

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	In FinnGen, genotyping of the samples was performed using Illumina and Affymetrix arrays. Chip genotyped samples were pre-phased with Eagle 2.3.5. Genotype imputation was carried out by using the Finnish population specific SiSu v3 reference panel with Beagle 4.1. The imputation protocol has been described at (dx.doi.org/10.17504/protocols.io.nmndc5e).
Data analysis	In FinnGen and in EstBB, GWAS using an additive genetic model was performed using the Regenie (version 2.2.4) program, adjusting each phenotype for age, sex, and the first 10 genetic principal components. The sensitivity analyses with stricter case definitions conducted in FinnGen were performed using Regenie and the same covariates as above. 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 favored. The SNP-based heritability, genetic correlations, and genomic inflation factor were estimated using LDSC regression implemented in LDSC software (v.1.0.1). Conditional analyses and fastBAT gene prioritization were conducted using GCTA (1.93 beta Linux). The functional annotations of the GWAS results were completed using a web-based platform, FUMA (accessed November 21, 2022). 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).Cmprsk (2.2–10), survminer (0.4.9), and survival (3.2–7) R-libraries were used to create Kaplan-Meier plots and for statistical analysis of the plots. To test for causal inferences between lumbar disc herniations (LDH) and traits genetically correlated with LDH, we performed bi-directional two-sample Mendelian randomization using TwoSampleMR (0.5.8) R-library. 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.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

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 Fingenious portal (<https://site.fingenious.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 inferences are available at the MRC-IEU GWAS database.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

In the quality control phase of genotyping, individuals whose sex did not match the genotype data were removed from the analysis. Sex was also one of the covariates used in the GWAS analyses. Otherwise, sex was not taken into account in our study, but the results are based on data that includes both sexes.

Reporting on race, ethnicity, or other socially relevant groupings

Our sample is limited to European ancestry only.

Population characteristics

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. The FinnGen project has the necessary ethical and prior permits for biobank research (Supplementary Note; can be seen in the supplementary materials of our manuscript), 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. The Estonian Biobank (www.genomics.ut.ee/en) cohort is a volunteer-based sample of the Estonian resident adult population (aged ≥18 years). 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. 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 this study.

Recruitment

We have used registry-based data from FinnGen: the FinnGen project utilizes genome information from a nationwide network of Finnish biobanks that are linked with digital health records from national hospital discharge (available from 1968), death (1969-), cancer (1953-), and medication reimbursement (1995-) registries.

Ethics oversight

FinnGen DF9 Ethics statement

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 came into effect (in September 2013) and 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.

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: VRK/43431/2017-3, 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, Auri Biobank AB17-5154 and amendment #1 (August 17 2020), AB20-5926 and amendment #1 (April 23 2020) and it's 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.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Genome-wide meta-analysis data used in the study contains 80 724 cases and 748 975 controls. Cases were required to have an entry of ICD-10:M51 (M51.1-51.9), and participants who had no records of these entries were deemed as controls. Meta-analysis data includes data from three large biobanks: FinnGen (37,636 cases and 270,964 controls), Estonian Biobank (34,035 cases and 66,533 controls), and UK Biobank (9,053 cases and 411,478 controls). We also performed two sensitivity GWAS analyses in FinnGen using stricter case definitions. The first of the two 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. The second additional case definition included LDH patients who had undergone an LDH-related operation (NOMESCO version 1.15, ABC07, ABC16 & ABC26); this analysis included 7347 cases and controls of 270 964. LDH cases that had not been operated on 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.
Data exclusions	In FinnGen, sample quality control (QC) was performed to exclude individuals with high genotype missingness (>5%), ambiguous gender, excess heterozygosity (4SD) and non-Finnish ancestry. Regarding variant QC, all variants with low Hardy-Weinberg equilibrium (HWE) p-value (<1e-6), high missingness (>2%), and minor allele count (MAC) <3 were excluded. In post-imputation QC, variants with imputation info <0.6 were excluded.
Replication	We report meta-analysed GWAS findings that were obtained using data from three large biobanks, namely FinnGen, Estonian Biobank, and UK Biobank. We found that the results are consistent in the studied data sets, with very little to no heterogeneity, indicative of successful replication.
Randomization	In most part, not relevant of our study- this is a genetic association study conducted in a case-control setting and randomization is not applicable to this methodology. Regarding Mendelian randomization analyses, the random distribution of genotypes at meiosis is considered comparable to exposure randomization in RTC according to the principles of the methodology.
Blinding	Not relevant to our study - this is a genetic association study conducted using registry-based health data and blinding is not applicable to this methodology.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>