nature portfolio

Corresponding author(s): Ville Salo

Last updated by author(s): 2025/11/14

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency 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 (*n*) 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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	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. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	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 *d*, Pearson's *r*), 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 v4.2 reference panel with Beagle 4.1. The imputation protocol has been described at https://docs.finngen.fi/finngen-data-specifics/red-library-data-individual-level-data/genotype-data/imputation-panel/sisu-v4.2-reference-panel. In the Estonian Biobank, Illumina Human CoreExome, OmniExpress, 370CNV BeadChip and GSA arrays were used for genotyping. SHAPEIT2 was used for prephasing, while the imputation was performed with the Estonian-specific reference panel and IMPUTE2 with default settings. In the UK Biobank, Applied Biosystems UK BiLEVE Axiom Array by Affymetrix and Applied Biosystems UK Biobank Axiom Arrays were used for genotyping. Phasing on the autosomes was carried out using SHAPEIT3, with the 1000 Genomes phase 3 dataset used as a reference panel. For imputation the Haplotype Reference Consortium (HRC) dataset was used as the main imputation reference panel but also merged UK10K and 1000 Genomes phase 3 reference panels were utilized. Imputation was carried out with the IMPUTE4 program (https://jmarchini.org/software/) which is a re-coded version of the haploid imputation functionality implemented in IMPUTE2.

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, the first 10 genetic principal components and genotyping arrays. In UKBB, GWAS was performed using an additive genetic model and the SAIGE pipeline, modifying each phenotype for age, sex, and the first 10 genetic principal components. A more detailed description of GWAS pipeline and implementation can be found here: https://github.com/atgu/ukbb_pan_ancestry/wiki/Batch-pipeline. The sensitivity analyses conducted in FinnGen and Estonian Biobank were performed using Regenie and the same covariates as above. An inverse variance-weighted fixed-effect meta-analysis of the GWAS results from FinnGen, EstBB, and UK Biobank was conducted using the METAL software. Variant data from the Estonian and UK Biobanks were converted from hg19 to hg38 before being meta-analysed using the Picard liftover (http://broadinstitute.io/picard). The METAL software was used to do an inverse variance-weighted fixed-effect meta-analysis of the surgical GWAS results from FinnGen and EstBB. In order to account for heterogeneity that was observed for some lead variants, we

additionally conducted a random effect meta-analysis. The lead variants observed in the main meta-analysis were used in the analysis that were conducted in same cohorts as out main meta-analysis. The analysis was conducted using the metafor R package using the "REML" model. We applied the FinnGen fine-mapping pipeline (available at https://github.com/FINNGEN/finemapping-pipeline) with default parameters after first extracting the summary statistics of each locus associated in the meta-analysis. Exception was made for variants in HLA region that were left out. The SNP-based heritability, genetic correlations, and genomic inflation factor were estimated using LDSC regression implemented in LDSC software (v.1.0.1). 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). 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, MAGMA (accessed March 12, 2025). In FinnGen, we also conducted a PheWAS analysis for lead variants associated with LSS. The PheWAS analysis excluded lead variants that located in the HLA region. In the analysis we used 124 clinician curated core endpoints to test the association of lead variants, and results with p<5x10-8 were considered significant. All endpoints and endpoint definitions are available here: https://www.finngen.fi/en/researchers/clinical-endpoints. Whole PheWAS pipeline has been described here: https://github.com/FINNGEN/pheweb/. We investigated into the colocalizations of LSS association signals using various QTL data and other FinnGen R12 endpoints. For this the FinnGen SuSiE colocalization pipeline was utilized (https://github.com/FINNGEN/coloc.susie.direct). Colocalizations between 571 resources are carried out by this pipeline, including all GWAS endpoints from UKBB, FinnGen, and the eQTL catelogue, the Generisk project, and the proteomics study from INTERVAL, UKBB, and FinnGen. 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. In addition, we conducted a GATE analysis in FinnGen, which examines the association between the variants and the time-to-event rather than the presence of the disease. In the analysis, we attempted to detect variants associated with LSS disease onset by extracting diagnosis ages for LSS cases and ages at the end of the follow-up or at death for controls. Sex and the first 4 genetic principal components were used as covariates in the analysis. To test for causal inferences between LSS and other traits, we performed bi-directional two-sample Mendelian randomization using TwoSampleMR (0.6.17) R-library. Also, MRPRESSO with default settings (NbDistribution=10000, P-value threshold=0.05) was utilised. The Locuszooms (https://github.com/Geeketics/LocusZooms) R package was used to generate the Locuszoom plots, and the Ensembl archive (https://jul2023.archive.ensembl.org) was used for the gene list. 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 meta-analysis summary statistics data generated in this study have been deposited in the FinnGen public cloud: https://storage.googleapis.com/fg-publication-green-public/F_2021_048_20251105/LSS_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.finngen.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	Only participants with European ancestry were included in our study
Population characteristics	The main goal of FinnGen (www.finngen.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 (see, Ethics oversight), 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). 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 R12 Ethics statement

Study 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: 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/4432/14.06/2020, THL/5189/14.06/2020, 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/4235/14.06.00/2021, 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 12 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, BB2021_65, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020, HUS/430/2021 §28, §29, HUS/150/2022 §12, §13, §14, §15, §16, §17, §18, §23, §58, §59, HUS/128/2023 §18, Auria Biobank AB17-5154 and amendment #1 (August 17 2020) and amendments BB_2021-0140, BB_2021-0156 (August 26 2021, Feb 2 2022), BB_2021-0169, BB_2021-0179, BB_2021-0161, AB20-5926 and amendment #1 (April 23 2020) and it´s modifications (Sep 22 2021), BB_2022-0262, BB_2022-0256, Biobank Borealis of Northern Finland_2017_1013, 2021_5010, 2021_5010 Amendment, 2021_5018, 2021_5018 Amendment, 2021_5015, 2021_5015 Amendment, 2021_5015 Amendment_2, 2021_5023, 2021_5023 Amendment, 2021_5023 Amendment_2, 2021_5017, 2021_5017 Amendment, 2022_6001, 2022_6001 Amendment, 2022_6006 Amendment, 2022_6006 Amendment, 2022_6006 Amendment_2, BB22-0067, 2022_0262, 2022_0262 Amendment, Biobank of Eastern Finland 1186/2018 and amendment 22§/2020, 53§/2021, 13§/2022, 14§/2022, 15§/2022, 27§/2022, 28§/2022, 29§/2022, 33§/2022, 35§/2022, 36§/2022, 37§/2022, 39§/2022, 7§/2023, 32§/2023, 33§/2023, 34§/2023, 35§/2023, 36§/2023, 37§/2023, 38§/2023, 39§/2023, 40§/2023, 41§/2023, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), BB2021-0140 8§/2021, 9§/2021, §9/2022, §10/2022, §12/2022, 13§/2022, §20/2022, §21/2022, §22/2022, §23/2022, 28§/2022, 29§/2022, 30§/2022, 31§/2022, 32§/2022, 38§/2022, 40§/2022, 42§/2022, 1§/2023, Central Finland Biobank 1-2017, BB_2021-0161, BB_2021-0169, BB_2021-0179, BB_2021-0170, BB_2022-0256, BB_2022-0262, BB22-0067, Decision allowing to continue data processing until 31st Aug 2024 for projects: BB_2021-0179, BB22-0067,BB_2022-0262, BB_2021-0170, BB_2021-0164, BB_2021-0161, and BB_2021-0169, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020, Finnish Hematological Registry and Clinical Biobank decision 18th June 2021, Arctic biobank P0844: ARC_2021_1001.

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 consists of 40 303 LSS cases and 741 468 controls, in total there were 781 771 participants. 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 the study. 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). Meta-analysis data includes data from three large biobanks: FinnGen (26 090 cases and 353 224 controls), Estonian Biobank (6019 cases and 65 258 controls), and UK Biobank (8194 cases and 322 986 controls).

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. In the Estonian Biobank, 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 (> mean ± 3SD), 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. In the UK Biobank, individuals that were identified as outliers for heterozygosity and missing rate and individuals whose sex did not match the genotype data were removed.
Replication	We report meta-analysed GWAS findings that were obtained using data from three large biobanks, 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	This is case-control association study, so randomization is not applicable to this methodology. In 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	This is genetic association study and blinding is not relevant for this kind of study

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	*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.*
Novel plant genotypes	*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.*
Authentication	*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, mosiacism, off-target gene editing) were examined.*