

Genome-wide Mendelian Randomization Identifies Potential Drug Targets for Dorsopathies

1 Yu Cui^{1,2,†}, Jiarui Guo^{1,2,†}, Yuanxi Lu^{1,2,†}, Mengting Hu^{1,2}, He Zhou^{1,2}, Wancong Zhang^{2,3,4,*},
2 Shijie Tang^{2,3,4,*}

3 †These authors contributed equally to this work and share first authorship

4

5 ¹Shantou University Medical College, Shantou, Guangdong 515041, China

6 ²Department of Plastic Surgery and Burns Center, Second Affiliated Hospital, Shantou University
7 Medical College, Shantou, Guangdong 515051, China

8 ³Plastic Surgery Institute of Shantou University Medical College, Shantou, Guangdong 515051,
9 China

10 ⁴Shantou Plastic surgery Clinical Research Center, Shantou, Guangdong 515051, China

11 * Correspondence:

12 Wancong Zhang, Shijie Tang

13 Martine2007@sina.com, sjtang3@stu.edu.cn

14 **Keywords:** Mendelian randomization, Genetics, Drug target, Dorsopathies, Colocalization
15 analysis, Genome-wide association studies, Expression quantitative trait locus.

16 **Abstract**

17 **Background** Dorsopathies are a group of musculoskeletal disorders affecting the spinal column and
18 related structures, contributing significantly to global disability rates and healthcare costs. Despite
19 their prevalence, the genetic and biological mechanisms underlying dorsopathies are not fully
20 understood.

21 **Method** Summary-data-based Mendelian Randomization (SMR) and colocalization analysis were
22 employed, using data from genome-wide association studies (GWAS) and cis-expression quantitative
23 trait loci (cis-eQTLs) databases. Genes with a colocalization posterior probability (PP.H4) above 0.7
24 in SMR results were selected for additional analysis. These selected genes underwent MR analysis to
25 examine possible causal connections with dorsopathies, and sensitivity analyses were carried out to
26 ensure robustness. Additionally, two transcriptome-wide association studies (TWAS) were utilized to
27 confirm and screen for potential drug targets.

28 **Result** We identified four essential genes linked to dorsopathies: *NLRC4*, *CGREF1*, *KHK*, and
29 *RNF212*. Mendelian randomization (MR) analysis revealed a potential causal link between these
30 genes and dorsopathies. Elevated transcription levels of *NLRC4*, *CGREF1*, and *KHK* correlated with
31 reduced dorsopathies risk, while increased levels of *RNF212* were associated with heightened risk of
32 dorsopathies. Regarding methylation sites, an increase in *cg04686953* fully mediated the decreased
33 risk of dorsopathies by *RNF212*. Similarly, the risk effect of *cg26638505* and *cg18948125* was
34 entirely mediated by *NLRC4*, while *CGREF1* predominantly mediated the risk-increasing effect of
35 *cg06112415* and the decrease effect of *cg22740783*.

36 **Conclusion** Dorsopathies were associated with four pivotal genes: *NLRC4*, *CGREF1*, *KHK*, and
37 *RNF212*. Methylation analysis identified *cg04686953* and *cg22740783* as protective against
38 dorsopathies risk, while *cg26638505*, *cg18948125*, and *cg06112415* exhibited a risk-increasing
39 impact.

40

41 **1 Introduction**

42 Dorsopathies encompass a broad spectrum of musculoskeletal issues that impact the spinal column
43 and its related structures[1]. These conditions are prevalent among individuals across various age
44 groups and socioeconomic backgrounds, leading affected individuals to seek medical assistance.
45 They significantly contribute to the burden of illness, disability, and distress. Since 1990, there has
46 been a notable rise in the prevalence of these disorders, rendering them one of the primary causes of
47 global disability-adjusted life years [2]. However, the pathogenic genes and biological mechanisms
48 of the disease remain largely unknown. Research conducted on the general population has indicated a
49 potential link between certain dorsopathies and lifestyle choices including smoking [3], a higher body
50 mass index[4,5], and lack of physical activity [6]. Additionally, these dorsopathies have been found
51 to coincide with various other health complications such as anxiety, depression, diabetes,
52 cardiovascular, respiratory, and gastrointestinal ailments[7,8]. Furthermore, specific dorsopathies like
53 osteopenia, osteomalacia, and tuberculosis can be influenced by factors such as diet, living
54 environment, and psychological aspects. These particular disorders often occur alongside systemic
55 comorbidities such as endocrine dysfunction and infection [8]. Relevant limitations in observational
56 epidemiological studies include the potential for confounding, reverse causation, and diverse biases,
57 which hinder comprehensive understanding of disease pathogenesis and identification of treatment
58 targets.

59 However, the utilization of Mendelian randomization (MR) techniques, which rely on the random
60 allocation of genetic variations, allows for the simulation of randomized controlled trials and can
61 effectively mitigate the impact of confounding variables. In relation to research on dorsopathies, this
62 implies a more precise evaluation of the causal association between drug targets and disease while
63 excluding any factors that may potentially disrupt the findings [9]. MR utilizes genetic variants as
64 instrumental variables to assess the causal impact of an exposure on outcomes. This approach has
65 been extensively utilized in other research studies related to diseases and has effectively facilitated
66 the identification of potential therapeutic targets for diverse medical conditions.[10] However, there
67 is limited research on employing MR methods to explore potential drug targets for dorsopathies.
68 Therefore, to gain a deeper understanding of the pathogenesis of dorsopathies and identify more
69 effective treatment approaches, further MR studies are needed to evaluate the drug targets for
70 dorsopathies. This will help eliminate factors that could interfere with the results and offer new
71 perspectives and methods for the treatment of dorsopathies. In theory, single nucleotide
72 polymorphisms (SNPs) are distributed randomly and not impacted by environmental factors, making
73 them an ideal tool for establishing causality. MR is a type of instrumental variable analysis that
74 primarily employs SNPs as genetic instruments to determine the causal impact of an exposure (in this
75 instance, circulatory proteins) on outcomes.[11]. Previous studies have successfully utilized MR to
76 identify biomarkers and treatment targets for various diseases, such as aortic aneurysms [12],
77 multiple sclerosis [13], and breast cancer [14].

78 However, in previous studies, the use of MR To analyze drug targets for dorsopathies is very rare. To
79 address this research gap, our study focuses on utilizing genes as factors of exposure in drug target

80 research, investigating the role of genes in dorsopathies development and their potential as viable
81 drug targets. Genes have stable genetic characteristics, and they are unaffected by environmental
82 factors, allowing for a more accurate assessment of the causal relationship between genes and
83 dorsopathies [15]. By exploring dorsopathies' potential drug targets through genes as the exposure
84 factor, we aim to offer new perspectives and methods for the treatment and prevention of
85 dorsopathies.

86 To investigate potential pathogenic genes related to dorsopathies, we employed a comprehensive
87 analytical approach including summary-data-based Mendelian randomization (SMR) analysis,
88 colocalization analysis, genome-wide association study (GWAS), and transcriptome-wide association
89 study (TWAS). We utilized a meta-analysis dataset of cis-expression quantitative trait loci (cis-
90 eQTLs) from peripheral blood samples as exposure data, along with results from the extensive
91 FinnGen database. Following preliminary analysis, we identified candidate genes and established
92 causal inference using MR methods. Additionally, we conducted mediation analysis of gene-
93 mediated methylation sites to explore disease related methylation sites.

94 **2 Method**

95 **2.1 Datasets**

96 Summary-level data for the GWAS on dorsopathies were obtained from the FinnGen consortium,
97 comprising 117,411 cases and 294,770 controls. This dataset represents the most recent GWAS
98 findings and comprises the largest cohort of dorsopathies cases documented to date. The primary
99 objective of FinnGen is to accumulate and rigorously analyze genomic and national health register
100 data from 500,000 Finnish individuals.[16] In the context of drug development studies, we prioritized
101 cis-eQTLs that were in closer proximity to the target gene. These cis-eQTLs used for SMR analyses
102 were sourced from the eQTLGen Consortium [17] and the eQTL meta-analyses conducted on
103 peripheral blood samples from a cohort of 31,684 individuals. For the TWAS analyses of eQTL data,
104 we utilized the GTEx v8 European whole blood dataset. Given our specific emphasis on dorsopathies,
105 we meticulously extracted comprehensive eQTL results exclusively from the whole blood samples
106 within the GTEx dataset [18].

107

108 **2.2 SMR analyses**

109 We conducted SMR and heterogeneity in dependent instruments (HEIDI) tests analyses on cis
110 regions using the SMR software (version 1.03) [19]. The methodologies for SMR analyses are
111 detailed in the original work. In brief, SMR analyses employs a well-established MR approach. This
112 technique employs a SNP at a prominent cis-eQTL as an instrumental variable (IV). The summary-
113 level eQTL data serve as the exposure variable, and the GWAS data for a specific trait serve as the
114 outcome variable. The primary objective is to explore a potential causal or pleiotropic association,
115 wherein the same causal variant influences both gene expression and the trait. It is crucial to
116 acknowledge that the SMR method lacks the ability to distinguish between a causal association, in
117 which gene expression causally influences the trait, and a pleiotropic association, in which the same
118 SNP affects both gene expression and the trait. This limitation arises because of the single
119 instrumental variable (IV) in the MR method, which cannot differentiate between causality and
120 pleiotropy. Nevertheless, the HEIDI test can make this distinction by discerning causality and
121 pleiotropy from linkage. Linkage refers to cases in which two different SNPs in linkage
122 disequilibrium (LD) independently influence gene expression and the trait. Although less biologically

123 intriguing than causality and pleiotropy, the HEIDI test provides clarity in such scenarios. For the
124 HEIDI test, a p-value below 0.05 was considered significant, suggesting that the observed association
125 was attributable to linkage.

126

127 **2.3 GWAS analyses**

128 Multimarker analyses of genomic annotation (MAGMA) employs a multiple regression model to
129 assess the cumulative effect of multiple SNPs within a specific gene region (± 10 kb)[20]. The
130 reference panel for calculating LD was derived from Phase 3 of the 1000 Genomes European
131 population. Significance thresholds for GWAS analyses using both SMR and MAGMA were set at a
132 false discovery rate (FDR) below 0.05, corrected using the Benjamini-Hochberg method.[21]

133

134 **2.4 TWAS analyses**

135 We conducted validations to integrate dorsopathies GWAS and eQTL data of whole blood from
136 GTEx using the FUSION and UTMOST, widely utilized tools in prior TWAS investigations [22,23].
137 FUSION constructs predictive models using various penalized linear models, such as GBLUP,
138 LASSO, Elastic Net, for the significant cis-heritability genes estimated from SNPs within 500 kb on
139 either side of the gene boundary. Subsequently, it selects the optimal model based on the coefficient
140 of determination (R^2) calculated through a fivefold cross-validation. For UTMOST, we performed
141 repeated 49 single-tissue association tests for each tissue. TWAS significance for both single-tissue
142 analyses was determined with a Benjamini-Hochberg corrected FDR value below 0.05.

143

144 **2.5 MR analyses**

145 In conducting the two-sample MR analyses [24], we utilized the TwoSampleMR R package. The
146 utilization of the two-sample MR framework requires employing two distinct datasets. In this study,
147 genetic instruments, specifically cis-eQTL, served as exposures, while GWAS were utilized to
148 determine outcome traits. The MR methodology investigates the relationship between gene
149 expression and diseases or traits by employing genetic variants associated with gene expression as
150 instrumental variables (exposure) and GWAS for the outcome measures. Mendelian Randomization
151 facilitates exploration into whether alterations in gene expression causally impact diseases or traits.
152 For instruments represented by a single SNP, we employed the Wald ratio. In cases where
153 instruments consisted of multiple SNPs, we implemented the inverse-variance-weighted MR
154 approach. When selecting SNPs, the significance thresholds were defined as $P < 5 \times 10^{-8}$ for
155 genome-wide significance, with a linkage disequilibrium parameter (r^2) set to 0.1, and a genetic
156 distance set to 10 MB.

157

158 **2.6 Colocalization analyses**

159 Conducted colocalization analyses using the coloc package in the R software environment (version
160 4.0.3). Colocalization analyses aims to assess the potential shared causality between SNPs associated
161 with both gene expression and phenotype at a specific locus, thereby indicating the "colocalization"

162 of these genetic signals. The analyses calculates posterior probabilities (PPs) for five hypotheses: H0
163 denotes no association with either gene expression or phenotype; H1 signifies an association solely
164 with gene expression; H2 indicates an association exclusively with the phenotype; H3 suggests an
165 association with both gene expression and phenotype through independent SNPs; and H4 implies an
166 association with both gene expression and phenotype through shared causal SNPs. A substantial PP
167 for H4 (PP.H4 above 0.70) strongly suggests the presence of shared causal variants influencing both
168 gene expression and phenotype [25].

169

170 **2.7 Methylation & Mediation analysis**

171 We hypothesize that methylation sites exert influence on the pathogenic risk of dorsopathies through
172 three primary pathways: a) they indirectly modulate the pathogenic risk by influencing gene
173 expression; b) the methylation sites themselves have a direct impact on the pathogenic risk of
174 dorsopathies; and c) methylation sites may affect the pathogenic risk of dorsopathies by influencing
175 other confounders. It is crucial to underscore that the overall impact of methylation on the pathogenic
176 risk of dorsopathies can be conceptualized as the combined action of these three pathways, expressed
177 as $T=a+b+c$, where a denotes the indirect modulation of pathogenic risk by influencing gene
178 expression, b denotes the direct effect of methylation sites on pathogenic risk, and c represents their
179 influence on pathogenic risk through impacting confounding factors. Significance tests are conducted
180 separately for T, x, y, and a. If all these parameters show significance, it can be inferred that the gene
181 plays a role in the intermediate pathway, thereby substantiating the existence of the a) pathway.

182

183 **3 Result**

184 **3.1 SMR analyses and colocalization for Preliminary Identifying Potential Genes**

185 In our initial analyses phase, we utilized the eQTLGen dataset for SMR analyses to pinpoint genes
186 significantly linked to dorsopathies. Employing the FDR method set our p-value significance
187 threshold. The eQTLGen dataset provided extensive data, yielding 15,633 candidate genes.

188 We identified 269 genes significantly associated with dorsopathies ($FDR_P < 0.05$). Subsequent
189 scrutiny uncovered four potential drug targets (*NLRC4*, *CGREF1*, *KHK*, *RNF212*) affecting
190 dorsopathies through various intersecting methods. For clarity, Manhattan plots illustrated SMR
191 results. Following gene identification, the HEIDI heterogeneity test ($p_HEIDI > 0.05$) sifted out
192 genes lacking horizontal pleiotropy. SMR locus plots further elucidated gene outcomes.

193 We proceeded with colocalization analyses to merge GWAS and blood eQTL data for genes passing
194 the SMR test. This aimed to determine if these genes colocalized with the dorsopathies trait. The
195 colocalization test results robustly supported colocalization between the trait and all four genes
196 (*NLRC4*: PP.H4 = 0.993, *CGREF1*: PP.H4 = 0.905, *KHK*: PP.H4 = 0.773, *RNF212*: PP.H4 = 0.923)
197 meeting both SMR and HEIDI test criteria. Consequently, we identify these genes as top-priority
198 candidates for subsequent functional studies.

199

200 **3.2 MR analyses Validates Potential Gene Causal Relationships**

201 Using cis-eQTL data from the eQTLGen Consortium, we conducted two-sample MR analyses on
202 European summary statistics of individuals with dorsopathies. The discovery cohort, consisting of
203 117,411 cases and 294,770 controls from the FinnGen cohort, underwent inverse variance weighted
204 (IVW) MR analyses to combine effect estimates from each genetic instrument. The analysis revealed
205 associations between the genetically predicted expression of 956 genes and dorsopathies risk
206 following multiple testing adjustments (FDR correction). Notably, *NLRC4* (OR = 0.886, 95% CI =
207 0.855-0.917, FDR_P = 1.09e-11), *CGREF1* (OR = 0.682, 95% CI = 0.585-0.794, FDR_P = 8.74e-07),
208 *KHK* (OR = 0.936, 95% CI = 0.921-0.951, FDR_P = 3.67e-16), and *RNF212* (OR = 1.145, 95% CI =
209 1.056-1.241, FDR_P = 1.01e-03) were among the genes identified. To ensure the reliability of our
210 findings, we conducted tests for horizontal pleiotropy, which did not reveal any evidence of its
211 presence in the dataset. These additional analyses confirmed the absence of horizontal pleiotropy,
212 bolstering the robustness and validity of our MR genetics findings.

213

214 **3.3 Validation**

215 **3.3.1 TWAS & UTMOST Validates Transcriptome-level Causal Relationships**

216 In our quest to enhance causal inference and gain deeper insights into genetic associations with
217 dorsopathies, we conducted a TWAS analyses on the four genes identified in previous analyses,
218 utilizing Fusion and UTMOST software. This comprehensive approach aimed to elucidate the
219 transcriptional associations of these genes with dorsopathies and bolster the evidence supporting their
220 potential causal role.

221 The results revealed significant transcriptional associations for the four genes with dorsopathies, with
222 all colocalization probabilities (PP.H4) exceeding 0.7. Specifically, elevated transcription levels of
223 *NLRC4* (Z score = -5.33068, P = 9.78E-08), *CGREF1* (Z score = -4.7644, P = 1.89E-06), and *KHK*
224 (Z score = -4.72939, P = 2.25E-06) were significantly correlated with a decreased risk of
225 dorsopathies, whereas an increased transcription level of *RNF212* (Z score = 3.751506, P =
226 0.000176) was significantly associated with an increased risk of dorsopathies. Notably, these TWAS
227 findings were consistent with those from the SMR analyses, providing robust support for the
228 transcriptional associations of these genes with dorsopathies.

229 **3.3.2 MAGMA Validates Genome-level Causal Relationships**

230 We also conducted a GWAS analyses using MAGMA software on the quartet of genes highlighted in
231 preceding studies. This approach aimed to illuminate the associations of these genes with
232 dorsopathies at the genome level, thereby reinforcing the evidence underpinning their potential causal
233 involvement.

234 Based on the MAGMA analyses, we identified 855 significant genes that passed the FDR test. The
235 outcomes unveiled noteworthy genome-level associations for the aforementioned genes with
236 dorsopathies. Specifically, heightened transcription levels of *NLRC4* (FDR_P = 0.01851868),
237 *CGREF1* (P = 0.002068503), and *KHK* (P = 0.009814817) exhibited significant correlations with
238 reduced risk of dorsopathies, whereas increased transcription levels of *RNF212* (P = 0.00202954)
239 were significantly linked with elevated risk of dorsopathies. Importantly, these GWAS findings
240 corroborated those from the SMR analyses, thereby fortifying the robustness of the transcriptional
241 associations of these genes with dorsopathies.

242

243 3.4 Methylation analysis & Mediation analysis

244 We evaluated the influence of *cg23387401* on *RNF212* expression ($\beta = -0.52$, $P = 2.08e-14$) and its
245 link to dorsopathies risk ($\beta = 0.19$, $P = 1.20E-07$). From these findings, we identified how gene-
246 mediated methylation impacts dorsopathies risk ($\beta = -0.10$, $P = 1.35E-05$), with an observed overall
247 effect ($\beta = -0.10$, $P = 2.68E-06$) indicating a 93.79% intermediate effect proportion. This suggests
248 that the rise in *cg23387401* is wholly mediated by *RNF212*, leading to reduced dorsopathies risk.

249 Likewise, we assessed *cg26638505*'s impact on *NLRC4* expression ($\beta = -0.81$, $P = 1.25E-09$) and its
250 association with dorsopathies risk ($\beta = -0.13$, $P = 7.87E-09$). These calculations revealed the
251 influence of gene-mediated methylation sites on dorsopathies risk ($\beta = 0.11$, $P = 2.87E-05$), with an
252 overall effect ($\beta = 0.09$, $P = 1.43E-03$) and an intermediate effect proportion of approximately
253 120.00%. This suggests that *cg26638505* elevation is entirely mediated by *NLRC4*, reducing
254 dorsopathies risk.

255 We also analyzed *cg18948125*'s effect on *NLRC4* expression ($\beta = -0.93$, $P = 1.72E-08$) and its
256 association with dorsopathies risk ($\beta = -0.13$, $P = 7.87E-09$). From this, we derived the impact of
257 gene-mediated methylation sites on dorsopathies risk ($\beta = 0.12$, $P = 5.50E-05$), with an overall
258 effect ($\beta = 0.12$, $P = 5.59E-04$) and an intermediate effect proportion of 104.41%. This suggests that
259 *cg18948125* elevation primarily contributes to increased dorsopathies risk, mediated by *NLRC4*.

260 Similarly, *cg22740783*'s influence on *CGREF1* expression ($\beta = 0.32$, $P = 7.71E-07$) and its
261 association with dorsopathies risk ($\beta = -0.38$, $P = 2.12E-05$) was calculated. This allowed us to
262 identify the impact of gene-mediated methylation sites on dorsopathies risk ($\beta = -0.12$, $P = 1.27E-$
263 03), with an overall effect ($\beta = -0.13$, $P = 1.63E-04$) and an intermediate effect proportion of
264 92.24%. Hence, we propose that the elevation in *cg22740783* is fully mediated by *CGREF1*, leading
265 to decreased dorsopathies risk.

266 Lastly, we evaluated *cg06112415*'s impact on *CGREF1* expression ($\beta = -0.34$, $P = 1.75E-06$) and its
267 association with dorsopathies risk ($\beta = -0.38$, $P = 2.12E-05$). Through these assessments, we
268 identified the effect of gene-mediated methylation sites on dorsopathies risk ($\beta = 0.13$, $P = 1.49E-$
269 03), with an overall effect ($\beta = 0.12$, $P = 5.59E-04$) and an intermediate effect proportion of
270 110.63%. This suggests that *cg06112415* elevation primarily contributes to increased dorsopathies
271 risk, mediated by *CGREF1*.

272

273 4 Discussion

274 This study aimed to offer new perspectives and methods for the treatment and prevention of
275 dorsopathies. In our research, the identification of candidate genes and the establishment of causal
276 inference were conducted via MR methods, while the mediation analysis was used to explore disease
277 related gene-mediated methylation sites. As a result, our study revealed four genes as potential
278 therapeutic targets for dorsopathies, including *NLRC4*, *CGREF1*, *KHK* and *RNF212*. Among these
279 genes, elevated transcription levels of the *NLRC4*, *CGREF1* and *KHK* were significantly associated
280 with a decreased risk of dorsopathies, whereas a raised transcription level of *RNF212* was
281 significantly related a higher risk of dorsopathies.

282 In order to find the novel drug targets for dorsopathies, we employed an integrative analysis that
283 combined colocalization with MR to assess causal genes for dorsopathies. Later, we conducted tests
284 for horizontal pleiotropy to ensure the reliability of our MR genetics findings. In addition, TWAS
285 analyses on *NLRC4*, *CGREF1*, *KHK* and *RNF212* was conducted through Fusion and UTMOST
286 software, which further illuminated the transcriptional associations of these genes with dorsopathies
287 and bolstered the evidence supporting their potential causal role. Moreover, we conducted a GWAS
288 analyses whose results were consistent with those from the SMR analyses, which reinforcing the
289 evidence underpinning the potential causal association between these four genes with dorsopathies.
290 Finally, we evaluated the influence of methylation sites on those genes expression and their links to
291 dorsopathies. The results exhibited that the rise in *cg23387401* was entirely mediated by *RNF212*, the
292 elevations of *cg26638505* and *cg18948125* were primarily mediated by *NLRC4*, and the up-
293 regulation of *cg06112415* was fully mediated by *CGREF1*, leading to increased dorsopathies risk;
294 while the rise of *cg22740783* was totally mediated by *CGREF1*, resulting in decreased dorsopathies
295 risk.

296 NOD-like receptors (NLRs) family caspase activation and recruitment domain-containing protein 4
297 (*NLRC4*) gene encodes NLRC4, which is mainly expressed in macrophages, neutrophils, dendritic
298 cells and glial cells[26]. As a member of the caspase recruitment domain-containing *NLR* family,
299 NLRC4 partners with NLR family of apoptosis inhibitory proteins (NAIP) to assemble
300 inflammasome complexes, which termed NAIP/NLRC4 inflammasomes[27]. Inflammasomes
301 function as intracellular multi-protein platforms that are activated by pathogen-associated molecular
302 patterns or damage-associated molecular patterns, triggering innate immune reactions and
303 inflammatory caspase-activation to defense pathogens and danger signals and maintain the
304 homeostasis[27–29]. The activation of caspases results in the proteolytic activation of the pro-
305 inflammatory cytokines interleukin-1 β (IL-1 β) and/or interleukin-18 (IL-18)[27]. Meanwhile,
306 activated caspase-1 also cleaves gasdermin D, whose N-terminal fragments become inserted into cell
307 membranes, thereby facilitating the release of more IL-1 β and IL-18[28,29]. In particular,
308 inflammasomes were classified into canonical inflammasomes, which activating caspase-1, and non-
309 canonical inflammasomes, which primarily activate caspase-4 and/or caspase-5 in human cells[29].

310 Specifically, the NAIP/NLRC4 inflammasome serves as a as part of the innate immune response and
311 senses a range of intracellular bacteria and bacterial components in the host cell cytosol, such as *S.*
312 *typhimurium*, *Legionella pneumophila*, *flagellin* and components of the virulence-associated type III
313 secretion apparatus, thereby to mediate host defense against bacterial pathogens[26,27,30]. NAIP
314 proteins function as specific cytosolic receptors for a variety of bacterial protein ligands[31], while
315 NLRC4 interacts directly with caspase-1 to induce cell death, such as the pyroptosis of macrophage,
316 and to cause rapidly initiate inflammation and vascular fluid loss[26]. The localized effects of
317 NAIP/NLRC4 inflammasome could defend against bacterial pathogens. However, the aberrant
318 activity of this cluster also leads to multiple clinical manifestations, including macrophage activation
319 syndrome, neonatal enterocolitis and autoimmune disorders[28,31], and promotes the process of
320 some diseases, such as gliomas[29], premature rupture of membranes[26], ulcerative colitis[28],
321 rheumatoid arthritis[30] and so on. Sim et al. found that non-canonical pathway molecules of
322 NAIP/NLRC4 inflammasomes, including caspase-4, caspase-5, and N-cleaved GSDMD, were
323 significantly increased with each glioma grade[29]. Additionally, Zhu et al. reported that NLRC4 was
324 upregulated in the membranes of patients with premature rupture of fetal membranes and recruited
325 more caspase-1, which promotes the process of rupture of fetal membranes by inducing apoptosis
326 and degrading extracellular matrix[26]. The study of An et al. presented that the persistent activation
327 of NAIP/NLRC4 inflammasome induces macrophage pyroptosis mediated by caspase1-dependent
328 cleavage of GSDMD and releases proinflammatory cytokines including IL-1b and IL-18, facilitating

329 the occurrence and progression of UC[28]. In addition, research conducted by Delgado-Arévalo et al.
330 indicated that *NLRC4* as an inflammasome sensor differentially upregulated in CD1c+ cDC from
331 patients with rheumatoid arthritis, and this sensor seems to be nonredundantly involved in the
332 detection of intracellular dsDNA[30]. Nevertheless, our research observed the significantly inhibitory
333 effect of up-regulated *NLRC4* expression on dorsopathies. Thus, further research on the roles and
334 mechanisms of *NLRC4* and inflammasome in the progress of dorsopathies is needed.

335 Cell growth regulator with EF-hand domain 1 (*CGREF1*) gene encodes a novel secretory protein
336 with 2 Ca²⁺-binding EF-hand domains which plays an important role in eukaryotic cellular
337 signaling[32]. *CGREF1* mRNA has high expressions in HCT116, H1299 and HepG2 cells while
338 expressing at low levels in other cell lines including Raji, Jurkat, BT325, PC12[32]. Mechanistically,
339 *CGREF1* is regulated by p53[33], and the overexpression of *CGREF1* significantly inhibits the
340 transcriptional activity of AP-1, reduces the phosphorylation of ERK (extracellular signal-regulated
341 kinases) and p38 MAPK (mitogen-activated protein kinases), and suppresses the proliferation of
342 HEK293T and HCT116 cells[32]. Furthermore, *CGREF1* can decrease the percent of G2/M and S
343 phase and repress cell proliferation while overexpressing[32]. We found that the increased expression
344 of *CGREF1* is significantly related to the decreased risk of dorsopathies. In the research of Xiang et
345 al., the colorectal cancer patients with higher expression of *CGREF1* were found to have significantly
346 better overall survival than patients with lower expression, and the role of *CGREF1* in the prognosis
347 of Early-onset colorectal cancer was reported[33]. Furthermore, Xie et al. observed that *CGREF1*
348 expression were high in the tissue of osteosarcoma patients while it were lowly expressed in normal
349 tissue, and speculated that *CGREF1* can predict drug resistance to osteosarcoma[34]. However, the
350 biological function of *CGREF1* is poorly explored, and further study for its mechanisms and roles in
351 these diseases is warranted.

352 Ketohexokinase (*KHK*) is an enzyme that phosphorylates fructose to produce fructose-1-phosphate
353 (F-1-P) at the first rate-limiting step in fructose metabolism[35,36]. The gene *KHK* encodes two
354 isoforms, KHK-C and KHK-A. Considered as the primary enzyme in fructose metabolism, KHK-C is
355 exclusively expressed in a few tissues, especially in the liver, and drives the aforementioned reaction
356 rapidly, resulting in the accumulation of uric acid and transient depletion of intracellular phosphate
357 and ATP[35–38]. In the contrary, KHK-A has a much weaker affinity and a higher Michaelis
358 constant for fructose[38]. However, KHK-A can directly phosphorylate phosphoribosyl
359 pyrophosphate synthetase 1 (PRPS1) by prevention of inhibitory nucleotide binding and facilitation
360 of ATP binding[37]. As a result, KHK-A improves nucleic acid synthesis and promotes the G1/S
361 phase transition in the cell cycle, accelerating cell proliferation[36,38].

362 Recent studies have highlighted the importance of *KHK* as the key mechanism stimulating the
363 various adverse metabolic effects of fructose, such as impaired insulin sensitivity,
364 hypertriglyceridemia, and oxidative stress[35]. In addition, fructose metabolism seems to provide
365 cancer cells with the supplementary fuel required for proliferation and metastasis in colon cancer, and
366 glioma[36,38]. Moreover, alternated expression of *KHK* is also related to several diseases. For
367 instance, knocked down the expression of *KHK* dramatically reduced fructose-induced production of
368 reactive oxidative species (ROS) in proximal tubular cells[35], indicating the association between
369 *KHK* and ROS. Furthermore, KHK-A activates and upregulates PRPS1, which could lead to
370 enhanced nucleic acid synthesis for tumorigenesis[38]. Kim et al. reported that most cancer cell
371 lines predominantly expressed KHK-A rather than KHK-C, and KHK-A overexpression could
372 augment cell invasion[36]. Moreover, in their study, upon fructose stimulation, KHK-A acted as a
373 nuclear protein kinase and triggered Epithelial-mesenchymal transition in breast cancer, promoting
374 breast cancer metastasis[36]. Besides, loss-of-function variants in *KHK* can cause essential

375 fructosuria, an autosomal recessive disease characterized by intermittent appearance of fructose in the
376 urine[39]. Yang et al. found the the overexpression of *KHK-A* in cell lines of oesophageal squamous
377 cell carcinoma, which may finally promote the proliferation, tumourigenicity and motility of ESCC
378 cells[38]. Similarly, *KHK-A* was considered to play an instrumental role in promoting de novo
379 nucleic acid synthesis and hepatocellular carcinoma development[37]. In our study, the raised
380 expression of *KHK* is found to be positively related to decreased risk of dorsopathies. In oder to
381 understand the mechanism of changed expression level of *KHK* in dorsopathies, further research are
382 needed.

383 Regulatory factor X2 (*RFX2*) gene is essential for maintaining normal spermatogenesis and involved
384 in spermatogenesis impairment and male infertility in mice[40]. In particular, Ring finger protein 212
385 (*RNF212*) gene is important for crossing over and chiasma formation during meiosis[41]. Mouse
386 *Rnf212* has a central role in designating crossover sites and coupling chromosome synapsis to the
387 formation of crossover-specific recombination complexes. In humans, *RNF212* has been associated
388 with variation in the genome-wide recombination rate[41]. The protein encoded by *RNF212* has
389 homology to two meiotic procrossover factors: Zip3 and ZHP-3[42]. It functions to couple
390 chromosome synapsis to the formation of crossover-specific recombination complexes[43].
391 Moreover, with the symbolic RING-finger domains, *RNF212* protein is a RING-family E3-ligase for
392 Small ubiquitin-like protein (SUMO), and the latter plays an important role in assembly and
393 disassembly of synaptonemal complex by regulating protein–protein inter action during
394 meiosis[42,43,43]. In addition, *RNF212* stabilizes association of a subset of MutS γ complexes with
395 recombination sites[42,44]. MutS γ complex is a kind of meiosis-specific recombination factors,
396 working as an attractive target for non-crossover/crossover differentiation[42]. It binds and stabilizes
397 DNA strand-exchange intermediates to promote both homolog synapsis and crossingover[44]. In
398 mammals, every pair of chromosomes obtains at least one crossover, while the majority of
399 recombination sites yield non-crossovers. Non-crossovers are inferred to arise from the disassembly
400 of D-loops and annealing of DNA double-strand breaks ends in a process termed synthesis dependent
401 strand annealing[42]. Designation of crossovers involves the formation of metastable joint molecules
402 and selective localization of SUMO-ligase *RNF212* to a minority of recombination sites where it
403 stabilizes pertinent factors, such as MutS γ [42,44]. Furthermore, this differential *RNF212*-dependent
404 stabilization of key recombination proteins at precrossover sites is thought to be the basic feature of
405 crossover/non-crossover differentiation[42]. It is suggested that *RNF212*-mediated SUMOylation
406 may stabilizes the association of MutS γ with nascent crossover CO intermediates in a number of
407 ways, such as promoting protein-protein interactions, altering ATP binding and hydrolysis (which
408 modulate the binding and dissociation of MutS γ complexes) or antagonizing ubiquitin-dependent
409 protein turnover[41,42].

410 Insufficient *RNF212* accumulating at recombination sites can lead to crossing-over stochastically
411 fails[43]. Fujiwara et al. reported that *Rnf212* knock out (KO) in mice leads to infertility of male and
412 female due to the loss of SPCs(spermatocytes) at post-anaphase stage. Moreover, crossing over is
413 diminished by $\geq 90\%$ in *Rnf212*^{-/-} mice[43]. In particular, the *Rnf212* KO spermatocytes lack
414 chiasmata and exhibit depletion of spermatids and mature spermatozoa[41]. A nonsense mutation in
415 the *Rnf212* gene was discovered in repro57 mutant mice, and this mutant mice exhibited male
416 infertility, arrest of spermatogenesis in meiosis, and defects in cytological markers of recombination
417 and chiasma formation, which is similar to the *Rnf212* KO phenotype[41]. In humans, the rate of
418 crossing-over varies significantly between individuals, and higher maternal crossover rates have been
419 associated with greater fecundity[42]. However, in the absence of *RNF212*, designation of crossover
420 sites fails because no MutS γ complexes are stabilized beyond early pachynema[44]. Yu et al.
421 detected that the frequencies of allele C and the genotype CC at the *rs4045481* locus in *RNF212* gene

422 were significantly higher in patients with azoospermia in comparison with controls. Furthermore,
423 they reported that homozygous of allele C (genotype CC) may decrease the activity of pre-mRNA
424 due to the disappearance of the binding motifs of SRSF5, leading to the reduced expression of
425 *RNF212* and influencing normal spermatogenesis, consequently increasing the risk of
426 azoospermia[40]. Nevertheless, we found that the increased expression of *RNF212* exhibited a
427 positive correlation with dorsopathies. The influence of raised expression of *RNF212* is poor
428 classified, and further researches are required to illustrate the mechanism of increased *RNF212* level
429 and dorsopathies.

430 To summary, our study revealed the causal associations between the genetically predicted expression
431 of four genes (*NLRC4*, *CGREF1*, *KHK* and *RNF212*) and dorsopathies risk, which offers new
432 perspectives and strategies for the improved diagnosis, treatment and prevention of dorsopathies.
433 Based on our results, we hypothesized that increased expression level of *NLRC4*, *CGREF1* and *KHK*
434 are associated with decreased risk of dorsopathies, while the increased expression level of *RNF212*
435 has the opposite effect. Moreover, further study focusing on the mechanism of these expression
436 changes in dorsopathies are needed, and future clinical studies should be conducted.

437

438 **5 Acknowledgements**

439 The authors appreciate the publicly available data of the FinnGen consortium, the eQTLGen
440 consortium, the GTEx project, and the MRC IEU OpenGWAS database.

441

442 **6 Conflict of interest**

443 The authors declare that the research was conducted in the absence of any commercial or financial
444 relationships that could be construed as a potential conflict of interest.

445

446 **7 Funding**

447 The study and publishing of this article were not supported by any funding.

448

449 **8 Reference**

450 [1] L.B. Connelly, A. Woolf, P. Brooks, Cost-Effectiveness of Interventions for Musculoskeletal
451 Conditions, in: D.T. Jamison, J.G. Breman, A.R. Measham, G. Alleyne, M. Claeson, D.B.
452 Evans, P. Jha, A. Mills, P. Musgrove (Eds.), Disease Control Priorities in Developing Countries,
453 2nd ed., The International Bank for Reconstruction and Development / The World Bank,
454 Washington (DC), 2006. <http://www.ncbi.nlm.nih.gov/books/NBK11713/> (accessed April 30,
455 2024).

- 456 [2] L. March, E.U.R. Smith, D.G. Hoy, M.J. Cross, L. Sanchez-Riera, F. Blyth, R. Buchbinder, T.
457 Vos, A.D. Woolf, Burden of disability due to musculoskeletal (MSK) disorders, *Best Pract Res*
458 *Clin Rheumatol* 28 (2014) 353–366. <https://doi.org/10.1016/j.berh.2014.08.002>.
- 459 [3] R. Shiri, J. Karppinen, P. Leino-Arjas, S. Solovieva, E. Viikari-Juntura, The association
460 between smoking and low back pain: a meta-analysis, *Am J Med* 123 (2010) 87.e7–35.
461 <https://doi.org/10.1016/j.amjmed.2009.05.028>.
- 462 [4] R. Shiri, J. Karppinen, P. Leino-Arjas, S. Solovieva, E. Viikari-Juntura, The association
463 between obesity and low back pain: a meta-analysis, *Am J Epidemiol* 171 (2010) 135–154.
464 <https://doi.org/10.1093/aje/kwp356>.
- 465 [5] A. Koyanagi, A. Stickley, N. Garin, M. Miret, J.L. Ayuso-Mateos, M. Leonardi, S. Koskinen,
466 A. Galas, J.M. Haro, The association between obesity and back pain in nine countries: a cross-
467 sectional study, *BMC Public Health* 15 (2015) 123. <https://doi.org/10.1186/s12889-015-1362-9>.
- 468 [6] M. Smuck, M.-C.J. Kao, N. Brar, A. Martinez-Ith, J. Choi, C.C. Tomkins-Lane, Does physical
469 activity influence the relationship between low back pain and obesity?, *Spine J* 14 (2014) 209–
470 216. <https://doi.org/10.1016/j.spinee.2013.11.010>.
- 471 [7] T.W. Strine, J.M. Hootman, US national prevalence and correlates of low back and neck pain
472 among adults, *Arthritis Rheum* 57 (2007) 656–665. <https://doi.org/10.1002/art.22684>.
- 473 [8] M. Scherer, H. Hansen, J. Gensichen, K. Mergenthal, S. Riedel-Heller, S. Weyerer, W. Maier,
474 A. Fuchs, H. Bickel, G. Schön, B. Wiese, H.-H. König, H. van den Bussche, I. Schäfer,
475 Association between multimorbidity patterns and chronic pain in elderly primary care patients:
476 a cross-sectional observational study, *BMC Fam Pract* 17 (2016) 68.
477 <https://doi.org/10.1186/s12875-016-0468-1>.
- 478 [9] G. Davey Smith, G. Hemani, Mendelian randomization: genetic anchors for causal inference in
479 epidemiological studies, *Human Molecular Genetics* 23 (2014) R89-98.
480 <https://doi.org/10.1093/hmg/ddu328>.
- 481 [10] M.K. Georgakis, D. Gill, Mendelian Randomization Studies in Stroke: Exploration of Risk
482 Factors and Drug Targets With Human Genetic Data, *Stroke* 52 (2021) 2992–3003.
483 <https://doi.org/10.1161/STROKEAHA.120.032617>.
- 484 [11] M.V. Holmes, M. Ala-Korpela, G.D. Smith, Mendelian randomization in cardiometabolic
485 disease: challenges in evaluating causality, *Nature Reviews. Cardiology* 14 (2017) 577–590.
486 <https://doi.org/10.1038/nrcardio.2017.78>.
- 487 [12] Y. Chen, X. Xu, L. Wang, K. Li, Y. Sun, L. Xiao, J. Dai, M. Huang, Y. Wang, D.W. Wang,
488 Genetic insights into therapeutic targets for aortic aneurysms: A Mendelian randomization
489 study, *eBioMedicine* 83 (2022) 104199. <https://doi.org/10.1016/j.ebiom.2022.104199>.
- 490 [13] J. Lin, J. Zhou, Y. Xu, Potential drug targets for multiple sclerosis identified through Mendelian
491 randomization analysis, *Brain* 146 (2023) 3364–3372. <https://doi.org/10.1093/brain/awad070>.

- 492 [14] Y. Wang, F. Liu, L. Sun, Y. Jia, P. Yang, D. Guo, M. Shi, A. Wang, G.-C. Chen, Y. Zhang, Z.
493 Zhu, Association between human blood metabolome and the risk of breast cancer, *Breast*
494 *Cancer Res* 25 (2023) 9. <https://doi.org/10.1186/s13058-023-01609-4>.
- 495 [15] S. Shah, A. Henry, C. Roselli, H. Lin, G. Sveinbjörnsson, G. Fatemifar, Å.K. Hedman, J.B.
496 Wilk, M.P. Morley, M.D. Chaffin, A. Helgadottir, N. Verweij, A. Dehghan, P. Almgren, C.
497 Andersson, K.G. Aragam, J. Ärnlöv, J.D. Backman, M.L. Biggs, H.L. Bloom, J. Brandimarto,
498 M.R. Brown, L. Buckbinder, D.J. Carey, D.I. Chasman, X. Chen, X. Chen, J. Chung, W.
499 Chutkow, J.P. Cook, G.E. Delgado, S. Denaxas, A.S. Doney, M. Dörr, S.C. Dudley, M.E. Dunn,
500 G. Engström, T. Esko, S.B. Felix, C. Finan, I. Ford, M. Ghanbari, S. Ghasemi, V. Giedraitis, F.
501 Giulianini, J.S. Gottdiener, S. Gross, D.F. Guðbjartsson, R. Gutmann, C.M. Haggerty, P. van
502 der Harst, C.L. Hyde, E. Ingelsson, J.W. Jukema, M. Kavousi, K.-T. Khaw, M.E. Kleber, L.
503 Køber, A. Koekemoer, C. Langenberg, L. Lind, C.M. Lindgren, B. London, L.A. Lotta, R.C.
504 Lovering, J. Luan, P. Magnusson, A. Mahajan, K.B. Margulies, W. März, O. Melander, I.R.
505 Mordi, T. Morgan, A.D. Morris, A.P. Morris, A.C. Morrison, M.W. Nagle, C.P. Nelson, A.
506 Niessner, T. Niiranen, M.L. O'Donoghue, A.T. Owens, C.N.A. Palmer, H.M. Parry, M. Perola,
507 E. Portilla-Fernandez, B.M. Psaty, Regeneron Genetics Center, K.M. Rice, P.M. Ridker, S.P.R.
508 Romaine, J.I. Rotter, P. Salo, V. Salomaa, J. van Setten, A.A. Shalaby, D.T. Smelser, N.L.
509 Smith, S. Stender, D.J. Stott, P. Svensson, M.-L. Tammesoo, K.D. Taylor, M. Teder-Laving, A.
510 Teumer, G. Thorgeirsson, U. Thorsteinsdottir, C. Torp-Pedersen, S. Trompet, B. Tyl, A.G.
511 Uitterlinden, A. Veluchamy, U. Völker, A.A. Voors, X. Wang, N.J. Wareham, D. Waterworth,
512 P.E. Weeke, R. Weiss, K.L. Wiggins, H. Xing, L.M. Yerges-Armstrong, B. Yu, F. Zannad, J.H.
513 Zhao, H. Hemingway, N.J. Samani, J.J.V. McMurray, J. Yang, P.M. Visscher, C. Newton-Cheh,
514 A. Malarstig, H. Holm, S.A. Lubitz, N. Sattar, M.V. Holmes, T.P. Cappola, F.W. Asselbergs,
515 A.D. Hingorani, K. Kuchenbaecker, P.T. Ellinor, C.C. Lang, K. Stefansson, J.G. Smith, R.S.
516 Vasani, D.I. Swerdlow, R.T. Lumbard, Genome-wide association and Mendelian randomisation
517 analysis provide insights into the pathogenesis of heart failure, *Nature Communications* 11
518 (2020) 163. <https://doi.org/10.1038/s41467-019-13690-5>.
- 519 [16] M.I. Kurki, J. Karjalainen, P. Palta, T.P. Sipilä, K. Kristiansson, K.M. Donner, M.P. Reeve, H.
520 Laivuori, M. Aavikko, M.A. Kaunisto, A. Loukola, E. Lahtela, H. Mattsson, P. Laiho, P. Della
521 Briotta Parolo, A.A. Lehisto, M. Kanai, N. Mars, J. Rämö, T. Kiiskinen, H.O. Heyne, K.
522 Veerapen, S. Rüeger, S. Lemmelä, W. Zhou, S. Ruotsalainen, K. Pärn, T. Hiekkalinna, S.
523 Koskelainen, T. Pajunen, V. Llorens, J. Gracia-Tabuenca, H. Siirtola, K. Reis, A.G. Elnahas,
524 B. Sun, C.N. Foley, K. Aalto-Setälä, K. Alasoo, M. Arvas, K. Auro, S. Biswas, A. Bizaki-
525 Vallaskangas, O. Carpen, C.-Y. Chen, O.A. Dada, Z. Ding, M.G. Ehm, K. Eklund, M. Färkkilä,
526 H. Finucane, A. Ganna, A. Ghazal, R.R. Graham, E.M. Green, A. Hakanen, M. Hautalahti, Å.K.
527 Hedman, M. Hiltunen, R. Hinttala, I. Hovatta, X. Hu, A. Huertas-Vazquez, L. Huilaja, J.
528 Hunkapiller, H. Jacob, J.-N. Jensen, H. Joensuu, S. John, V. Julkunen, M. Jung, J. Junttila, K.
529 Kaarniranta, M. Kähönen, R. Kajanne, L. Kallio, R. Kälviäinen, J. Kaprio, FinnGen, N.
530 Kerimov, J. Kettunen, E. Kilpeläinen, T. Kilpi, K. Klinger, V.-M. Kosma, T. Kuopio, V. Kurra,
531 T. Laisk, J. Laukkanen, N. Lawless, A. Liu, S. Longrich, R. Mägi, J. Mäkelä, A. Mäkitie, A.
532 Malarstig, A. Mannermaa, J. Maranville, A. Matakidou, T. Meretoja, S.V. Mozaffari, M.E.K.
533 Niemi, M. Niemi, T. Niiranen, C.J. O'Donnell, M.E. Obeidat, G. Okafo, H.M. Ollila, A.
534 Palomäki, T. Palotie, J. Partanen, D.S. Paul, M. Pelkonen, R.K. Pendergrass, S. Petrovski, A.
535 Pitkäranta, A. Platt, D. Pulford, E. Punkka, P. Pussinen, N. Raghavan, F. Rahimov, D. Rajpal,
536 N.A. Renaud, B. Riley-Gillis, R. Rodosthenous, E. Saarentaus, A. Salminen, E. Salminen, V.
537 Salomaa, J. Schleutker, R. Serpi, H.-Y. Shen, R. Siegel, K. Silander, S. Siltanen, S. Soini, H.
538 Soininen, J.H. Sul, I. Tachmazidou, K. Tasanen, P. Tienari, S. Toppila-Salmi, T. Tukiainen, T.

- 539 Tuomi, J.A. Turunen, J.C. Ulirsch, F. Vaura, P. Virolainen, J. Waring, D. Waterworth, R. Yang,
540 M. Nelis, A. Reigo, A. Metspalu, L. Milani, T. Esko, C. Fox, A.S. Havulinna, M. Perola, S.
541 Ripatti, A. Jalanko, T. Laitinen, T.P. Mäkelä, R. Plenge, M. McCarthy, H. Runz, M.J. Daly, A.
542 Palotie, FinnGen provides genetic insights from a well-phenotyped isolated population, *Nature*
543 613 (2023) 508–518. <https://doi.org/10.1038/s41586-022-05473-8>.
- 544 [17] U. Võsa, A. Claringbould, H.-J. Westra, M.J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha,
545 R. Kreuzhuber, S. Yazar, H. Brugge, R. Oelen, D.H. de Vries, M.G.P. van der Wijst, S. Kasela,
546 N. Pervjakova, I. Alves, M.-J. Favé, M. Agbessi, M.W. Christiansen, R. Jansen, I. Seppälä, L.
547 Tong, A. Teumer, K. Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R. Sönmez Flitman, A.
548 Brown, V. Kukushkina, A. Kalnapenkis, S. Rüeger, E. Porcu, J. Kronberg, J. Kettunen, B. Lee,
549 F. Zhang, T. Qi, J.A. Hernandez, W. Arindrarto, F. Beutner, BIOS Consortium, i2QTL
550 Consortium, J. Dmitrieva, M. Elansary, B.P. Fairfax, M. Georges, B.T. Heijmans, A.W. Hewitt,
551 M. Kähönen, Y. Kim, J.C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U.M. Marigorta, H.
552 Mei, Y. Momozawa, M. Müller-Nurasyid, M. Nauck, M.G. Nivard, B.W.J.H. Penninx, J.K.
553 Pritchard, O.T. Raitakari, O. Rotzschke, E.P. Slagboom, C.D.A. Stehouwer, M. Stumvoll, P.
554 Sullivan, P.A.C. 't Hoen, J. Thiery, A. Tönjes, J. van Dongen, M. van Itersen, J.H. Veldink, U.
555 Völker, R. Warmerdam, C. Wijmenga, M. Swertz, A. Andiappan, G.W. Montgomery, S.
556 Ripatti, M. Perola, Z. Kutalik, E. Dermitzakis, S. Bergmann, T. Frayling, J. van Meurs, H.
557 Prokisch, H. Ahsan, B.L. Pierce, T. Lehtimäki, D.I. Boomsma, B.M. Psaty, S.A. Gharib, P.
558 Awadalla, L. Milani, W.H. Ouwehand, K. Downes, O. Stegle, A. Battle, P.M. Visscher, J.
559 Yang, M. Scholz, J. Powell, G. Gibson, T. Esko, L. Franke, Large-scale cis- and trans-eQTL
560 analyses identify thousands of genetic loci and polygenic scores that regulate blood gene
561 expression, *Nature Genetics* 53 (2021) 1300–1310. [https://doi.org/10.1038/s41588-021-00913-](https://doi.org/10.1038/s41588-021-00913-z)
562 [z](https://doi.org/10.1038/s41588-021-00913-z).
- 563 [18] GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human
564 tissues, *Science (New York, N.Y.)* 369 (2020) 1318–1330.
565 <https://doi.org/10.1126/science.aaz1776>.
- 566 [19] Z. Z, Z. F, H. H, B. A, R. Mr, P. Je, M. Gw, G. Me, W. Nr, V. Pm, Y. J, Integration of summary
567 data from GWAS and eQTL studies predicts complex trait gene targets, *Nature Genetics* 48
568 (2016). <https://doi.org/10.1038/ng.3538>.
- 569 [20] C.A. de Leeuw, J.M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set analysis
570 of GWAS data, *PLoS Comput Biol* 11 (2015) e1004219.
571 <https://doi.org/10.1371/journal.pcbi.1004219>.
- 572 [21] The 1000 Genomes Project Consortium, Corresponding authors, A. Auton, G.R. Abecasis,
573 Steering committee, D.M. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, A.
574 Chakravarti, A.G. Clark, P. Donnelly, E.E. Eichler, P. Flicek, S.B. Gabriel, R.A. Gibbs, E.D.
575 Green, M.E. Hurles, B.M. Knoppers, J.O. Korbel, E.S. Lander, C. Lee, H. Lehrach, E.R.
576 Mardis, G.T. Marth, G.A. McVean, D.A. Nickerson, J.P. Schmidt, S.T. Sherry, J. Wang, R.K.
577 Wilson, Production group, Baylor College of Medicine, R.A. Gibbs, E. Boerwinkle, H.
578 Doddapaneni, Y. Han, V. Korchina, C. Kovar, S. Lee, D. Muzny, J.G. Reid, Y. Zhu, BGI-
579 Shenzhen, J. Wang, Y. Chang, Q. Feng, X. Fang, X. Guo, M. Jian, H. Jiang, X. Jin, T. Lan, G.
580 Li, J. Li, Y. Li, S. Liu, X. Liu, Y. Lu, X. Ma, M. Tang, B. Wang, G. Wang, H. Wu, R. Wu, X.
581 Xu, Y. Yin, D. Zhang, W. Zhang, J. Zhao, M. Zhao, X. Zheng, Broad Institute of MIT and
582 Harvard, E.S. Lander, D.M. Altshuler, S.B. Gabriel, N. Gupta, Coriell Institute for Medical

583 Research, N. Gharani, L.H. Toji, N.P. Gerry, A.M. Resch, European Molecular Biology
584 Laboratory, European Bioinformatics Institute, P. Flicek, J. Barker, L. Clarke, L. Gil, S.E. Hunt,
585 G. Kelman, E. Kulesha, R. Leinonen, W.M. McLaren, R. Radhakrishnan, A. Roa, D. Smirnov,
586 R.E. Smith, I. Streeter, A. Thormann, I. Toneva, B. Vaughan, X. Zheng-Bradley, Illumina, D.R.
587 Bentley, R. Grocock, S. Humphray, T. James, Z. Kingsbury, Max Planck Institute for Molecular
588 Genetics, H. Lehrach, R. Sudbrak, M.W. Albrecht, V.S. Amstislavskiy, T.A. Borodina, M.
589 Lienhard, F. Mertes, M. Sultan, B. Timmermann, M.-L. Yaspo, McDonnell Genome Institute at
590 Washington University, E.R. Mardis, R.K. Wilson, L. Fulton, R. Fulton, US National Institutes
591 of Health, S.T. Sherry, V. Ananiev, Z. Belaia, D. Beloslyudtsev, N. Bouk, C. Chen, D. Church,
592 R. Cohen, C. Cook, J. Garner, T. Hefferon, M. Kimelman, C. Liu, J. Lopez, P. Meric, C.
593 O'Sullivan, Y. Ostapchuk, L. Phan, S. Ponomarov, V. Schneider, E. Shekhtman, K. Sirotkin, D.
594 Slotta, H. Zhang, University of Oxford, G.A. McVean, Wellcome Trust Sanger Institute, R.M.
595 Durbin, S. Balasubramaniam, J. Burton, P. Danecek, T.M. Keane, A. Kolb-Kokocinski, S.
596 McCarthy, J. Stalker, M. Quail, Analysis group, Affymetrix, J.P. Schmidt, C.J. Davies, J.
597 Gollub, T. Webster, B. Wong, Y. Zhan, Albert Einstein College of Medicine, A. Auton, C.L.
598 Campbell, Y. Kong, A. Marcketta, Baylor College of Medicine, R.A. Gibbs, F. Yu, L. Antunes,
599 M. Bainbridge, D. Muzny, A. Sabo, Z. Huang, BGI-Shenzhen, J. Wang, L.J.M. Coin, L. Fang,
600 X. Guo, X. Jin, G. Li, Q. Li, Y. Li, Z. Li, H. Lin, B. Liu, R. Luo, H. Shao, Y. Xie, C. Ye, C. Yu,
601 F. Zhang, H. Zheng, H. Zhu, Bilkent University, C. Alkan, E. Dal, F. Kahveci, Boston College,
602 G.T. Marth, E.P. Garrison, D. Kural, W.-P. Lee, W. Fung Leong, M. Stromberg, A.N. Ward, J.
603 Wu, M. Zhang, Broad Institute of MIT and Harvard, M.J. Daly, M.A. DePristo, R.E. Handsaker,
604 D.M. Altshuler, E. Banks, G. Bhatia, G. Del Angel, S.B. Gabriel, G. Genovese, N. Gupta, H. Li,
605 S. Kashin, E.S. Lander, S.A. McCarroll, J.C. Nemes, R.E. Poplin, Cold Spring Harbor
606 Laboratory, S.C. Yoon, J. Lihm, V. Makarov, Cornell University, A.G. Clark, S. Gottipati, A.
607 Keinan, J.L. Rodriguez-Flores, European Molecular Biology Laboratory, J.O. Korbel, T.
608 Rausch, M.H. Fritz, A.M. Stütz, European Molecular Biology Laboratory, European
609 Bioinformatics Institute, P. Flicek, K. Beal, L. Clarke, A. Datta, J. Herrero, W.M. McLaren,
610 G.R.S. Ritchie, R.E. Smith, D. Zerbino, X. Zheng-Bradley, Harvard University, P.C. Sabeti, I.
611 Shlyakhter, S.F. Schaffner, J. Vitti, Human Gene Mutation Database, D.N. Cooper, E.V. Ball,
612 P.D. Stenson, Illumina, D.R. Bentley, B. Barnes, M. Bauer, R. Keira Cheetham, A. Cox, M.
613 Eberle, S. Humphray, S. Kahn, L. Murray, J. Peden, R. Shaw, Icahn School of Medicine at
614 Mount Sinai, E.E. Kenny, Louisiana State University, M.A. Batzer, M.K. Konkel, J.A. Walker,
615 Massachusetts General Hospital, D.G. MacArthur, M. Lek, Max Planck Institute for Molecular
616 Genetics, R. Sudbrak, V.S. Amstislavskiy, R. Herwig, McDonnell Genome Institute at
617 Washington University, E.R. Mardis, L. Ding, D.C. Koboldt, D. Larson, K. Ye, McGill
618 University, S. Gravel, National Eye Institute, NIH, A. Swaroop, E. Chew, New York Genome
619 Center, T. Lappalainen, Y. Erlich, M. Gymrek, T. Frederick Willems, Ontario Institute for
620 Cancer Research, J.T. Simpson, Pennsylvania State University, M.D. Shriver, Rutgers Cancer
621 Institute of New Jersey, J.A. Rosenfeld, Stanford University, C.D. Bustamante, S.B.
622 Montgomery, F.M. De La Vega, J.K. Byrnes, A.W. Carroll, M.K. DeGorter, P. Lacroute, B.K.
623 Maples, A.R. Martin, A. Moreno-Estrada, S.S. Shringarpure, F. Zakharia, Tel-Aviv University,
624 E. Halperin, Y. Baran, The Jackson Laboratory for Genomic Medicine, C. Lee, E. Cerveira, J.
625 Hwang, A. Malhotra, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, Thermo Fisher
626 Scientific, F.C.L. Hyland, Translational Genomics Research Institute, D.W. Craig, A.
627 Christoforides, N. Homer, T. Izatt, A.A. Kurdoglu, S.A. Sinari, K. Squire, US National
628 Institutes of Health, S.T. Sherry, C. Xiao, University of California, San Diego, J. Sebat, D.
629 Antaki, M. Gujral, A. Noor, K. Ye, University of California, San Francisco, E.G. Burchard,
630 R.D. Hernandez, C.R. Gignoux, University of California, Santa Cruz, D. Haussler, S.J.
631 Katzman, W. James Kent, University of Chicago, B. Howie, University College London, A.

632 Ruiz-Linares, University of Geneva, E.T. Dermitzakis, University of Maryland School of
633 Medicine, S.E. Devine, University of Michigan, G.R. Abecasis, H. Min Kang, J.M. Kidd, T.
634 Blackwell, S. Caron, W. Chen, S. Emery, L. Fritsche, C. Fuchsberger, G. Jun, B. Li, R. Lyons,
635 C. Scheller, C. Sidore, S. Song, E. Sliwerska, D. Taliun, A. Tan, R. Welch, M. Kate Wing, X.
636 Zhan, University of Montréal, P. Awadalla, A. Hodgkinson, University of North Carolina at
637 Chapel Hill, Y. Li, University of North Carolina at Charlotte, X. Shi, A. Quitadamo, University
638 of Oxford, G. Lunter, G.A. McVean, J.L. Marchini, S. Myers, C. Churchhouse, O. Delaneau, A.
639 Gupta-Hinch, W. Kretschmar, Z. Iqbal, I. Mathieson, A. Menelaou, A. Rimmer, D.K. Xifara,
640 University of Puerto Rico, T.K. Oleksyk, University of Texas Health Sciences Center at
641 Houston, Y. Fu, X. Liu, M. Xiong, University of Utah, L. Jorde, D. Witherspoon, J. Xing,
642 University of Washington, E.E. Eichler, B.L. Browning, S.R. Browning, F. Hormozdiari, P.H.
643 Sudmant, Weill Cornell Medical College, E. Khurana, Wellcome Trust Sanger Institute, R.M.
644 Durbin, M.E. Hurles, C. Tyler-Smith, C.A. Albers, Q. Ayub, S. Balasubramaniam, Y. Chen, V.
645 Colonna, P. Danecek, L. Jostins, T.M. Keane, S. McCarthy, K. Walter, Y. Xue, Yale University,
646 M.B. Gerstein, A. Abyzov, S. Balasubramaniam, J. Chen, D. Clarke, Y. Fu, A.O. Harmanci, M.
647 Jin, D. Lee, J. Liu, X. Jasmine Mu, J. Zhang, Y. Zhang, Structural variation group, BGI-
648 Shenzhen, Y. Li, R. Luo, H. Zhu, Bilkent University, C. Alkan, E. Dal, F. Kahveci, Boston
649 College, G.T. Marth, E.P. Garrison, D. Kural, W.-P. Lee, A.N. Ward, J. Wu, M. Zhang, Broad
650 Institute of MIT and Harvard, S.A. McCarroll, R.E. Handsaker, D.M. Altshuler, E. Banks, G.
651 Del Angel, G. Genovese, C. Hartl, H. Li, S. Kashin, J.C. Nemes, K. Shakir, Cold Spring
652 Harbor Laboratory, S.C. Yoon, J. Lihm, V. Makarov, Cornell University, J. Degenhardt,
653 European Molecular Biology Laboratory, J.O. Korbel, M.H. Fritz, S. Meiers, B. Raeder, T.
654 Rausch, A.M. Stütz, European Molecular Biology Laboratory, European Bioinformatics
655 Institute, P. Flicek, F. Paolo Casale, L. Clarke, R.E. Smith, O. Stegle, X. Zheng-Bradley,
656 Illumina, D.R. Bentley, B. Barnes, R. Keira Cheetham, M. Eberle, S. Humphray, S. Kahn, L.
657 Murray, R. Shaw, Leiden University Medical Center, E.-W. Lammeijer, Louisiana State
658 University, M.A. Batzer, M.K. Konkel, J.A. Walker, McDonnell Genome Institute at
659 Washington University, L. Ding, I. Hall, K. Ye, Stanford University, P. Lacroute, The Jackson
660 Laboratory for Genomic Medicine, C. Lee, E. Cerveira, A. Malhotra, J. Hwang, D.
661 Plewczynski, K. Radew, M. Romanovitch, C. Zhang, Translational Genomics Research
662 Institute, D.W. Craig, N. Homer, US National Institutes of Health, D. Church, C. Xiao,
663 University of California, San Diego, J. Sebat, D. Antaki, V. Bafna, J. Michaelson, K. Ye,
664 University of Maryland School of Medicine, S.E. Devine, E.J. Gardner, University of Michigan,
665 G.R. Abecasis, J.M. Kidd, R.E. Mills, G. Dayama, S. Emery, G. Jun, University of North
666 Carolina at Charlotte, X. Shi, A. Quitadamo, University of Oxford, G. Lunter, G.A. McVean,
667 University of Texas MD Anderson Cancer Center, K. Chen, X. Fan, Z. Chong, T. Chen,
668 University of Utah, D. Witherspoon, J. Xing, University of Washington, E.E. Eichler, M.J.
669 Chaisson, F. Hormozdiari, J. Huddleston, M. Malig, B.J. Nelson, P.H. Sudmant, Vanderbilt
670 University School of Medicine, N.F. Parrish, Weill Cornell Medical College, E. Khurana,
671 Wellcome Trust Sanger Institute, M.E. Hurles, B. Blackburne, S.J. Lindsay, Z. Ning, K. Walter,
672 Y. Zhang, Yale University, M.B. Gerstein, A. Abyzov, J. Chen, D. Clarke, H. Lam, X. Jasmine
673 Mu, C. Sisu, J. Zhang, Y. Zhang, Exome group, Baylor College of Medicine, R.A. Gibbs, F.
674 Yu, M. Bainbridge, D. Challis, U.S. Evani, C. Kovar, J. Lu, D. Muzny, U. Nagaswamy, J.G.
675 Reid, A. Sabo, J. Yu, BGI-Shenzhen, X. Guo, W. Li, Y. Li, R. Wu, Boston College, G.T. Marth,
676 E.P. Garrison, W. Fung Leong, A.N. Ward, Broad Institute of MIT and Harvard, G. Del Angel,
677 M.A. DePristo, S.B. Gabriel, N. Gupta, C. Hartl, R.E. Poplin, Cornell University, A.G. Clark,
678 J.L. Rodriguez-Flores, European Molecular Biology Laboratory, European Bioinformatics
679 Institute, P. Flicek, L. Clarke, R.E. Smith, X. Zheng-Bradley, Massachusetts General Hospital,
680 D.G. MacArthur, McDonnell Genome Institute at Washington University, E.R. Mardis, R.

681 Fulton, D.C. Koboldt, McGill University, S. Gravel, Stanford University, C.D. Bustamante,
682 Translational Genomics Research Institute, D.W. Craig, A. Christoforides, N. Homer, T. Izatt,
683 US National Institutes of Health, S.T. Sherry, C. Xiao, University of Geneva, E.T. Dermitzakis,
684 University of Michigan, G.R. Abecasis, H. Min Kang, University of Oxford, G.A. McVean,
685 Yale University, M.B. Gerstein, S. Balasubramanian, L. Habegger, Functional interpretation
686 group, Cornell University, H. Yu, European Molecular Biology Laboratory, European
687 Bioinformatics Institute, P. Flicek, L. Clarke, F. Cunningham, I. Dunham, D. Zerbino, X.
688 Zheng-Bradley, Harvard University, K. Lage, J. Berg Jespersen, H. Horn, Stanford University,
689 S.B. Montgomery, M.K. DeGorter, Weill Cornell Medical College, E. Khurana, Wellcome
690 Trust Sanger Institute, C. Tyler-Smith, Y. Chen, V. Colonna, Y. Xue, Yale University, M.B.
691 Gerstein, S. Balasubramanian, Y. Fu, D. Kim, Chromosome Y group, Albert Einstein College of
692 Medicine, A. Auton, A. Marcketta, American Museum of Natural History, R. Desalle, A.
693 Narechania, Arizona State University, M.A. Wilson Sayres, Boston College, E.P. Garrison,
694 Broad Institute of MIT and Harvard, R.E. Handsaker, S. Kashin, S.A. McCarroll, Cornell
695 University, J.L. Rodriguez-Flores, European Molecular Biology Laboratory, European
696 Bioinformatics Institute, P. Flicek, L. Clarke, X. Zheng-Bradley, New York Genome Center, Y.
697 Erlich, M. Gymrek, T. Frederick Willems, Stanford University, C.D. Bustamante, F.L. Mendez,
698 G. David Poznik, P.A. Underhill, The Jackson Laboratory for Genomic Medicine, C. Lee, E.
699 Cerveira, A. Malhotra, M. Romanovitch, C. Zhang, University of Michigan, G.R. Abecasis,
700 University of Queensland, L. Coin, H. Shao, Virginia Bioinformatics Institute, D. Mittelman,
701 Wellcome Trust Sanger Institute, C. Tyler-Smith, Q. Ayub, R. Banerjee, M. Cerezo, Y. Chen,
702 T.W. Fitzgerald, S. Louzada, A. Massaia, S. McCarthy, G.R. Ritchie, Y. Xue, F. Yang, Data
703 coordination center group, Baylor College of Medicine, R.A. Gibbs, C. Kovar, D. Kalra, W.
704 Hale, D. Muzny, J.G. Reid, BGI-Shenzhen, J. Wang, X. Dan, X. Guo, G. Li, Y. Li, C. Ye, X.
705 Zheng, Broad Institute of MIT and Harvard, D.M. Altshuler, European Molecular Biology
706 Laboratory, European Bioinformatics Institute, P. Flicek, L. Clarke, X. Zheng-Bradley,
707 Illumina, D.R. Bentley, A. Cox, S. Humphray, S. Kahn, Max Planck Institute for Molecular
708 Genetics, R. Sudbrak, M.W. Albrecht, M. Lienhard, McDonnell Genome Institute at
709 Washington University, D. Larson, Translational Genomics Research Institute, D.W. Craig, T.
710 Izatt, A.A. Kurdoglu, US National Institutes of Health, S.T. Sherry, C. Xiao, University of
711 California, Santa Cruz, D. Haussler, University of Michigan, G.R. Abecasis, University of
712 Oxford, G.A. McVean, Wellcome Trust Sanger Institute, R.M. Durbin, S. Balasubramanian,
713 T.M. Keane, S. McCarthy, J. Stalker, Samples and ELSI group, A. Chakravarti, B.M. Knoppers,
714 G.R. Abecasis, K.C. Barnes, C. Beiswanger, E.G. Burchard, C.D. Bustamante, H. Cai, H. Cao,
715 R.M. Durbin, N.P. Gerry, N. Gharani, R.A. Gibbs, C.R. Gignoux, S. Gravel, B. Henn, D. Jones,
716 L. Jorde, J.S. Kaye, A. Keinan, A. Kent, A. Kerasidou, Y. Li, R. Mathias, G.A. McVean, A.
717 Moreno-Estrada, P.N. Ossorio, M. Parker, A.M. Resch, C.N. Rotimi, C.D. Royal, K. Sandoval,
718 Y. Su, R. Sudbrak, Z. Tian, S. Tishkoff, L.H. Toji, C. Tyler-Smith, M. Via, Y. Wang, H. Yang,
719 L. Yang, J. Zhu, Sample collection, British from England and Scotland (GBR), W. Bodmer,
720 Colombians in Medellín, Colombia (CLM), G. Bedoya, A. Ruiz-Linares, Han Chinese South
721 (CHS), Z. Cai, Y. Gao, J. Chu, Finnish in Finland (FIN), L. Peltonen, Iberian Populations in
722 Spain (IBS), A. Garcia-Montero, A. Orfao, Puerto Ricans in Puerto Rico (PUR), J. Dutil, J.C.
723 Martinez-Cruzado, T.K. Oleksyk, African Caribbean in Barbados (ACB), K.C. Barnes, R.A.
724 Mathias, A. Hennis, H. Watson, C. McKenzie, Bengali in Bangladesh (BEB), F. Qadri, R.
725 LaRocque, P.C. Sabeti, Chinese Dai in Xishuangbanna, China (CDX), J. Zhu, X. Deng, Esan in
726 Nigeria (ESN), P.C. Sabeti, D. Asogun, O. Folarin, C. Happi, O. Omoniwa, M. Stremlau, R.
727 Tariyal, Gambian in Western Division – Mandinka (GWD), M. Jallow, F. Sisay Joof, T. Corrah,
728 K. Rockett, D. Kwiatkowski, Indian Telugu in the UK (ITU) and Sri Lankan Tamil in the UK
729 (STU), J. Kooner, Kinh in Ho Chi Minh City, Vietnam (KHV), T. Tjinh Hiên, S.J. Dunstan, N.

- 730 Thuy Hang, Mende in Sierra Leone (MSL), R. Fonnies, R. Garry, L. Kanneh, L. Moses, P.C.
731 Sabeti, J. Schieffelin, D.S. Grant, Peruvian in Lima, Peru (PEL), C. Gallo, G. Poletti, Punjabi in
732 Lahore, Pakistan (PIL), D. Saleheen, A. Rasheed, Scientific management, L.D. Brooks, A.L.
733 Felsenfeld, J.E. McEwen, Y. Vaydylevich, E.D. Green, A. Duncanson, M. Dunn, J.A. Schloss,
734 J. Wang, H. Yang, Writing group, A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H. Min
735 Kang, J.O. Korbel, J.L. Marchini, S. McCarthy, G.A. McVean, G.R. Abecasis, A global
736 reference for human genetic variation, *Nature* 526 (2015) 68–74.
737 <https://doi.org/10.1038/nature15393>.
- 738 [22] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B.W.J.H. Penninx, R. Jansen, E.J.C. de Geus,
739 D.I. Boomsma, F.A. Wright, P.F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A.J. Lusis, T.
740 Lehtimäki, E. Raitoharju, M. Kähönen, I. Seppälä, O.T. Raitakari, J. Kuusisto, M. Laakso, A.L.
741 Price, P. Pajukanta, B. Pasaniuc, Integrative approaches for large-scale transcriptome-wide
742 association studies, *Nat Genet* 48 (2016) 245–252. <https://doi.org/10.1038/ng.3506>.
- 743 [23] Y. Hu, M. Li, Q. Lu, H. Weng, J. Wang, S.M. Zekavat, Z. Yu, B. Li, J. Gu, S. Muchnik, Y. Shi,
744 B.W. Kunkle, S. Mukherjee, P. Natarajan, A. Naj, A. Kuzma, Y. Zhao, P.K. Crane, H. Lu, H.
745 Zhao, A statistical framework for cross-tissue transcriptome-wide association analysis, *Nat*
746 *Genet* 51 (2019) 568–576. <https://doi.org/10.1038/s41588-019-0345-7>.
- 747 [24] G. Hemani, J. Zheng, B. Elsworth, K.H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess,
748 J. Bowden, R. Langdon, V.Y. Tan, J. Yarmolinsky, H.A. Shihab, N.J. Timpson, D.M. Evans, C.
749 Relton, R.M. Martin, G. Davey Smith, T.R. Gaunt, P.C. Haycock, The MR-Base platform
750 supports systematic causal inference across the human phenome, *eLife* 7 (2018) e34408.
751 <https://doi.org/10.7554/eLife.34408>.
- 752 [25] C. Giambartolomei, D. Vukcevic, E.E. Schadt, L. Franke, A.D. Hingorani, C. Wallace, V.
753 Plagnol, Bayesian test for colocalisation between pairs of genetic association studies using
754 summary statistics, *PLoS Genetics* 10 (2014) e1004383.
755 <https://doi.org/10.1371/journal.pgen.1004383>.
- 756 [26] J. Zhu, C. Ma, L. Zhu, J. Li, F. Peng, L. Huang, X. Luan, A role for the NLRC4 inflammasome
757 in premature rupture of membrane, *PLoS ONE* 15 (2020) e0237847.
758 <https://doi.org/10.1371/journal.pone.0237847>.
- 759 [27] D. Zheng, T. Liwinski, E. Elinav, Inflammasome activation and regulation: toward a better
760 understanding of complex mechanisms, *Cell Discov* 6 (2020) 36.
761 <https://doi.org/10.1038/s41421-020-0167-x>.
- 762 [28] Y. An, Z. Zhai, X. Wang, Y. Ding, L. He, L. Li, Q. Mo, C. Mu, R. Xie, T. Liu, W. Zhong, B.
763 Wang, H. Cao, Targeting *Desulfovibrio vulgaris* flagellin-induced NAIP/NLRC4
764 inflammasome activation in macrophages attenuates ulcerative colitis, *Journal of Advanced*
765 *Research* 52 (2023) 219–232. <https://doi.org/10.1016/j.jare.2023.08.008>.
- 766 [29] J. Sim, J.W. Ahn, J. Park, Y.J. Kim, J.-Y. Jeong, J.M. Lee, K. Cho, H.J. Ahn, K.S. Sung, J.-S.
767 Moon, J.H. Moon, J. Lim, Non-canonical NLRC4 inflammasomes in astrocytes contribute to
768 glioma malignancy, *Inflamm. Res.* 72 (2023) 813–827. [https://doi.org/10.1007/s00011-023-](https://doi.org/10.1007/s00011-023-01710-6)
769 [01710-6](https://doi.org/10.1007/s00011-023-01710-6).

- 770 [30] C. Delgado-Arévalo, M. Calvet-Mirabent, A. Triguero-Martínez, E. Vázquez De Luis, A.
771 Benguría-Filippini, R. Largo, D. Calzada-Fraile, O. Popova, I. Sánchez-Cerrillo, I. Tsukalov, R.
772 Moreno-Vellisca, H. De La Fuente, G. Herrero-Beaumont, A. Ramiro, F. Sánchez-Madrid, S.
773 Castañeda, A. Dopazo, I. González Álvaro, E. Martín-Gayo, NLRC4-mediated activation of
774 CD1c+ DC contributes to perpetuation of synovitis in rheumatoid arthritis, *JCI Insight* 7 (2022)
775 e152886. <https://doi.org/10.1172/jci.insight.152886>.
- 776 [31] R.E. Vance, The NAIP/NLRC4 inflammasomes, *Current Opinion in Immunology* 32 (2015)
777 84–89. <https://doi.org/10.1016/j.coi.2015.01.010>.
- 778 [32] W. Deng, L. Wang, Y. Xiong, J. Li, Y. Wang, T. Shi, D. Ma, The novel secretory protein
779 CGREF1 inhibits the activation of AP-1 transcriptional activity and cell proliferation, *The*
780 *International Journal of Biochemistry & Cell Biology* 65 (2015) 32–39.
781 <https://doi.org/10.1016/j.biocel.2015.05.019>.
- 782 [33] M. Xiang, Y. Gao, Y. Zhou, M. Wang, X. Yao, A novel nomogram based on cell cycle-related
783 genes for predicting overall survival in early-onset colorectal cancer, *BMC Cancer* 23 (2023)
784 595. <https://doi.org/10.1186/s12885-023-11075-y>.
- 785 [34] M. Xie, H. Dai, Q. Gu, C. Xiao, H. Wang, Y. Lei, C. Wu, X. Li, B. Lin, S. Li, Identification of
786 genes contributing to cisplatin resistance in osteosarcoma cells, *FEBS Open Bio* 13 (2023) 164–
787 173. <https://doi.org/10.1002/2211-5463.13524>.
- 788 [35] M.T. Le, M.A. Lanaspá, C.M. Cicerchi, J. Rana, J.D. Scholten, B.L. Hunter, C.J. Rivard, R.K.
789 Randolph, R.J. Johnson, Bioactivity-Guided Identification of Botanical Inhibitors of
790 Ketoheokinase, *PLoS ONE* 11 (2016) e0157458.
791 <https://doi.org/10.1371/journal.pone.0157458>.
- 792 [36] J. Kim, J. Kang, Y.-L. Kang, J. Woo, Y. Kim, J. Huh, J.-W. Park, Ketoheokinase-A acts as a
793 nuclear protein kinase that mediates fructose-induced metastasis in breast cancer, *Nat Commun*
794 11 (2020) 5436. <https://doi.org/10.1038/s41467-020-19263-1>.
- 795 [37] X. Li, X. Qian, L.-X. Peng, Y. Jiang, D.H. Hawke, Y. Zheng, Y. Xia, J.-H. Lee, G. Cote, H.
796 Wang, L. Wang, C.-N. Qian, Z. Lu, A splicing switch from ketoheokinase-C to
797 ketoheokinase-A drives hepatocellular carcinoma formation, *Nat Cell Biol* 18 (2016) 561–571.
798 <https://doi.org/10.1038/ncb3338>.
- 799 [38] J. Yang, S. Yang, Q. Wang, J. Pang, Y. Wang, H. Wang, X. Fu, KHK-A promotes the
800 proliferation of oesophageal squamous cell carcinoma through the up-regulation of PRPS1,
801 *Arab Journal of Gastroenterology* 22 (2021) 40–46. <https://doi.org/10.1016/j.ajg.2020.08.007>.
- 802 [39] J.A. Johnston, D.R. Nelson, P. Bhatnagar, S.E. Curtis, Y. Chen, J.G. MacKrell, Prevalence and
803 cardiometabolic correlates of ketoheokinase gene variants among UK Biobank participants,
804 *PLoS ONE* 16 (2021) e0247683. <https://doi.org/10.1371/journal.pone.0247683>.
- 805 [40] C.-H. Yu, T. Xie, R.-P. Zhang, Z.-C. A, Association of the common SNPs in RNF212, STAG3
806 and RFX2 gene with male infertility with azoospermia in Chinese population, *European Journal*
807 *of Obstetrics & Gynecology and Reproductive Biology* 221 (2018) 109–112.
808 <https://doi.org/10.1016/j.ejogrb.2017.12.030>.

- 809 [41] A. Riera-Escamilla, A. Enguita-Marruedo, D. Moreno-Mendoza, C. Chianese, E. Sleddens-
 810 Linkels, E. Contini, M. Benelli, A. Natali, G.M. Colpi, E. Ruiz-Castañé, M. Maggi, W.M.
 811 Baarends, C. Krausz, Sequencing of a ‘mouse azoospermia’ gene panel in azoospermic men:
 812 identification of RNF212 and STAG3 mutations as novel genetic causes of meiotic arrest,
 813 Human Reproduction 34 (2019) 978–988. <https://doi.org/10.1093/humrep/dez042>.
- 814 [42] A. Reynolds, H. Qiao, Y. Yang, J.K. Chen, N. Jackson, K. Biswas, J.K. Holloway, F. Baudat, B.
 815 De Massy, J. Wang, C. Höög, P.E. Cohen, N. Hunter, RNF212 is a dosage-sensitive regulator of
 816 crossing-over during mammalian meiosis, Nat Genet 45 (2013) 269–278.
 817 <https://doi.org/10.1038/ng.2541>.
- 818 [43] Y. Fujiwara, H. Matsumoto, K. Akiyama, A. Srivastava, M. Chikushi, M. Ann Handel, T.
 819 Kunieda, An ENU-induced mutation in the mouse Rnf212 gene is associated with male meiotic
 820 failure and infertility, REPRODUCTION 149 (2015) 67–74. <https://doi.org/10.1530/REP-14-0122>.
- 822 [44] H. Qiao, H.B.D. Prasada Rao, Y. Yang, J.H. Fong, J.M. Cloutier, D.C. Deacon, K.E. Nagel,
 823 R.K. Swartz, E. Strong, J.K. Holloway, P.E. Cohen, J. Schimenti, J. Ward, N. Hunter,
 824 Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic
 825 recombination, Nat Genet 46 (2014) 194–199. <https://doi.org/10.1038/ng.2858>.

826

827 9 Tables

828 9.1 Table 1

Gene	Ch	Position	Z_TWAS	P_TWAS	P_TWAS (FDR)	Tissue	topSNP	B_SMR	SE_SMR	P_SMR	P_SMR (FDR)	P_HEIDI	N_HEIDI
NLRC4	2	32265853	-5.33	9.8E-08	5.1E-05	GTEXv8.Whole_Blood	rs385076	-0.13	0.02	7.9E-09	9.5E-06	3.6E-01	20
CGREF1	2	27119114	-4.76	1.9E-06	5.0E-04	GTEXv8.Whole_Blood	rs2304681	-0.38	0.09	2.1E-05	4.6E-03	4.4E-01	20
KHK	2	27086746	-4.73	2.3E-06	5.7E-04	GTEXv8.Whole_Blood	rs2119026	-0.06	0.01	2.3E-06	1.0E-03	2.5E-01	20
RNF212	4	1113561	3.75	1.8E-04	0.01	GTEXv8.Whole_Blood	rs4974591	0.19	0.04	1.2E-07	9.4E-05	5.5E-02	20

829 **Table 1** TWAS / SMR/HEIDI results of the GWAS data on Dorsopathies, blood eQTL
 830 data. TWAS/SMR/HEIDI results of the GWAS data on Dorsopathies, blood eQTL data. Ch
 831 represents chromosome; position indicates the gene's position, Z_TWAS, P_TWAS are the Z-score
 832 and p-value of the TWAS test; Tissue represents the tissue source used in the TWAS analyses.
 833 topSNP represents the top SNP in the SMR analyses. P_SMR is the p-value for the SMR test;
 834 B_SMR is the effect size from the SMR test; SE_SMR is the standard error of B_SMR; P_HEIDI is

835 the p-value for the HEIDI test; N_HEIDI is the number of SNPs used in the HEIDI test; FDR
 836 represents false discovery rate.

837

838 **9.2 Table 2**

839

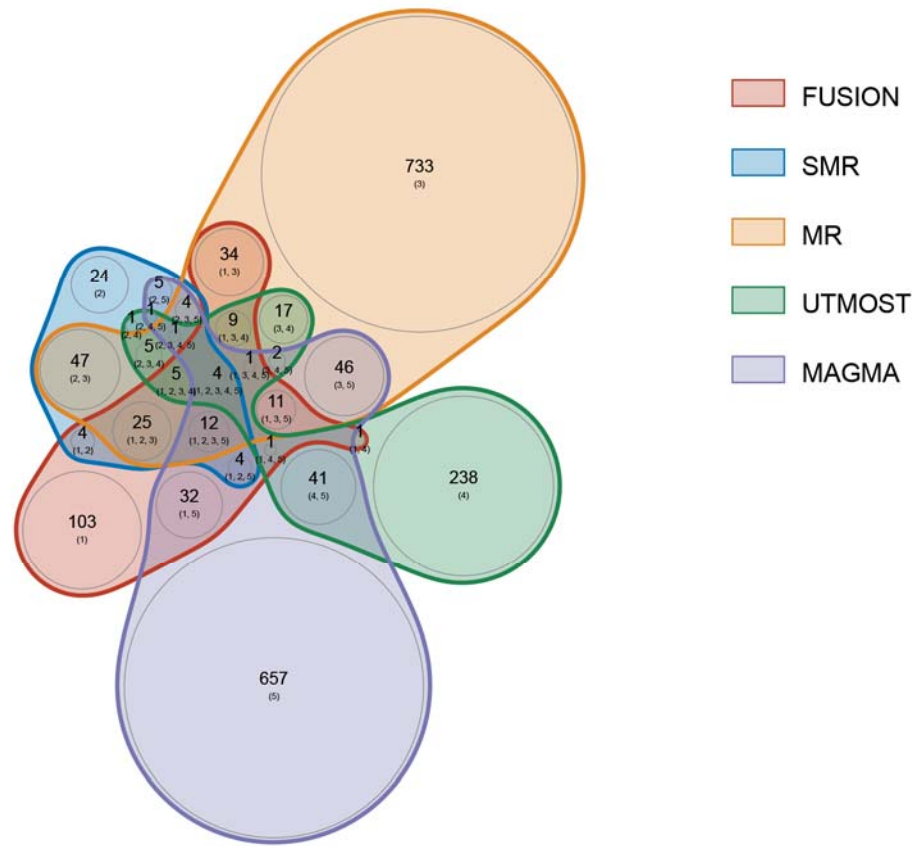
Marker	Gene	Methylation on Gene			Methylation on Dorsopathies			Gene on Dorsopathies			Intermediary analyses			prop mediated	prop mediated se
		b	se	p	b	se	p	b	se	p	indirect	indirect_se	indirect_p		
cg04686953	RNF212	-0.52	0.07	2.08E-14	-0.12	0.02	2.68E-06	0.19	0.04	1.20E-07	-0.1	0.02	1.35E-05	93.79%	0.29
cg26638505	NLRC4	-0.81	0.13	1.25E-09	0.09	0.03	1.43E-03	-0.13	0.02	7.87E-09	0.11	0.03	2.87E-05	120.00%	0.47
cg18948125	NLRC4	-0.93	0.17	1.72E-08	0.12	0.03	5.59E-04	-0.13	0.02	7.87E-09	0.12	0.03	5.50E-05	104.41%	0.4
cg22740783	CGREF1	0.32	0.06	7.71E-07	-0.13	0.04	1.63E-04	-0.38	0.09	2.12E-05	-0.12	0.04	1.27E-03	92.24%	0.38
cg06112415	CGREF1	-0.34	0.07	1.75E-06	0.12	0.03	5.59E-04	-0.38	0.09	2.12E-05	0.13	0.04	1.49E-03	110.63%	0.47

840 **Table 2** Main Results of Intermediary analyses. b is the effect size, se is the standard error of effect
 841 size, and p is the P-value. prop_mediated represents the proportion of indirect effect in the total
 842 effect.

843

844 **10 Figures**

845 **10.1 Figure 1**

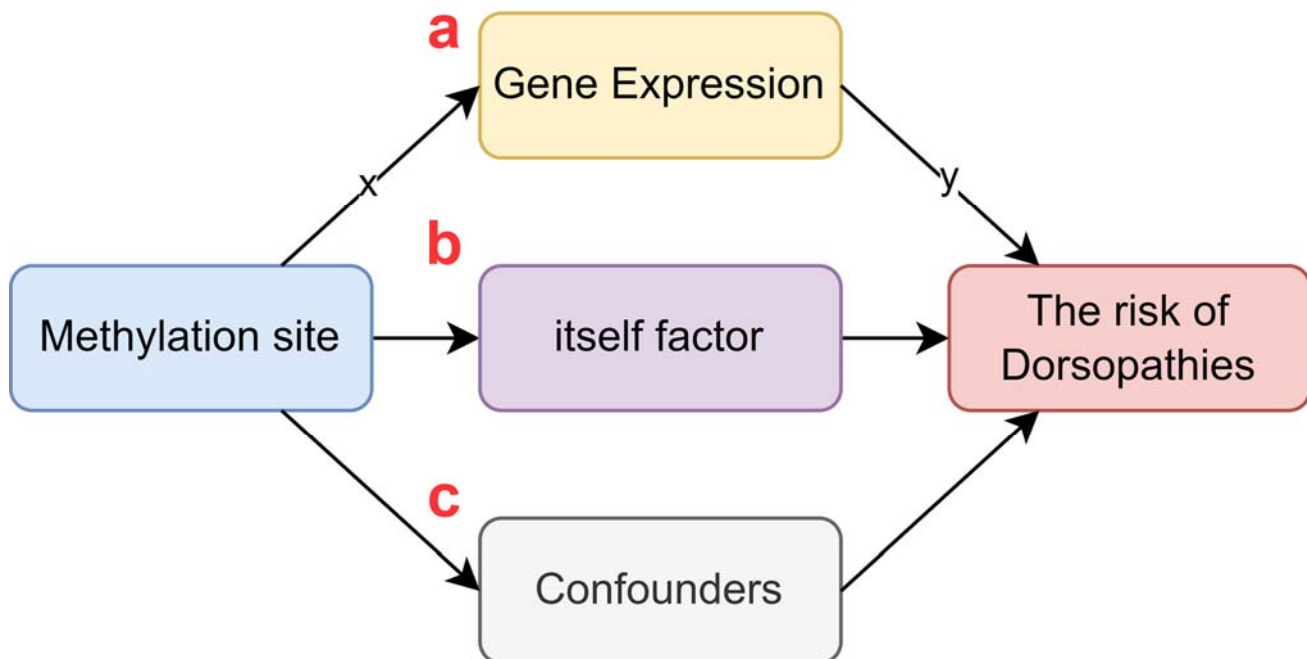


846

847 **Fig.1** Venn Plot of the main results of multi analysis

848

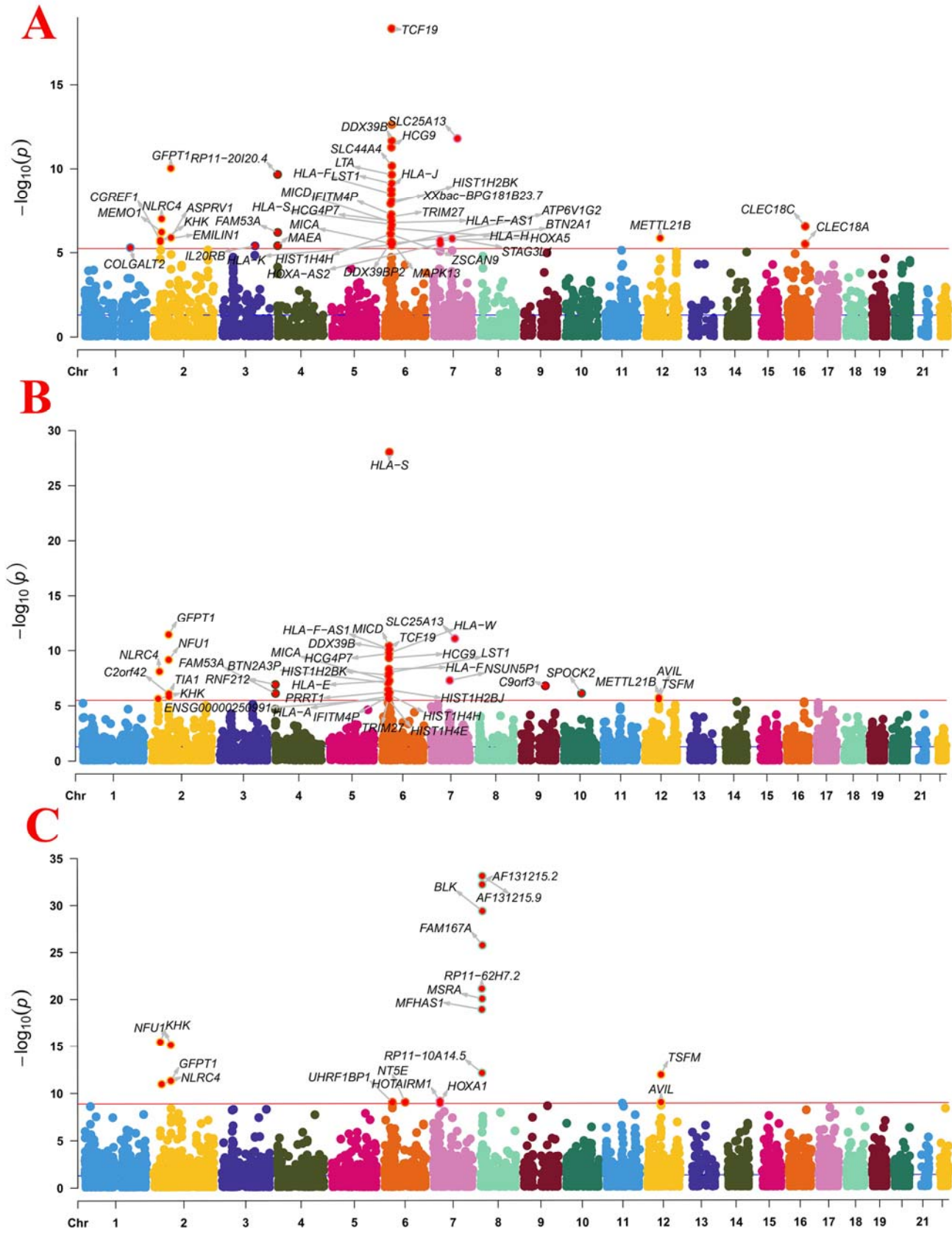
849 **10.2 Figure 2**



850

851 **Fig.2** Methylation sites affect disease pathways.

852 **10.3 Figure 3**



854 **Fig.3** Manhattan plot of the MR/SMR/FUSION analysis results using QTLs and Dorsopathies
 855 GWAS summary statistics. The red dashed line represents the Bonferroni-corrected significance
 856 threshold.

857

858 **10.4 Figure 4**

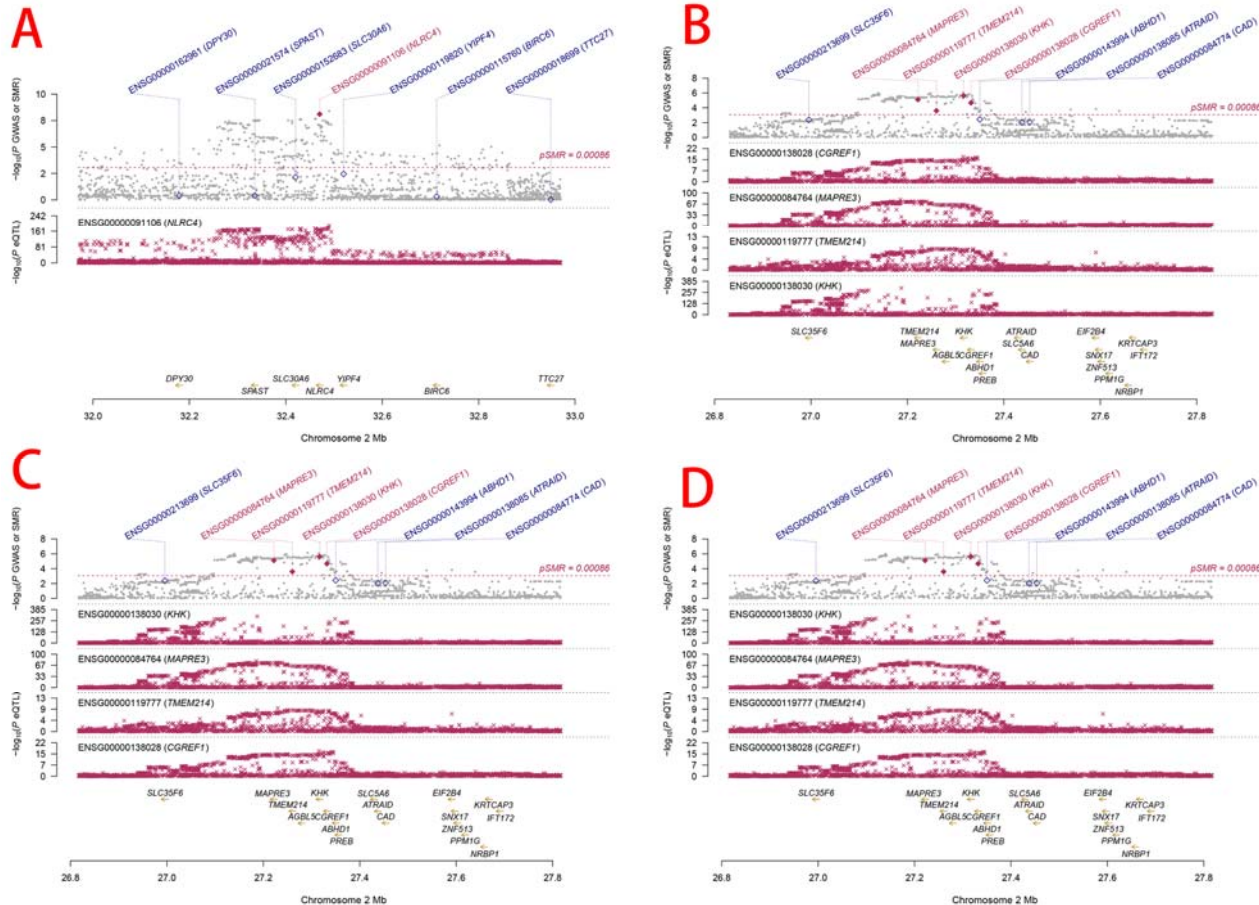
Gene	method	nSNP	pval	OR (95% CI)
NLRC4	Inverse variance weighted	21	1.09e-11	0.886 (0.855 - 0.917)
CGREF1	Wald ratio	1	8.74e-07	0.682 (0.585 - 0.794)
KHK	Inverse variance weighted	32	3.67e-16	0.936 (0.921 - 0.951)
RNF212	Inverse variance weighted	11	1.01e-03	1.145 (1.056 - 1.241)

859

860 **Fig.4** MR results for genes expression significantly associated with dorsopathies after FDR
 861 correction.

862

863 **10.5 Figure 5**

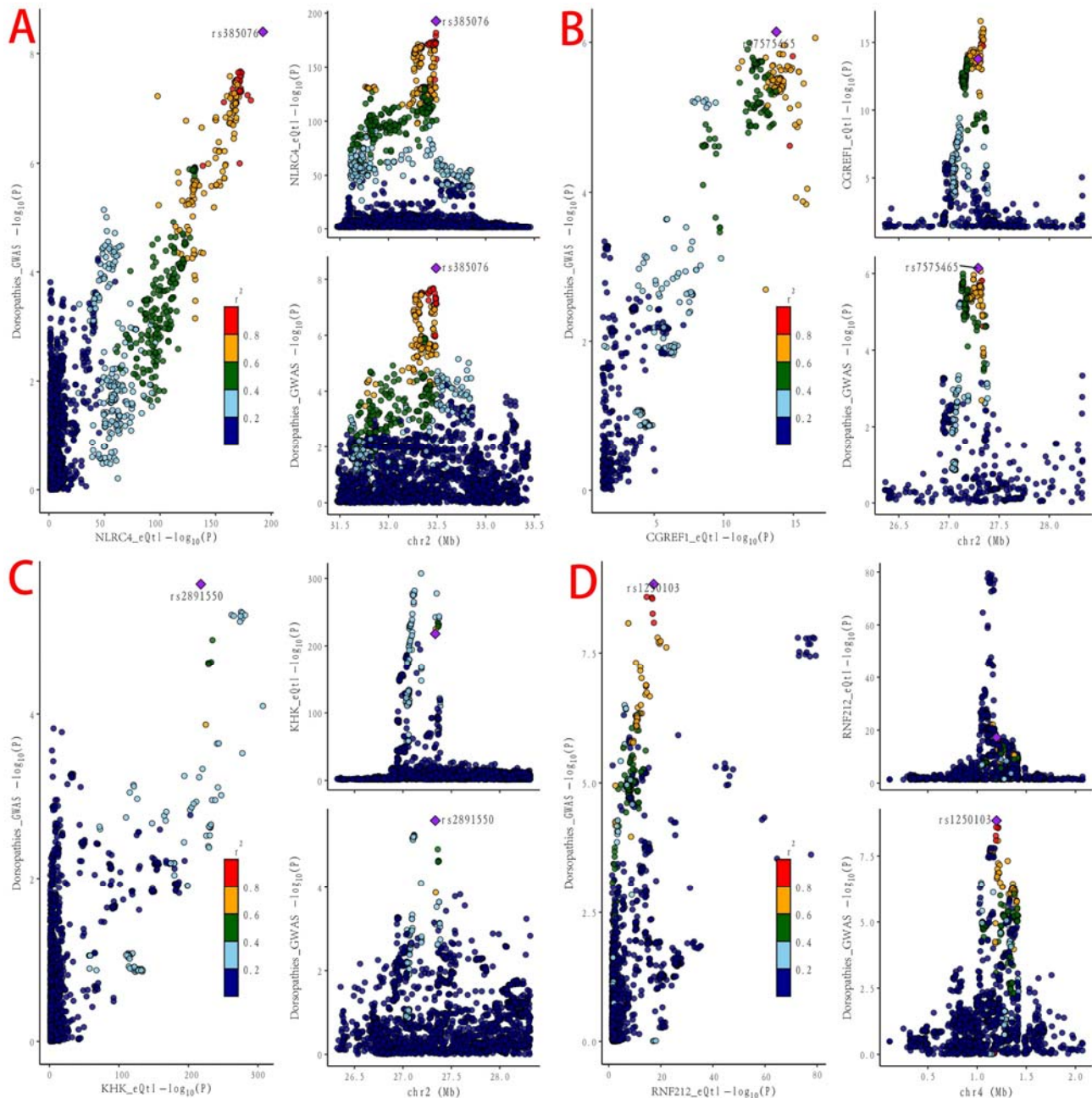


864

865 **Fig. 5** SMR locus plot illustrating dorsopathies at the gene locus utilizing blood eQTL data. In the
866 upper plot, each gray dot represents a SNP identified through GWAS on dorsopathies. A red
867 diamond indicates passage of the SMR test for the probe, while a solid diamond signifies successful
868 completion of both the SMR and HEIDI tests. In the lower plot, each red cross represents an SNP
869 identified in the eQTL study corresponding to each gene. The x-axis displays the genomic positions
870 (Mb, GRCh37) of SNPs, probes, and genes on the chromosome. The y-axis displays the negative
871 logarithm (base 10) of p-values for SNPs identified in the GWAS on dorsopathies, SMR test, and
872 eQTL study corresponding to each gene.

873

874 **10.6 Figure 6**



875

876 **Fig. 6** Locus comparison plot displays the results of colocalization analysis for SNPs associated with
877 gene expression in blood and datasets related to dorsopathies. Each dot represents a specific SNP,
878 with the color indicating its linkage disequilibrium (LD) value (r^2) with the lead GWAS variant,
879 marked by a purple diamond. In the right panel, the x-axis represents genomic positions in
880 megabases (GRCh37) along the chromosome, while the y-axis shows the $-\log_{10}$ p-values for SNPs
881 from the dorsopathies GWAS (top) and the gene expression eQTL study (bottom). The left panel
882 compares the p-values from the dorsopathies GWAS with those from the gene expression eQTL
883 study.

884