medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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-NC-ND 4.0 International license .

Research Letter: Proteome-wide Mendelian randomization identifies OLFM1 as a potential mediator of the risk effect of myxoedema on haemorrhoidal disease

Shifang Li[#]*, Meijiao Gong[#]

Laboratory of Immunology and Vaccinology, FARAH, ULiège, Liège 4000, Belgium.

[#]Shifang Li and Meijiao Gong contributed equally to this work

*Correspondence:

Shifang Li, fruceslee@gmail.com

Laboratory of Immunology and Vaccinology, FARAH, ULiège, Liège 4000, Belgium

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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/10.1101/2023.06.19.23291373; this version posted June 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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-NC-ND 4.0 International license .

Abstract

Human haemorrhoidal disease (HEM) is a common anorectal pathology. Being one of the diseases that affect a wide range of people, the etiology of HEM, as well as its molecular mechanism, remains largely unclear. In this study, we applied a two-sample bi-direction Mendelian randomization (MR) analysis to estimate the causal effects of 4677 plasma proteins on HEM outcomes and investigated the mediating impacts of plasma proteins on HEM risk factors to uncover potential HEM treatment targets by integrating GWASs statistics of HEM and plasma protein levels. Following MR analysis, our study identified 13 probable causal proteins associated with HEM. Particularly, genetically predicted OLFM1 levels were linked to an increased risk of HEM. In addition, it was discovered that a genetically greater risk of myxoedema, diverticular disease, and ulcerative colitis was linked to an elevated risk of HEM ($FDR<0.05$). However, there was no evidence that dorsalgia, hernia, and ankylosing spondylitis were causally associated with HEM. Interestingly, a higher risk of myxoedema was associated with higher genetically predicted OLFM1 levels. Finally, mediation analysis suggested that the effect of myxoedema on HEM via OLFM1 might explain 32.8% of the mediation effect. Overall, this study identified some causal associations of circulating proteins and risk factors with HEM by integrating the largest-to-date plasma proteome and GWASs of HEM. The findings could provide further insight into understanding biological mechanisms for HEM.

Keywords

Haemorrhoidal disease, Mendelian randomization, OLFM1, myxoedema

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

Human haemorrhoidal disease (HEM) is a common anorectal disorders. Recently, Zhang *et al.* reported the first and largest genome-wide association study (GWAS) with haemorrhoidal disease (HEM), and these data offered us a resource for understanding the genetic risk factors for HEM.[1] However, being one of the diseases that affect a wide range of people, the etiology of HEM, as well as its molecular mechanism, remains primarily unclear.[2] In addition, identification of genes with therapeutic effects needs to be conducted. Here, we applied a two-sample bi-direction Mendelian randomization (MR) analysis to estimate the causal effects of 4677 plasma proteins on HEM outcomes and investigated the mediating impacts of plasma proteins on HEM risk factors to uncover potential HEM treatment targets by integrating GWASs statistics of HEM and plasma protein levels (**Figure 1**).

As stated in the **Supplementary Methods**, we used 4677 protein quantitative trait loci (pQTL) of proteins as instrumental variables for exposure and HEM as the outcome to estimate the causal effect of plasma protein levels on HEM in a proteome-wide context using MR analysis.[3-6] Following a stringent filtering approach, our study identified 13 probable causal proteins (*FDR*<0.05), including 4 negative and 9 positive associations (**Figure 2A**). MR analysis, for example, revealed that genetically predicted OLFM1 levels were linked to an increased risk of HEM (p=1.27e-06) (**Figure 2B**). Notably, only two (GUK1 and IL1RL2) of the 13 causative associations with HEM have been proposed to be driven by *cis*-regulatory variations, with the remaining significant associations assumed to be driven by *trans*-SNPs.

Following that, we attempted to identify causal risk factors for HEM. 6 clinical

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

traits that genetically correlated with HEM were selected (**Supplementary Methods**), with instrumental variables generated from GWAS results confined to European populations. It was discovered that a genetically greater risk of myxoedema, diverticular disease, and ulcerative colitis was linked to an elevated risk of HEM (*FDR*<0.05) (**Figure 2C**). Although genetic correlations with HEM were reported,[1] there was no evidence that dorsalgia, hernia, and ankylosing spondylitis were causally associated (p>0.05). Subsequently, MR analyses were further conducted using myxoedema genetic data from the Neale lab Consortium to assess whether this causal estimation could be repeated. A consistent causal effect of myxoedema on the risk of HEM utilizing these MR analyses was found (p<0.05) (**Figure 2D**).

In order to identify the protein related to HEM risk factors, we conducted MR analysis again on 13 plasma proteins impacting HEM with 3 risk factors. After adjusting for multiple testing, we discovered that only myxoedema exhibited a substantial causal association with OLFM1 (**Figure 2E**). Particularly, a higher risk of myxoedema was associated with higher genetically predicted OLFM1 levels (p=9.5e-06) and this causal effect could also be replicated when using the genetic data of myxoedema from the Neale lab Consortium (p=4.9e-05) (**Figure 2E**). Notably, the direction of action of myxoedema with OLFM1 was consistent with the link of proteins with HEM, suggesting that these proteins may be potential mediators of the myxoedema-HEM association. To study the indirect effects of protein-mediated risk variables on HEM outcomes further, a mediation analysis using two-step MR and coefficient product methods was performed to understand the extent to which plasma

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

OLFM1 levels mediate the association between myxoedema and HEM (**Figure 2F**). It was found that nearly one-third (32.8%) of myxoedema-induced effects were mediated by OLFM1.

Overall, this study identified some causal associations of circulating proteins and risk factors with HEM by integrating the largest-to-date plasma proteome and GWAS of HEM. OLFM1 was found to be a potential mediator of the risk effect of myxoedema on HEM. The findings could provide further insight into developing potential targets for HEM.

Competing interests

None declared.

Contributors

SF was involved in conceptualization. SF and MJ were involved in the formal analysis. SF was involved in writing, reviewing, and editing.

Acknowledgments

The authors would like to thank all of the researchers who contributed to the GWAS datasets used in this study for making them available for research purposes.

References

1 Zheng T, Ellinghaus D, Juzenas S, *et al*. Genome-wide analysis of 944 133 individuals provides insights into the etiology of haemorrhoidal disease. *Gut* 2021;70:1538-49.

2 EAM Festen & RK Weersma. Large-scale genetic analyses in an understudied disease: haemorrhoidal disease. *Gut 2021*;70:1429-1430.

3 Zheng J, Haberland V, Baird D, *et al*. Phenome-wide Mendelian randomization

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

mapping the influence of the plasma proteome on complex diseases. *Nat Genet* 2020;52:1122-31.

4 Chen L, Peters JE, Prins B, *et al*. Systematic Mendelian randomization using the human plasma proteome to discover potential therapeutic targets for stroke. *Nat Commun* 2022;13:1-14.

5 Yoshiji S, Butler-Laporte G, Lu T, *et al*. Proteome-wide Mendelian randomization implicates nephronectin as an actionable mediator of the effect of obesity on COVID-19 severity. *Nat Metab* 2023;5:248-64.

6 Chen J, Xu F, Ruan X, Sun J, *et al*. Therapeutic targets for inflammatory bowel disease: proteome-wide Mendelian randomization and colocalization analyses. *EBioMedicine* 2023;89:104494.

Figure Legends

Figure 1 An overview of MR in this study. A two-step MR approach was used to determine the mediate effect of myxoedema on HEM via plasma protein: First, we employ MR to quantify the effect of 4677 plasma protein levels on HEM, which provided 13 plasma proteins linked with HEM after a strict filtering approach (step 1 MR). Subsequently, we used MR to quantify the effect of the risk factors on HEM outcomes (step 2 MR). The effect of the risk variables on HEM-related proteins was next investigated using MR (step 3 MR). Finally, mediation analysis was carried out to assess the proportion mediated by plasma protein levels for the effect of risk factors on HEM outcomes (step 4 mediation analysis). In steps 2 and 3, the effect of causality was validated using GWAS of myxoedema from the Neale lab Consortium. IVW,

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

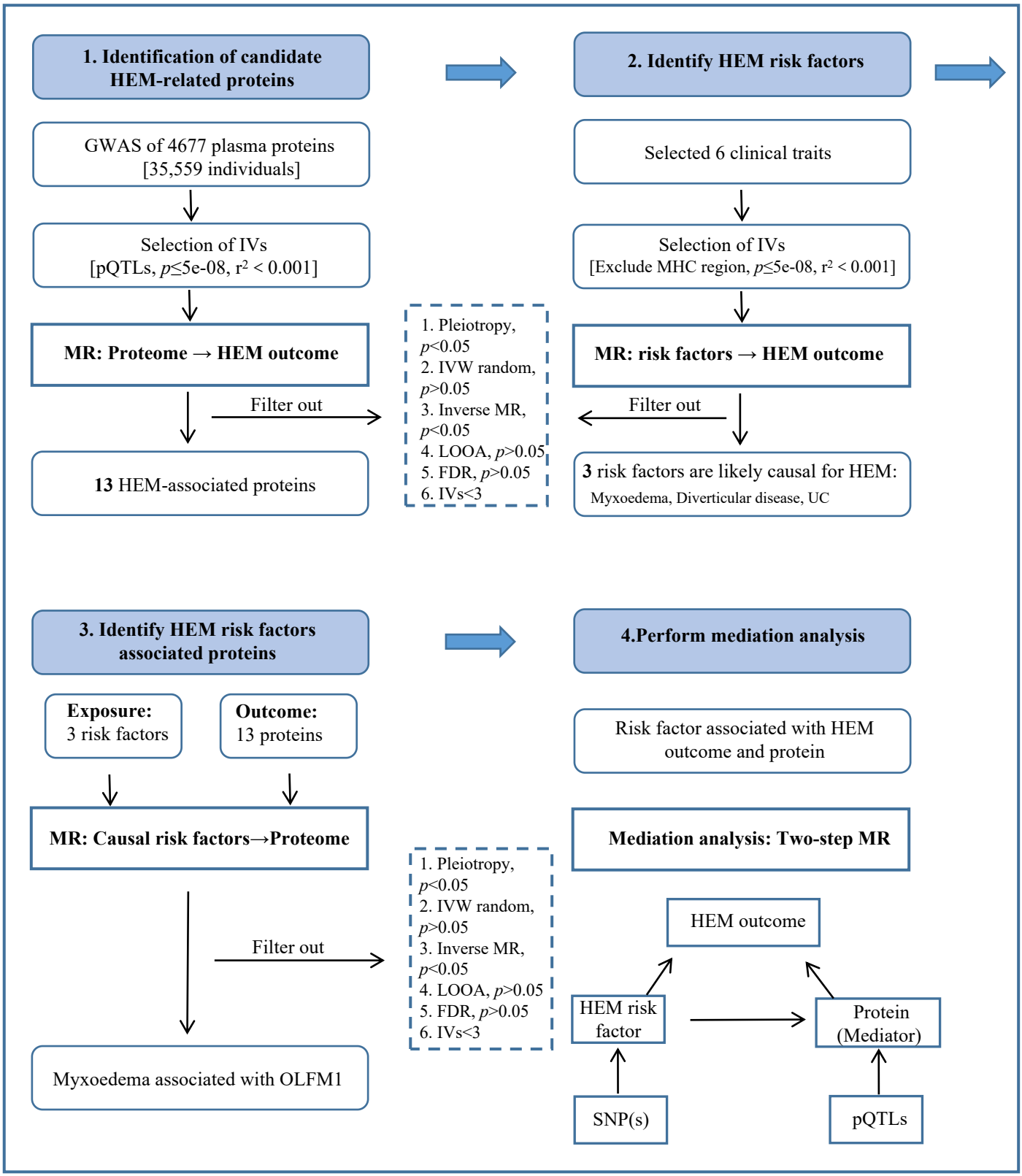
inverse variance weighted; LOOA, leave-one-out analysis; FDR, false discovery rate; HEM, haemorrhoidal disease; IV, instrumental variable; MR, Mendelian randomization; UC, ulcerative colitis; FDR, false discovery rate.

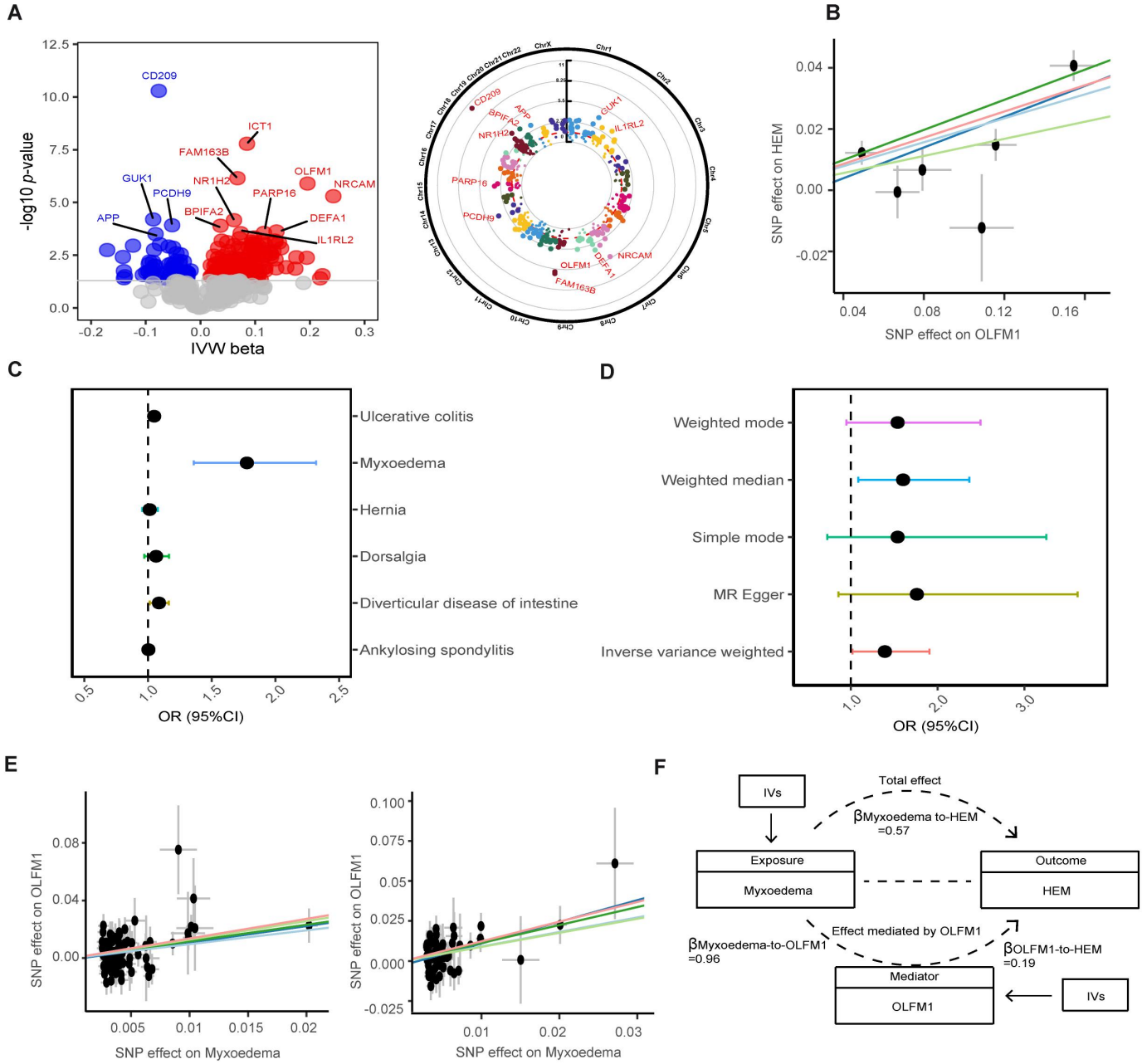
Figure 2 Mendelian randomization results. (**A**) The effect of plasma protein levels on HEM. (Left) Volcano plot indicating the effect of plasma protein on HEM using MR analysis using the inverse variance weighted approach. Red represents the positive association of plasma proteins with a high risk of HEM, whereas blue represents the negative correlation of plasma proteins with HEM. (Right) Manhattan plot depicting the distribution of plasma protein linked with HEM across chromosomes. (**C**) Forest plots showing the causal effect of chosen risk factors on HEM. (**D**) Forest plots illustrating the causative effect of myxoedema on HEM using myxoedema GWAS from the Neale lab Consortium. (**E**) Dot plots indicated the odd ratio (OR) estimates and 95% confidence interval (CI) values of the causal effects of myxoedema on OLFM1, calculated using myxoedema GWAS from the Medical Research Council-Integrative Epidemiology Unit (MRC-IEU) consortium (Left) and the Neale lab Consortium (Right), respectively. (**F**) Assess the effect of myxoedema on HEM via OLFM1 by mediation analysis. To estimate the mediate effect of OLFM1, the product of coefficients method was used, where $\beta_{\text{myxoedema-to-OLFM1}}$ is the effect of myxoedema on OLFM1, $\beta_{\text{OLFM1-to-HEM}}$ is the effect of OLFM1 on HEM, and $\beta_{\text{myxoedema-to-HEM}}$ is the total effect of myxoedema on HEM. The x-axis for B and E reflects SNP exposure effects, while the y-axis shows SNP outcome effects. HEM, haemorrhoidal disease.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

Supplementary Methods The statistics method used in the study.

Supplementary Tables1 The significant MR summary statistics obtained in this study.


1. Identification of candidate HEM-related proteins
GWAS of 4677 plasma proteins [35,559 individuals]
Selection of IVs [pQTLs, p≤5e-08, r² < 0.001]
MR: Proteome → HEM outcome
Filter out
13 HEM-associated proteins
1. Pleiotropy, p<0.05
2. IVW random, p>0.05
3. Inverse MR, p<0.05
4. LOOA, p>0.05
5. FDR, p>0.05
6. IVs<3
2. Identify HEM risk factors
Selected 6 clinical traits
Selection of IVs [Exclude MHC region, p≤5e-08, r² < 0.001]
MR: risk factors → HEM outcome
Filter out
3 risk factors are likely causal for HEM:
Myxoedema, Diverticular disease, UC
3. Identify HEM risk factors associated proteins
Exposure: 3 risk factors
Outcome: 13 proteins
MR: Causal risk factors→Proteome
Filter out
Myxoedema associated with OLFM1
1. Pleiotropy, p<0.05
2. IVW random, p>0.05
3. Inverse MR, p<0.05
4. LOOA, p>0.05
5. FDR, p>0.05
6. IVs<3
4.Perform mediation analysis
Risk factor associated with HEM outcome and protein
Mediation analysis: Two-step MR
HEM outcome
HEM risk factor
Protein (Mediator)
SNP(s)
pQTLs


A
-log10 p-value
IVW beta
CD209
ICT1
FAM163B
OLFM1
NRCAM
GUK1
NR1H2
PCDH9
PARP16
APP
BPIFA2
DEFA1
IL1RL2
B
SNP effect on HEM
SNP effect on OLFM1
C
Ulcerative colitis
Myxoedema
Hernia
Dorsalgia
Diverticular disease of intestine
Ankylosing spondylitis
OR (95%CI)
D
Weighted mode
Weighted median
Simple mode
MR Egger
Inverse variance weighted
OR (95%CI)
E
SNP effect on OLFM1
SNP effect on Myxoedema
F
IVs
Total effect
βMyxoedema to-HEM =0.57
Exposure
Myxoedema
Outcome
HEM
Effect mediated by OLFM1
βMyxoedema-to-OLFM1 =0.96
βOLFM1-to-HEM =0.19
Mediator
OLFM1
IVs

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

Supplementary Methods

GWASs of haemorrhoidal disease and risk factors

We used recently published large-scale genome-wide associations (GWASs) for haemorrhoidal disease (HEM).[1] This GWAS summary statistics were derived from 944,133 European ancestry individuals (Ncase = 218,920 and Ncontrol = 725,213) from 5 cohorts and downloaded from the GWAS Catalog (https://www.ebi.ac.uk/gwas/, access ID: EFO_0009552). Diverticular disease of the intestine, ankylosing spondylitis (AS), ulcerative colitis (US), dorsalgia, and hernia were evaluated as potential causal risk factors associated with HEM in order to determine the probable causal risk factors. All GWASs for the six risk factors were obtained from the ieu open gwas project (https://gwas.mrcieu.ac.uk/datasets/). The summary statistics of the large GWAS (14,357 cases and 182,423 controls) were used for diverticular disease of the intestine (access ID: finn-b-K11_DIVERTIC). The GWAS for AS (access ID: finn-b-M13_ANKYLOSPON) have a sample size of 1,462 cases and 164,682 controls. The GWAS for myxoedema (access ID: ukb-b-19732) have a sample size of 22,687 cases and 440,246 controls. In the GWAS for the US (access ID: finn-b-K11_ULCER), a sample size of 4,320 cases and 210,300 controls were reported. The GWAS for dorsalgia (access ID: finn-b-M13_DORSALGIA) included 193467 individuals, with 28,785 cases and 164,682 controls. A total of 218792 individuals were reported with GWAS of hernia (access ID: finn-b-K11_HERNIA), including 28,235 cases and 190,557 controls. The GWAS for myxoedema (access ID: ukb-a-77), which included 337,159 individuals, including

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

16,376 cases and 320,783 controls, were utilized to validate MR results.

Plasma protein quantitative trait loci (pQTL) data

To conduct Proteome-wide Mendelian randomization (MR), we first obtained genetic instrumental variables using the protein quantitative trait loci (pQTL) data generated by Ferkingstad *et al*.[2] The largest-to-date pQTL analysis on plasma proteome (a total of 4907 proteins) in 35,559 Icelanders was performed in their study, and an amount of 18,084 pQTL associations between genetic variation and protein levels in plasma were identified. A total of 4667 pQTLs were successfully downloaded from the deCODE study (access date 03/06/2023) using aria2c.[3] In the following analysis, *cis*- and *trans*-pQTLs from each corresponding GWAS were employed as genetic instruments for each protein.

Mendelian randomization analysis

MR analysis is an analytical method that uses genetic variation as an instrumental variable (IV) to estimate causal effects. It overcomes the limitations of measurement error and confounding factors that are common in observational studies and is widely used to assess causal relationships.[4] In this study, the TwoSampleMR package (v0.5.6, https://mrcieu.github.io/TwoSampleMR/) was used for MR analysis.[5] The instrumental variables that determined the exposure in each MR study were specified as genome-wide significant ($p \leq 5e\text{-}08$) and independent SNPs. SNPs in the human major histocompatibility complex (MHC) region at chromosome 6: 28,477,797-33,448,354 (GRCh37) were excluded from the analysis of inferring the causal effect of risk factors on protein levels and HEM due to its complex linkage

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

disequilibrium (LD) structure. Using the 1000 Genomes Project European reference panel and an LD threshold of r^2 0.001 with a clumping window of 10,000 kb, PLINK v.1.9 (http://pngu.mgh.harvard.edu/purcell/plink/) was employed to derive instrumental variables.[6-7] F-statistics were used to determine the strength of each SNP's association with exposure, and F-statistics of more than 10 were considered strong. The inverse-variance weighted (IVW) approach with a fixed effects model was employed for the main MR analysis to assess the causal influence of exposure on outcome. In addition, four additional MR methods (weighted median, simple mode, weighted mode, and MR-Egger method) were used to assess the reliability of the primary results. For exposures with multiple IVs, we additionally investigated heterogeneity across variant-level MR estimations with the "mr_heterogeneity()" function in the TwoSampleMR package (Cochrane's Q test). In addition, a pleiotropy test was performed using MR Egger analysis to determine whether there is horizontal pleiotropy among IVs.

Finally, in the event there are more than two IVs in exposure, a leave-one-out analysis is performed, and the MR findings of the remaining IVs are calculated by deleting the IVs one by one to ensure the robustness of the MR data. To acquire robust evidence for the casual estimation, MR findings that meet all of the following criteria were chosen: (1) no pleiotropy was found using MR-Egger regression (p>0.05); (2) consistent results were obtained using the IVW method and weight median model (p<0.05); (3) IVW random effect model p<0.05; (4) leave-one-out analysis p<0.05 after removing outliers; and (5) reverse MR p>0.05; (6) more than 2 IVs were used in

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

MR analysis. The same procedure as mentioned above was utilized to explore the causal effect of the given exposure and associated outcome in the reverse MR analysis. For multiple testing, the *p*-values were adjusted using the Benjamini-Hochberg (BH) method. To determine whether the significant relationships found in our analysis were caused by cis-regulatory variations, we selected *cis*-SNPs (SNPs located within ±5,000 kb window from the gene body of the target gene) from all important plasma proteins and carried out Mendelian randomization studies with the HEM GWASs using IVs.

Mediation analysis

Using two-step MR, a mediation analysis was performed to assess the proportion of the effect of myxoedema on HEM that was mediated by OLFM1. In brief, the effect of myxoedema on OLFM1 was first assessed, then multiplied by the effect of OLFM1 on HEM. Following that, the proportion of the entire effect of myxoedema on HEM mediated by OLFM1 was determined by dividing the OLFM1-mediated effect (OLFM1-to-HEM) by the whole effect (myxoedema-to-HEM), as previously stated.[8-10]

References

1 Zheng T, Ellinghaus D, Juzenas S, *et al*. Genome-wide analysis of 944 133 individuals provides insights into the etiology of haemorrhoidal disease. *Gut* 2021;70:1538-49.

2 Ferkingstad E, Sulem P, Atlason BA, *et al*. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet* 2021;53:1712-21.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291373; this version posted June 27, 2023. 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-NC-ND 4.0 International license .

3 Aria2c Multi-souorce Download Utilily. Available: http://aria2.sourceforge.net/

4 Skrivankova VW, Richmond RC, Woolf BAR, *et al*. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): Explanation and elaboration. BMJ 2021;375. doi:10.1136/bmj.n2233

5 Hemani G, Zheng J, Elsworth B, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife* 2018;7:1-29.

6 Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature* 2015;526:68-74.

7 Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75.

8 Yoshiji S, Butler-Laporte G, Lu T, *et al*. Proteome-wide Mendelian randomization implicates nephronectin as an actionable mediator of the effect of obesity on COVID-19 severity. *Nat Metab* 2023;5:248-64.

9 Burgess S, Daniel RM, Butterworth AS, *et al*. Network Mendelian randomization: Using genetic variants as instrumental variables to investigate mediation in causal pathways. *Int J Epidemiol* 2015;44:484-95.

10 Carter AR, Gill D, Davies NM, *et al*. Understanding the consequences of education inequality on cardiovascular disease: Mendelian randomisation study. *BMJ* 2019;365:1-12.