

Genetic architecture of lumbar spinal stenosis

Ville Salo^{1,2}, Juhani Määttä^{1,3}, Jasmin Takala^{2,3}, Anni Heikkilä^{2,3}, FinnGen⁵, Ene Reimann⁶, Reedik Mägi⁶, Estonian Biobank Research Team^{6,7}, Kadri Reis⁶, Abdelrahman G.Elhanas⁶, Anu Reigo⁶, Priit Palta^{6,8}, Tõnu Esko⁶, Ville Leinonen⁹, Jaro Karppinen^{1,3,10}, Eeva Sliz^{2,3*}, Johannes Kettunen^{2,3*}

¹ Research Unit of Health Sciences and Technology, Faculty of Medicine, University of Oulu, Oulu, Finland

² Biocenter Oulu, Oulu, Finland

³ Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland

⁴ Research unit of Population Health, Faculty of Medicine, and Biocenter Oulu, University of Oulu, Oulu, Finland

⁵ A list of FinnGen authors and their affiliations available at the end of the paper

⁶ Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia

⁷ A list of Estonian Biobank Research Team authors and their affiliations available at the end of the paper

⁸ Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki

⁹ Neurosurgery, Institute of Clinical Medicine, School of Medicine, Faculty of Health Sciences, University of Eastern Finland and Department of Neurosurgery, Kuopio University Hospital, Kuopio, Finland

¹⁰ Rehabilitation Services of Wellbeing Services County of South Karelia, Lappeenranta, Finland

* These authors contributed equally to this work

Introductory paragraph

Over 100 million people worldwide suffer from lumbar spinal stenosis (LSS) with increasing incidence with ageing population, yet little is known about the LSS genetic background. Given the high cost of treating LSS, a deeper understanding of LSS pathogenesis may eventually result in the development of novel preventative and treatment methods, potentially leading to reductions in related societal costs. Our aim is to gain a better understanding of the genetic components underlying LSS. In the FinnGen, Estonian, and UK biobanks, we conduct a genome-wide association study (GWAS) of LSS and merge the results in the genome-wide meta-analysis. In addition to the seven known risk loci, our meta-analysis reveals 47 loci that have not been associated with LSS in previous studies. Many downstream analyses and multiple candidate genes discovered from the LSS-associated loci suggest that spinal degeneration plays a major role in the pathogenesis of LSS.

Main text

Lumbar spinal stenosis (LSS) narrows the lumbar central spinal canal or foramina causing symptoms mainly to buttocks and lower extremities^{1,2}. Term central stenosis is used if the whole central spinal canal is compressed, and foraminal or lateral stenosis if the single nerve root exiting the neural foramina is compressed¹. LSS usually develops with ageing when lumbar structures degenerate, including changes in intervertebral discs, facet joints, ligamentum flavum, and vertebrae. In addition, spondylolisthesis can cause or aggravate LSS¹. Even though the majority of LSS cases have a degenerative origin, in a smaller subgroup of patients, LSS arises from developmental or congenital factors^{2,3}. Naturally, developmental LSS can predispose individuals to degenerative LSS as they age³. Developmental causes include a congenitally small central canal and Paget disease, which leads to atypical bone growth^{1,3}.

The prevalence of LSS has been estimated to be around 11% in the general population and 20% among people over 60 years, varying greatly with age, and with the majority of cases being asymptomatic^{1,4}. LSS becomes clinically significant when it causes symptoms affecting daily life. The main symptoms include discomfort, pain, fatigue, paresthesia, and/or numbness in the lower extremities or the buttocks, which exacerbate during activities such as walking, frequently called as neurological claudication^{1,2,4}. Low back pain can be present or absent². The symptoms are typically relieved during sitting or forward flexion^{1,2}. Diagnosis is based on patient history and symptoms, and clinical examination, with confirmation through imaging, usually magnetic resonance imaging (MRI) or computed tomography (CT)¹. Non-surgical management, such as physiotherapy, oral medication, and activity modification, is usually the first-line treatment. Surgical treatment is warranted if the symptoms are severe or non-surgical care does not improve the symptoms enough^{1,2}.

In this study, we used the International Classification of Diseases (ICD)-10 code M48.0 to characterize LSS in the genome-wide meta-analysis. Data from FinnGen, the Estonian Biobank, and the UK Biobank were analyzed, comprising 30 269 cases and 739 414 controls (Table S1). A total of 54 loci with at least one genome-wide significant ($P < 5 \times 10^{-8}$) variant were associated with LSS in the meta-analysis. Forty-seven of the loci were previously unreported (Table 1, Supplementary Data), and we also successfully replicated seven loci previously associated with LSS (Fig. 1A, Table S2-S4 Fig. S1)⁵⁻⁷. However, 17 of the newly identified loci had been previously associated with highly related phenotypes, such as back pain (BP) or lumbar disc herniations (LDH) (Table 1). In the conditional analysis, we also detected secondary signals at four loci (Table S2). At locus 12p11.22, we observed several secondary signals, and the association signal extended beyond the 2MB window used to define loci (Supplementary Data 29).

Table 1. The lead variants at 47 novel LSS-associated ($p < 5 \times 10^{-8}$) loci. In the meta-analysis, there were a total of 30 269 cases and 739 414 controls from FinnGen, the Estonian Biobank, and the UK Biobank. All genome-wide significantly associated loci observed in the meta-analysis are presented in Table S2.

Locus	Candidate gene	CHR: POS	rsid	EA	OA	OR (95% CI)	p-value	EAF	Het pval	Fin Enr.	Ref.
Meta-analysis (FinnGen, EstBB, UKBB)											
1p36.11	<i>RUNX3</i>	1:24894533	rs9438857	G	A	0.90 (0.88-0.92)	4.92e-08	0.31	0.005	0.86	
1p13.2	<i>NGF</i>	1:115287365	rs6327	T	C	0.93 (0.91-0.95)	5.4e-16	0.47	0.26	1.06	LDH ⁹
1q21.2	<i>OTUD7B</i>	1:149934520	rs11205303	C	T	0.92 (0.90-0.94)	1.76e-19	0.38	0.68	0.75	
1q21.3	<i>ZNF687</i>	1:151032308	rs147042690	T	C	1.21 (1.14-1.27)	8.5e-09	0.01	0.69	34.3	
1q32.1	<i>PTPRC</i>	1:198819354	rs72739729	C	G	0.91 (0.88-0.93)	2.18e-15	0.15	0.74	1.19	LDH ^{9,10}
2q35	<i>USP37</i>	2:218486683	rs60362609	G	C	0.94 (0.92-0.96)	1.47e-08	0.61	0.71	0.98	
3p21.31	<i>DAG1</i>	3:49498696	rs138073019	TT TA	T	0.95 (0.93-0.97)	9.31e-09	0.35	0.84	1.48	LDH ⁹ , BP ¹¹
4p15.33	<i>RAB28</i>	4:12893286	rs3057070	A	AG AG	0.92 (0.90-0.94)	3.22e-15	0.24	0.44	1.08	
5p13.1	<i>DAB2</i>	5:39844572	rs148346348	AT AT G	A	1.08 (1.05-1.10)	7.56e-10	0.17	0.22	1.29	
5q11.2	<i>ITGA2</i>	5:53038840	rs11747379	A	C	1.08 (1.05-1.10)	9.21e-11	0.18	0.68	1.07	BP ⁵
5q33.3	<i>EBF1</i>	5:159017122	rs200608668	G	GC	1.06 (1.04-1.09)	3.28e-08	0.22	0.94	0.66	
6p24.3	<i>BMP6</i>	6:7797607	rs10498672	G	C	1.08 (1.06-1.11)	9.09e-14	0.20	0.82	1.24	SP ¹²
6p22.2	<i>TRIM38</i>	6:26239176	rs9358913	G	A	0.94 (0.93-0.96)	8.72e-10	0.32	0.41	1.62	LDH ⁹
6p22.1	<i>HLA</i>	6:30106665	rs62390186	A	T	0.94 (0.92-0.96)	3.99e-08	0.18	0.07	1.83	LDH ⁹
6p21.33	<i>HLA</i>	6:31133806	rs4294047	A	G	0.91 (0.88-0.93)	2.93e-13	0.11	0.20	1.97	LDH ⁹
6p21.32	<i>HLA</i>	6:32634706	rs17205170	T	G	0.92 (0.90-0.94)	3.33e-13	0.20	0.68	1.34	LDH ⁹
6p21.1	<i>CDC5L</i>	6:44479182	rs7770012	A	C	1.07 (1.05-1.08)	1.02e-12	0.53	0.21	0.99	LDH ¹⁰ , OPLL ¹³ , ₁₄
6q21	<i>LIN28B</i>	6:105005542	rs314261	C	A	1.05 (1.03-1.07)	2.18e-08	0.57	0.55	1.03	
6q23.2	<i>EYAA</i>	6:132913577	rs2764217	A	G	1.06 (1.05-1.08)	1.36e-10	0.43	0.97	0.97	
7p22.1	<i>FOXL1</i>	7:4638590	rs28658109	C	T	0.93 (0.91-0.96)	1.08e-08	0.19	0.45	1.35	
7p21.1	<i>TWIST1</i>	7:19488615	rs2390116	T	C	0.95 (0.93-0.97)	2.93e-08	0.49	0.43	0.97	LDH ^{9,10}
7p15.1	<i>JAZF1</i>	7:28156794	rs849135	A	G	0.95 (0.93-0.97)	7.32e-09	0.49	0.51	0.90	
7q11.22	<i>Empty</i>	7:69064281	rs139203647	G	GT	1.08 (1.05-1.11)	1.01e-08	0.15	0.68	1.22	
10p12.1	<i>MKY</i>	10:27612430	rs2637326	T	G	1.06 (1.05-1.08)	3.44e-12	0.47	0.10	0.73	LDH ^{9,10}
11p14.1	<i>MPPED2</i>	11:30788512	rs607165	C	T	1.05 (1.04-1.07)	1.94e-09	0.44	0.85	1.19	
12p12.1	<i>SOX5</i>	12:23821470	rs12308843	C	G	1.08 (1.06-1.10)	3.3e-15	0.27	0.65	1.28	LDH ^{9,10} , BP ¹⁵
12p11.22	<i>CCDC91</i>	12:27844985	rs7958415	T	C	0.93 (0.91-0.95)	1.07e-13	0.35	0.97	0.72	OPLL ¹³ , ₁₄
12q14.1	<i>GLII</i>	12:57749071	rs2069506	A	C	0.94 (0.93-0.96)	1.56e-09	0.33	0.40	1.19	LDH ⁹
12q14.1	<i>USP15</i>	12:62218131	rs11174377	A	G	1.07 (1.05-1.10)	2.47e-09	0.17	0.69	0.97	
12q24.31	<i>ZCCHC8</i>	12:122480161	rs10846978	C	T	1.06 (1.04-1.09)	4.49e-08	0.77	0.48	1.07	
13q14.11	<i>TNFSF11</i>	13:42423131	rs9594751	T	C	0.95 (0.93-0.97)	3.09e-08	0.25	0.13	1.13	
14q11.2	<i>MMP14</i>	14:22843385	rs1042704	A	G	0.93 (0.91-0.96)	4.02e-10	0.22	0.77	1.00	
14q32.33	<i>GPR132</i>	14:105137078	rs1022431	A	C	0.91 (0.88-0.95)	3.76e-08	0.09	0.67	0.78	
15q21.3	<i>ALDH1A2</i>	15:58050948	rs4646570	C	G	1.06 (1.05-1.08)	5.47e-12	0.49	0.86	1.07	
15q22.31	<i>PPIB</i>	15:64138226	rs148027154	A	T	1.28 (1.20-1.37)	3.52e-08	0.006	0.20	5.12	
15q25.3	<i>PDE8A</i>	15:84986797	rs11629962	T	C	0.94 (0.92-0.96)	2.36e-08	0.25	0.29	1.14	
16q23.2	<i>GAN</i>	16:81482778	rs2966084	G	A	0.94 (0.92-0.96)	4.3e-11	0.69	0.78	0.95	
17q12	<i>SOCS7</i>	17:38474030	rs113353816	C	T	1.06 (1.04-1.09)	4.25e-08	0.83	0.30	0.90	
17q24.1	<i>AXIN2</i>	17:65278900	rs16961974	C	T	1.06 (1.04-1.07)	2.75e-08	0.28	0.46	1.05	
17q24.3	<i>Empty</i>	17:71476022	rs12941198	C	A	1.06 (1.04-1.08)	3.6e-08	0.78	0.25	1.02	
18q21.2	<i>DCC</i>	18:53220821	rs6508210	C	A	1.08 (1.06-1.09)	2.64e-16	0.52	0.49	0.85	LDH ⁹ , BP ^{11,15} , ₁₆
19p13.3	<i>CACTIN</i>	19:3537186	rs77733715	G	A	1.14 (1.09-1.18)	1.44e-08	0.03	0.42	5.24	
19p13.11	<i>CRLF1</i>	19:18616479	rs2074177	G	A	0.93 (0.91-0.95)	1.03e-12	0.25	0.38	1.05	
20p13	<i>DDR3</i>	20:2800795	rs11697820	T	C	0.95 (0.93-0.96)	5.15e-10	0.58	0.40	0.98	
21q22.2	<i>ETS2</i>	21:39183865	rs111901809	T	TA GA GG GG CA GG GG GC G	1.05 (1.03-1.07)	3.94e-08	0.49	0.18	1.32	
22q12.2	<i>OSM</i>	22:30026719	rs2285645	T	C	0.83 (0.78-0.89)	4.19e-12	0.02	0.95	15.0	
Xq23	<i>CHRD1</i>	X:110450046	rs67648651	C	T	1.05 (1.04-1.07)	1.13e-11	0.40	0.49	0.87	LDH ⁹

Candidate gene, a gene at a novel locus the biological function of which is likely to explain the LSS association; CHR: POS, chromosome and position (genome build hg38); rsid, SNP markers identification number; EA, effect allele; OA, other allele; OR, odds ratio; 95% CI, odds ratio 95% confidence interval; EAF, effect allele frequency; Het pval, p-value for heterogeneity; Fin Enr., enrichment in Finns (calculated FIN AF/NFEE AF in the Genome Aggregation Database [gnomAD, v.3.1.2], FIN AF is the allele frequency in Finns and NFEE AF is the allele frequency in Europeans (does not include Finns or Estonians)); Ref., reference article in which a back related disease association other than LSS was observed +/- 1 Mb in the vicinity of the lead variant, BP=back pain, LDH=lumbar disc herniation, OPLL= ossification of the posterior longitudinal ligament, SP=spondylosis

Table 2. The lead variants at 20 loci associated ($p < 5 \times 10^{-8}$) with LSS requiring surgical treatment. The surgery GWAS involved patients with LSS who had undergone surgery (*NOMESCO version 1.15*, ABC36, ABC56, ABC66, ABC99, NAG61, NAG62, NAG63, NAG66, NAG67, and NAG99). A total of 8627 cases and 270 773 controls were included in the surgery GWAS. LSS cases that lacked documentation of the surgical codes were excluded from the analysis. Patients who had included undergone surgery but did not have a LSS diagnosis were excluded, as they likely had an acute injury.

Locus	Candidate gene	CHR: POS	rsid	EA	OA	OR (95% CI)	EAF	p-value	p-value (meta)
Surgery GWAS (FinnGen)									
1p13.2	<i>NGF</i>	1:115287365	rs6327	T	C	0.90 (0.87-0.93)	0.47	5.54e-12	5.40e-16
1q21.2	<i>OTUD7B</i>	1:149934520	rs11205303	C	T	0.89 (0.86-0.92)	0.33	8.88e-12	1.76e-19
1q32.1	<i>PTPRC</i>	1:198821974	rs7537080	G	A	0.89 (0.84-0.93)	0.16	2.69e-08	2.74e-15
2p13.3	<i>GFPT1</i>	2:69358044	rs6724567	G	A	0.89 (0.86-0.93)	0.55	2.09e-12	1.09e-24
2p13.3	<i>GFPT1</i>	2:70358644	rs10192480	C	T	0.91 (0.87-0.94)	0.30	8.10e-09	1.04e-07
2p13.2	<i>GFPT1</i>	2:71726556	rs6749644	T	G	0.91 (0.88-0.94)	0.61	4.07e-08	6.15e-07
4p15.33	<i>RAB28</i>	4:12915441	rs1861418	A	G	0.88 (0.84-0.92)	0.25	7.96e-12	4.03e-15
6p24.3	<i>BMP6</i>	6:7792936	rs4371882	G	A	1.15 (1.12-1.19)	0.22	9.85e-15	1.55e-13
6p22.2	<i>TRIM38</i>	6:26295698	rs9393698	A	G	0.91 (0.88-0.94)	0.42	1.05e-09	1.33e-07
6p22.1	<i>HLA</i>	6:30107095	rs7383537	T	C	0.91 (0.88-0.94)	0.51	6.60e-09	3.97e-07
6p22.1	<i>COL10A1</i>	6:115984474	rs12662901	C	T	0.91 (0.88-0.94)	0.60	9.31e-09	1.34e-04
7p21.2	<i>AGMO</i>	7:15344063	rs75674922	G	A	1.14 (1.10-1.18)	0.17	4.87e-11	6.72e-12
12p11.22	<i>CCDC91</i>	12:28121303	rs5797253	C	CT	0.89 (0.85-0.92)	0.25	1.03e-10	NA
12q14.1	<i>GLII</i>	12:57723862	rs701008	C	T	0.88 (0.84-0.94)	0.41	1.61e-08	9.03e-09
13q14.11	<i>FOXO1</i>	13:40132854	rs12872985	T	C	1.11 (1.08-1.14)	0.46	9.33e-11	5.91e-07
13q14.11	<i>TNFSF11</i>	13:42423131	rs9594751	T	C	0.91 (0.87-0.94)	0.27	3.74e-08	3.09e-08
15q21.3	<i>ALDH1A2</i>	15:58050948	rs4646570	C	G	1.11 (1.08-1.14)	0.49	3.99e-11	5.47e-12
15q22.33	<i>SMAD3</i>	15:67078051	rs12901071	G	A	0.90 (0.87-0.94)	0.37	1.10e-09	3.63e-28
22q12.2	<i>OSM</i>	22:29236519	rs371011194	T	C	0.77 (0.68-0.86)	0.04	2.09e-08	2.31e-06
22q12.2	<i>OSM</i>	22:30243756	rs192914111	A	G	0.71 (0.61-0.81)	0.03	1.52e-11	2.78e-11

Candidate gene, a gene at a novel locus the biological function of which is likely to explain the LSS requiring surgery association; CHR: POS, chromosome and position (genome build hg38); rsid; SNP markers identification number; EA, effect allele; OA, other allele; OR, odds ratio; 95% CI, odds ratio 95% confidence interval; EAF, effect allele frequency; p-value (meta), lead variants p-value in meta-analysis, NA, not available (*CCDC91* variant detected only in FinnGen, as meta-analysis consists only about variants that have been detected in at least two of the study populations)

LSS is suggested to be caused by degenerative changes in the spine¹⁷, and we identified several potential genes that may account for these degenerative changes. In our PheWAS analysis, we found that variants at LSS-associated loci were also associated with numerous other spine-related diseases, such as spondylosis and lumbar disc herniations (Table S3).

Of particular interest was the locus close to *BMP6* (*bone morphogenetic protein 6*). *BMP6* has previously been associated with spondylosis¹² and osteoarthritis¹⁸ and has been found to play an important role in bone and cartilage formation and repair¹⁹. At this locus, the effect estimate was significantly larger in the LSS surgery GWAS compared to the primary LSS meta-analysis (Fig. 1B). Other examples of interesting LSS-associated loci include the ones near *CDC5L* (*cell division cycle 5 like*) and *CCDC91* (*coiled-coil domain containing 91*), which have previously been associated with the ossification of the posterior longitudinal ligament (OPLL)^{13,14}. OPLL is characterized by ectopic bone formation within the posterior longitudinal ligament in the spinal canal. OPLL may also contribute to the pathogenesis of LSS, as ossified ligament can compress the spinal cord or nerve roots^{20,21}. Another interesting LSS-associated locus is the one near *MMP14* (*matrix metalloproteinase 14*). The lead variant is a missense variant (rs1042704, p.D273N) known to significantly reduce collagen catabolism²², which might have an impact on spinal degeneration, where modifications in collagen structures are often observed²³. However, based on our PheWAS results (Table S3) and the previously reported associations with Dupuytren contracture²⁴ and adhesive capsulitis²⁵, the *MMP14* gene may not be a suitable target for translational research, as the rs1042704 variant appears to protect from LSS while increasing the risk of the aforementioned disorders.

The loci we observed in surgical GWAS were largely overlapping with the loci we observed in the meta-analysis. Only the locus near *FOXO1* (*forkhead box O1*) was specific to the surgical GWAS (Table S3).

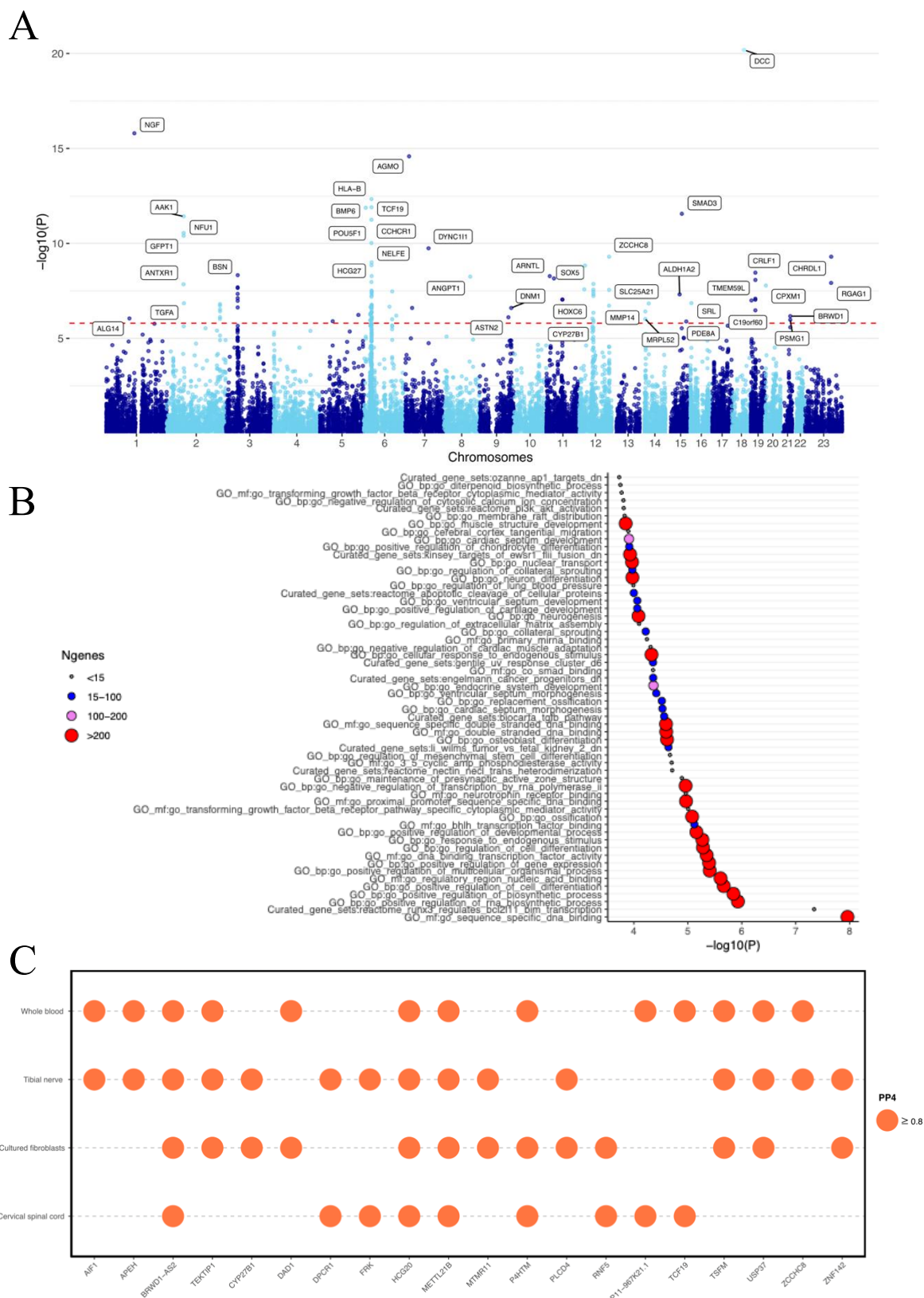


Fig. 2. A) Results of the MAGMA²⁶ gene-based test in a Manhattan plot. X-axis chromosomes, y-axis $-\log_{10}(p\text{-value})$. **B)** MAGMA gene-set enrichment analysis. The plot shows significantly enriched pathways ($p\text{FDR} < 0.05$), curated gene sets, and GO-annotations ranked by significance ($-\log_{10}(P\text{-value})$). The circle size refers to the size of the gene set. Small gray <15 , blue $15\text{--}100$, violet $100\text{--}200$, and red >200 genes. The analysis was done using FUMA²⁷, and the gene sets and GO annotations included in the analysis are from MSigDB²⁸. **C)** Colocalizations between LSS GWAS and eQTL signals were estimated with approximate Bayesian factor analyses, implemented with the ‘coloc.abf’ function available in the ‘coloc’ R-library (Table S7). The ‘biomaRt’ R-library was used to extract genes from Ensembl archives²⁹ within a 1 MB (± 500000 bases) window surrounding each lead variant. Colocalizations were investigated for 754 genes that were available in GTEx. Variant-gene expression associations for selected tissues (whole blood, cultured fibroblasts, tibial nerve, and cervical spinal cord) were downloaded from GTEx v8. Colocalizations with posterior probabilities ≥ 0.8 for the variant were considered significant³⁰. Only genes with at least two significant colocalizations are shown in the plot; all significant colocalizations are listed in Table S7.

Multiple genes that we identified as potential candidate genes at the novel LSS-associated loci were also highlighted by the GCTA-fastBAT analysis and the MAGMA gene-based test (Table 1, Fig. 2A, Table S6). The pathway responsible for sequence-specific DNA binding showed the most significant enrichments in the MAGMA gene-set analysis (Fig. 2B). Gene-set analysis also highlighted pathways associated with ossification and pathways controlling biosynthetic processes and differentiation. Additionally, enrichments were found in many gene-sets related to the nervous system, such as the pathway related to the binding of neurotrophin receptors. There were no tissues identified in the MAGMA tissue expression analysis (Fig. S3) that exhibited a positive correlation between associations with LSS and tissue-specific gene expression profiles. The lack of the relevant tissues—namely bone and cartilage—in the Genotype-Tissue Expression (GTEx) dataset utilized for these analyses is likely the reason for the null result.

Colocalizations between LSS GWAS and gene expression signals (eQTL) were observed in each selected tissue, namely whole blood, cultured fibroblasts, tibial nerve, and cervical spinal cord. We found evidence of the colocalization of the expression of the 85 genes and LSS signals in at least one tissue (posterior probability [PP] for the shared variant ≥ 0.8 , Table S7), and 20 colocalizations were observed in at least two or more of the tissues investigated (Fig. 2C). For example, significant colocalizations between LSS association signals and *BRWD1-AS2* (*BRWD1 antisense RNA2*), *HCG20* (*HLA complex group 20*) and *METTL21B* (*EEF1A lysine methyltransferase 3*, also known as *EEF1AKMT3*) in all studied tissues (Fig. 2C, Table S7).

LSS usually manifests later in life¹. Before our study, an age-of-onset analysis for genetic variants was not possible since the genetic background of the LSS has not been investigated extensively and there have been limitations in the availability of large, longitudinal datasets. At each locus associated with LSS, we examined the differences in the cumulative incidences between homozygotes for the lead variants and calculated the cumulative incidence of LSS diagnoses. Around age 40, the number of LSS diagnoses started to accumulate, with the majority of diagnoses occurring between the ages of 50 and 70. All lead variants showed a similar trend in terms of the accumulation of LSS diagnoses, with the lead variant genotypes showing only slight variation and cumulative morbidities following the sample prevalence value (6.5% for LSS in FinnGen, Table S8). We observed that for five LSS-associated variants the cumulative incidence differed between homozygotes already at the age of 40 (Table S8). These lead variants located near the *OTUD7B* (*OTU deubiquitinase 7B*), *USP37* (*ubiquitin specific peptidase 37*), *CDC5L* (*cell division cycle 5 like*), *SOX5* (*SRY-box transcription factor 5*) and *SMAD3* (*SMAD family member 3*) genes (Fig. S4). With the majority of the other variants, the difference in the cumulative LSS diagnoses between homozygotes manifested at a later age, but with some variants no significant difference was observed at all (Table S8). We carried out the same analysis in the Estonian Biobank and discovered that diagnoses were accumulating for the variants in a similar manner to that observed in FinnGen (Fig. S4).

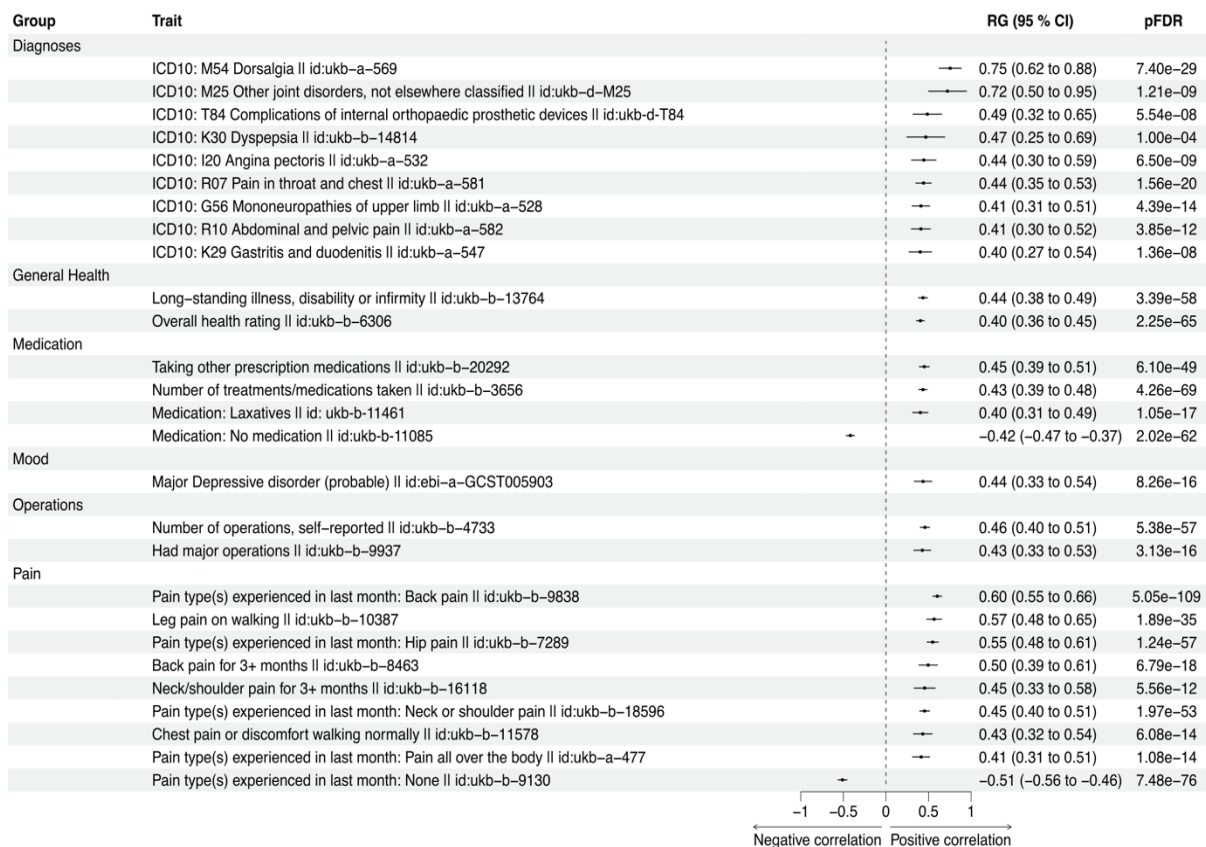


Fig. 3. Genetic correlations were calculated using LDSC-software^{8,31}. All traits were extracted from the GWAS database provided by the MRC Integrative Epidemiology Unit (IEU). Only the strongest observed ($rg < -0.4$ & $rg > 0.4$) correlations with a significant false discovery corrected p-value ($p_{FDR} < 0.05$) are shown in the figure. RG, genetic correlation coefficient value; pFDR, false discovery rate-corrected p-value. Genetic correlations for all 517 phenotypes can be seen in Table S9.

Significant genetic correlations were found between 517 traits and LSS (Fig. 3, Table S9). Back pain ($rg=0.60$, $p_{FDR}=5.05 \times 10^{-109}$) showed the most significant positive genetic correlation in terms of the lowest p-value, and the diagnosis of M54 Dorsalgia ($rg=0.75$, $p_{FDR}=7.40 \times 10^{-29}$) showed the largest significant genetic correlation. There was also a positive genetic correlation between LSS and many pain-related endpoints, including leg pain on walking ($rg=0.57$, $p_{FDR}=1.89 \times 10^{-35}$). Positive genetic correlations were also observed for general health- and mood-related traits, such as overall health rating ($rg=0.40$, $p_{FDR}=2.25 \times 10^{-65}$) and probable major depressive disorder ($rg=0.44$, $p_{FDR}=8.26 \times 10^{-16}$).

A

Method	Exposure	nsnp		OR (95% CI)	pFDR
IVW	Whole body fat-free mass II id:ukb-b-13354	481	■—	1.78 (1.59 to 1.99)	2.22e-21
	Body mass index (BMI) II id:ukb-b-19953	387	■—	1.55 (1.42 to 1.70)	2.22e-21
	Hip circumference II id:ukb-b-15590	369	■—	1.51 (1.38 to 1.64)	2.62e-18
	Whole body fat mass II id:ukb-b-19393	383	■—	1.48 (1.35 to 1.63)	2.82e-15
MR Egger	Whole body fat-free mass II id:ukb-b-13354	481	■—	1.53 (1.17 to 2.00)	4.07e-03
	Body mass index (BMI) II id:ukb-b-19953	387	■—	1.60 (1.26 to 2.03)	3.76e-04
	Hip circumference II id:ukb-b-15590	369	■—	1.60 (1.26 to 2.03)	4.49e-04
	Whole body fat mass II id:ukb-b-19393	383	■—	1.66 (1.28 to 2.16)	4.62e-04

B

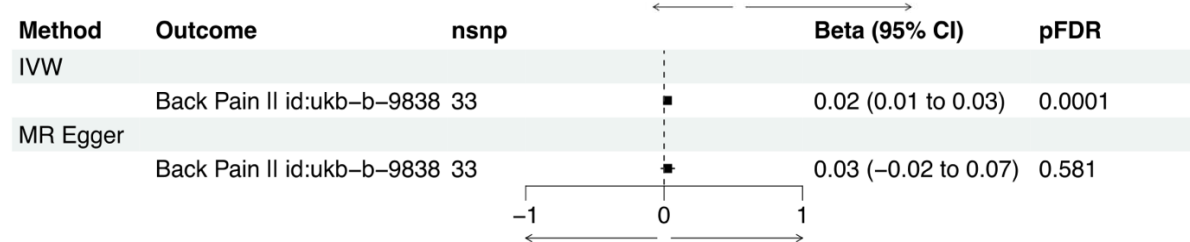


Fig. 4. A) Exposures potentially causal for LSS, and **B)** outcomes that LSS was potentially found to be causal for. The analysis was performed using the TwoSampleMR R library and data from the present study and MRC-IEU database. The inverse variance weighted (IVW) model was our primary analysis, for which statistical significance was considered at false discovery rate corrected p-value (pFDR < 0.05). We also performed a sensitivity analysis using MR Egger. nsnp, number of SNPs; 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.

In Mendelian randomization, we uncovered potential causal relationships between several factors and LSS (Fig. 4A, Fig. S5.1-4, Table S10, Table S11). The most significant potential causal relationships were observed between whole body fat-free mass and higher LSS risk (OR=1.78, pFDR=2.22e-21, Fig. S5.1) and between body mass index (BMI) and higher LSS risk (OR=1.55, pFDR=2.22e-21, Fig. S5.2). The only outcome for which we found LSS to be potentially causal was back pain (Beta=0.02, pFDR=0.0001, Fig. 4B, Fig. S6, Table S10). Pain is the main reason for seeking care and often also the main symptom². Although pleiotropy was not observed, some causal estimates were heterogeneous (Table S10), so these results should be interpreted cautiously. It appears that individual variants do not drive the observed causal relationships, as all causal estimates in the leave-out analyses consistently point in the same direction (Fig. S7.1-4, Fig. S8).

Our study has strengths and limitations. Multiple genome-wide significant associations with LSS were discovered, facilitated by integrating data from three large biobanks, which resulted in a substantial sample size. Divergences in the proportional frequency of LSS cases among the sample groups incorporated in the meta-analysis indicate possible inconsistencies in the ability of biobanks to recognize patients with LSS. It is also important to note that only people with European ancestry were included in our study. Both the frequency of LSS surgical treatment and the type of surgery performed can differ greatly between countries². In our research, we have used surgical codes that are used in Finland for the surgical treatment of LSS patients³². The eQTL analyses would benefit from incorporating bone and cartilage data, which are crucial tissues in LSS pathogenesis. Since our work only uses computational methods, future functional studies would enhance our results.

In conclusion, by identifying a significant number of novel loci and performing multiple downstream analyses, we were able to expand the current understanding of the genetic basis of LSS. Our findings highlighted the significance of spinal degeneration as a major risk factor for lumbar spinal stenosis. Since we found that the number of diagnoses for LSS increased only after the age of 50, our study's age-of-onset analyses lend credence to the notion that LSS is an age-related condition. However, our analysis of cumulative LSS incidence patterns revealed variants that significantly impact the risk of LSS morbidity as early as 40 years of age.

Additionally, numerous loci associated with LSS requiring surgery underlined pathways that can lead to a more severe LSS. Taken together, our work contributes to our understanding of the genetic origins of LSS and may open the door to future research into novel therapies and preventative strategies for LSS.

References

1. Katz, J. N., Zimmerman, Z. E., Mass, H. & Makhni, M. C. Diagnosis and Management of Lumbar Spinal Stenosis. *JAMA* **327**, 1688 (2022).
2. Lurie, J. & Tomkins-Lane, C. Management of lumbar spinal stenosis. *BMJ* h6234 (2016) doi:10.1136/bmj.h6234.
3. Lai, M. K. L., Cheung, P. W. H. & Cheung, J. P. Y. A systematic review of developmental lumbar spinal stenosis. *European Spine Journal* **29**, 2173–2187 (2020).
4. Jensen, R. K., Jensen, T. S., Koes, B. & Hartvigsen, J. Prevalence of lumbar spinal stenosis in general and clinical populations: a systematic review and meta-analysis. *European Spine Journal* vol. 29 2143–2163 Preprint at <https://doi.org/10.1007/s00586-020-06339-1> (2020).
5. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415–1424 (2021).
6. Suri, P. *et al.* Genome-wide association studies of low back pain and lumbar spinal disorders using electronic health record data identify a locus associated with lumbar spinal stenosis. *Pain* **162**, 2263–2272 (2021).
7. Bovonratwet, P. *et al.* Identification of Novel Genetic Markers for the Risk of Spinal Pathologies: A Genome-Wide Association Study of 2 Biobanks. *J Bone Joint Surg Am* (2023) doi:10.2106/JBJS.22.00872.
8. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
9. Salo, V. *et al.* Genome-wide meta-analysis conducted in three large biobanks expands the genetic landscape of lumbar disc herniations. doi:10.1101/2023.10.15.23296916.
10. Bjornsdottir, G. *et al.* Rare SLC13A1 variants associate with intervertebral disc disorder highlighting role of sulfate in disc pathology. *Nat Commun* **13**, 634 (2022).
11. Johnston, K. J. A. *et al.* Genome-wide association study of multisite chronic pain in UK biobank. *PLoS Genet* **15**, (2019).
12. Zhang, Y. *et al.* Genome-wide Association Analysis Across 16,956 Patients Identifies a Novel Genetic Association Between BMP6, NIPAL1, CNGA1 and Spondylosis. *Spine (Phila Pa 1976)* **46**, E625–E631 (2021).
13. Koike, Y. *et al.* Genetic insights into ossification of the posterior longitudinal ligament of the spine. *Elife* **12**, (2023).
14. Nakajima, M. *et al.* A genome-wide association study identifies susceptibility loci for ossification of the posterior longitudinal ligament of the spine. *Nat Genet* **46**, 1012–1016 (2014).
15. 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).
16. 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).
17. Katz, J. N., Zimmerman, Z. E., Mass, H. & Makhni, M. C. Diagnosis and Management of Lumbar Spinal Stenosis: A Review. *JAMA* vol. 327 1688–1699 Preprint at <https://doi.org/10.1001/jama.2022.5921> (2022).

18. McDonald, M.-L. N. *et al.* Novel genetic loci associated with osteoarthritis in multi-ancestry analyses in the Million Veteran Program and UK Biobank. *Nat Genet* **54**, 1816–1826 (2022).
19. Ebisawa, T. *et al.* Characterization of bone morphogenetic protein-6 signaling pathways in osteoblast differentiation. *J Cell Sci* **112 (Pt 20)**, 3519–27 (1999).
20. Matsunaga, S. *et al.* Pathogenesis of myelopathy in patients with ossification of the posterior longitudinal ligament. *J Neurosurg* **96**, 168–72 (2002).
21. Abiola, R., Rubery, P. & Mesfin, A. Ossification of the Posterior Longitudinal Ligament: Etiology, Diagnosis, and Outcomes of Nonoperative and Operative Management. *Global Spine J* **6**, 195–204 (2016).
22. Itoh, Y. *et al.* A common SNP risk variant MT1-MMP causative for Dupuytren’s disease has a specific defect in collagenolytic activity. *Matrix Biology* **97**, 20–39 (2021).
23. Kosaka, H. *et al.* Pathomechanism of Loss of Elasticity and Hypertrophy of Lumbar Ligamentum Flavum in Elderly Patients With Lumbar Spinal Canal Stenosis. *Spine (Phila Pa 1976)* **32**, 2805–2811 (2007).
24. Ng, M. *et al.* A Genome-wide Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis. *Am J Hum Genet* **101**, 417–427 (2017).
25. Green, H. D. *et al.* A genome-wide association study identifies 5 loci associated with frozen shoulder and implicates diabetes as a causal risk factor. *PLoS Genet* **17**, (2021).
26. 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).
27. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826 (2017).
28. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–40 (2011).
29. Martin, F. J. *et al.* Ensembl 2023. *Nucleic Acids Res* **51**, D933–D941 (2023).
30. Sliz, E. *et al.* Evidence of a causal effect of genetic tendency to gain muscle mass on uterine leiomyomata. *Nat Commun* **14**, 542 (2023).
31. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236–1241 (2015).
32. Salmenkivi, J., Sund, R., Paavola, M., Ruuth, I. & Malmivaara, A. Mortality Caused by Surgery for Degenerative Lumbar Spine. *Spine (Phila Pa 1976)* **42**, 1080–1087 (2017).
33. 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).
34. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
35. 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).
36. 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).
37. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* **53**, 1097–1103 (2021).
38. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

39. Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. GenBank. *Nucleic Acids Res* **44**, D67–D72 (2016).
40. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506–D515 (2019).
41. Stanfill, A. G. & Cao, X. Enhancing Research Through the Use of the Genotype-Tissue Expression (GTEx) Database. *Biol Res Nurs* **23**, 533–540 (2021).
42. 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).
43. 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).
44. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).

Methods

Study populations

FinnGen

FinnGen strives to advance our understanding of disease mechanisms and the efficacy of medical care. The studies aim is to identify links between individual genetic differences and diseases. The FinnGen project has the necessary ethical and prior permits for biobank research (Supplementary Note), and all persons who have provided a research sample are aware of the intended use of the samples and have given their written consent to biobank research either in connection with sample donation or when participating in older research projects, the materials of which have been transferred to Finnish biobanks with the written consent of Fimea. We used *The International Classification of Diseases Tenth Revision* (ICD-10) code M48.0 and *Ninth Revision* (ICD-9) codes 7230A, 7240A, and 7240B (Table S1) to characterize cases in FinnGen. Patients without a record of these ICD codes were classified as controls, and participants who were diagnosed with other dorsopathies were excluded (M40-M54, M76). Patient data were obtained from the Hospital Discharge Registry and the Cause of Death Registry however, patients whose registration data was only available in the primary care registry were excluded from the analysis. The FinnGen (R9-version) data used in the study included 18 989 LSS cases and 270 964 controls.

Estonian Biobank (EstBB)

The Estonian Biobank cohort is a volunteer-based sample of the Estonian resident adult population (aged ≥ 18 years)³³. 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, 72 679 participants from the Estonian Biobank were included; 6019 of these were LSS cases, and 65 258 were controls.

UK Biobank

The UK biobank³⁴ cohort comprises of samples acquired between 2006-2010. Hundreds of thousands of people aged 40-69 from across Britain provided samples. In our research, we used a subset of European ancestry, consisting of 5134 LSS patients and 401 917 controls from the UK biobank, more precisely summary statistics from the PanUKBB project.

Genotyping, imputation & quality control

FinnGen

Genotyping was performed using Illumina and Affymetrix DNA microarrays. 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. Eagle 2.3.5 was used in prephasing, where 20000 was used as a threshold of conditioning haplotypes. Genotype imputation was done with Beagle 4.1 (version 08Jun17.d8b), and Finnish SISu v3 was used as a reference panel. Whole imputation protocol has been described at doi.org/10.17504/protocols.io.nmndc5e. For post-imputation quality control, we excluded variants with imputation information less than 0.6.

EstBB

Illumina Human CoreExome, OmniExpress, 370CNV BeadChip and GSA arrays were used for genotyping. 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 ($> \text{mean} \pm 3\text{SD}$), genotype and phenotype sex discordance, cryptic relatedness ($\text{IBD} > 20\%$) 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 ($P < 1 \times 10^{-4}$). SHAPEIT2 was used for prephasing, while the imputation was performed with the Estonian-specific reference panel³⁵ and IMPUTE2³⁶ with default settings. 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

The Regenie (version 2.2.4) program³⁷ was used to conduct GWAS using an additive genetic model in FinnGen and EstBB, modifying each phenotype for age, sex, the first 10 genetic principal components and genotyping arrays. The sensitivity analysis was performed using Regenie and the same covariates as above.

FinnGen-based sensitivity analysis on surgical LSS

Additionally, we conducted sensitivity analysis in FinnGen, with the aim of identifying genomic regions associated with more severe LSS that required surgical treatment. The sensitivity analysis was done using patients with LSS who had undergone surgery (*NOMESCO version 1.15*, ABC36, ABC56, ABC66, ABC99, NAG61, NAG62, NAG63, NAG66, NAG67, and NAG99)³², with the control group defined similarly to the meta-analysis. A total of 8627 cases and 270 773 controls were included in this analysis. LSS cases that lacked documentation of studied surgical codes were not included in the analysis. The same policy was followed in cases of patients who underwent surgery but did not have a LSS diagnosis because it is likely that these patients had an acute injury.

Meta-analysis (FinnGen, EstBB, UKBB)

The inverse-variance weighted meta-analysis was carried out using Python software (https://github.com/FINNGEN/META_ANALYSIS/). Variant data from the Estonian and UK Biobanks were converted from hg19 to hg38 before being meta-analyzed using the Picard liftover (<http://broadinstitute.io/picard>). Meta-analysis data consists only SNPs that have been observed in at least two of the datasets. In total, there were 769 683 participants in the meta-analysis, of which there were 30 269 cases and 739 414 controls.

FinnGen+EstBB meta-analysis

The METAL software³⁸ was used to do an inverse variance-weighted fixed-effect meta-analysis of the GWAS results from FinnGen and EstBB. Prior to the meta-analysis, variant data from the Estonian Biobank were converted from the hg19 to the hg38 prior to the meta-analysis. Summary statistics from the FinnGen+EstBB meta-analysis results were used in the Mendelian randomization, in order to increase statistical power for detecting potential causal relationships.

Candidate gene characterization

A locus was defined as a window of 2MB ($\pm 1,000,000$ bases) that contains at least one variant associated with LSS at $P < 5 \times 10^{-8}$. Using literature and databases (Genbank³⁹, UniProt⁴⁰, GTEX-Portal⁴¹), and identified a potential candidate gene with a relevant biological function for the loci that had not been reported in association with LSS in previous studies.

Conditional analyses

With help of GCTA software (1.93.0 beta Linux)⁴², we performed conditional analyzes for the loci to identify possible secondary signals. We used lead variants observed from the loci as a covariate and utilized FinnGen as the reference sample to estimate linkage disequilibrium (LD) corrections. The associations were first conditioned on the most significant variant at each locus using the --cojo-cond option with the default settings. For those loci where secondary signals were detected, the analysis was repeated using the first conditional analysis results. Conditioning was continued until no variant reached genome-wide significance ($p < 5 \times 10^{-8}$).

Downstream analyses

Heritability

The LD score regression-derived SNP-based heritability estimate was computed using LDSC software (version 1.0.1)⁸. Heritability estimation was performed using the liability scale, with a sample prevalence of 0.039 and a population prevalence of 0.11 as estimated by Jensen et al. (2020)⁴. For heritability estimation we used the HapMap3, a European-only reference panel provided by Alkes Price's group (<https://alkesgroup.broadinstitute.org/LDSCORE/>). Summary statistics included 1202708 SNPs, and after merging with the LD reference panel 1177076 SNPs remained. FinnGen data showed a sample prevalence of 0.065. The reference panel remained unchanged; 1191433 SNPs were included in the FinnGen summary statistics, and after merging with the LD reference panel, 1168512 SNPs remained.

Functional annotations

FUMA²⁷ was used for the functional annotations that were performed based on the meta-analysis summary statistics. We selected that the analysis would use functional information for the mapping. Positional mapping was also carried out, with SNP markers chosen from regions of exons or introns that affect post-transcriptional modifications and are involved in gene regulation. Gene expression data were also used for mapping, as well as eQTL mapping. We used whole blood, cultured fibroblasts, tibial nerve, and cervical spinal cord cervical (GTEx v8) tissues in the analysis, and the focused on the protein-coding genes only. MAGMA analysis²⁶, a functional association test, was also performed, it focuses on gene-level

information, unlike GWAS, where associations are reported at the variant level. MAGMA uses curated gene sets and GO annotations from MSigDB²⁸ in the analyses. We used 10kb gene window and selected GTEx v8 tissue variants in the analysis. We also used fastBAT from the GCTA software package (1.93.0 beta Linux)⁴² for gene prioritization. re calculates 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 regions using LD correlations between SNPs from a reference sample with individual-level genotypes and summary data from GWAS⁴³. We utilized FinnGen as a LD reference and used 2MB window for the loci. For LD pruning 0.9 r2 threshold was used alongside default methods. We considered results with pfastBAT < 5×10^{-8} and pGWAS < 5×10^{-8} significant⁴³. Lead variants that located in the HLA region were left out from the annotation analyses.

eQTL Colocalizations

We investigated eQTL colocalizations between LSS association signals and gene expression using the 'coloc.abf' function from the R-library 'coloc' (version 5.2.3)⁴⁴. The 'biomaRt' (version 2.58.2) R-library was used to extract genes from Emsembl archives²⁹ within a 1 MB (± 500000) window surrounding each lead variant. Colocalizations were investigated for 754 genes that were available in GTEx. We chose tissues relevant to LSS for analysis, including the tibial nerve, cervical spinal cord, and whole blood. For the selected tissues, data on variant-gene expression associations (GTEx v8) were downloaded from GTExportal⁴¹. Colocalizations with a posterior probability of ≥ 0.8 for the variant were deemed significant³⁰.

Survival analysis

Our aim was to observe how the LSS diagnoses accumulate for different variants according to age and to evaluate whether there are early onset variants, by exploring the cumulative events. 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, of the variants where significant differences in the accumulation of diagnoses were observed. 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 used was: $diff = ('survival\ rate\ EA/EA' - 'survival\ rate\ OA/OA') / \sqrt{('se_survival_rate\ EA/EA2' + 'se_survival_rate\ OA/OA2')}$, $p = 2 * (1 - diff)$. A significant difference was defined as a P-value < 0.05. Furthermore, we calculated cumulative morbidity for each variant to determine whether some variants result in more diagnoses.

Genetic correlations

Genetic correlations were calculated between meta-analysed LSS summary statistics and 517 other phenotypes extracted from the GWAS database provided by the MRC Integrative Epidemiology Unit (IEU) (<https://gwas.mrcieu.ac.uk/>). These calculations were performed using the LDSC software (version 1.0.1)^{8,31}. We set the threshold for significant correlations at a false discovery rate (FDR)-corrected p-value (pFDR) < 0.05. Genetic correlations were performed using HapMap3 European-only reference panel provided by Alkes Price's group (<https://alkesgroup.broadinstitute.org/LDSCORE/>). For genetic correlations, we used our LSS 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

In order to investigate the causal relationships between LSS and its associated risk factors, we used the Two-Sample MR (version 0.5.8) R library to perform a bi-directional Mendelian

randomization. Risk factors related to body composition, pain, and inflammation were included in the analysis (Table S11). We were able to assess if LSS is causal for risk factors and, conversely, if risk factors are causal for LSS due to a bi-directional study approach. To ensure that there was no overlap between the study populations. We obtained the LSS instruments from the FinnGen+EstBB meta-analysis results since many of the GWAS data provided by the MRC-IEU are UKBB-based. Variants that were correlated were removed from the data, leaving only independent variants for analysis. We used the default clumping settings in the analysis (clumping window 10 000kb, r^2 0.001). The principal analysis method we employed was the Inverse Variance Weighted (IVW) model. In the sensitivity analyses, we obtained MR Egger estimates and performed Cochran's Q-test and the MR Egger intercept test to evaluate the heterogeneity and pleiotropy of the instruments. In order to determine whether a particular SNP is responsible for a potentially detectable causal relationship, a leave-one-out analysis was also carried out.

Data availability

The individual-level data are available under restricted access for legal and ethical reasons. Formal approval for the researchers is needed to access the data: please see https://www.finnngen.fi/en/access_results for more details. Individual-level data access and genotype data is managed by the Finnish Biobank Cooperative at the Fingenious portal [<https://site.fingenious.fi/en/>]). We will make the summary statistics of this study available through the NHGRI-EBI GWAS Catalog upon publication.

Code availability

The citations and URLs listed in the Methods section contain the code for every tool used in the analyses. All tools used in our study were open source.

Acknowledgements

E.S. was funded by Academy of Finland (grant number: 338229) and Orion Research Foundation sr. J.K. was funded by Sigrid Juselius foundation. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources. We want to acknowledge the participants and investigators of 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/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (<https://www.ppshep.fi/Tutkimus-ja-opetus/Biopankki/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-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta), Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/) 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 Fingenious® 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 Center 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 [1.1-12/624] from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application [N04] from the Estonian Biobank.

Consortia

FinnGen

Aarno Palotie^{11,12,13}, Mark Daly^{11,12,13}, Bridget Riley-Gills¹⁴, Howard Jacob¹⁴, Dirk Paul¹⁵, Slavé Petrovski¹⁵, Heiko Runz¹⁶, Sally John¹⁶, George Okafo¹⁷, Nathan Lawless¹⁷, Heli Salminen-Mankonen¹⁷, Robert Plenge¹⁸, Joseph Maranville¹⁸, Mark McCarthy¹⁹, Margaret G. Ehm²⁰, Kirsi Auro²¹, Simonne Longrich²², Anders Mälarstig²³, Katherine Klinger²⁴, Clement Chatelain²⁴, Matthias Gossel²⁴, Karol Estrada²⁵, Robert Graham²⁵, Robert Yang²⁶, Chris O'Donnell²⁷, Tomi P. Mäkelä¹¹, Jaakko Kaprio¹¹, Petri Virolainen²⁸, Antti Hakanen²⁸, Terhi Kilpi²⁹, Markus Perola²⁹, Jukka Partanen³⁰, Anne Pitkäranta³¹, Taneli Raivio³¹, Jani Tikkanen³², Raisa Serpi³², Tarja Laitinen³³, Veli-Matti Kosma³⁴, Jari Laukkanen^{35,36}, Marco Hautalahti³⁷, Outi Tuovila³⁸, Raimo Pakkanen³⁸, Jeffrey Waring¹⁴, Fedik Rahimov¹⁴, Ioanna Tachmazidou¹⁵, Chia-Yen Chen¹⁶, Zhihao Ding¹⁷, Marc Jung¹⁷, Shameek Biswas¹⁸, Rion Pendergrass¹⁹, David Pulford³⁹, Neha Raghavan²², Adriana Huertas-Vazquez²², Jae-Hoon Sul²², Xinli Hu²³, Åsa Hedman²³, Manuel Rivas^{25,40}, Dawn Waterworth⁴¹, Nicole Renaud²⁷, Ma'en Obeidat²⁷, Samuli Ripatti¹¹, Johanna Schleutker²⁸, Mikko Arvas³⁰, Olli Carpén³¹, Reetta Hinttala³², Arto Mannermaa³⁴, Katriina Aalto-Setälä⁴², Mika Kähönen³³, Johanna Mäkelä³⁷, Reetta Kälviäinen⁴³, Valtteri Julkunen⁴³, Hilka Soininen⁴³, Anne Remes⁴⁴, Mikko Hiltunen⁴⁵, Jukka Peltola⁴⁶, Minna Raivio⁴⁷, Pentti Tienari⁴⁷, Juha Rinne⁴⁸, Roosa Kallionpää⁴⁸, Juulia Partanen¹¹, Ali Abbasi¹⁴, Adam Ziemann¹⁴, Nizar Smaoui¹⁴, Anne Lehtonen¹⁴, Susan Eaton¹⁶, Sanni Lahdenperä¹⁶, Natalie Bowers¹⁹, Edmond Teng¹⁹, Fanli Xu⁴⁹, Laura Addis⁴⁹, John Eicher⁴⁹, Qingqin S Li⁵⁰, Karen He⁴¹, Ekaterina Khramtsova⁴¹, Martti Färkkilä⁴⁷, Jukka Koskela⁴⁷, Sampsa Pikkarainen⁴⁷, Airi Jussila⁴⁶, Katri Kaukinen⁴⁶, Timo Blomster⁴⁴, Mikko Kiviniemi⁴³, Markku Voutilainen⁴⁸, Tim Lu¹⁹, Linda McCarthy⁴⁹, Amy Hart⁴¹, Meijian Guan⁴¹, Jason Miller²², Kirsi Kalpala²³, Melissa Miller²³, Kari Eklund⁴⁷, Antti Palomäki⁴⁸, Pia Isomäki⁴⁶, Laura Pirilä⁴⁸, Oili Kaipainen-Seppänen⁴³, Johanna Huhtakangas⁴⁴, Nina Mars¹¹, Apinya Lertratanakul¹⁴, Coralie Viollet¹⁵, Marla Hochfeld¹⁸, Jorge Esparza Gordillo⁴⁹, Fabiana Farias²², Nan Bing²³, Margit Pelkonen⁴³, Paula Kauppi⁴⁷, Hannu Kankaanranta^{42,51,52}, Terttu Harju⁴⁴, Riitta Lahesmaa⁴⁸, Hubert Chen¹⁹, Joanna Betts⁴⁹, Rajashree Mishra⁴⁹, Majd Mouded⁵³, Debby Ngo⁵³, Teemu Niiranen⁵⁴, Felix Vaura⁵⁴, Veikko Salomaa⁵⁴, Kaj Metsärinne⁴⁸, Jenni Aittokallio⁴⁸, Jussi Hernesniemi⁴⁶, Daniel Gordin⁴⁷, Juha Sinisalo⁴⁷, Marja-Riitta Taskinen⁴⁷, Tiinamaija Tuomi⁴⁷, Timo Hiltunen⁴⁷, Amanda Elliott^{11,13,55}, Mary Pat Reeve¹¹, Sanni Ruotsalainen¹¹, Audrey Chu⁴⁹, Dermot Reilly⁵⁶, Mike Mendelson⁵⁷, Jaakko Parkkinen²³, Tuomo Meretoja⁴⁷, Heikki Joensuu⁴⁷, Johanna Mattson⁴⁷, Eveliina Salminen⁴⁷, Annika Auranen⁴⁶, Peeter Karihtala⁴⁴, Päivi Auvinen⁴³, Klaus Elenius⁴⁸, Esa Pitkänen¹¹, Relja Popovic¹⁴, Margarete Fabre¹⁵, Jennifer Schutzman¹⁹, Diptee Kulkarni⁴⁹, Alessandro Porello⁴¹, Andrey Loboda²², Heli Lehtonen²³, Stefan McDonough²³, Sauli Vuoti⁵⁸, Kai Kaarniranta^{43,59}, Joni A. Turunen^{60,61}, Terhi Ollila⁴⁷, Hannu Uusitalo⁴⁶, Juha Karjalainen¹¹, Mengzhen Liu¹⁴,

Stephanie Loomis¹⁶, Erich Strauss¹⁹, Hao Chen¹⁹, Kaisa Tasanen⁴⁴, Laura Huilaja⁴⁴, Katariina Hannula-Jouppi⁴⁷, Teea Salmi⁴⁶, Sirkku Peltonen⁴⁸, Leena Koulu⁴⁸, David Choy¹⁹, Ying Wu²³, Pirkko Pussinen⁴⁷, Aino Salminen⁴⁷, Tuula Salo⁴⁷, David Rice⁴⁷, Pekka Nieminen⁴⁷, Ulla Palotie⁴⁷, Maria Siponen⁴³, Liisa Suominen⁴³, Päivi Mäntylä⁴³, Ulvi Gursoy⁴⁸, Vuokko Anttonen⁴⁴, Kirsi Sipilä^{62,63}, Hannele Laivuori¹¹, Venla Kurra⁴⁶, Laura Kotaniemi-Talonen⁴⁶, Oskari Heikinheimo⁴⁷, Ilkka Kalliala⁴⁷, Lauri Aaltonen⁴⁷, Varpu Jokimaa⁴⁸, Marja Vääräsmäki⁴⁴, Outi Uimari⁴⁴, Laure Morin-Papunen⁴⁴, Maarit Niinimäki⁴⁴, Terhi Piltonen⁴⁴, Katja Kivinen¹¹, Elisabeth Widen¹¹, Taru Tukiainen¹¹, Niko Välimäki⁶⁴, Eija Laakkonen⁶⁵, Jaakko Tyrmi^{42,62}, Heidi Silven⁶², Riikka Arffman⁶², Susanna Savukoski⁶², Triin Laisk⁶, Natalia Pujol⁶, Janet Kumar²⁰, Iris Hovatta⁶⁴, Erkki Isometsä⁴⁷, Hanna Ollila¹¹, Jaana Suvisaari⁵⁴, Thomas Damm Als⁶⁶, Antti Mäkitie⁶⁷, Argyro Bizaki-Vallaskangas⁴⁶, Sanna Toppila-Salmi⁶⁸, Tytti Willberg⁴⁸, Elmo Saarentaus¹¹, Antti Aarnisalo⁴⁷, Elisa Rahikkala⁴⁴, Kristiina Aittomäki⁶⁹, Fredrik Åberg⁷⁰, Mitja Kurki^{11,55}, Aki Havulinna^{11,54}, Juha Mehtonen¹¹, Shabbeer Hassan¹¹, Pietro Della Briotta Parolo¹¹, Wei Zhou⁵⁵, Mutaamba Maasha⁵⁵, Susanna Lemmelä¹¹, Aoxing Liu¹¹, Arto Lehisto¹¹, Andrea Ganna¹¹, Vincent Llorens¹¹, Henrike Heyne¹¹, Joel Rämö¹¹, Rodos Rodosthenous¹¹, Satu Strausz¹¹, Tuula Palotie^{47,64}, Kimmo Palin⁶⁴, Javier Gracia-Tabuenca⁴², Harri Siirtola⁴², Tuomo Kiiskinen¹¹, Jiwoo Lee^{11,55}, Kristin Tsuo^{11,55}, Kati Kristiansson²⁹, Kati Hyvärinen⁷¹, Jarmo Ritari⁷¹, Katri Pylkäs⁶², Minna Karjalainen⁶², Tuomo Mantere³², Eeva Kangasniemi³³, Sami Heikkinen⁴⁵, Nina Pitkänen²⁸, Samuel Lessard²⁴, Lila Kallio²⁸, Tiina Wahlfors²⁹, Eero Punkka³¹, Sanna Siltanen³³, Teijo Kuopio^{35,36}, Anu Jalanko¹¹, Huei-Yi Shen¹¹, Risto Kajanne¹¹, Mervi Aavikko¹¹, Helen Cooper¹¹, Denise Öller¹¹, Rasko Leinonen^{11,72}, Henna Palin³³, Malla-Maria Linna³¹, Masahiro Kanai⁵⁵, Zhili Zheng⁵⁵, L. Elisa Lahtela¹¹, Mari Kaunisto¹¹, Elina Kilpeläinen¹¹, Timo P. Sipilä¹¹, Oluwaseun Alexander Dada¹¹, Awaisa Ghazal¹¹, Anastasia Kytölä¹¹, Rigbe Weldatsadik¹¹, Kati Donner¹¹, Anu Loukola³¹, Päivi Laiho²⁹, Tuuli Sistonen²⁹, Essi Kaiharju²⁹, Markku Laukkanen²⁹, Elina Järvensivu²⁹, Sini Lähteenmäki²⁹, Lotta Männikkö²⁹, Regis Wong²⁹, Auli Toivola²⁹, Minna Brunfeldt²⁹, Hannele Mattsson²⁹, Sami Koskelainen²⁹, Tero Hiekkalinna²⁹, Teemu Paajanen²⁹, Kalle Pärn¹¹, Mart Kals¹¹, Shuang Luo¹¹, Shanmukha Sampath Padmanabhuni¹¹, Marianna Niemi⁴², Mika Helminen⁴², Tiina Luukkaala⁴², Iida Vähätalo⁴², Jyrki Tammerluoto¹¹, Sarah Smith³⁷, Tom Southerington³⁷, Petri Lehto³⁷

¹¹ Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland, ¹² Broad Institute of MIT and Harvard, Cambridge, MA, United States, ¹³ Massachusetts General Hospital, Boston, MA, United States, ¹⁴ Abbvie, Chicago, IL, United States, ¹⁵ Astra Zeneca, Cambridge, United Kingdom, ¹⁶ Biogen, Cambridge, MA, United States, ¹⁷ Boehringer Ingelheim, Ingelheim am Rhein, Germany, ¹⁸ Bristol Myers Squibb, New York, NY, United States, ¹⁹ Genentech, San Francisco, CA, United States, ²⁰ GlaxoSmithKline, Collegeville, PA, United States, ²¹ GlaxoSmithKline, Espoo, Finland, ²² Merck, Kenilworth, NJ, United States, ²³ Pfizer, New York, NY, United States, ²⁴ Translational Sciences, Sanofi R&D, Framingham, MA, USA, ²⁵ Maze Therapeutics, San Francisco, CA, United States, ²⁶ Janssen Biotech, Beerse, Belgium, ²⁷ Novartis Institutes for BioMedical Research, Cambridge, MA, United States, ²⁸ Auria Biobank, University of Turku, Hospital District of Southwest Finland, Turku, Finland, ²⁹ THL Biobank, Finnish Institute for Health and Welfare (THL), Helsinki, Finland, ³⁰ Finnish Red Cross Blood Service, Finnish Hematology Registry and Clinical Biobank, Helsinki, Finland, ³¹ Helsinki Biobank, Helsinki University and Hospital District of Helsinki and Uusimaa, Helsinki, ³² Northern Finland Biobank Borealis, University of Oulu, Northern Ostrobothnia Hospital District, Oulu, Finland, ³³ Finnish Clinical Biobank Tampere, University of Tampere, Pirkanmaa Hospital District, Tampere, Finland, ³⁴ Biobank of Eastern Finland, University of Eastern Finland, Northern Savo Hospital District, Kuopio, Finland, ³⁵ Central Finland Biobank, University of Jyväskylä, Central Finland Health Care District, Jyväskylä, Finland, ³⁶ Central Finland Health Care District, Jyväskylä, Finland, ³⁷ FINBB - Finnish biobank cooperative, ³⁸ Business Finland, Helsinki, Finland, ³⁹ GlaxoSmithKline, Stevenage, United Kingdom, ⁴⁰ University of Stanford, Stanford, CA, United States, ⁴¹ Janssen Research & Development, LLC, Spring House, PA, United States, ⁴² University of Tampere, Tampere, Finland, ⁴³ Northern Savo Hospital District, Kuopio, Finland, ⁴⁴ Northern Ostrobothnia Hospital District, Oulu, Finland, ⁴⁵ University of Eastern Finland, Kuopio, Finland, ⁴⁶ Pirkanmaa Hospital District, Tampere, Finland, ⁴⁷ Hospital District of Helsinki and Uusimaa, Helsinki, Finland, ⁴⁸ Hospital District of Southwest Finland, Turku, Finland, ⁴⁹ GlaxoSmithKline, Brentford, United Kingdom, ⁵⁰ Janssen Research &

Development, LLC, Titusville, NJ 08560, United States, ⁵¹ University of Gothenburg, Gothenburg, Sweden, ⁵² Seinäjoki Central Hospital, Seinäjoki, Finland, ⁵³ Novartis, Basel, Switzerland, ⁵⁴ Finnish Institute for Health and Welfare (THL), Helsinki, Finland, ⁵⁵ Broad Institute, Cambridge, MA, United States, ⁵⁶ Janssen Research & Development, LLC, Boston, MA, United States, ⁵⁷ Novartis, Boston, MA, United States, ⁵⁸ Janssen-Cilag Oy, Espoo, Finland, ⁵⁹ Department of Molecular Genetics, University of Lodz, Lodz, Poland, ⁶⁰ Helsinki University Hospital and University of Helsinki, Helsinki, Finland, ⁶¹ Eye Genetics Group, Folkhälsan Research Center, Helsinki, Finland, ⁶² University of Oulu, Oulu, Finland, ⁶³ Medical Research Center, Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland, ⁶⁴ University of Helsinki, Helsinki, Finland, ⁶⁵ University of Jyväskylä, Jyväskylä, Finland, ⁶⁶ Aarhus University, Denmark, ⁶⁷ Department of Otorhinolaryngology - Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland, ⁶⁸ University of Eastern Finland and Kuopio University Hospital, Department of Otorhinolaryngology, Kuopio, Finland and Department of Allergy, Helsinki University Hospital and University of Helsinki, Finland, ⁶⁹ Department of Medical Genetics, Helsinki University Central Hospital, Helsinki, Finland, ⁷⁰ Transplantation and Liver Surgery Clinic, Helsinki University Hospital, Helsinki University, Helsinki, Finland, ⁷¹ Finnish Red Cross Blood Service, Helsinki, Finland, ⁷² European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK

Estonian Biobank Research Team

Andres Metspalu⁶, Mari Nelis⁶, Lili Milani⁶, Georgi Hudjashov⁶