

Genome-wide meta-analysis conducted in three large biobanks expands the genetic landscape of lumbar disc herniations

Received: 6 November 2023

Accepted: 11 October 2024

Published online: 07 November 2024



Ville Salo¹✉, Juhani Määttä^{1,2}, Eeva Sliz³, FinnGen*, Ene Reimann⁴, Reedik Mägi⁴, Estonian Biobank Research Team*, Kadri Reis⁴, Abdelrahman G. Elhanas⁴, Anu Reigo⁴, Priit Palta^{4,5}, Tõnu Esko⁴, Jaro Karppinen^{1,2,6} & Johannes Kettunen³

Given that lumbar disc herniation (LDH) is a prevalent spinal condition that causes significant individual suffering and societal costs, the genetic basis of LDH has received relatively little research. Our aim is to increase understanding of the genetic factors influencing LDH. We perform a genome-wide association analysis (GWAS) of LDH in the FinnGen project and in Estonian and UK biobanks, followed by a genome-wide meta-analysis to combine the results. In the meta-analysis, we identify 41 loci that have not been associated with LDH in prior studies on top of the 23 known risk loci. We detect LDH-associated loci in the vicinity of genes related to inflammation, disc-related structures, and synaptic transmission. Overall, our research contributes to a deeper understanding of the genetic factors behind LDH, potentially paving the way for the development of new therapeutics, prevention methods, and treatments for symptomatic LDH in the future.

Lumbar disc herniation (LDH) is one of the most frequent structural findings in the lumbar spine to cause specific symptoms. Disc herniation is a general term that includes different types of disc displacements such as disc protrusion, extrusion, and sequestration¹. Disc protrusions, for instance, are rarely symptomatic, and they are prevalent even among asymptomatic subjects. The overall prevalence of LDH in the whole population has been estimated to be 14% in a large cross-sectional study, and herniations are found most frequently at the L4/5 and L5/S1 segments of the spine².

LDH has a clinical relevance if it causes radicular symptoms in the lower extremity. LDH is, indeed, the most frequent condition that causes lumbar radicular pain, i.e., sciatica, or radiculopathy³. The mechanisms underlying LDH-induced radicular pain are manifold. While mechanical compression of the nerve root is a contributing factor, inflammatory mediators and autoimmune

responses also play a considerable role in eliciting radicular pain⁴. When the nerve root is perturbed, typical symptoms include radicular pain, paraesthesia, or numbness in the area of the nerve, and possible weakness in the muscles innervated by the affected nerve root⁵.

Typically, patients with symptomatic LDH are treated conservatively, but surgery is required if there is a sudden or progressive neurological deficit or unmanageable pain despite appropriate conservative treatment⁵. Overall, surgery has not been proven to have superior outcomes in the long term, even though surgery can have better pain relief in the short term^{6,7}.

While most patients tend to recover from symptoms rather quickly, some studies have reported conflicting evidence. Psychosocial factors, such as the patient's own beliefs of recovery, can also have a role in prognosis⁸.

¹Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland. ²Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland. ³Research unit of Population Health, Faculty of Medicine, and Biocenter Oulu, University of Oulu, Oulu, Finland. ⁴Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. ⁵Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. ⁶Rehabilitation Services of Wellbeing Services County of South Karelia, Lappeenranta, Finland. *Lists of authors and their affiliations appear at the end of the paper. ✉e-mail: ville.salo@oulu.fi

Genetic influence on LDH and sciatica has been established through research that has shown certain loci to be associated with these conditions in genome-wide association studies (GWASs)^{9,10}. Symptomatic LDH could be caused by factors affecting either disc-related structures, such as collagen¹¹, or other factors, such as those related to nerves, inflammation, or autoimmunity⁴. Previous studies have associated several pathways with LDH pathogenesis, encompassing inflammation, chondroitin sulfation, collagen synthesis, and chondrogenic differentiation^{9,10}. Some of these genes have also been associated with back pain and the regulation of pain sensations^{12–14}. The aetiological factors behind LDH are quite well understood¹⁵. However, the genetic factors behind these distinct features warrant more research. The aim of this study was to explore different genetic features behind LDH by conducting a GWAS using data from three large biobanks: FinnGen, Estonian Biobank, and UK Biobank.

Results and discussion

64 loci associate with LDH risk

In this genome-wide meta-analysis of LDH, as defined by International Classification of Diseases (ICD)–10 code M51 (M51.1–M51.9), we utilised data from FinnGen, the Estonian Biobank, and the UK Biobank, comprising a total of 829 699 participants (80 724 cases, and 748 975 controls; Supplementary Table 1). In the meta-analysis, we identified 41 previously unreported (Supplementary Figs 1–41, Table 1) and replicated 23 known loci located more than 1MB apart (Supplementary Fig 42, Supplementary Table 2), each containing at least one variant associated with LDH at the genome level (M51.1–51.9, Fig. 1, Supplementary Data 1, Supplementary Data 2). Out of the previously unreported loci, 4 loci had previously been associated with other back-related diseases^{12,13,16}; however, our research demonstrates that these loci are also associated with LDH. In the conditional analysis, we observed secondary signals at 5 of the loci (Supplementary Table 2). We estimated LD score regression-derived SNP-based heritability of LDH to be 0.081 (standard error [SE] = 0.003), suggesting that genetic factors account for 8.1% of the common variation in LDH risk. This estimate aligns with the one available at the UKB-Heritability Browser (0.077, SE = 0.026; https://nealelab.github.io/UKBB_ldsc/h2_summary_M51.html). The genomic inflation factor lambda (1.47) suggested inflation in the test statistics. Given the intercept value of 1.12, the inflation observed may be attributed to a polygenic signal and population stratification¹⁷. In FinnGen, the SNP-based heritability estimate was 14.3% (SE = 0.0083), and lambda 1.39 with an intercept of 1.13.

Five previously unreported loci associated with LDH requiring surgical intervention

In two sets of sensitivity analyses conducted in FinnGen, more strict case definitions were used by limiting LDH cases to the ones with radiculopathy (M51.1) and to those who have undergone surgery (Supplementary Table 1). No statistically significant differences in the effect sizes of the lead variants were observed between the GWAS specific to the M51.1 endpoint and the original meta-analysis. When comparing the GWAS on LDH patients who underwent surgical treatment with the original meta-analysis, however, the effect sizes were significantly larger for nine variants (Fig. 1, Supplementary Fig 43, Supplementary Table 3). In this analysis, we also identified five previously unreported loci associated with LDH-related surgical operations and replicated three previously associated loci (Supplementary Fig 44, Table 1, Supplementary Table 4)¹⁰.

Description of the LDH-associated loci

As reported previously, numerous LDH-associated loci were in the vicinity of genes related to inflammation or disc-related structures (Supplementary Table 2, Supplementary Data 2), indicating that these pathways play a central role in LDH pathogenesis. In addition, we detected unreported loci near genes related to the Wnt/ β -catenin

pathway, such as *PDZRN3* (*PDZ domain containing ring finger 1*, locus 3p13), activation of the Wnt/ β -catenin pathway has been found to be associated with endplate degeneration, increased intervertebral disc (IVD) cell senescence, and extracellular matrix degradation^{18,19}. LDH associations were also observed in loci near *NGF* (*nerve growth factor*), *DCC* (*DCC netrin-1 receptor*), and *NCK1* (*NCK adaptor protein 1*). These genes are involved in the growth and regulation of nerve axons^{20,21}, suggesting a potential connection between genes affecting nerves and the nervous system and LDH pathogenesis. The dysfunction of these genes could lead to IVD innervation²⁰, and increase sensation of pain. Notably, some of these genes have already been associated with pain sensation in previous studies^{12,20}.

In addition, we observed previously unreported LDH associations near *CA10* (*carbonic anhydrase 10*) and *DNMI* (*dynamitin 1*) that are involved in synaptic transmission. *CA10* blocks the binding of heparan sulphate to neuexin, which possibly affects the function of neuexins²². Neuexins are pre-synaptic cell adhesion molecules that play a role in connecting neurons at synapses. Heparan sulphate has been found to potentially expand the interactome of neuexins, and they also play a role in fine-tuning synaptic transmission²³. *CA10* is expressed especially in the central nervous system, and it has been associated with chronic pain in previous studies^{12,24}. *DNMI* plays a central role in the transmitting nociceptive messages within the nociceptive circuits in the dorsal horn of the spinal cord, where *DNMI*-mediated endocytosis of synaptic vesicles enables sustained neurotransmission²⁵. While further studies are needed, the association of genes involved in synaptic transmission with LDH suggests that differences in synaptic transmission may influence differences in pain perception in patients with morphologically similar LDHs. Moreover, these genes could also contribute to the chronification of pain in patients with symptomatic LDH. The findings above underline the relevance of the physiological factors of the nervous system in addition to the disc-related structures behind symptomatic LDH.

Connective tissue related traits were enriched in functional analyses

The MAGMA gene-based test and GCTA-fastBAT analysis highlighted multiple genes that we had deemed as potential candidate genes at the previously unreported LDH-associated loci (Table 1, Fig. 2A, Supplementary Table 5), thus providing supportive evidence for our findings. In the MAGMA gene-set analysis (Fig. 2B), we observed the most significant enrichments for pathways responsible for chondrocyte differentiation. Significant gene set enrichments were also observed for cartilage and connective tissue development pathways, as well as for pathways related to chromatin organization and modification. Among the gene sets related to the nervous system, we also observed significant enrichments in gene sets related to the structure and active zone organization of presynapses. In the MAGMA tissue expression analysis (Supplementary Fig 45), we found no tissues showing a positive correlation between tissue-specific gene expression profiles and LDH associations. The likely reason for the null result is the absence of the relevant tissues, namely cartilage and bone, in the Genotype-Tissue Expression (GTEx) dataset used in these analyses. Colocalizations between LDH GWAS and eQTL signals were observed in each selected tissue, namely whole blood, cultured fibroblasts, tibial nerve, and cervical spinal cord. In one or more of the four investigated tissues, we discovered evidence of the colocalization of LDH signals and the expression of twenty genes (posterior probability [PP] for the shared variant ≥ 0.8 , Fig. 2C, Supplementary Table 6). At 6p21.31, the LDH association signal colocalized with the expression of *ILRUN* (*inflammation and lipid regulation with UBA-like and NBRI-like domains*, formerly known as *C6orf106*) in three tissues (PP_{Whole Blood} = 0.967, PP_{Cultured fibroblasts} = 0.966, PP_{Tibial Nerve} = 0.996). In addition to its potential involvement in the development of chronic pain in the brain, *ILRUN*, an inhibitor of proinflammatory and antiviral cytokine

Table 1 | The list of previously unreported lead variants at the 41 genome-wide significant ($p < 5 \times 10^{-8}$) loci that were associated with LDH in the meta-analysis (80 724 cases, 748 975 controls) and five previously unreported loci associated with LDH related surgical operations in sensitivity analysis that were performed in FinnGen (7347 cases, 270 964 controls)

Locus	Candidate gene	CHR: POS	rsid	EA	OA	OR (95CI)	p-value	EAF	Het pval	Fin En.	Ref
Meta-analysis (FinnGen, EstBB, UKBB)											
1p36.12	ALPL	1:21559185	rs150211890	G	T	1.07 (1.05–1.10)	2.5e–08	0.05	0.18	2.6	
1p13.2	NGF	1:115310363	rs4644491	A	G	0.96 (0.95–0.98)	6.78e–09	0.63	0.22	1.0	
1q41	TGFB2	1:218924545	rs779040	C	G	0.97 (0.95–0.98)	7.13e–09	0.55	0.02	1.0	
1q41	HHLPL2	1:222541797	rs35455442	A	C	0.96 (0.94–0.97)	3.96e–13	0.33	0.54	1.3	
2q33.3	GPR1	2:206137590	rs78826721	G	A	0.94 (0.92–0.96)	2.85e–09	0.11	0.17	1.2	
3p21.31	HYAL2	3:50380254	rs41308273	A	T	0.92 (0.89–0.94)	1.02e–10	0.05	0.50	2.4	
3p13	PDZRN3	3:73200973	rs11914834	T	C	0.97 (0.96–0.98)	4.85e–08	0.44	0.98	1.1	
3q21.1	ADCY5	3:123568959	rs1965290	C	T	0.97 (0.95–0.98)	2.67e–08	0.55	0.04	1.1	
3q22.3	NCK1	3:136490708	rs13321721	G	A	1.04 (1.03–1.06)	4.49e–09	0.24	0.82	1.1	BP ¹²
3q25.32	SHOX2	3:158478549	rs5853827	ATCC	A	0.96 (0.95–0.98)	1.55e–08	0.33	0.49	0.9	
4q22.1	IBSP	4:87779677	rs10019020	A	G	1.04 (1.02–1.05)	4.07e–09	0.49	0.44	0.9	
5q31.3	HDAC3	5:141735121	rs5871786	G	GT	0.97 (0.96–0.98)	4.71e–08	0.44	0.73	1.0	
5q35.1	FGF18	5:171413500	rs4302608	G	A	0.97 (0.95–0.98)	2.61e–08	0.55	0.90	1.0	
6p22.2	TRIM38	6:26276422	rs9393692	G	A	0.97 (0.95–0.98)	4.46e–08	0.58	0.83	1.2	
6p22.1	HILA	6:29873925	rs1611653	C	G	1.04 (1.03–1.06)	1.98e–09	0.58	0.22	0.9	
6p21.33	HILA	6:31279637	rs2844608	T	C	0.96 (0.94–0.97)	1.58e–13	0.39	0.02	1.1	
6q14.3	TBX18	6:84938145	rs2224214	T	C	1.04 (1.03–1.06)	2.33e–12	0.37	0.67	1.0	
7q11.23	ELN	7:73714641	rs10227463	T	C	0.97 (0.95–0.98)	1.82e–08	0.41	0.77	1.2	
7q32.2	ZC3HC1	7:130023656	rs11556924	T	C	1.04 (1.02–1.05)	3.48e–08	0.36	0.29	0.8	
7q33	STMP1	7:135418700	rs2551776	T	C	0.97 (0.95–0.98)	2.48e–08	0.64	0.32	1.1	
9p21.3	TUSC1	9:25398495	rs7019841	G	A	0.96 (0.95–0.98)	3.34e–09	0.54	0.70	1.0	
9q31.3	LPAR1	9:110930238	rs10980637	T	C	1.05 (1.03–1.06)	3.28e–08	0.13	0.40	1.9	
9q34.11	DNM1	9:128236873	rs9644952	A	C	1.04 (1.03–1.05)	3.3e–08	0.22	0.81	1.1	
10p15.1	AKR1C1	10:4989436	rs536435747	AC	A	0.94 (0.93–0.96)	5.02e–11	0.13	0.51	1.2	
10q26.13	HTRA1	10:122475088	rs2672590	C	A	0.95 (0.94–0.97)	1.2e–10	0.24	0.52	0.9	BP ¹⁶
11q13.3	MYEOV	11:69208032	rs144549742	A	T	0.91 (0.88–0.94)	3.66e–11	0.04	0.50	2.8	
11q23.1	SIK2	11:11459420	rs77651758	T	C	0.92 (0.90–0.95)	1.32e–10	0.05	0.67	2.1	
12q14.1	GLI1	12:57825898	rs871871	A	G	0.96 (0.95–0.97)	7.98e–10	0.35	0.64	1.3	
12q24.31	KMT5A	12:123226288	rs1626703	C	A	1.04 (1.03–1.06)	1.55e–09	0.75	0.58	1.0	
13q21.2	DIAPH3	13:59904471	rs340208	T	A	1.04 (1.02–1.05)	3.98e–08	0.70	0.45	1.1	
17q22	CA10	17:52164544	rs59704663	A	G	1.48 (1.34–1.61)	1.29e–08	0.003	0.16	0.1	BP ¹²
18q11.2	NPC1	18:23557478	rs1788760	G	A	0.97 (0.95–0.98)	4.62e–08	0.67	0.89	1.0	
18q12.3	SETBP1	18:44571296	rs8088824	T	C	0.95 (0.94–0.96)	3.76e–12	0.78	0.19	1.1	
18q21.2	DCC	18:53189047	rs17487130	T	C	1.06 (1.04–1.07)	4.16e–14	0.40	0.86	0.9	BP ^{12,13,16,54}
19p13.12	TECR	19:14533044	rs11671111	T	C	0.96 (0.95–0.98)	1.56e–08	0.25	0.77	1.6	
19q13.42	LENG8	19:54471384	rs2287822	A	G	1.04 (1.02–1.05)	1.05e–08	0.30	0.39	1.3	
21q22.11	SLC5A3	21:33662166	rs3827180	A	G	1.04 (1.03–1.05)	2.48e–09	0.28	0.10	1.3	
Xp22.11	PCYT1B	23:24653216	rs5944665	A	G	0.97 (0.96–0.98)	3.36e–08	0.60	0.93	NA	

Table 1 (continued) | The list of previously unreported lead variants at the 41 genome-wide significant ($p < 5 \times 10^{-8}$) loci that were associated with LDH in the meta-analysis (80 724 cases, 748 975 controls) and five previously unreported loci associated with LDH related surgical operations in sensitivity analysis that were performed in FinnGen (7347 cases, 270 964 controls)

Locus	Candidate gene	CHR: POS	rsid	EA	OA	OR (95CI)	p-value	EAF	Het pval	Fin En.	Ref
Xq21.1	ITIM2A	23:79455466	rs191015078	T	C	1.05 (1.03–1.07)	4.85e–08	0.12	0.54	NA	
Xq21.1	Empty	23:82687578	rs11872003	A	T	0.92 (0.89–0.95)	3.11e–08	0.04	0.18	NA	
Xq23	CHRD1	23:110640115	rs7884700	G	A	1.04 (1.03–1.05)	9.64e–12	0.40	0.44	NA	
Surgical GWAS (FinnGen)											
1p21.1	COL11A1	1:102875067	rs1318756	C	T	1.10 (1.07–1.13)	2.41e–08	0.53	-	1.1	
2p13.3	TGFA	2:70465425	rs3732247	T	C	0.88 (0.85–0.92)	2.69e–11	0.35	-	1.2	
7p21.1	TWIST1	7:19508326	rs6944632	G	A	0.91 (0.88–0.94)	2.05e–08	0.61	-	0.9	
12p12.1	SOX5	12:23823019	rs11834104	T	G	1.14 (1.10–1.18)	5.82e–09	0.23	-	0.9	
17q24.3	SOX9	17:71514369	rs7225015	C	A	0.89 (0.86–0.93)	8.67e–10	0.29	-	0.9	

Candidate gene, a gene at a previously unreported locus the biological function of which is likely to explain the LDH association; CHR: POS, chromosome and position (genome build hg38), rsid, SNP markers identification number; EA effect allele, OA other allele; OR (95 CI), odds ratio and its 95% confidence interval; p-value, EAF effect allele frequency; Het pval, p-value for heterogeneity; Fin En. enrichment in Finns (calculated FIN AF/NFEE AF in the Genome Aggregation Database [gnomAD], FIN AF is the allele frequency in Finns and NFEE AF is the allele frequency in Europeans (Does not include Finns or Estonians)), NA not available; Ref, reference allele in which a back related disease association other than LDH was observed +/- 1 Mb in the vicinity of the lead variant, BP back pain.

transcription, has also been linked to peripheral pain perception^{26,27}. Colocalization of the LDH association signal and gene expression for ten additional genes was observed in nervous system-related tissue, supporting the notion that the nervous system has a role in the pathophysiology of LDH (Fig. 2C, Supplementary Table 6).

Cumulative incidence of LDH diagnoses uncovers early-onset variants

We investigated the cumulative incidence of LDH diagnoses and assessed the differences in the cumulative incidences between homozygotes for the lead variants at each LDH-associated locus. As observed previously^{28,29}, LDH diagnoses accumulated greatly between the ages of 40 and 50 until circa 70 years old, after which the accumulation of diagnoses was very low (Fig. 3). Overall, the variants exhibited a similar pattern in terms of the accumulation of the LDH diagnoses, with a few exceptions. The cumulative incidence of LDH differed between *GSDMC* (*gasdermin C*), rs7814941 G/G and A/A homozygotes already at the age of 26 ($p = 0.0005$; Fig. 3A) and *CHST3* (*carbohydrate sulfotransferase 3*), rs4284332 C/C and T/T homozygotes at the age of 25 ($p = 2.22e^{-5}$; Fig. 3B). For the other variants, the differences became statistically significant at a later age, *IGFBP3* (*insulin like growth factor binding protein 3*), rs788747 G/G and T/T at the 35 years of age ($p = 0.001$; Fig. 3C) and *SOX6* (*SRY-box transcription factor 6*), rs9787942 C/C and T/T at the 36 years of age ($p = 0.03$; Fig. 3D). While these genes have already been associated with LDH in previous studies, we were also able to describe the age at which diagnoses begin to accumulate for the variants under investigation. Regarding the other variants, the difference in the cumulative LDH diagnoses between homozygotes manifested at a later age, or there was no significant difference (Supplementary Table 8). Other variants cumulative morbidities followed the sample prevalence values (12.2% for LDH diagnoses and 2.6% for surgical patients, Supplementary Table 8). We also repeated the analysis restricted to radiculopathy cases (M51.1), where all variants behaved identically in the analyses compared to the original analysis (Supplementary Fig 47). The results obtained in the Estonian Biobank were similar to those observed in FinnGen (Supplementary Fig 48).

Numerous strong genetic correlations were observed for LDH

We found specific significant genetic correlations between LDH and 437 traits. The most significant positive genetic correlation in terms of the smallest p -value was observed with the number of treatments/medications taken (Fig. 4: $rg = 0.52$, $pFDR = 2.27^{-100}$), while the largest significant genetic correlation was observed with the diagnosis of dorsalgia (Fig. 4: $rg = 0.83$, $pFDR = 1.23^{-45}$). LDH was also positively genetically correlated with other pain-related endpoints, such as neck and shoulder pain (Fig. 4: $rg = 0.58$, $pFDR = 8.60^{-95}$) and knee pain (Fig. 4: $rg = 0.46$, $pFDR = 2.07^{-51}$). The most significant negative genetic correlation of LDH was with a higher level of education (Fig. 4: $rg = -0.41$, $pFDR = 2.10^{-88}$).

Evidence of a causal relationship between LDH and back pain

In Mendelian randomization, we uncovered a potential causal relationship between several factors and LDH (Fig. 5, Supplementary Table 9, Supplementary Table 10). Results suggested a causal relationship between being overweight and higher LDH risk ($OR = 1.15$, $pFDR = 0.002$, Fig. 5, Supplementary Fig 49). Similarly, a potential causal relationship was observed between lumbar spine bone mineral density (LSBMD) and higher LDH risk ($OR = 1.15$, $pFDR = 2.29^{-5}$, Fig. 5, Supplementary Fig 50). Both overweight and LSBMD are well-known risk factors for LDH, and both have been found to cause increased mechanical loading on the lumbar discs and vertebral endplates, affecting the pathogenesis of LDH^{30–34}. A possible causal relationship was also observed between a higher level of education and lower LDH risk ($OR = 0.34$, $pFDR = 4.35^{-23}$, Fig. 5, Supplementary Fig 51). In

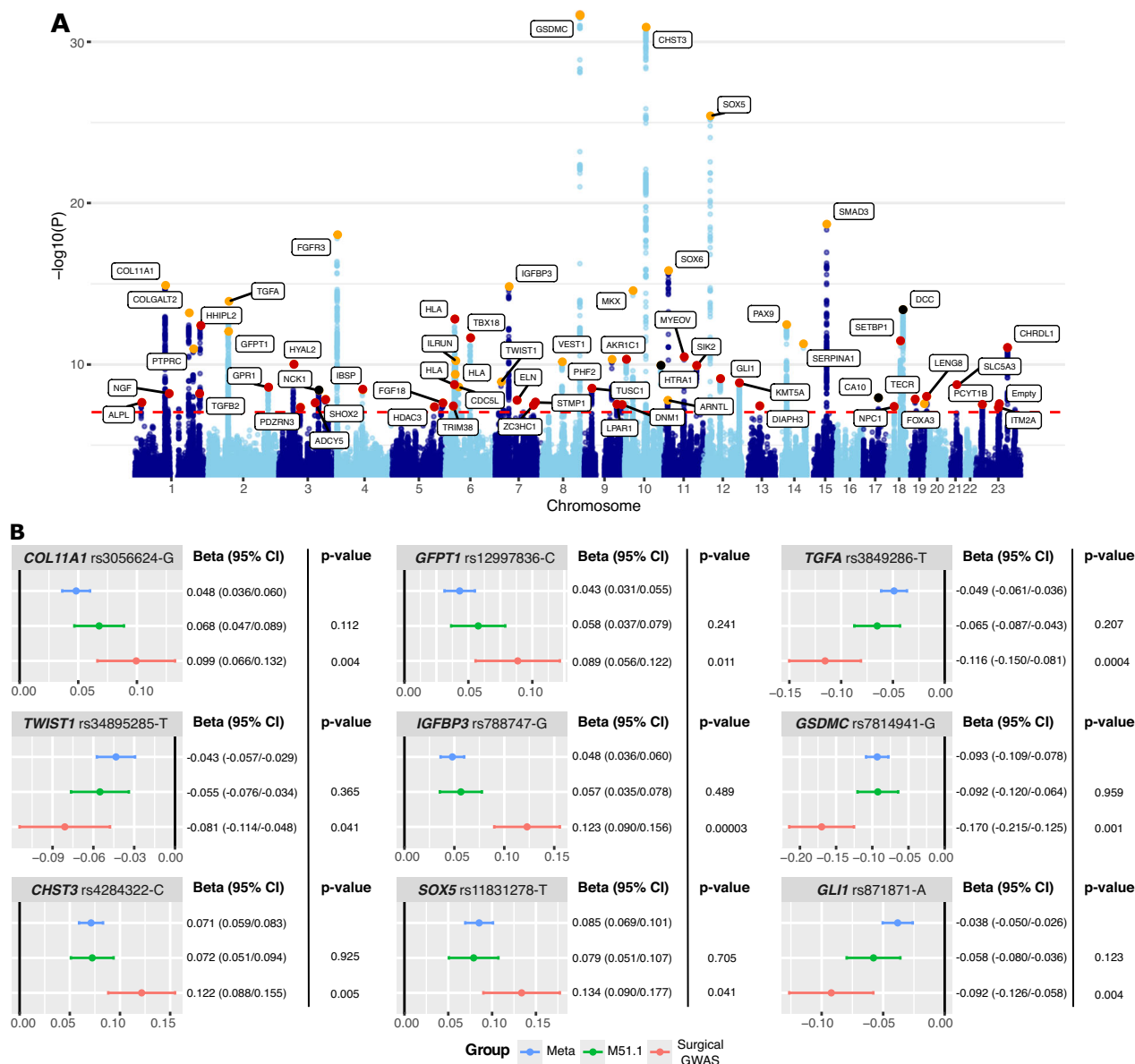
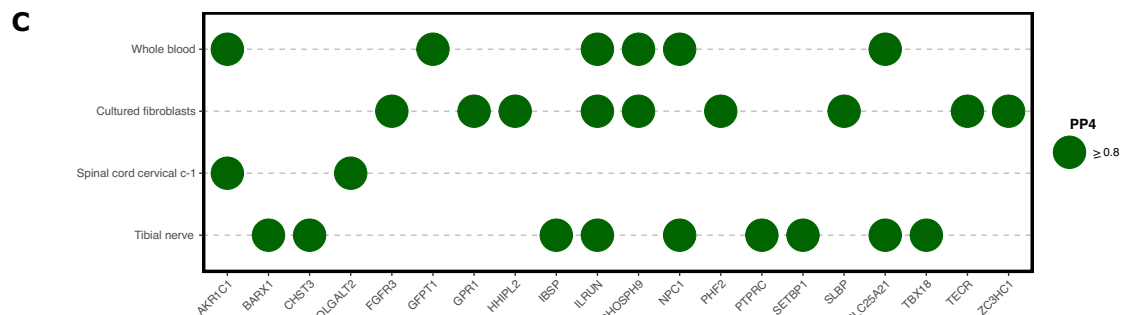


Fig. 1 | Manhattan plot and lead variant effect size comparison with sensitivity analyses. A The Manhattan plot of the genome-wide meta-analysis of LDH with 80 724 cases and 748 975 controls. Previously reported loci are indicated with orange colour, 37 previously unreported loci are highlighted in red, and 4 unreported loci for LDH, but previously associated with back pain are indicated with black. Candidate genes possibly explaining the LDH associations were used as loci identifiers. The red dashed line depicts the genome-wide significance limit ($p < 5 \times 10^{-8}$). **B** A comparison of the effect sizes of the lead variants, discovered in the original meta-analysis (ICD-10:M51: 80 724 cases, 748 975 controls, [blue]), and in sensitivity analyses with more strict case definitions (ICD-10:M51.1: 18 857 cases, 270 964 controls, [green]) or a surgery: 7347 cases, 270 964 controls, [red]). Dots indicate

effect size and vertical lines are the corresponding 95% confidence intervals. For effect differences statistical comparison, we used a two-tailed test, using group-specific effect estimates of the variants and the corresponding standard errors $\text{diff} = ((\text{Effect_Meta} - \text{Effect_M51.1}) / \sqrt{(\text{standarderror_Meta}^2 + \text{standarderror_M51.1}^2)})$, $P\text{-value} = 2 \times (1 - \text{diff})$. $P\text{-value} < 0.05$ was considered the limit of a significant effect size difference. Significant effect size differences were observed for nine variants in surgical GWAS compared to the original meta-analysis. No statistically significant effect differences were found between the meta-analysis and M51.1. Other variants can be seen in Supplementary Fig 43. Source data are provided in the Source Data file.

previous studies, patients with lower socioeconomic status tend to have more pronounced symptoms, experience more severe pain, and report higher rates of depression³⁵. Also, individuals with lower socioeconomic status are more likely to work in physically demanding manual labour occupations. It has been previously determined that physically demanding work poses a risk, particularly for recurrent LDH³⁶. Conversely, individuals with higher education usually have a better income level, they tend to seek treatment at an earlier stage after the onset of LDH, and they often have better pain management methods and generally healthier lifestyles^{35,37}. Additionally, we noted a potential causal relationship between LDH and the frequency of

tiredness in the last two weeks ($\beta = 0.02$, $p\text{FDR} = 0.046$, Fig. 5, Supplementary Fig 52) and back pain ($\beta = 0.05$, $p\text{FDR} = 6.20 \times 10^{-13}$, Fig. 5, Supplementary Fig 53). The role of LDH to cause back pain is well known³, and the possible causal relationship we observed between LDH, and increased tiredness likely arises from sleeping problems, which are commonly reported by patients suffering from radicular pain⁸. These findings should be interpreted with caution; while we did not observe pleiotropy, some causal estimates exhibited heterogeneity. In the leave-out analyses, all causal estimates were consistently in the same direction, so individual variants do not seem to drive the observed causal relationships (Supplementary Figs 54–58).



Discussion

Although the incorporation of data from three extensive biobanks enabled a large sample size and facilitated discoveries of multiple genome-wide significant associations with LDH, our sample is limited to European ancestry only. Variations in the relative prevalence of LDH cases across the sample populations included in the meta-

analysis suggest potential discrepancies in how biobanks can identify LDH patients. The eQTL analyses would benefit from incorporating bone and cartilage data, which are crucial tissues in LDH pathogenesis. Given this paper solely employs computational techniques, our findings would be enhanced by further functional research in the future.

Fig. 2 | Results of functional analyses. **A** MAGMA²⁹ gene-based test in a Manhattan plot. X-axis chromosomes, y-axis $-\log_{10}(p\text{-value})$. **B** MAGMA gene-set enrichment analysis. The plot shows significantly ($pFDR < 0.05$) enriched pathways ($pFDR < 0.05$), curated gene sets, and GO-annotations ranked by $p\text{-value} - \log_{10}(P)$. The size of the circles refers to the size of the gene set. Small grey < 15 , blue 15–100, violet 100–200, and red > 200 genes. The analysis was done using FUMA⁴⁷. The gene sets and GO annotations are from MSigDB⁴⁹. **C** Colocalizations between LDH GWAS and eQTL signals were estimated with approximate Bayesian factor analyses,

implemented with the ‘coloc.abf’ function available in the ‘coloc’ R-library⁵¹. Candidate genes and genes closest to the association signal were selected for the analysis if the closest gene was a different gene than the candidate gene (Supplementary Table 7). Variant-gene expression associations for selected tissues (whole blood, cultured fibroblasts, tibial nerve, and cervical spinal cord), were downloaded from GTEx v8⁴⁵. Colocalizations with posterior probabilities ≥ 0.8 for the variant were considered significant⁵². Source data are provided in Source Data file.

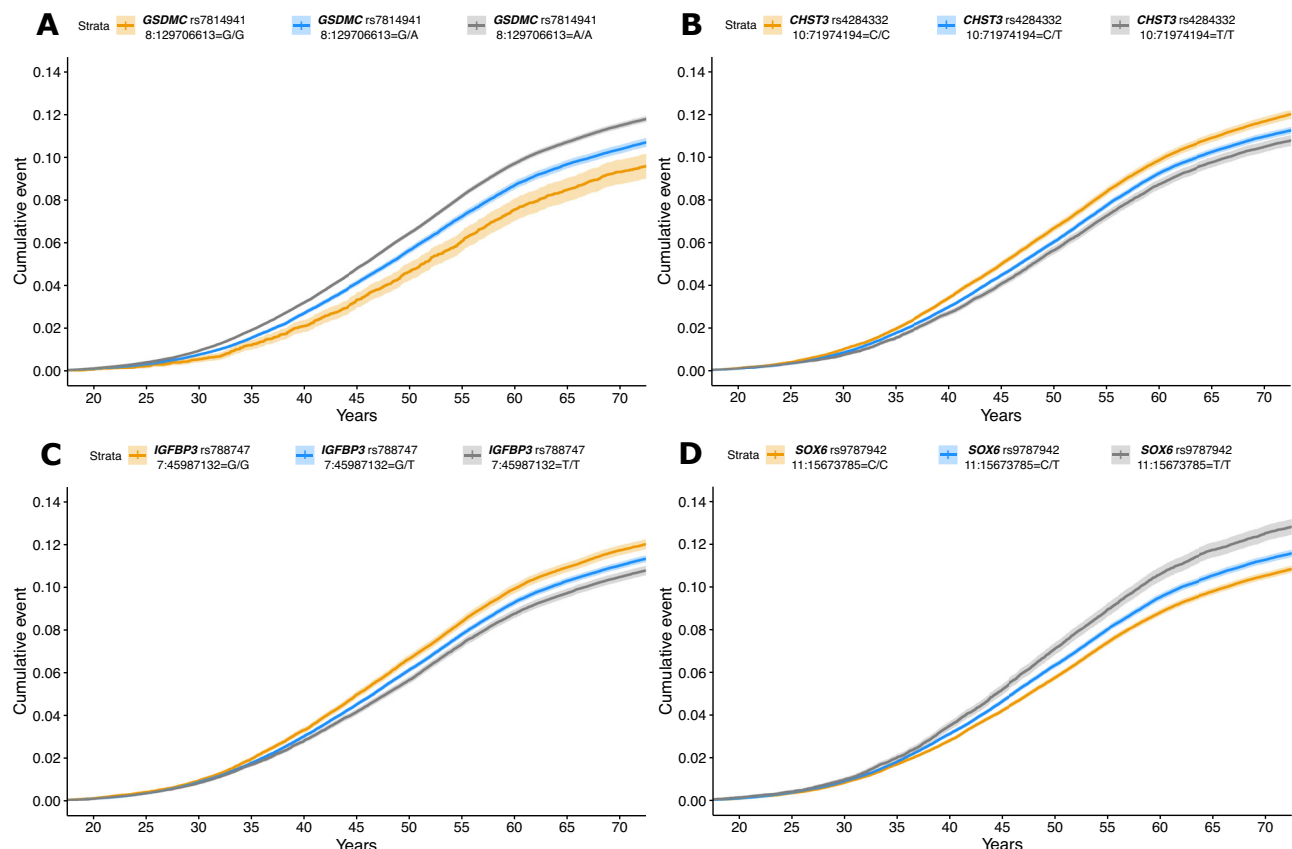


Fig. 3 | Cumulative LDH diagnoses. Cumulative LDH diagnoses for **(A)** *GSDMC*, **(B)** *CHST3*, **(C)** *IGFBP3*, **(D)** *SOX6* variants in the Kaplan-Meier plots. Age in years is on the x-axis, and the corresponding fraction of the cases is on the y-axis. The orange line depicts homozygotes for the effect allele, the grey homozygotes for the other allele, and the blue line corresponds to heterozygotes having one copy of each allele. Additionally, candidate genes, rsid, CHR:POS, and alleles were used to identify variants in the plots. Analyses were conducted in FinnGen ($n = 308\,600$);

the exact survival rates and ranges are listed in the Source data. Cumulative diagnoses for the **(A)** *GSDMC* and **(B)** *CHST3* variants before the age of 30 can be seen in more detail in Supplementary Fig 46. Corresponding plots obtained using the subset of cases with radiculopathy (M51.1) and data from Estonian Biobank are given in the supplementary information (Supplementary Fig 47, Supplementary Fig 48). CHR:POS, chromosome and position (genome build hg38). Source data are provided in Source Data file.

In conclusion, the previously unreported LDH risk loci that we found expand the understanding of the hereditary causes of LDH. While changes in disc-related structures and inflammation-related factors play a major role in the aetiology of LDH, our results from multiple downstream analyses suggest that nervous system-related mechanisms may also be implicated. Through sensitivity analyses, we also discovered genetic factors associated with a more severe form of LDH that required surgical interventions. We also recognized that certain variants exhibit a statistically significant increase in LDH risk even before the age of 30, as evidenced by the patterns of cumulative LDH incidence. All things considered, our study advances our knowledge of the genetic causes of LDH and has the potential to pave the way for the development of novel medications, preventative strategies, and therapies for LDH in the future.

Methods

Our research complies with all necessary ethical regulations (Supplementary Note 1). Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act coming into effect (in September 2013) and the start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017.

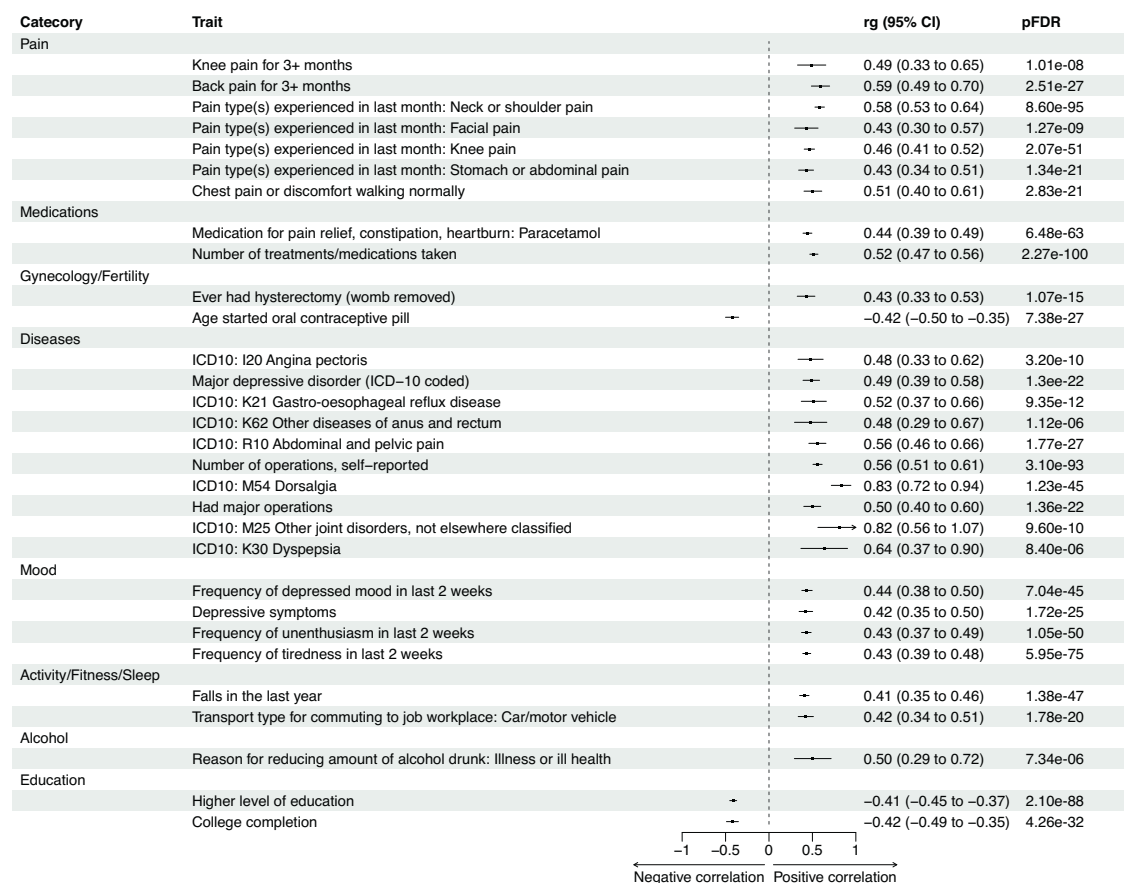


Fig. 4 | Genetic correlations (rg) between LDH and other traits. Analyses were conducted using the LDSC-software, and analysis was done utilizing our meta-analysis summary statistics ($N = 829\,699$). All traits were extracted from the GWAS database provided by the MRC-IEU database (n ranges from 7637 to 575 601; the trait-specific sample sizes are provided in Supplementary Data 3). Only the strongest observed ($rg < -0.4$ & $rg > 0.45$) correlations with a significant false discovery

corrected p -value than ($p_{FDR} < 0.05$) are shown in the figure. rg, genetic correlation coefficient value; pFDR, false discovery rate-corrected p -value. The forest plot visually represents observed rg (95% CI) values. Genetic correlations for all 437 phenotypes can be seen in Supplementary Data 3. Source data are provided in the Source Data file.

The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020), Findata permit numbers THL/2364/14.02.2020, THL/4055/14.06.00/2020, THL/3433/14.06.00/2020, THL/4432/14.06.00/2020, THL/5189/14.06.00/2020, THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021, THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021, THL/5546/14.02.00/2020, THL/2658/14.06.00/2021, THL/4235/14.06.00/202, Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20) TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases permission/extract from the meeting minutes on 4th July 2019.

The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 9 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, Auria Biobank AB17-5154 and amendment #1 (August 17 2020), AB20-5926 and amendment #1 (April 23 2020) and its modification (Sep 22

2021), Biobank Borealis of Northern Finland 2017_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020.

Study populations

FinnGen. The main goal of FinnGen (www.finnngen.fi/en) is a better understanding of disease mechanisms by combining genomic and health data from up to over 500,000 Finns, with the aim of making healthcare and medical care more efficient. The aim of the studies is to find connections between individual genetic differences and diseases. *Tenth Revision* (ICD-10) codes were used to characterize phenotype M51 (M51.0-M51.9: Thoracic, thoracolumbar, and lumbosacral intervertebral disc disorders, Supplementary Table 1). Patients without a record of these ICD codes were categorized as controls, and from the controls participants that had been diagnosed with other dorsopathies were excluded (M40-M50, M52-M54, M76). The Hospital Discharge Registry and the Cause of Death Registry served as patient data sources; the analysis did not include patients whose registration data was only available in the primary care registry. The FinnGen (R9-version) data used in the study contains 37 636 LDH cases and 270 964 controls.

Sensitivity analyses. We also performed two sensitivity GWAS analyses in FinnGen using stricter case definitions. The first of the two

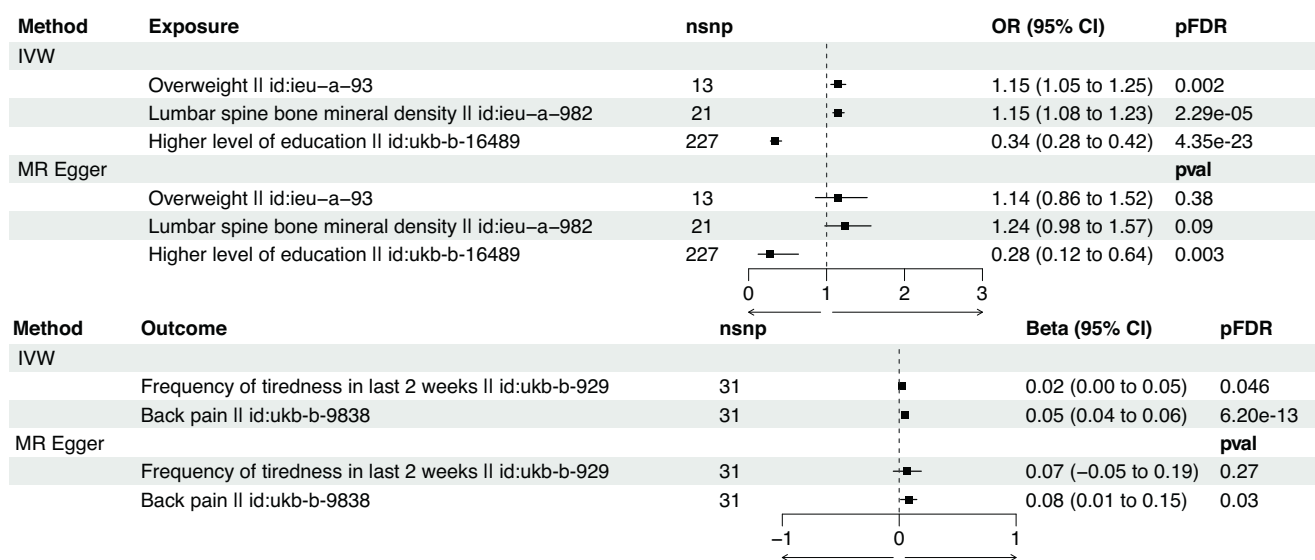


Fig. 5 | Exposures potentially causal for LDH (above, Supplementary Table 9), and outcomes that LDH was potentially found to be causal for (below, Supplementary Table 10). The analysis was performed using the TwoSampleMR R library, and data from FinnGen LDH-GWAS results ($N = 308\,600$) and MRC-IEU database (n ranges from 28 498 to 461 857; the trait-specific sample sizes are provided in Supplementary Table 11). Inverse variance weighted model was our primary analysis, for which statistical significance was considered at false discovery

rate corrected p -value ($pFDR$) < 0.05 . As a sensitivity analysis, we also performed analysis by using MR Egger. nsnp, number of SNP's; OR (95% CI), odds ratio and its 95% confidence interval; Beta (95% CI), beta estimate and its 95% confidence interval; $pFDR$, false discovery rate-corrected p -value. The forest plot visually represents observed OR (95% CI) and Beta (95% CI). Source data are provided in Source Data file.

additional case definitions included only LDH patients with the M51.1 code. There were 18 857 cases and 270 964 controls in the GWAS; LDH patients who also had other LDH codes were excluded from the analysis (Supplementary Table 1). The second additional case definition included LDH patients who had undergone an LDH-related operation (NOMESCO version 1.15, ABC07, ABC16 & ABC26)³⁸; this analysis included 7347 cases and controls of 270 964. LDH cases that had not been operated were excluded from the analysis. The same was done for operated cases that didn't have a LDH diagnosis, as these patients were probably operated as a result of acute injury.

Estonian biobank (EstBB). The Estonian Biobank (www.genomics.ut.ee/en) cohort is a volunteer-based sample of the Estonian resident adult population (aged ≥ 18 years)³⁹. Estonians represent 83%, Russians 14%, and other nationalities 3% of all participants. The current number of participants is $> 205,000$ and represents a large proportion, $> 15\%$ of the Estonian adult population, making it ideally suited to population-based studies. General practitioners (GPs) and medical personnel in the special recruitment offices have recruited participants throughout the country. At baseline, the GPs performed a standardized health examination of the participants, who also donated blood samples for DNA, white blood cells and plasma tests and filled out a 16-module questionnaire on health-related topics such as lifestyle, diet and clinical diagnoses described in WHO ICD-10. A significant part of the cohort has whole genome sequencing (3000), whole exome sequencing (2500), genome-wide single nucleotide polymorphism (SNP) array data (200 000) and/or NMR metabolome data (200 000) available. In the meta-analysis, there were 34 035 LDH cases and 66 533 controls from the Estonian Biobank.

UK biobank (UKBB). The UK biobank (<https://pan.ukbb.broadinstitute.org/>) material consists of samples collected during the years 2006–2010. Samples were collected from hundreds of thousands of people aged 40–69 from all over Great Britain. We utilized the summary statistics from the PanUKBB project and used

subset of European ancestry in the analysis that contained 9053 LDH patients and 411 478 controls from the UK biobank.

Genotyping, imputation & quality control

FinnGen. Illumina and Affymetrix DNA microarrays were used to determine genotypes. Genotype data were quality controlled to exclude variants with a low Hardy-Weinberg equilibrium (HWE) p -value ($< 1 \times 10^{-6}$), minor allele count (MAC) below three, and high missingness (cut-off 2%), as well as individuals with high genotype missingness (cut-off 5%), high levels of heterozygosity (± 4 SD), non-Finnish ancestry, and individuals whose sex did not match the genotype data. Samples were pre phased using Eagle 2.3.5, with the number of conditioning haplotypes set to 20 000. Beagle 4.1 (version 08Jun17.d8b) was used for genotype imputation. The reference panel was Finnish SISu v4, and the imputation protocol has been described at (<https://www.protocols.io/view/genotype-imputation-workflow-v3-0-e6nvw78dlmkj/v2>). Finally, post-imputation quality control was carried out to exclude variants with imputation information less than 0.6.

EstBB. For genotyping, Illumina Human CoreExome, OmniExpress, 370CNV BeadChip and GSA arrays were used. Individuals were excluded from the analysis if their call-rate was $< 95\%$ or sex defined using X chromosome heterozygosity estimates didn't match phenotypic data. Quality control included filtering on the basis of sample call rate ($< 98\%$), heterozygosity ($> \text{mean} \pm 3$ SD), genotype and phenotype sex discordance, cryptic relatedness ($IBD > 20\%$) and outliers from the European descent based on the MDS plot in comparison with HapMap reference samples. SNP quality filtering included call rate ($< 99\%$), MAF ($< 1\%$) and extreme deviation from Hardy-Weinberg equilibrium ($P < 1 \times 10^{-4}$). Imputation was performed using SHAPEIT2 for pre-phasing, the Estonian-specific reference panel⁴⁰ and IMPUTE2⁴¹ with default parameters. Association testing was carried out with snptest-2.5.2, adjusting for 4 PCs, arrays, current age, and sex (when relevant). Variants with call-rate $< 95\%$, MAF $< 1\%$ or HWE p -value $< 1e-4$ (autosomal variants only) and indels were excluded.

GWAS. In FinnGen and in EstBB, GWAS using an additive genetic model was performed using the Regenie (version 2.2.4) programme⁴², adjusting each phenotype for age, sex, the first 10 genetic principal components, and genotyping arrays. The sensitivity analyses with stricter case definitions conducted in FinnGen were performed using Regenie and the same covariates as above. The goal of the sensitivity analyses was to evaluate whether general degeneration of IVD has an effect on the effect estimates of the lead variants observed in the meta-analysis, and if the effect estimates obtained in the original meta-analysis differ from the ones obtained using stricter case definitions. The aim was also to identify variants that could underlie LDH cases requiring surgery.

Meta-analysis. A Python-based software was used for inverse-variance weighted meta-analysis (https://github.com/FINNGEN/META_ANALYSIS/). Variant data from the Estonian and UK biobanks were converted with from hg19 to hg38 prior to meta-analysis using the Picard liftover (<http://broadinstitute.io/picard>). In case of no exact match, matching was tried by flipping strand and or switching (EA > OA/OA > EA) for the EstBB and UKBB variants. If there were multiple variants in the same position, the exact match was favoured. In total, there were 829 699 participants in the meta-analysis, of which there were 80 724 cases and 748 975 controls.

Candidate gene characterization. We defined a locus as a window of 2MB ($\pm 1,000,000$ bases) containing at least one variant associated with LDH at $P < 5 \times 10^{-8}$. For the loci that had not been reported in association with LDH in prior studies, we determined a potential candidate gene with a relevant biological function with the help of literature and databases (Genbank⁴³, Uniprot⁴⁴, GTEx-Portal⁴⁵) and identified variants affecting gene regulation (eQTL).

Conditional analysis. We performed conditional analyses for the loci to identify possible secondary signals. The analyses were performed using the GCTA software (1.93.0 beta Linux)⁴⁶ and the meta-analysed summary statistics. FinnGen was used as the reference sample to estimate linkage disequilibrium corrections. The associations were first conditioned on the most significant variant at each locus using the `--cojo-cond` option and default settings. Conditioning persisted until no variant achieved genome-wide significance ($p < 5 \times 10^{-8}$).

Heritability. The LDSC software (version 1.0.1)¹⁷ was used to calculate the SNP-based heritability estimate. Heritability estimation was performed using the liability scale, with a sample prevalence of 0.097 and a population prevalence of 0.14, as estimated by Zhang et al. (2016)². HapMap3, a European-only reference panel provided by Alkes Price's group (<https://alkesgroup.broadinstitute.org/LDSCORE/>) was used in heritability estimation. Summary statistics included 1,202,124 SNPs, and after merging with the LD reference panel 1,176,449 SNPs remained. In FinnGen's data, the sample prevalence was 0.122. The reference panel remained the same, 1,166,621 SNPs were included from the summary statistics based on FinnGen, and after merging with the LD reference panel 1,157,245 SNPs remained.

Functional annotations. Functional annotations of the results of the meta-analysis were completed using FUMA⁴⁷. Functional settings were selected for the analysis, in which case the programme uses functional information for mapping. Positional mapping was also performed, and for that, SNP markers were selected for the region of exons or introns affecting post-transcriptional modifications and involved in gene regulation. Filtering of SNP markers based on CADD results was also done, which provided additional information on the possible harmful effects of SNP markers. In addition, filtering of SNP markers was performed based on the RegulomeDB results, and in turn, information was obtained based on gene expression data and epigenomics about the

possible functions of SNP markers affecting gene regulation. In the mapping, gene expression data were also utilized, and eQTL mapping was performed. For this, we used whole blood, cultured fibroblasts, tibial nerve, and cervical spinal cord cervical (GTEx v8) as tissues, and the focus was only on genes involved in protein-coding. A MAGMA analysis⁴⁸, a functional association test, was also performed in the run, which focuses on gene-level information, unlike GWAS, where associations are reported at the variant level. MAGMA uses curated gene sets and GO annotations from MSigDB⁴⁹ in the analyses. A 10 kb gene window and selected GTEx v8 tissue variants were put into the analysis; the HLA region was also left out of the annotations. For gene prioritization we also utilized fastBAT, from the GCTA software package (1.93.0 beta Linux)⁴⁶. By utilizing LD correlations between SNPs from a reference sample with individual-level genotypes and summary data from GWAS, fastBAT computes the association p -value for a set of SNPs from an approximated distribution of the sum of χ^2 -statistics over the SNPs at gene regions⁵⁰. We analysed the loci using a 2MB window and FinnGen as a reference for LD estimation. Default methods and a 0.9 r^2 threshold for LD pruning were used in the analysis. Results with $p_{\text{fastBAT}} < 5 \times 10^{-8}$ and $p_{\text{GWAS}} < 5 \times 10^{-8}$ were considered significant⁵⁰.

Colocalizations. We investigated colocalizations between LDH association signals and gene expression: approximate Bayesian factor analyses were performed with the 'coloc.abf' function found in the 'coloc' R-library (version 5.2.3)⁵¹. Colocalizations were investigated for a total of 78 genes (Supplementary Table 7). Candidate genes and genes closest to the association signal were selected for the analysis if the closest gene was a different gene than the candidate gene. For the analysis, we selected tissues that could be relevant for LDH, namely whole blood, cultured fibroblasts, tibial nerve, and cervical spinal cord. Variant-gene expression associations (GTEx v8), for tissues were downloaded (02/22/2024) from GTEx-Portal⁴⁵. Colocalizations with posterior probabilities ≥ 0.8 for the variant were considered significant⁵².

Survival analysis. By exploring the cumulative events, our aim was to observe how the LDH diagnoses accumulate for different variants according to age, and to evaluate whether there are early onset variants. Of the variants where differences in the accumulation of diagnoses were observed on the basis of the plots, we determined the diagnosis age in years when the curve of the effect allele (EA) homozygotes statistically differed from the curve of other allele (OA) homozygotes of the same variant. The calculation was performed with the two-tailed test by using the survival rates and survival rate standard errors obtained with the 'survfit' function, which is part of the 'survival' (<https://github.com/therneau/survival>, version 3.2-7) R library. The equation is given below:

$$\text{diff} = (\text{'survival rate EA/OA'} - \text{'survival rate OA/OA'}) / \sqrt{(\text{'se_survival_rate_EA/OA'}^2 + \text{'se_survival_rate_OA/OA'}^2)},$$

$p = 2 \times (1 - \text{diff})$. $P < 0.05$ was used as the statistical cut-off value for a significant difference. Additionally, we calculated cumulative morbidity for every variant to clarify further whether some variants accumulate more diagnoses.

Genetic correlations. Genetic correlations were calculated between our meta-analysed LDH summary statistics and 437 other phenotypes extracted from the GWAS database provided by the MRC Integrative Epidemiology Unit (IEU) (<https://gwas.mrcieu.ac.uk/>). The LDSC software (version 1.0.1)^{17,53} was used for these calculations. We used a false discovery rate (FDR)-corrected p -value ($p_{\text{FDR}} < 0.05$) as the limit for significant correlations. HapMap3 European-only reference panel provided by Alkes Price's group (<https://alkesgroup.broadinstitute.org/LDSCORE/>) were used in genetic correlations. For genetic correlations, we used our LDH

summary statistics and the GWAS summary statistics that were similar to those that have been previously available in LDhub (<https://ldsc.broadinstitute.org>).

Mendelian randomization. For Mendelian randomization, we used the Two-Sample MR (<https://github.com/MRCIEU/TwoSampleMR>, version 0.5.8) R library to conduct a bi-directional Mendelian randomization to examine the causal relationships between LDH and its associated risk factors. Risk factors included in the analysis (Supplementary Table 11). Due to a bi-directional study approach, we were able to evaluate whether risk factors are causal for LDH and, consequently, if LDH is causal for risk factors. We obtained the LDH instruments from the FinnGen GWAS results since many of the GWAS data provided by the MRC-IEU are UKBB-based. This ensured that there was no overlap between the study populations. Variants correlated with each other were removed from the data so that only independent variants would be included in the analysis. For this, we used the default clumping settings (clumping window 10 kb, $r^2 = 0.001$). We run our primary analysis using the Inverse Variance Weighted (IVW) model. In the sensitivity analyses, we obtained MR Egger estimates. We also performed Cochran's Q-test and the MR Egger intercept test to evaluate the heterogeneity and pleiotropy of the instruments. A leave-one-out analysis was also performed to see if there is a specific SNP driving a potentially observable causal relationship.

Reporting summary

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

Data availability

The summary statistics from the meta-analysis and sensitivity analyses are available in FinnGen's public cloud bucket (https://storage.googleapis.com/fg-publication-green-public/F_2021_048_20240909/LDH_summary_statistics.zip). The individual-level data are available under restricted access for legal and ethical reasons. Formal approval for the researchers is required to access the data: please see https://www.finnngen.fi/en/access_results for more details. To get access to FinnGen GWAS summary statistics, fill out an online form at: <https://elomake.helsinki.fi/lomakkeet/124935/lomake.html>. Access to individual-level data and genotype data is managed by the Finnish Biobank Cooperative at the Fingenius portal [<https://site.fingenius.fi/en/>]. The expected response time for access requests to individual-level data is 1–2 months, with an account termination date of December 31, 2028. The genome-wide data used in calculating genetic correlations and causal relationships are available at the MRC-IEU GWAS database (<https://gwas.mrcieu.ac.uk/>). Source data are provided with this paper.

Code availability

While we did not develop any original code, all tools used in our study were open source. The citations and URLs listed in the Methods section contain the code for every tool used in the analyses.

References

- Fardon, D. F. et al. Lumbar disc nomenclature: version 2.0. *Spine J.* **14**, 2525–2545 (2014).
- Zhang, J. et al. Identification of lumbar disc disease hallmarks: a large cross-sectional study. *Springerplus* **5**, (2016).
- Dower, A., Davies, M. A. & Ghahreman, A. Pathologic basis of lumbar radicular pain. *World Neurosurg.* **128**, 114–121 (2019).
- Cosamalón-Gan, I. et al. Inflammation in the intervertebral disc herniation. *Neurocirugia (Engl. Ed.)* **32**, 21–35 (2021).
- Patel, E. A. & Perloff, M. D. Radicular pain syndromes: cervical, lumbar, and spinal stenosis. *Semin Neurol.* **38**, 634–639 (2018).
- Jacobs, W. C. H. et al. Surgery versus conservative management of sciatica due to a lumbar herniated disc: a systematic review. *Eur. Spine J.* **20**, 513–22 (2011).
- Liu, C. et al. Surgical versus non-surgical treatment for sciatica: systematic review and meta-analysis of randomised controlled trials. *BMJ* e070730 <https://doi.org/10.1136/bmj-2022-070730> (2023).
- Konstantinou, K. et al. Prognosis of sciatica and back-related leg pain in primary care: the ATLAS cohort. *Spine J.* **18**, 1030–1040 (2018).
- Björnsdóttir, G. et al. Rare SLC13A1 variants associate with intervertebral disc disorder highlighting role of sulfate in disc pathology. *Nat. Commun.* **13**, 634 (2022).
- Björnsdóttir, G. et al. Sequence variant at 8q24.21 associates with sciatica caused by lumbar disc herniation. *Nat. Commun.* **8**, 14265 (2017).
- Theodore, N. et al. Genetic predisposition to symptomatic lumbar disk herniation in pediatric and young adult patients. *Spine (Phila. Pa 1976)* **44**, E640–E649 (2019).
- Johnston, K. J. A. et al. Genome-wide association study of multisite chronic pain in UK biobank. *PLoS Genet* **15**, e1008164 (2019).
- Suri, P. et al. Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. *PLoS Genet* **14**, e1007601 (2018).
- Naureen, Z. et al. Genetics of pain: from rare Mendelian disorders to genetic predisposition to pain. *Acta Biomed.* **91**, e2020010 (2020).
- Parreira, P., Maher, C. G., Steffens, D., Hancock, M. J. & Ferreira, M. L. Risk factors for low back pain and sciatica: an umbrella review. *Spine J.* **18**, 1715–1721 (2018).
- Freidin, M. B. et al. Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals. *Pain* **160**, 1361–1373 (2019).
- Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet* **47**, 291–295 (2015).
- Wu, Z. L. et al. Role of the Wnt pathway in the formation, development, and degeneration of intervertebral discs. *Pathology Research and Practice.* **220**, 153366 (2021).
- Honda, T., Yamamoto, H., Ishii, A. & Inui, M. PDZRN3 negatively regulates BMP-2-induced osteoblast differentiation through inhibition of Wnt signaling. *Mol. Biol. Cell* **21**, 3269–3277 (2010).
- Freemont, A. J. et al. Nerve growth factor expression and innervation of the painful intervertebral disc. *J. Pathol.* **197**, 286–292 (2002).
- Li, X. et al. The adaptor protein Nck-1 couples the netrin-1 receptor DCC (deleted in colorectal cancer) to the activation of the small GTPase Rac1 through an atypical mechanism. *J. Biol. Chem.* **277**, 37788–37797 (2002).
- Montoliu-Gaya, L. et al. CA10 regulates neurexin heparan sulfate addition via a direct binding in the secretory pathway. *EMBO Rep* **22**, (2021).
- Noborn, F. & Sterky, F. H. Role of neurexin heparan sulfate in the molecular assembly of synapses—expanding the neurexin code? *FEBS Journal* **290**, 252–265 (2021).
- Sterky, F. H. et al. Carbonic anhydrase-related protein CA10 is an evolutionarily conserved pan-neurexin ligand. *Proceedings of the National Academy of Sciences* **114**, E1253–E1262 (2017).
- Tonello, R. et al. The contribution of endocytosis to sensitization of nociceptors and synaptic transmission in nociceptive circuits. <https://doi.org/10.1097/j.pain.0000000000002826> (2023).
- Ambrose, R. L. et al. Molecular characterisation of ILRUN, a novel inhibitor of proinflammatory and antimicrobial cytokines. *Heliyon* **6**, e04115 (2020).
- Johnston, K. J. A., Cote, A. C., Hicks, E., Johnson, J. & Huckins, L. M. Genetically regulated gene expression in the brain associated with chronic pain: relationships with clinical traits and potential for drug

- repurposing. *Biol. Psychiatry* <https://doi.org/10.1016/j.biopsych.2023.08.023> (2023).
28. Ma, D. et al. Trend of the incidence of lumbar disc herniation: Decreasing with aging in the elderly. *Clin. Inter. Aging* **8**, 1047–1050 (2013).
 29. Roberto Vialle, L., Neves Vialle, E., Esteban Suárez Henao, J. & Giraldo, G. Lumbar disc herniation. *Rev. Bras Ortop.* **45**, (2010).
 30. Fine, N. et al. Intervertebral disc degeneration and osteoarthritis: a common molecular disease spectrum. *Nat Rev. Rheumatol.* <https://doi.org/10.1038/s41584-022-00888-z> (2023).
 31. Hangai, M. et al. Factors associated with lumbar intervertebral disc degeneration in the elderly. *Spine J.* **8**, 732–40 (2008).
 32. Videman, T., Levälähti, E. & Battié, M. C. The effects of anthropometrics, lifting strength, and physical activities in disc degeneration. *Spine (Philos. Pa 1976)* **32**, 1406–13 (2007).
 33. Zhou, X. et al. Trans-ethnic polygenic analysis supports genetic overlaps of lumbar disc degeneration with height, body mass index, and bone mineral density. *Front Genet* **9**, 267 (2018).
 34. Livshits, G. et al. Evidence that bone mineral density plays a role in degenerative disc disease: The UK twin spine study. *Ann. Rheum. Dis.* **69**, 2102–2106 (2010).
 35. Farrell, S. F. et al. A shared genetic signature for common chronic pain conditions and its impact on biopsychosocial traits. <https://doi.org/10.1101/2022.03.13.22272317>.
 36. Louis, R., Kollannur, L. J., A. S. U. & G. R. C. Risk factors for recurrent lumbar disc herniation. *Intl. J. Acad. Med. Pharm.* <https://doi.org/10.47009/jamp.2023.5.4.264> (2016).
 37. Olson, P. R. et al. Lumbar disc herniation in the spine patient outcomes research trial: does educational attainment impact outcome? *Spine (Philos. Pa 1976)* **36**, 2324–2332 (2011).
 38. Salmenkivi, J., Sund, R., Paavola, M., Ruuth, I. & Malmivaara, A. Mortality caused by surgery for degenerative lumbar spine. *Spine (Philos. Pa 1976)* **42**, 1080–1087 (2017).
 39. Leitsalu, L., Alavere, H., Tammesoo, M.-L., Leego, E. & Metspalu, A. Linking a population biobank with national health registries—the estonian experience. *J. Pers. Med* **5**, 96–106 (2015).
 40. Mitt, M. et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet* **25**, 869–876 (2017).
 41. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529 (2009).
 42. Mbatchou, J. et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat. Genet* **53**, 1097–1103 (2021).
 43. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res* **44**, D67–D72 (2016).
 44. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506–D515 (2019).
 45. Stanfill, A. G. & Cao, X. Enhancing Research Through the Use of the Genotype-Tissue Expression (GTEx) Database. *Biol. Res Nurs.* **23**, 533–540 (2021).
 46. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet* **88**, 76–82 (2011).
 47. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
 48. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* **11**, e1004219 (2015).
 49. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–40 (2011).
 50. Bakshi, A. et al. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Sci Rep* **6**, (2016).
 51. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
 52. Sliz, E. et al. Evidence of a causal effect of genetic tendency to gain muscle mass on uterine leiomyomata. *Nat. Commun.* **14**, 542 (2023).
 53. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet* **47**, 1236–1241 (2015).
 54. Bortsov, A. V. et al. Brain-specific genes contribute to chronic but not to acute back pain. *Pain. Rep.* **7**, E1018 (2022).

Acknowledgements

E.S. was funded by the Academy of Finland (grant number: 338229) and Orion Research Foundation Sr. J.K. was funded by the Sigrid Juselius Foundation. The authors wish to acknowledge CSC – IT Centre for Science, Finland, for computational resources. We want to acknowledge the participants and investigators of the FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis Pharma AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biobankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiobankki.fi), Biobank Borealis of Northern Finland (<https://www.ppsph.fi/Tutkimus-ja-opetus/Biobankki/Pages/Biobank-Borealis-briefly-in-English.aspx>), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiobankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biobankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biobankkitoiminta), Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biobankki/Biobankki/) and Arctic Biobank (<https://www oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank>). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbMRI.fi). Finnish Biobank Cooperative -FINBB (<https://finbb.fi/>) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Ingenious® services (<https://site.fingenious.fi/en/>) managed by FINBB. This study was funded by European Union through the European Regional Development Fund Project No. 2014-2020.4.01.15-0012 GENTRANSMED, the Estonian Research Council Grant PUTs (PRG1911, PRG1291) and by Estonian Ministry of Education and Research Funding (TK214). Data analysis was carried out in part in the High-Performance Computing Centre of University of Tartu. The activities of the EstBB are regulated by the Human Genes Research Act, which was adopted in 2000 specifically for the operations of the EstBB. Individual level data analysis in the EstBB was carried out under ethical approval [1.12/624] from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application [NO4] from the Estonian Biobank.

Author contributions

V.S., J.M., E.S., J.K., J.Ke. conceptualized the study and contributed to the analysis plan. V.S. and E.R. analysed data and generated results. V.S., J.M., E.S., J.K., and J.Ke. interpreted the results. V.S., J.M., and E.S. wrote the original manuscript. FinnGen and the Estonian Biobank Research Team provided data for the study. E.S. and J.K. supervised the study and

obtained funding. J.K., R.M., K.R., A.G.E., A.R., P.P., and T.E. provided additional study resources. All authors contributed to revising the content of the manuscript and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-53467-z>.

Correspondence and requests for materials should be addressed to Ville Salo.

Peer review information *Nature Communications* thanks Jing Hua Zhao and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

© The Author(s) 2024

FinnGen

Aarno Palotie^{5,7,8}, Mark Daly^{5,7,8}, Bridget Riley-Gills⁹, Howard Jacob⁹, Dirk Paul¹⁰, Slavé Petrovski¹⁰, Heiko Runz¹¹, Sally John¹¹, George Okafo¹², Nathan Lawless¹², Heli Salminen-Mankonen¹², Robert Plenge¹³, Joseph Maranville¹³, Mark McCarthy¹⁴, Margaret G. Ehm¹⁵, Kirsi Auro¹⁶, Simonne Longereich¹⁷, Anders Mälarstig¹⁸, Katherine Klinger¹⁹, Clement Chatelain¹⁹, Matthias Gossel¹⁹, Karol Estrada²⁰, Robert Graham²⁰, Robert Yang²¹, Chris O'Donnell²², Tomi P. Mäkelä⁵, Jaakko Kaprio⁵, Petri Virolainen²³, Antti Hakanen²³, Terhi Kilpi²⁴, Markus Perola²⁴, Jukka Partanen²⁵, Anne Pitkäranta²⁶, Taneli Raivio²⁶, Jani Tikkanen²⁷, Raisa Serpi²⁷, Tarja Laitinen²⁸, Veli-Matti Kosma²⁹, Jari Laukkanen^{30,31}, Marco Hautalahti³², Outi Tuovila³³, Raimo Pakkanen³³, Jeffrey Waring⁹, Fedik Rahimov⁹, Ioanna Tachmazidou¹⁰, Chia-Yen Chen¹¹, Zhihao Ding¹², Marc Jung¹², Shameek Biswas¹³, Rion Pendergrass¹⁴, David Pulford³⁴, Neha Raghavan¹⁷, Adriana Huertas-Vazquez¹⁷, Jae-Hoon Sul¹⁷, Xinli Hu¹⁸, Åsa Hedman¹⁸, Manuel Rivas^{20,35}, Dawn Waterworth³⁶, Nicole Renaud²², Ma'en Obeidat²², Samuli Ripatti⁵, Johanna Schleutker²³, Mikko Arvas²⁵, Olli Carpén²⁶, Reetta Hinttala²⁷, Arto Mannermaa²⁹, Katriina Aalto-Setälä³⁷, Mika Kähönen²⁸, Johanna Mäkelä³², Reetta Kälviäinen³⁸, Valtteri Julkunen³⁸, Hilka Soininen³⁸, Anne Remes³⁹, Mikko Hiltunen⁴⁰, Jukka Peltola⁴¹, Minna Raivio⁴², Pentti Tienari⁴², Juha Rinne⁴³, Roosa Kallionpää⁴³, Juulia Partanen⁵, Ali Abbasi⁹, Adam Ziemann⁹, Nizar Smaoui⁹, Anne Lehtonen⁹, Susan Eaton¹¹, Sanni Lahdenperä¹¹, Natalie Bowers¹⁴, Edmond Teng¹⁴, Fanli Xu⁴⁴, Laura Addis⁴⁴, John Eicher⁴⁴, Qingqin S. Li⁴⁵, Karen He³⁶, Ekaterina Khramtsova³⁶, Martti Färkkilä⁴², Jukka Koskela⁴², Sampsa Pikkarainen⁴², Airi Jussila⁴¹, Katri Kaukinen⁴¹, Timo Blomster³⁹, Mikko Kiviniemi³⁸, Markku Voutilainen⁴³, Tim Lu¹⁴, Linda McCarthy⁴⁴, Amy Hart³⁶, Meijian Guan³⁶, Jason Miller¹⁷, Kirsi Kalpala¹⁸, Melissa Miller¹⁸, Kari Eklund⁴², Antti Palomäki⁴³, Pia Isomäki⁴¹, Laura Pirilä⁴³, Oili Kaipainen-Seppänen³⁸, Johanna Huhtakangas³⁹, Nina Mars⁵, Apinya Lertratanakul⁹, Coralie Viollet¹⁰, Marla Hochfeld¹³, Jorge Esparza Gordillo⁴⁴, Fabiana Farias¹⁷, Nan Bing¹⁸, Margit Pelkonen³⁸, Paula Kauppi⁴², Hannu Kankaanranta^{37,46,47}, Terttu Harju³⁹, Riitta Lahesmaa⁴³, Hubert Chen¹⁴, Joanna Betts⁴⁴, Rajashree Mishra⁴⁴, Majd Mouded⁴⁸, Debby Ngo⁴⁸, Teemu Niiranen⁴⁹, Felix Vaura⁴⁹, Veikko Salomaa⁴⁹, Kaj Metsärinne⁴³, Jenni Aittokallio⁴³, Jussi Hernesniemi⁴¹, Daniel Gordin⁴², Juha Sinisalo⁴², Marja-Riitta Taskinen⁴², Tiinamaija Tuomi⁴², Timo Hiltunen⁴², Amanda Elliott^{5,8,50}, Mary Pat Reeve⁵, Sanni Ruotsalainen⁵, Audrey Chu⁴⁴, Dermot Reilly⁵¹, Mike Mendelson⁵², Jaakko Parkkinen¹⁸, Tuomo Meretoja⁴², Heikki Joensuu⁴², Johanna Mattson⁴², Eveliina Salminen⁴², Annika Auranen⁴¹, Peeter Karihtala³⁹, Päivi Auvinen³⁸, Klaus Elenius⁴³, Esa Pitkänen⁵, Relja Popovic⁹, Margarete Fabre¹⁰, Jennifer Schutzman¹⁴, Diptee Kulkarni⁴⁴, Alessandro Porello³⁷, Andrey Loboda¹⁷, Heli Lehtonen¹⁸, Stefan McDonough¹⁸, Sauli Vuoti⁵³, Kai Kaarniranta^{38,54}, Joni A. Turunen^{55,56}, Terhi Ollila⁴², Hannu Uusitalo⁴¹, Juha Karjalainen⁵, Mengzhen Liu⁹, Stephanie Loomis¹¹, Erich Strauss¹⁴, Hao Chen¹⁴, Kaisa Tasanen³⁹, Laura Huilaja³⁹, Katariina Hannula-Jouppi⁴², Teea Salmi⁴¹, Sirkku Peltonen⁴³, Leena Koulou⁴³, David Choy¹⁴, Ying Wu¹⁸, Pirkko Pussinen⁴², Aino Salminen⁴², Tuula Salo⁴², David Rice⁴², Pekka Nieminen⁴², Ulla Palotie⁴², Maria Siponen³⁸, Liisa Suominen³⁸, Päivi Mäntylä³⁸, Ulvi Gursoy⁴³, Vuokko Anttonen³⁹, Kirsi Sipilä⁵⁷, Hannele Laivuori⁵, Venla Kurra⁴¹, Laura Kotaniemi-Talonen⁴¹, Oskari Heikinheimo⁴²,

Ilkka Kalliala⁴², Lauri Aaltonen⁴², Varpu Jokimaa⁴³, Marja Väärasmäki³⁹, Outi Uimari³⁹, Laure Morin-Papunen³⁹, Maarit Niinimäki³⁹, Terhi Piltonen³⁹, Katja Kivinen⁵, Elisabeth Widen⁵, Taru Tukiainen⁵, Niko Välimäki⁵⁸, Eija Laakkonen⁵⁹, Jaakko Tyrmi^{37,57}, Heidi Silven⁵⁷, Riikka Arffman⁵⁷, Susanna Savukoski⁵⁷, Triin Laisk⁴, Natalia Pujol⁴, Janet Kumar¹⁵, Iiris Hovatta⁵⁸, Erkki Isometsä⁴², Hanna Ollila⁵, Jaana Suvisaari⁴⁹, Thomas Damm Als⁶⁰, Antti Mäkitie⁶¹, Argyro Bizaki-Vallaskangas⁴¹, Sanna Toppila-Salmi⁶², Tytti Willberg⁴³, Elmo Saarentaus⁵, Antti Aarnisalo⁴², Elisa Rahikkala³⁹, Kristiina Aittomäki⁶³, Fredrik Åberg⁶⁴, Mitja Kurki^{5,50}, Aki Havulinna^{5,49}, Juha Mehtonen⁵, Shabbeer Hassan⁵, Pietro Della Briotta Parolo⁵, Wei Zhou⁵⁰, Mutaamba Maasha⁵⁰, Susanna Lemmelä⁵, Aoxing Liu⁵, Arto Lehisto⁵, Andrea Ganna⁵, Vincent Llorens⁵, Henrike Heyne⁵, Joel Rämö⁵, Rodos Rodosthenous⁵, Satu Strausz⁵, Tuula Palotie^{42,58}, Kimmo Palin⁵⁸, Javier Gracia-Tabuenca³⁷, Harri Siirtola³⁷, Tuomo Kiiskinen⁵, Jiwoo Lee^{5,50}, Kristin Tsuo^{5,50}, Kati Kristiansson²⁴, Kati Hyvärinen⁶⁵, Jarmo Ritari⁶⁵, Katri Pyrkäs⁵⁷, Minna Karjalainen⁵⁷, Tuomo Mantere²⁷, Eeva Kangasniemi²⁸, Sami Heikkinen⁴⁰, Nina Pitkänen²³, Samuel Lessard¹⁹, Lila Kallio²³, Tiina Wahlfors²⁴, Eero Punkka²⁶, Sanna Siltanen²⁸, Teijo Kuopio^{30,31}, Anu Jalanko⁵, Huei-Yi Shen⁵, Risto Kajanne⁵, Mervi Aavikko⁵, Helen Cooper⁵, Denise Öller⁵, Rasko Leinonen^{5,66}, Henna Palin²⁸, Malla-Maria Linna²⁶, Masahiro Kanai⁵⁰, Zhili Zheng⁵⁰, L. Elisa Lahtela⁵, Mari Kaunisto⁵, Elina Kilpeläinen⁵, Timo P. Sipilä⁵, Oluwaseun Alexander Dada⁵, Awaisa Ghazal⁵, Anastasia Kytölä⁵, Rigbe Weldatsadik⁵, Kati Donner⁵, Anu Loukola²⁶, Päivi Laiho²⁴, Tuuli Sistonen²⁴, Essi Kaiharju²⁴, Markku Laukkanen²⁴, Elina Järvensivu²⁴, Sini Lähteenmäki²⁴, Lotta Männikkö²⁴, Regis Wong²⁴, Auli Toivola²⁴, Minna Brunfeldt²⁴, Hannele Mattsson²⁴, Sami Koskelainen²⁴, Tero Hiekkalinna²⁴, Teemu Paajanen²⁴, Kalle Pärn⁵, Mart Kals⁵, Shuang Luo⁵, Shanmukha Sampath Padmanabhuni⁵, Marianna Niemi³⁷, Mika Helminen³⁷, Tiina Luukkaala³⁷, Iida Vähätalo³⁷, Jyrki Tammerluoto⁵, Sarah Smith³², Tom Southerington³² & Petri Lehto³²

⁷Broad Institute of MIT and Harvard, Cambridge, MA, USA. ⁸Massachusetts General Hospital, Boston, MA, USA. ⁹Abbvie, Chicago, IL, USA. ¹⁰Astra Zeneca, Cambridge, United Kingdom. ¹¹Biogen, Cambridge, MA, USA. ¹²Boehringer Ingelheim, Ingelheim am Rhein, Ingelheim, Germany. ¹³Bristol Myers Squibb, New York, NY, USA. ¹⁴Genentech, San Francisco, CA, USA. ¹⁵GlaxoSmithKline, Collegeville, PA, USA. ¹⁶GlaxoSmithKline, Espoo, Finland. ¹⁷Merck, Kenilworth, NJ, USA. ¹⁸Pfizer, New York, NY, USA. ¹⁹Translational Sciences, Sanofi R&D, Framingham, MA, USA. ²⁰Maze Therapeutics, San Francisco, CA, USA. ²¹Janssen Biotech, Beerse, Belgium. ²²Novartis Institutes for Biomedical Research, Cambridge, MA, USA. ²³Auria Biobank, University of Turku, Hospital District of Southwest Finland, Turku, Finland. ²⁴THL Biobank, Finnish Institute for Health and Welfare (THL), Helsinki, Finland. ²⁵Finnish Red Cross Blood Service, Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland. ²⁶Helsinki Biobank, Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki, Finland. ²⁷Northern Finland Biobank Borealis, University of Oulu, Northern Ostrobothnia Hospital District, Oulu, Finland. ²⁸Finnish Clinical Biobank Tampere, University of Tampere, Pirkanmaa Hospital District, Tampere, Finland. ²⁹Biobank of Eastern Finland, University of Eastern Finland, Northern Savo Hospital District, Kuopio, Finland. ³⁰Central Finland Biobank, University of Jyväskylä, Central Finland Health Care District, Jyväskylä, Finland. ³¹Central Finland Health Care District, Jyväskylä, Finland. ³²FINBB - Finnish biobank cooperative, Turku, Finland. ³³Business Finland, Helsinki, Finland. ³⁴GlaxoSmithKline, Stevenage, United Kingdom. ³⁵University of Stanford, Stanford, CA, USA. ³⁶Janssen Research & Development, LLC, Spring House, PA, USA. ³⁷University of Tampere, Tampere, Finland. ³⁸Northern Savo Hospital District, Kuopio, Finland. ³⁹Northern Ostrobothnia Hospital District, Oulu, Finland. ⁴⁰University of Eastern Finland, Kuopio, Finland. ⁴¹Pirkanmaa Hospital District, Tampere, Finland. ⁴²Hospital District of Helsinki and Uusimaa, Helsinki, Finland. ⁴³Hospital District of Southwest Finland, Turku, Finland. ⁴⁴GlaxoSmithKline, Brentford, United Kingdom. ⁴⁵Janssen Research & Development, LLC, Titusville, NJ 08560, USA. ⁴⁶University of Gothenburg, Gothenburg, Sweden. ⁴⁷Seinäjäki Central Hospital, Seinäjoki, Finland. ⁴⁸Novartis, Basel, Switzerland. ⁴⁹Finnish Institute for Health and Welfare (THL), Helsinki, Finland. ⁵⁰Broad Institute, Cambridge, MA, USA. ⁵¹Janssen Research & Development, LLC, Boston, MA, USA. ⁵²Novartis, Boston, MA, USA. ⁵³Janssen-Cilag Oy, Espoo, Finland. ⁵⁴Department of Molecular Genetics, University of Lodz, Lodz, Poland. ⁵⁵Helsinki University Hospital and University of Helsinki, Helsinki, Finland. ⁵⁶Eye Genetics Group, Folkhälsan Research Center, Helsinki, Finland. ⁵⁷University of Oulu, Oulu, Finland. ⁵⁸University of Helsinki, Helsinki, Finland. ⁵⁹University of Jyväskylä, Jyväskylä, Finland. ⁶⁰Aarhus University, Aarhus, Denmark. ⁶¹Department of Otorhinolaryngology - Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. ⁶²University of Eastern Finland and Kuopio University Hospital, Department of Otorhinolaryngology, Kuopio, Finland and Department of Allergy, Helsinki University Hospital and University of Helsinki, Helsinki, Finland. ⁶³Department of Medical Genetics, Helsinki University Central Hospital, Helsinki, Finland. ⁶⁴Transplantation and Liver Surgery Clinic, Helsinki University Hospital, Helsinki University, Helsinki, Finland. ⁶⁵Finnish Red Cross Blood Service, Helsinki, Finland. ⁶⁶European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK.

Estonian Biobank Research Team

Andres Metspalu⁴, Mari Nelis⁴, Lili Milani⁴ & Georgi Hudjashov⁴