medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Genome-wide Mendelian Randomization Identifies Potential Drug Targets for Dorsopathies

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,*], Shijie Tang[2,3,4,*]

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

[1]Shantou University Medical College, Shantou, Guangdong 515041, China

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

[3]Plastic Surgery Institute of Shantou University Medical College, Shantou, Guangdong 515051, China

[4]Shantou Plastic surgery Clinical Research Center, Shantou, Guangdong 515051, China

*** Correspondence:**
Wancong Zhang, Shijie Tang
Martine2007@sina.com , sjtang3@stu.edu.cn

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

Abstract

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

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

1 Introduction

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

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

2 Method

2.1 Datasets

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

2.2 SMR analyses

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

2.3 GWAS analyses

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

2.4 TWAS analyses

We conducted validations to integrate dorsopathies GWAS and eQTL data of whole blood from GTEx using the FUSION and UTMOST, widely utilized tools in prior TWAS investigations [22,23]. FUSION constructs predictive models using various penalized linear models, such as GBLUP, LASSO, Elastic Net, for the significant cis-heritability genes estimated from SNPs within 500 kb on either side of the gene boundary. Subsequently, it selects the optimal model based on the coefficient of determination (R2) calculated through a fivefold cross-validation. For UTMOST, we performed repeated 49 single-tissue association tests for each tissue. TWAS significance for both single-tissue analyses was determined with a Benjamini-Hochberg corrected FDR value below 0.05.

2.5 MR analyses

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

2.6 Colocalization analyses

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

2.7 Methylation & Mediation analysis

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

3 Result

3.1 SMR analyses and colocalization for Preliminary Identifying Potential Genes

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

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

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

3.2 MR analyses Validates Potential Gene Causal Relationships

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

3.3 Validation

3.3.1 TWAS & UTMOST Validates Transcriptome-level Causal Relationships

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

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

3.3.2 MAGMA Validates Genome-level Causal Relationships

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

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

3.4 Methylation analysis & Mediation analysis

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

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

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

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

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

4 Discussion

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

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

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

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

5 Acknowledgements

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

6 Conflict of interest

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

7 Funding

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

8 Reference

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[2] L. March, E.U.R. Smith, D.G. Hoy, M.J. Cross, L. Sanchez-Riera, F. Blyth, R. Buchbinder, T. Vos, A.D. Woolf, Burden of disability due to musculoskeletal (MSK) disorders, Best Pract Res Clin Rheumatol 28 (2014) 353–366. https://doi.org/10.1016/j.berh.2014.08.002.

[3] R. Shiri, J. Karppinen, P. Leino-Arjas, S. Solovieva, E. Viikari-Juntura, The association between smoking and low back pain: a meta-analysis, Am J Med 123 (2010) 87.e7–35. https://doi.org/10.1016/j.amjmed.2009.05.028.

[4] R. Shiri, J. Karppinen, P. Leino-Arjas, S. Solovieva, E. Viikari-Juntura, The association between obesity and low back pain: a meta-analysis, Am J Epidemiol 171 (2010) 135–154. https://doi.org/10.1093/aje/kwp356.

[5] A. Koyanagi, A. Stickley, N. Garin, M. Miret, J.L. Ayuso-Mateos, M. Leonardi, S. Koskinen, A. Galas, J.M. Haro, The association between obesity and back pain in nine countries: a cross-sectional study, BMC Public Health 15 (2015) 123. https://doi.org/10.1186/s12889-015-1362-9.

[6] M. Smuck, M.-C.J. Kao, N. Brar, A. Martinez-Ith, J. Choi, C.C. Tomkins-Lane, Does physical activity influence the relationship between low back pain and obesity?, Spine J 14 (2014) 209–216. https://doi.org/10.1016/j.spinee.2013.11.010.

[7] T.W. Strine, J.M. Hootman, US national prevalence and correlates of low back and neck pain among adults, Arthritis Rheum 57 (2007) 656–665. https://doi.org/10.1002/art.22684.

[8] M. Scherer, H. Hansen, J. Gensichen, K. Mergenthal, S. Riedel-Heller, S. Weyerer, W. Maier, A. Fuchs, H. Bickel, G. Schön, B. Wiese, H.-H. König, H. van den Bussche, I. Schäfer, Association between multimorbidity patterns and chronic pain in elderly primary care patients: a cross-sectional observational study, BMC Fam Pract 17 (2016) 68. https://doi.org/10.1186/s12875-016-0468-1.

[9] G. Davey Smith, G. Hemani, Mendelian randomization: genetic anchors for causal inference in epidemiological studies, Human Molecular Genetics 23 (2014) R89-98. https://doi.org/10.1093/hmg/ddu328.

[10] M.K. Georgakis, D. Gill, Mendelian Randomization Studies in Stroke: Exploration of Risk Factors and Drug Targets With Human Genetic Data, Stroke 52 (2021) 2992–3003. https://doi.org/10.1161/STROKEAHA.120.032617.

[11] M.V. Holmes, M. Ala-Korpela, G.D. Smith, Mendelian randomization in cardiometabolic disease: challenges in evaluating causality, Nature Reviews. Cardiology 14 (2017) 577–590. https://doi.org/10.1038/nrcardio.2017.78.

[12] Y. Chen, X. Xu, L. Wang, K. Li, Y. Sun, L. Xiao, J. Dai, M. Huang, Y. Wang, D.W. Wang, Genetic insights into therapeutic targets for aortic aneurysms: A Mendelian randomization study, eBioMedicine 83 (2022) 104199. https://doi.org/10.1016/j.ebiom.2022.104199.

[13] J. Lin, J. Zhou, Y. Xu, Potential drug targets for multiple sclerosis identified through Mendelian randomization analysis, Brain 146 (2023) 3364–3372. https://doi.org/10.1093/brain/awad070.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[14] Y. Wang, F. Liu, L. Sun, Y. Jia, P. Yang, D. Guo, M. Shi, A. Wang, G.-C. Chen, Y. Zhang, Z. Zhu, Association between human blood metabolome and the risk of breast cancer, Breast Cancer Res 25 (2023) 9. https://doi.org/10.1186/s13058-023-01609-4.

[15] S. Shah, A. Henry, C. Roselli, H. Lin, G. Sveinbjörnsson, G. Fatemifar, Å.K. Hedman, J.B. Wilk, M.P. Morley, M.D. Chaffin, A. Helgadottir, N. Verweij, A. Dehghan, P. Almgren, C. Andersson, K.G. Aragam, J. Ärnlöv, J.D. Backman, M.L. Biggs, H.L. Bloom, J. Brandimarto, M.R. Brown, L. Buckbinder, D.J. Carey, D.I. Chasman, X. Chen, X. Chen, J. Chung, W. Chutkow, J.P. Cook, G.E. Delgado, S. Denaxas, A.S. Doney, M. Dörr, S.C. Dudley, M.E. Dunn, G. Engström, T. Esko, S.B. Felix, C. Finan, I. Ford, M. Ghanbari, S. Ghasemi, V. Giedraitis, F. Giulianini, J.S. Gottdiener, S. Gross, D.F. Guðbjartsson, R. Gutmann, C.M. Haggerty, P. van der Harst, C.L. Hyde, E. Ingelsson, J.W. Jukema, M. Kavousi, K.-T. Khaw, M.E. Kleber, L. Køber, A. Koekemoer, C. Langenberg, L. Lind, C.M. Lindgren, B. London, L.A. Lotta, R.C. Lovering, J. Luan, P. Magnusson, A. Mahajan, K.B. Margulies, W. März, O. Melander, I.R. Mordi, T. Morgan, A.D. Morris, A.P. Morris, A.C. Morrison, M.W. Nagle, C.P. Nelson, A. Niessner, T. Niiranen, M.L. O'Donoghue, A.T. Owens, C.N.A. Palmer, H.M. Parry, M. Perola, E. Portilla-Fernandez, B.M. Psaty, Regeneron Genetics Center, K.M. Rice, P.M. Ridker, S.P.R. Romaine, J.I. Rotter, P. Salo, V. Salomaa, J. van Setten, A.A. Shalaby, D.T. Smelser, N.L. Smith, S. Stender, D.J. Stott, P. Svensson, M.-L. Tammesoo, K.D. Taylor, M. Teder-Laving, A. Teumer, G. Thorgeirsson, U. Thorsteinsdottir, C. Torp-Pedersen, S. Trompet, B. Tyl, A.G. Uitterlinden, A. Veluchamy, U. Völker, A.A. Voors, X. Wang, N.J. Wareham, D. Waterworth, P.E. Weeke, R. Weiss, K.L. Wiggins, H. Xing, L.M. Yerges-Armstrong, B. Yu, F. Zannad, J.H. Zhao, H. Hemingway, N.J. Samani, J.J.V. McMurray, J. Yang, P.M. Visscher, C. Newton-Cheh, A. Malarstig, H. Holm, S.A. Lubitz, N. Sattar, M.V. Holmes, T.P. Cappola, F.W. Asselbergs, A.D. Hingorani, K. Kuchenbaecker, P.T. Ellinor, C.C. Lang, K. Stefansson, J.G. Smith, R.S. Vasan, D.I. Swerdlow, R.T. Lumbers, Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure, Nature Communications 11 (2020) 163. https://doi.org/10.1038/s41467-019-13690-5.

[16] M.I. Kurki, J. Karjalainen, P. Palta, T.P. Sipilä, K. Kristiansson, K.M. Donner, M.P. Reeve, H. Laivuori, M. Aavikko, M.A. Kaunisto, A. Loukola, E. Lahtela, H. Mattsson, P. Laiho, P. Della Briotta Parolo, A.A. Lehisto, M. Kanai, N. Mars, J. Rämö, T. Kiiskinen, H.O. Heyne, K. Veerapen, S. Rüeger, S. Lemmelä, W. Zhou, S. Ruotsalainen, K. Pärn, T. Hiekkalinna, S. Koskelainen, T. Paajanen, V. Llorens, J. Gracia-Tabuenca, H. Siirtola, K. Reis, A.G. Elnahas, B. Sun, C.N. Foley, K. Aalto-Setälä, K. Alasoo, M. Arvas, K. Auro, S. Biswas, A. Bizaki-Vallaskangas, O. Carpen, C.-Y. Chen, O.A. Dada, Z. Ding, M.G. Ehm, K. Eklund, M. Färkkilä, H. Finucane, A. Ganna, A. Ghazal, R.R. Graham, E.M. Green, A. Hakanen, M. Hautalahti, Å.K. Hedman, M. Hiltunen, R. Hinttala, I. Hovatta, X. Hu, A. Huertas-Vazquez, L. Huilaja, J. Hunkapiller, H. Jacob, J.-N. Jensen, H. Joensuu, S. John, V. Julkunen, M. Jung, J. Junttila, K. Kaarniranta, M. Kähönen, R. Kajanne, L. Kallio, R. Kälviäinen, J. Kaprio, FinnGen, N. Kerimov, J. Kettunen, E. Kilpeläinen, T. Kilpi, K. Klinger, V.-M. Kosma, T. Kuopio, V. Kurra, T. Laisk, J. Laukkanen, N. Lawless, A. Liu, S. Longerich, R. Mägi, J. Mäkelä, A. Mäkitie, A. Malarstig, A. Mannermaa, J. Maranville, A. Matakidou, T. Meretoja, S.V. Mozaffari, M.E.K. Niemi, M. Niemi, T. Niiranen, C.J. O Donnell, M.E. Obeidat, G. Okafo, H.M. Ollila, A. Palomäki, T. Palotie, J. Partanen, D.S. Paul, M. Pelkonen, R.K. Pendergrass, S. Petrovski, A. Pitkäranta, A. Platt, D. Pulford, E. Punkka, P. Pussinen, N. Raghavan, F. Rahimov, D. Rajpal, N.A. Renaud, B. Riley-Gillis, R. Rodosthenous, E. Saarentaus, A. Salminen, E. Salminen, V. Salomaa, J. Schleutker, R. Serpi, H.-Y. Shen, R. Siegel, K. Silander, S. Siltanen, S. Soini, H. Soininen, J.H. Sul, I. Tachmazidou, K. Tasanen, P. Tienari, S. Toppila-Salmi, T. Tukiainen, T.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Tuomi, J.A. Turunen, J.C. Ulirsch, F. Vaura, P. Virolainen, J. Waring, D. Waterworth, R. Yang, M. Nelis, A. Reigo, A. Metspalu, L. Milani, T. Esko, C. Fox, A.S. Havulinna, M. Perola, S. Ripatti, A. Jalanko, T. Laitinen, T.P. Mäkelä, R. Plenge, M. McCarthy, H. Runz, M.J. Daly, A. Palotie, FinnGen provides genetic insights from a well-phenotyped isolated population, Nature 613 (2023) 508–518. https://doi.org/10.1038/s41586-022-05473-8.

[17] U. Võsa, A. Claringbould, H.-J. Westra, M.J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber, S. Yazar, H. Brugge, R. Oelen, D.H. de Vries, M.G.P. van der Wijst, S. Kasela, N. Pervjakova, I. Alves, M.-J. Favé, M. Agbessi, M.W. Christiansen, R. Jansen, I. Seppälä, L. Tong, A. Teumer, K. Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R. Sönmez Flitman, A. Brown, V. Kukushkina, A. Kalnapenkis, S. Rüeger, E. Porcu, J. Kronberg, J. Kettunen, B. Lee, F. Zhang, T. Qi, J.A. Hernandez, W. Arindrarto, F. Beutner, BIOS Consortium, i2QTL Consortium, J. Dmitrieva, M. Elansary, B.P. Fairfax, M. Georges, B.T. Heijmans, A.W. Hewitt, M. Kähönen, Y. Kim, J.C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U.M. Marigorta, H. Mei, Y. Momozawa, M. Müller-Nurasyid, M. Nauck, M.G. Nivard, B.W.J.H. Penninx, J.K. Pritchard, O.T. Raitakari, O. Rotzschke, E.P. Slagboom, C.D.A. Stehouwer, M. Stumvoll, P. Sullivan, P.A.C. ’t Hoen, J. Thiery, A. Tönjes, J. van Dongen, M. van Iterson, J.H. Veldink, U. Völker, R. Warmerdam, C. Wijmenga, M. Swertz, A. Andiappan, G.W. Montgomery, S. Ripatti, M. Perola, Z. Kutalik, E. Dermitzakis, S. Bergmann, T. Frayling, J. van Meurs, H. Prokisch, H. Ahsan, B.L. Pierce, T. Lehtimäki, D.I. Boomsma, B.M. Psaty, S.A. Gharib, P. Awadalla, L. Milani, W.H. Ouwehand, K. Downes, O. Stegle, A. Battle, P.M. Visscher, J. Yang, M. Scholz, J. Powell, G. Gibson, T. Esko, L. Franke, Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression, Nature Genetics 53 (2021) 1300–1310. https://doi.org/10.1038/s41588-021-00913-z.

[18] GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues, Science (New York, N.Y.) 369 (2020) 1318–1330. https://doi.org/10.1126/science.aaz1776.

[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 data from GWAS and eQTL studies predicts complex trait gene targets, Nature Genetics 48 (2016). https://doi.org/10.1038/ng.3538.

[20] C.A. de Leeuw, J.M. Mooij, T. Heskes, D. Posthuma, MAGMA: generalized gene-set analysis of GWAS data, PLoS Comput Biol 11 (2015) e1004219. https://doi.org/10.1371/journal.pcbi.1004219.

[21] The 1000 Genomes Project Consortium, Corresponding authors, A. Auton, G.R. Abecasis, Steering committee, D.M. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, A. Chakravarti, A.G. Clark, P. Donnelly, E.E. Eichler, P. Flicek, S.B. Gabriel, R.A. Gibbs, E.D. Green, M.E. Hurles, B.M. Knoppers, J.O. Korbel, E.S. Lander, C. Lee, H. Lehrach, E.R. Mardis, G.T. Marth, G.A. McVean, D.A. Nickerson, J.P. Schmidt, S.T. Sherry, J. Wang, R.K. Wilson, Production group, Baylor College of Medicine, R.A. Gibbs, E. Boerwinkle, H. Doddapaneni, Y. Han, V. Korchina, C. Kovar, S. Lee, D. Muzny, J.G. Reid, Y. Zhu, BGI-Shenzhen, J. Wang, Y. Chang, Q. Feng, X. Fang, X. Guo, M. Jian, H. Jiang, X. Jin, T. Lan, G. Li, J. Li, Y. Li, S. Liu, X. Liu, Y. Lu, X. Ma, M. Tang, B. Wang, G. Wang, H. Wu, R. Wu, X. Xu, Y. Yin, D. Zhang, W. Zhang, J. Zhao, M. Zhao, X. Zheng, Broad Institute of MIT and Harvard, E.S. Lander, D.M. Altshuler, S.B. Gabriel, N. Gupta, Coriell Institute for Medical

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Thuy Hang, Mende in Sierra Leone (MSL), R. Fonnie, R. Garry, L. Kanneh, L. Moses, P.C. Sabeti, J. Schieffelin, D.S. Grant, Peruvian in Lima, Peru (PEL), C. Gallo, G. Poletti, Punjabi in Lahore, Pakistan (PJL), D. Saleheen, A. Rasheed, Scientific management, L.D. Brooks, A.L. Felsenfeld, J.E. McEwen, Y. Vaydylevich, E.D. Green, A. Duncanson, M. Dunn, J.A. Schloss, J. Wang, H. Yang, Writing group, A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H. Min Kang, J.O. Korbel, J.L. Marchini, S. McCarthy, G.A. McVean, G.R. Abecasis, A global reference for human genetic variation, Nature 526 (2015) 68–74. https://doi.org/10.1038/nature15393.

[22] A. Gusev, A. Ko, H. Shi, G. Bhatia, W. Chung, B.W.J.H. Penninx, R. Jansen, E.J.C. de Geus, D.I. Boomsma, F.A. Wright, P.F. Sullivan, E. Nikkola, M. Alvarez, M. Civelek, A.J. Lusis, T. Lehtimäki, E. Raitoharju, M. Kähönen, I. Seppälä, O.T. Raitakari, J. Kuusisto, M. Laakso, A.L. Price, P. Pajukanta, B. Pasaniuc, Integrative approaches for large-scale transcriptome-wide association studies, Nat Genet 48 (2016) 245–252. https://doi.org/10.1038/ng.3506.

[23] Y. Hu, M. Li, Q. Lu, H. Weng, J. Wang, S.M. Zekavat, Z. Yu, B. Li, J. Gu, S. Muchnik, Y. Shi, B.W. Kunkle, S. Mukherjee, P. Natarajan, A. Naj, A. Kuzma, Y. Zhao, P.K. Crane, H. Lu, H. Zhao, A statistical framework for cross-tissue transcriptome-wide association analysis, Nat Genet 51 (2019) 568–576. https://doi.org/10.1038/s41588-019-0345-7.

[24] G. Hemani, J. Zheng, B. Elsworth, K.H. Wade, V. Haberland, D. Baird, C. Laurin, S. Burgess, J. Bowden, R. Langdon, V.Y. Tan, J. Yarmolinsky, H.A. Shihab, N.J. Timpson, D.M. Evans, C. Relton, R.M. Martin, G. Davey Smith, T.R. Gaunt, P.C. Haycock, The MR-Base platform supports systematic causal inference across the human phenome, eLife 7 (2018) e34408. https://doi.org/10.7554/eLife.34408.

[25] C. Giambartolomei, D. Vukcevic, E.E. Schadt, L. Franke, A.D. Hingorani, C. Wallace, V. Plagnol, Bayesian test for colocalisation between pairs of genetic association studies using summary statistics, PLoS Genetics 10 (2014) e1004383. https://doi.org/10.1371/journal.pgen.1004383.

[26] J. Zhu, C. Ma, L. Zhu, J. Li, F. Peng, L. Huang, X. Luan, A role for the NLRC4 inflammasome in premature rupture of membrane, PLoS ONE 15 (2020) e0237847. https://doi.org/10.1371/journal.pone.0237847.

[27] D. Zheng, T. Liwinski, E. Elinav, Inflammasome activation and regulation: toward a better understanding of complex mechanisms, Cell Discov 6 (2020) 36. https://doi.org/10.1038/s41421-020-0167-x.

[28] Y. An, Z. Zhai, X. Wang, Y. Ding, L. He, L. Li, Q. Mo, C. Mu, R. Xie, T. Liu, W. Zhong, B. Wang, H. Cao, Targeting Desulfovibrio vulgaris flagellin-induced NAIP/NLRC4 inflammasome activation in macrophages attenuates ulcerative colitis, Journal of Advanced Research 52 (2023) 219–232. https://doi.org/10.1016/j.jare.2023.08.008.

[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. Moon, J.H. Moon, J. Lim, Non-canonical NLRC4 inflammasomes in astrocytes contribute to glioma malignancy, Inflamm. Res. 72 (2023) 813–827. https://doi.org/10.1007/s00011-023-01710-6.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[30] C. Delgado-Arévalo, M. Calvet-Mirabent, A. Triguero-Martínez, E. Vázquez De Luis, A. Benguría-Filippini, R. Largo, D. Calzada-Fraile, O. Popova, I. Sánchez-Cerrillo, I. Tsukalov, R. Moreno-Vellisca, H. De La Fuente, G. Herrero-Beaumont, A. Ramiro, F. Sánchez-Madrid, S. Castañeda, A. Dopazo, I. González Álvaro, E. Martin-Gayo, NLRC4-mediated activation of CD1c+ DC contributes to perpetuation of synovitis in rheumatoid arthritis, JCI Insight 7 (2022) e152886. https://doi.org/10.1172/jci.insight.152886.

[31] R.E. Vance, The NAIP/NLRC4 inflammasomes, Current Opinion in Immunology 32 (2015) 84–89. https://doi.org/10.1016/j.coi.2015.01.010.

[32] W. Deng, L. Wang, Y. Xiong, J. Li, Y. Wang, T. Shi, D. Ma, The novel secretory protein CGREF1 inhibits the activation of AP-1 transcriptional activity and cell proliferation, The International Journal of Biochemistry & Cell Biology 65 (2015) 32–39. https://doi.org/10.1016/j.biocel.2015.05.019.

[33] M. Xiang, Y. Gao, Y. Zhou, M. Wang, X. Yao, A novel nomogram based on cell cycle-related genes for predicting overall survival in early-onset colorectal cancer, BMC Cancer 23 (2023) 595. https://doi.org/10.1186/s12885-023-11075-y.

[34] M. Xie, H. Dai, Q. Gu, C. Xiao, H. Wang, Y. Lei, C. Wu, X. Li, B. Lin, S. Li, Identification of genes contributing to cisplatin resistance in osteosarcoma cells, FEBS Open Bio 13 (2023) 164–173. https://doi.org/10.1002/2211-5463.13524.

[35] M.T. Le, M.A. Lanaspa, C.M. Cicerchi, J. Rana, J.D. Scholten, B.L. Hunter, C.J. Rivard, R.K. Randolph, R.J. Johnson, Bioactivity-Guided Identification of Botanical Inhibitors of Ketohexokinase, PLoS ONE 11 (2016) e0157458. https://doi.org/10.1371/journal.pone.0157458.

[36] J. Kim, J. Kang, Y.-L. Kang, J. Woo, Y. Kim, J. Huh, J.-W. Park, Ketohexokinase-A acts as a nuclear protein kinase that mediates fructose-induced metastasis in breast cancer, Nat Commun 11 (2020) 5436. https://doi.org/10.1038/s41467-020-19263-1.

[37] X. Li, X. Qian, L.-X. Peng, Y. Jiang, D.H. Hawke, Y. Zheng, Y. Xia, J.-H. Lee, G. Cote, H. Wang, L. Wang, C.-N. Qian, Z. Lu, A splicing switch from ketohexokinase-C to ketohexokinase-A drives hepatocellular carcinoma formation, Nat Cell Biol 18 (2016) 561–571. https://doi.org/10.1038/ncb3338.

[38] J. Yang, S. Yang, Q. Wang, J. Pang, Y. Wang, H. Wang, X. Fu, KHK-A promotes the proliferation of oesophageal squamous cell carcinoma through the up-regulation of PRPS1, Arab Journal of Gastroenterology 22 (2021) 40–46. https://doi.org/10.1016/j.ajg.2020.08.007.

[39] J.A. Johnston, D.R. Nelson, P. Bhatnagar, S.E. Curtis, Y. Chen, J.G. MacKrell, Prevalence and cardiometabolic correlates of ketohexokinase gene variants among UK Biobank participants, PLoS ONE 16 (2021) e0247683. https://doi.org/10.1371/journal.pone.0247683.

[40] C.-H. Yu, T. Xie, R.-P. Zhang, Z.-C. A, Association of the common SNPs in RNF212, STAG3 and RFX2 gene with male infertility with azoospermia in Chinese population, European Journal of Obstetrics & Gynecology and Reproductive Biology 221 (2018) 109–112. https://doi.org/10.1016/j.ejogrb.2017.12.030.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[41] A. Riera-Escamilla, A. Enguita-Marruedo, D. Moreno-Mendoza, C. Chianese, E. Sleddens-Linkels, E. Contini, M. Benelli, A. Natali, G.M. Colpi, E. Ruiz-Castañé, M. Maggi, W.M. Baarends, C. Krausz, Sequencing of a 'mouse azoospermia' gene panel in azoospermic men: identification of RNF212 and STAG3 mutations as novel genetic causes of meiotic arrest, Human Reproduction 34 (2019) 978–988. https://doi.org/10.1093/humrep/dez042.

[42] A. Reynolds, H. Qiao, Y. Yang, J.K. Chen, N. Jackson, K. Biswas, J.K. Holloway, F. Baudat, B. De Massy, J. Wang, C. Höög, P.E. Cohen, N. Hunter, RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis, Nat Genet 45 (2013) 269–278. https://doi.org/10.1038/ng.2541.

[43] Y. Fujiwara, H. Matsumoto, K. Akiyama, A. Srivastava, M. Chikushi, M. Ann Handel, T. Kunieda, An ENU-induced mutation in the mouse Rnf212 gene is associated with male meiotic failure and infertility, REPRODUCTION 149 (2015) 67–74. https://doi.org/10.1530/REP-14-0122.

[44] H. Qiao, H.B.D. Prasada Rao, Y. Yang, J.H. Fong, J.M. Cloutier, D.C. Deacon, K.E. Nagel, R.K. Swartz, E. Strong, J.K. Holloway, P.E. Cohen, J. Schimenti, J. Ward, N. Hunter, Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination, Nat Genet 46 (2014) 194–199. https://doi.org/10.1038/ng.2858.

9 Tables

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

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

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

9.2 Table 2

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	-0.1	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

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

10 Figures

10.1 Figure 1

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

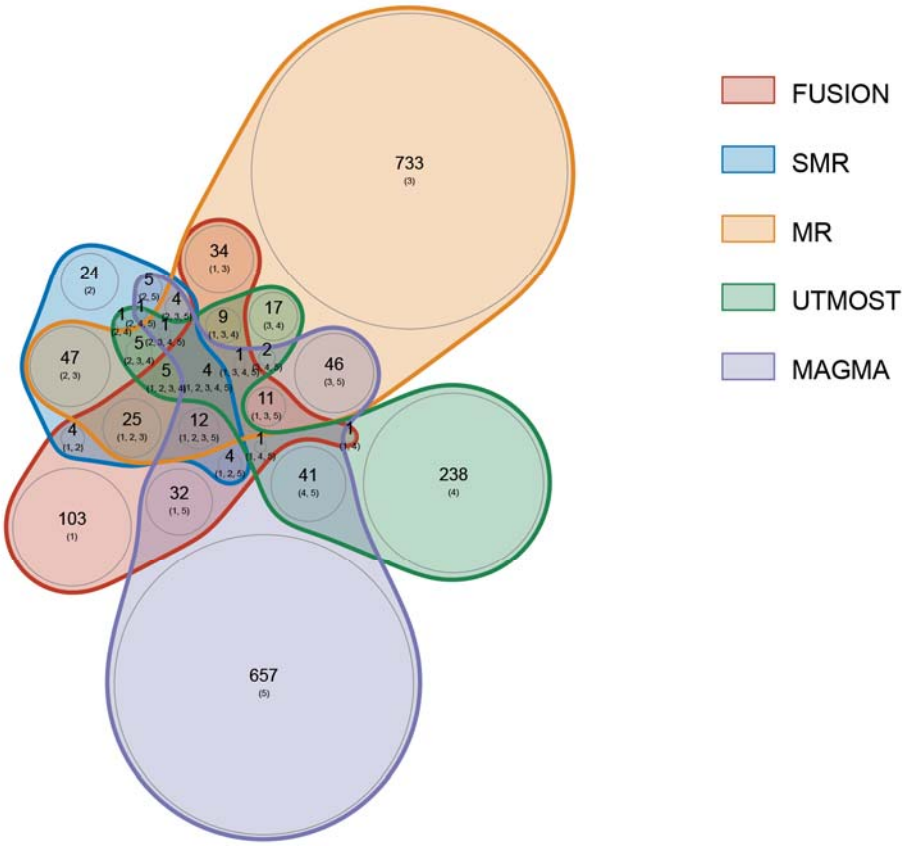
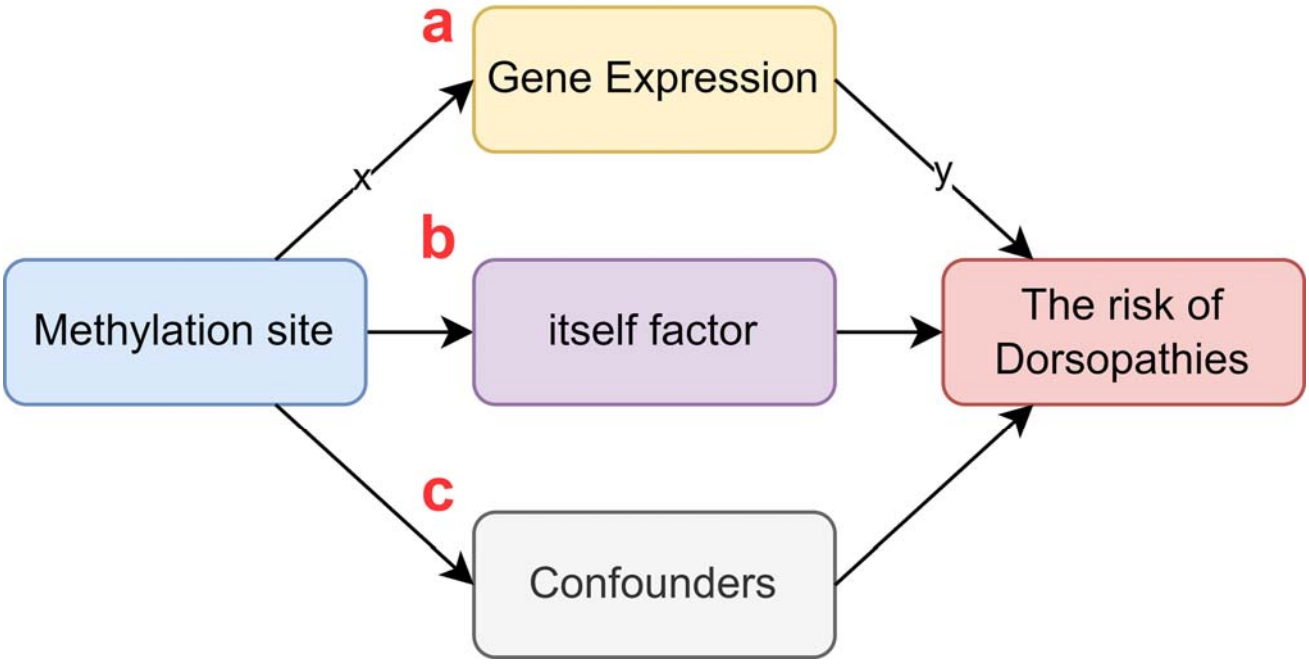


Fig.1 Venn Plot of the main results of multi analysis

10.2 Figure 2

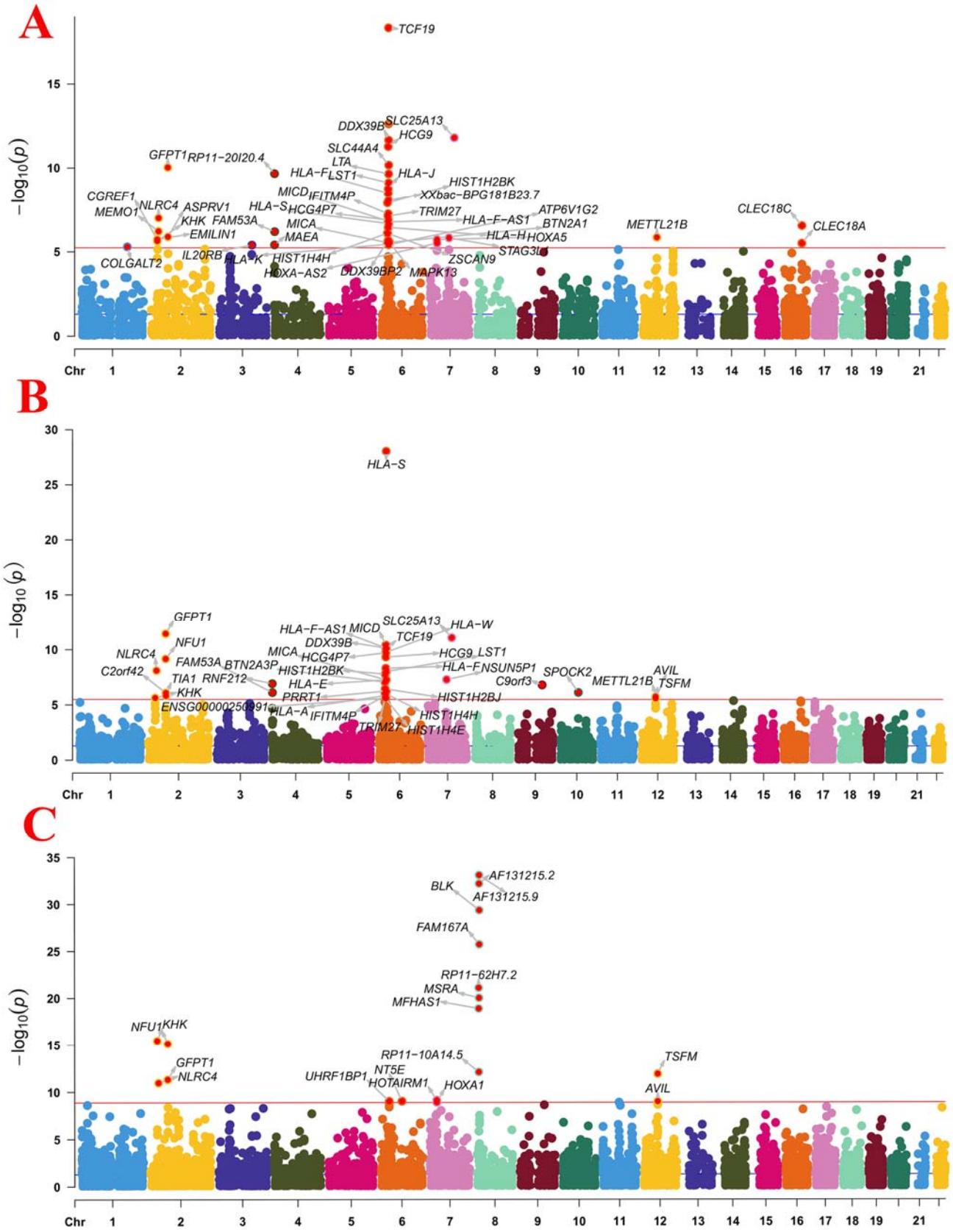


medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fig.2 Methylation sites affect disease pathways.

10.3 Figure 3

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



853

medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

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)

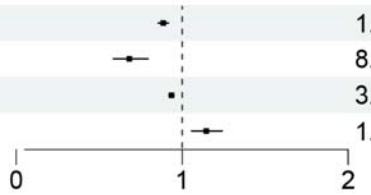
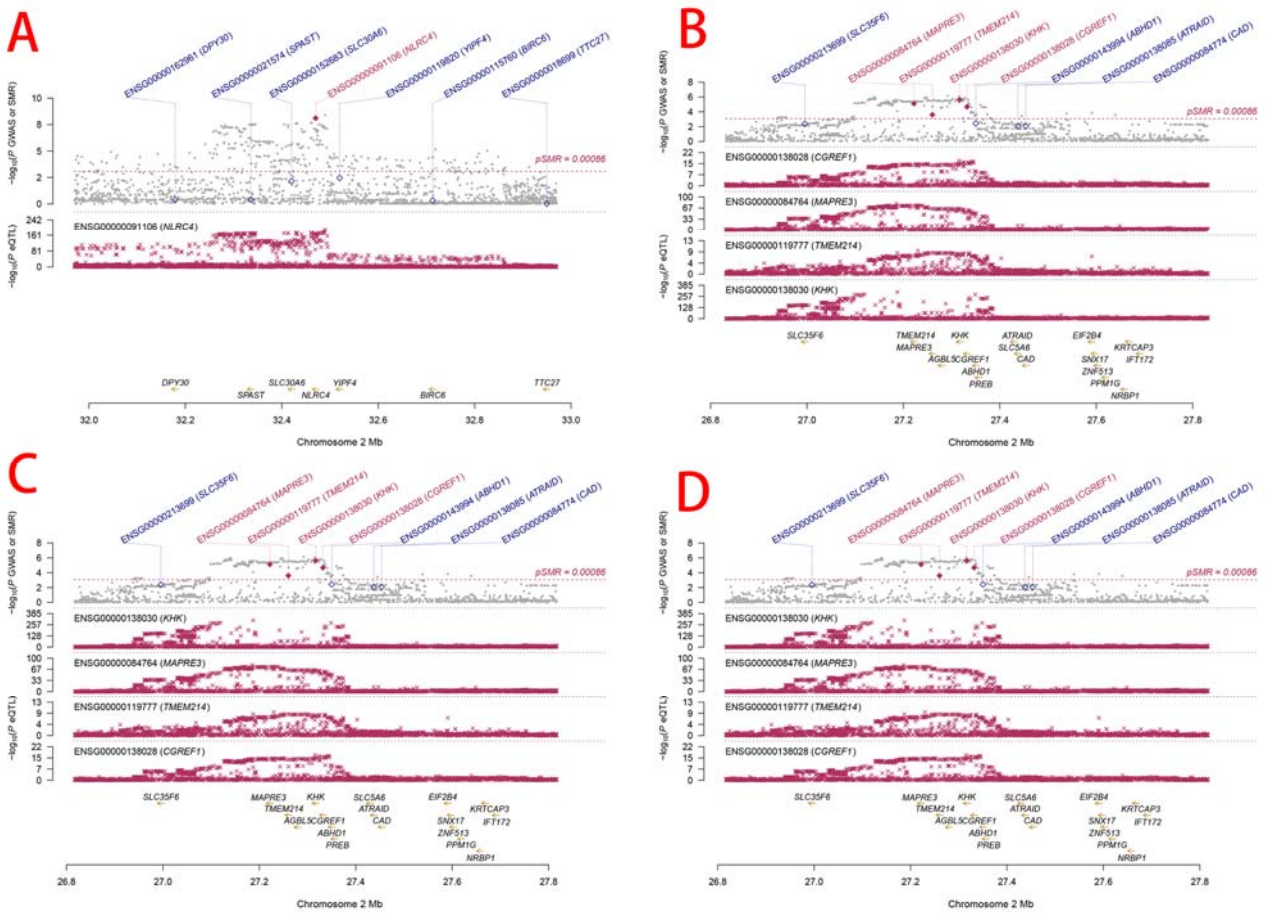


Fig.4 MR results for genes expression significantly associated with dorsopathies after FDR correction.

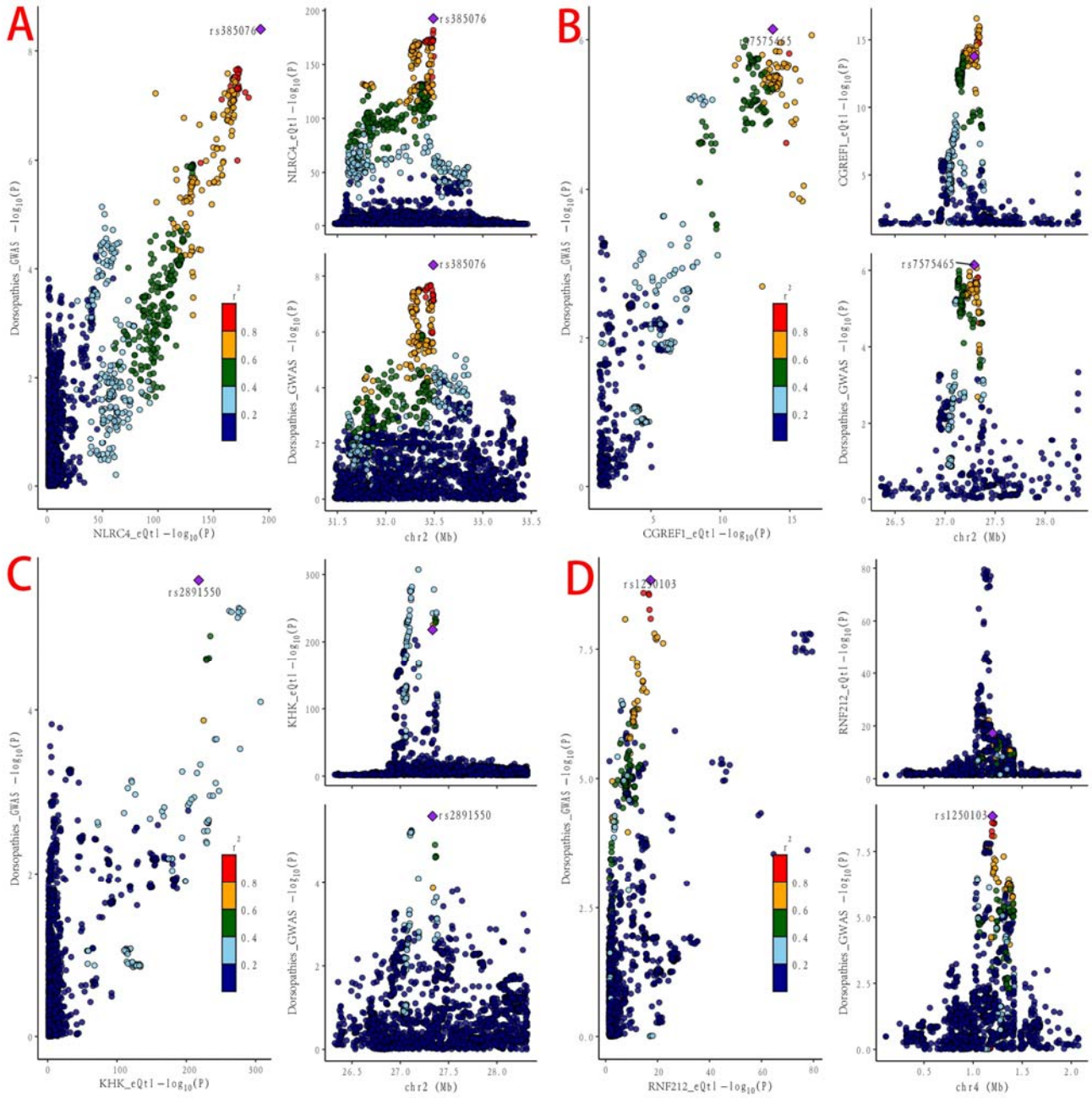
10.5 Figure 5



medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

10.6 Figure 6



medRxiv preprint doi: https://doi.org/10.1101/2024.05.01.24306675; this version posted May 20, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Fig. 6 Locus comparison plot displays the results of colocalization analysis for SNPs associated with gene expression in blood and datasets related to dorsopathies. Each dot represents a specific SNP, with the color indicating its linkage disequilibrium (LD) value (r2) with the lead GWAS variant, marked by a purple diamond. In the right panel, the x-axis represents genomic positions in megabases (GRCh37) along the chromosome, while the y-axis shows the -log10 p-values for SNPs from the dorsopathies GWAS (top) and the gene expression eQTL study (bottom). The left panel compares the p-values from the dorsopathies GWAS with those from the gene expression eQTL study.