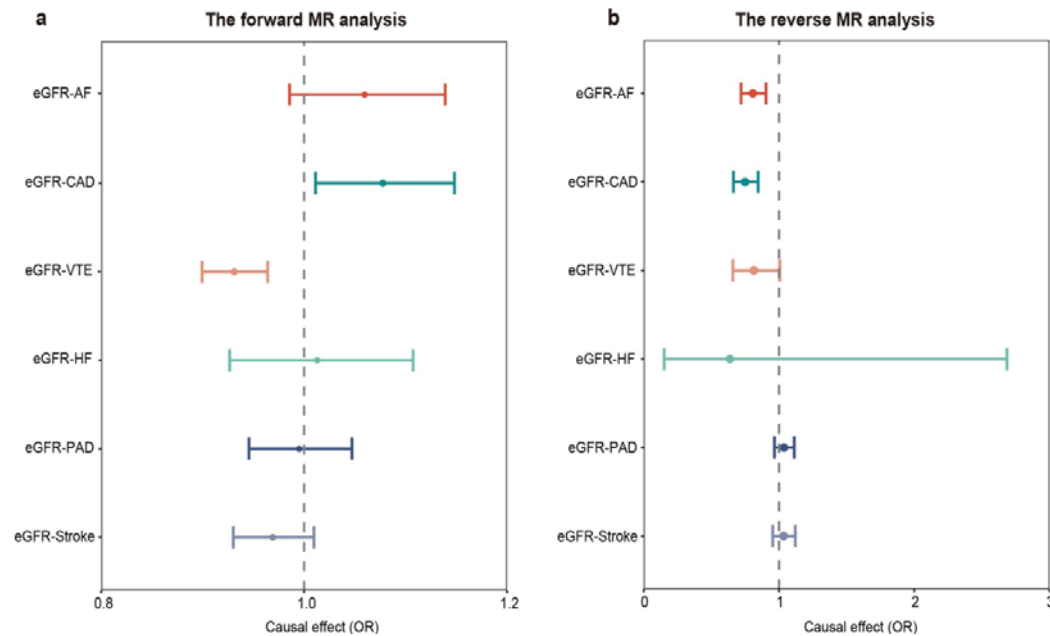
1110

Figure 3: Inference of causal relationship between estimated glomerular filtration rate and six major cardiovascular diseases.

Forest plots of the causal relationship between eGFR and six major CVDs using the LHC-MR method. a: indicates the estimated causal relationship of eGFR on CVDs. b: indicates the estimated causal relationship of CVDs on eGFR. Circles indicate odds ratio (OR) estimates, and error bars indicate 95% confidence intervals. OR > 1 indicates a positive association, and OR < 1 indicates a negative association. eGFR, estimated glomerular filtration rate; AF, atrial fibrillation; CAD, coronary artery disease; VTE, venous thromboembolism; HF, heart failure; PAD, peripheral arterial disease.
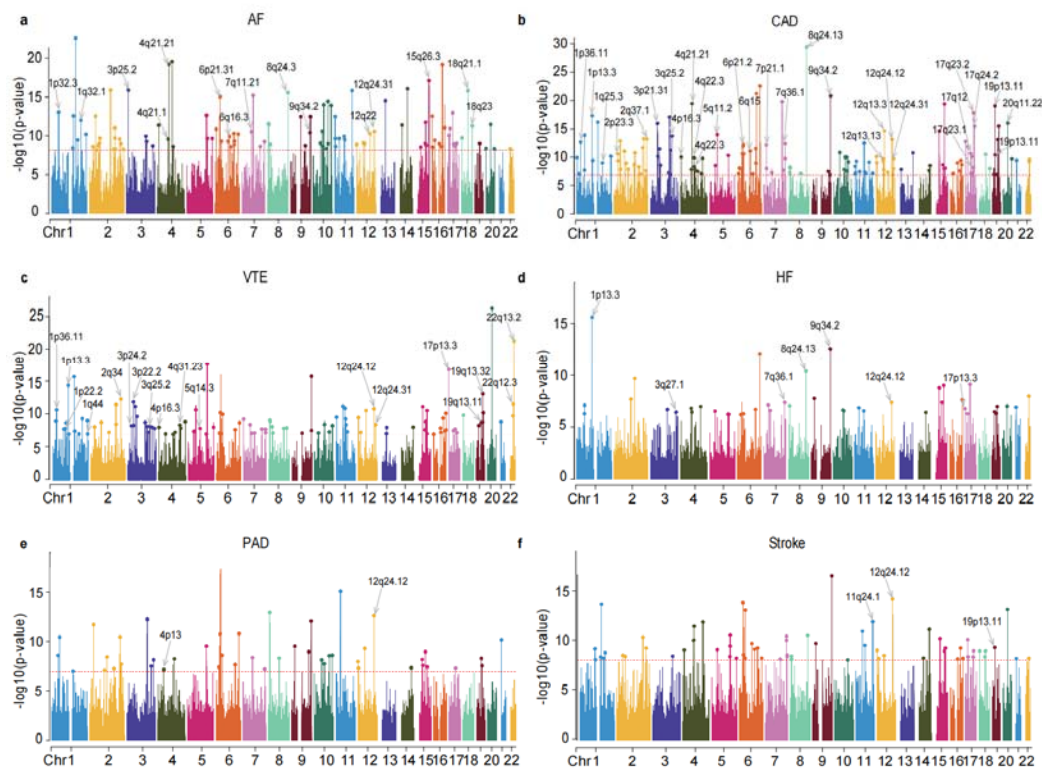
Figure 4: Manhattan plot of PLACO results for estimated glomerular filtration rate and six major cardiovascular diseases.

Manhattan plots reflect chromosomal position (x-axis) and negative log10-transformed *P*-values (y-axis) for each SNP. Horizontal lines indicate genome-wide significant P-values -log10 (5×10^{-8}). The r^2 threshold for defining independent significant SNPs was set to 0.2, and the maximum distance between LD blocks merged into one locus was set to 500 kb. *P*-values were derived using GWAS multi-trait analysis in discovery studies, and independent genome-wide significant associations with the smallest *P*-values (top SNPs) are circled in colored circles. Only SNPs that were common in all summary statistics were included. Labels are chromosomal regions where genomic risk loci with strong colocalization evidence (PP.H4 > 0.7) are located. eGFR, estimated glomerular filtration rate; AF, atrial fibrillation; CAD, coronary artery disease; VTE, venous thromboembolism; HF, heart failure; PAD, peripheral arterial disease.
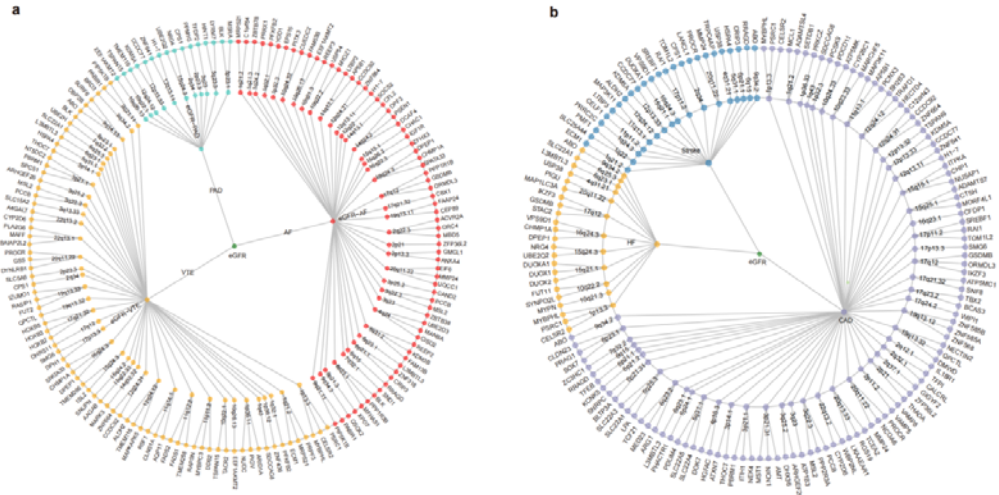
Figure 5: The overall situation of the pleiotropy association between estimated glomerular filtration rate and six major cardiovascular diseases

Based on the polygenic nature of the Mixer results, we separated the pleiotropic association between eGFR and CVDs into two network diagrams, showing the pleiotropic loci and genes identified in the six trait pairs. a: The circular dendrogram is centered on eGFR (inner circle) and branches out for the three CVDs (second circle, AF, VTE, and PAD). These three trait pairs show 91 pleiotropic loci (third circle) corresponding to 159 pleiotropic genes (outer circle). b: The circular dendrogram is centered on eGFR (the inner circle) and branches out for the three CVDs (second circle, CAD, HF, and stroke). Another three trait pairs show 80 pleiotropic loci (third circle) corresponding to 153 pleiotropic genes (outer circle). For trait pairs with three or more pleiotropic genes, we only show the top three pleiotropic genes according to priority, and these genes show statistical priority decay in a clockwise direction. For example, eGFR-AF-12q24.31-CCDC92 and eGFR-AF-12q24.31-ZNF664 indicate that 12q24.31 is a shared locus for this pair of specific diseases involving CCDC92 and ZNF664. eGFR, estimated glomerular filtration rate; AF, atrial fibrillation; CAD, coronary artery disease; VTE, venous thromboembolism; HF, heart failure; PAD, peripheral arterial disease.