

STUDY PROTOCOL

A protocol for recruiting and analyzing the disease-oriented Russian disc degeneration study (RuDDS) biobank for functional omics studies of lumbar disc degeneration

Olga N. Leonova^{1*}, Elizaveta E. Elgaeva^{2,3}, Tatiana S. Golubeva³, Alexey V. Peleganchuk¹, Aleksandr V. Krutko⁴, Yurii S. Aulchenko³, Yakov A. Tsepilov^{2,3*}

1 Novosibirsk Research Institute of Traumatology and Orthopedics, Novosibirsk, Russia, **2** Novosibirsk State University, Novosibirsk, Russia, **3** Institute of Cytology and Genetics, Novosibirsk, Russia, **4** Priorov National Medical Research Center of Traumatology and Orthopedics, Moscow, Russia

✉ These authors contributed equally to this work.

* oleonova@gmail.com (ONL); tsepilov@bionet.nsc.ru (YAT)



OPEN ACCESS

Citation: Leonova ON, Elgaeva EE, Golubeva TS, Peleganchuk AV, Krutko AV, Aulchenko YS, et al. (2022) A protocol for recruiting and analyzing the disease-oriented Russian disc degeneration study (RuDDS) biobank for functional omics studies of lumbar disc degeneration. PLoS ONE 17(5): e0267384. <https://doi.org/10.1371/journal.pone.0267384>

Editor: Walid Kamal Abdelbasset, Prince Sattam Bin Abdulaziz University, College of Applied Medical Sciences, SAUDI ARABIA

Received: May 26, 2021

Accepted: April 7, 2022

Published: May 13, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0267384>

Copyright: © 2022 Leonova et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Lumbar intervertebral disc degeneration (DD) disease is one of the main risk factors for low back pain and a leading cause of population absenteeism and disability worldwide. Despite a variety of biological studies, lumbar DD is not yet fully understood, partially because there are only few studies that use systematic and integrative approaches. This urges the need for studies that integrate different omics (including genomics and transcriptomics) measured on samples within a single cohort. This protocol describes a disease-oriented Russian disc degeneration study (RuDDS) biobank recruitment and analyses aimed to facilitate further omics studies of lumbar DD integrating genomic, transcriptomic and glycomic data. A total of 1,100 participants aged over 18 with available lumbar MRI scans, medical histories and biological material (whole blood, plasma and intervertebral disc tissue samples from surgically treated patients) will be enrolled during the three-year period from two Russian clinical centers. Whole blood, plasma and disc tissue specimens will be used for genotyping with genome-wide SNP-arrays, glycome profiling and RNA sequencing, respectively. Omics data will be further used for a genome-wide association study of lumbar DD with *in silico* functional annotation, analysis of plasma glycome and lumbar DD disease interactions and transcriptomic data analysis including an investigation of differential expression patterns associated with lumbar DD disease. Statistical tests applied in each of the analyses will meet the standard criteria specific to the attributed study field. In a long term, the results of the study will expand fundamental knowledge about lumbar DD development and contribute to the elaboration of novel personalized approaches for disease prediction and therapy. Additionally to the lumbar disc degeneration study, a RuDDS cohort could be used for other genetic studies, as it will have unique omics data.

Trial registration number [NCT04600544](https://clinicaltrials.gov/ct2/show/study/NCT04600544).

Funding: The study is covered by the Russian Science Foundation grant number 22-15-20037. The data analysis will be performed using computational resources of the “Bioinformatics” Joint Computational Center supported by the budget project № FWNR-2022-0020. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: YSA is a founder and co-owner of PolyOmics and PolyKnomics, private organisations that provide services, research, and development in the field of quantitative and statistical genetics and computational genomics. This does not alter our adherence to PLOS ONE policies on sharing data and materials. Other authors declare no conflicts of interest.

Introduction

Intervertebral disc degeneration is a normal aging process, but in some cases it causes lumbar disc degeneration disease (LDDD) [1,2]. Intervertebral disc degeneration (DD) often begins earlier than degenerative changes in the ligaments, cartilages and other tissues of spinal segment [3]. DD is a major contributor to low back pain [4], but can also proceed without back pain [5–8]. However, LDDD is associated with a lower health-related quality of life [9] and is the leading cause of population absenteeism and disability [10,11]. Intervertebral disc degeneration (DD) is an initiating factor in LDDD and is one of the major contributors to subsequent low back pain [4]. The prevalence of lumbar DD in the general population is extremely high: up to 50% in people aged 30–39 years [2,12]. There is not a single adult who has not at least some degree of degeneration in the intervertebral discs [13]. The list of well-known risk factors of LDDD includes female gender, advanced age [14], obesity, smoking [15], absence of or extreme physical activity [16] and genetic risk factors [17]. However, even in the absence of obvious risk factors, there are cases of LDDD among young people [1,14], as well as progressive severe LDDD requiring a number of surgical interventions.

The biology of intervertebral DD is extensively studied from different points of view. Besides various studies of degenerative disc morphology [18–22] and molecular biology [16,23–25], there is an increasing number of genetic [26,27], transcriptomic [28–30] and proteomic [31–33] studies of intervertebral DD. Thus, based on candidate-gene [34] and genome-wide associated studies (GWAS) [35,36], over 160 genes have now been considered to be potentially involved in intervertebral disc degeneration, although less than ten of them provide strong evidence for the association [34]. More than 500 genes have been highlighted as being expressed differently in degenerative disc tissue and healthy intervertebral discs [37,38]. Moreover, proteomic [31–33] and metabolomic [39] changes have also been detected in degenerative intervertebral discs. Despite a large number of LDDD studies, the exact mechanisms underlying this disease are not yet fully understood, partially because there are only few studies that use systematic and integrative approaches [40]. To our knowledge, there are no large-scale studies that integrate different genomics and other omics data within a single cohort. This complicates the extrapolation of previous findings on an integrative picture of LDDD pathology. Therefore, more integrative omics studies are needed to put all pieces of knowledge together, build a complete picture of the biology of lumbar DD and obtain a deeper understanding of the processes underlying this pathology.

Moreover, the low number of genes reliably and reproducibly associated with LDDD [34] does not explain its high heritability (up to 74%) [41]. It highlights that genetic and transcriptomic studies with greater sample sizes and replication samples are needed.

This protocol describes a study that provides broad opportunities for combining newly acquired and existing knowledge in various fields of genomics, transcriptomics and molecular biology due to its scale and the use of an integrative approach. Not only will this be beneficial for researchers as a resource of fundamentally new information on the mechanisms involved in disc degeneration pathology, it will enrich the list of reliable disease markers and drug targets, expanding the diagnostic and treatment approaches, which is crucial for the development of personalized, accurate and effective treatment of LDDD.

Research aim and objectives

The main aim of this study is to establish a disease-oriented biobank to facilitate research into the biology of lumbar disc degeneration. Different biological samples (whole blood, plasma, disc tissue) along with MRI imaging, clinical, socio-demographic and various omics data (e.g. genomic and transcriptomic) will be available for researchers and clinicians for a variety of

further multi-omics studies. It will lay the groundwork for the development of early diagnostics of LDDD and its personalized treatment.

Methods and analysis

Study design and settings

This disease-oriented biobank to study lumbar disc degeneration will be recruited from two centers: the Priorov National Medical Research Center of Traumatology and Orthopedics (Priorov CITO), Moscow, Russia, and the Novosibirsk Research Institute of Traumatology and Orthopedics (NRITO), Novosibirsk, Russia. The study will be performed according to the Helsinki Declaration; the study protocol was approved by the Local Ethical Committee of the Novosibirsk Research Institute of Traumatology and Orthopedics (№034/20 dated 02 Oct 2020) and by the Local Ethical Committee of the Priorov National Medical Research Center of Traumatology and Orthopedics (№1/21 dated 25 Feb 2021). It will be conducted for a period of three years starting from 2021. Patients aged over 18 years with available MRI of the lumbar spine, who signed informed consent and met eligibility criteria, will be recruited. Clinical data and specimens will be collected in three visits as described in [Table 1](#). At the first step, general information about a patient, his medical history and MRI scans of lumbar spine will be obtained. Patients' biological material (whole blood and plasma) will be sampled in the centers at baseline. Then, from those patients, who will undergo a spine surgery during the treatment disc tissue samples will be gained. Eventually, postoperative clinical data will be collected from operated patients during the follow-up. Over three years, we expect to collect information and biological material from a total of 1,100 patients. A full schematic of the study is presented in [Fig 1](#).

All biological samples from NRITO will then be transported to the Institute of Cytology and Genetics (ICG), Novosibirsk, Russia, for long-term storage, further processing and omics

Table 1. Scheduled procedures for clinical data and sample collection.

	Visit 1—Baseline	Visit 2—Day of surgery ¹	Visit 3—3 months after surgery ²
Visit Window (±Days)			Day 90 ±14 days
Informed Consent	X		
Eligibility Criteria ³	X		
Medical History	X		
Anamnesis Vitae	X		
Demographics	X		
Blood samples	X		
MRI	X		
ODI ⁴	X		X
VAS ⁵	X		X
Indications for surgery		X	
Surgical Procedure		X	
Intervertebral disc samples		X	
Adverse events/Serious adverse events	X	X	X

¹only for patients subjected to surgery

²optional

³see [Table 2](#) for details

⁴Oswestry Disability Index (ODI) scale

⁵Visual analogue scale (VAS) for back and leg pain.

<https://doi.org/10.1371/journal.pone.0267384.t001>

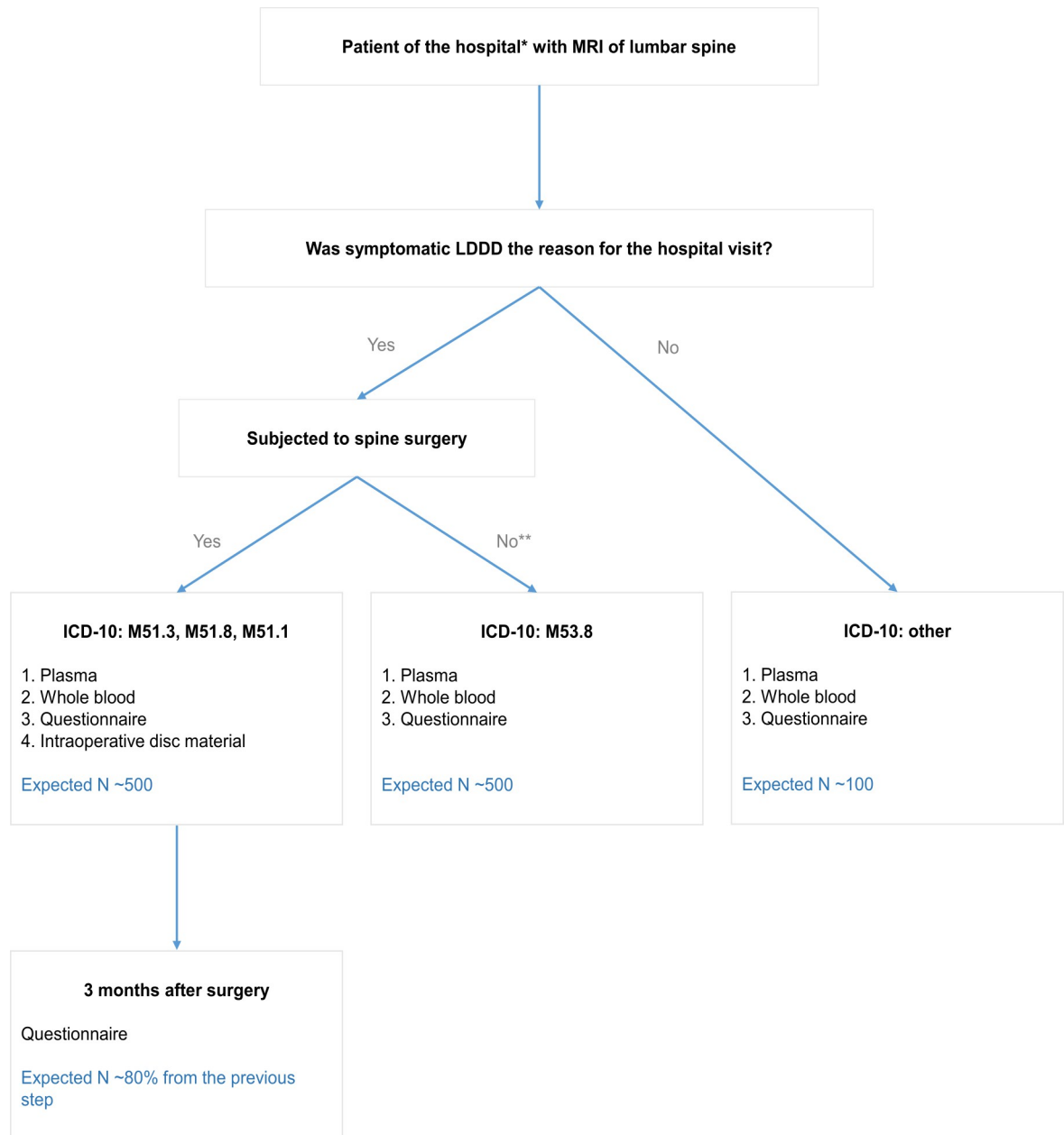


Fig 1. Patient selection. Sample size at various time-points of the project. Expected total sample size is 1,100. *Novosibirsk Research Institute of Traumatology and Orthopedics (NRITO) / Priorov National Medical Research Center of Traumatology and Orthopedics (Priorov CITO). **Conservative treatment.

<https://doi.org/10.1371/journal.pone.0267384.g001>

profiling, including genotyping of blood samples and RNA sequencing of disc tissue. Priorov CITO will store the specimens on at their place.

Patient selection and expected sample size

All patients of the two clinical centers aged over 18 with available MRI scans of the lumbar spine, who will sign an informed consent (see [S1 File](#)) and meet all the eligibility criteria ([Table 2](#)), will be included in the study. All participants will be assessed by the orthopedists and neurologists and grouped into three categories based on their diagnosis: patients with

Table 2. Eligibility criteria.

Inclusion criteria	Exclusion criteria
1. Age over 18; 2. Presence of lumbar MRI scans; 3. Signed informed consent for voluntary participation is provided.	1. Any contraindication or inability to undergo baseline procedures; 2. Prior surgeries at any level of the lumbar spine; 3. Other non-degenerative spinal conditions that may have an impact on subject safety, wellbeing or the intent and conduction of the study; 4. History or presence of HIV, hepatitis B, hepatitis C.

<https://doi.org/10.1371/journal.pone.0267384.t002>

spine surgery, patients with conservative treatment of symptomatic LDDD and other patients with MRI scans of the lumbar spine (see Fig 1). If the reason for the hospital visit is symptomatic LDDD, patients will be divided according to the type of treatment (conservative or surgery) and corresponding ICD-10 codes. If the reason for the hospital visit is other than symptomatic LDDD, the patient will be included in the group of other patients with MRI scans of the lumbar spine in the study, regardless of the ICD-10 diagnosis code. Restrictions on participation in the study for patients with bloodborne pathogens are determined by the requirements of laboratory biosafety.

We plan to collect whole blood and plasma samples from not less than 1,100 patients (see Fig 1). The SNP array based whole-genome genotyping and total plasma protein N-glycosylation profiles will be measured for at least 384 participants. Total RNA sequencing expression profiles of approximately 40 disc specimens will be measured.

MRI imaging, clinical and socio-demographic data

We require each patient from this study to have lumbar spine MRI images. MRI scanning will be performed on a 1.5 Tesla (or higher) tomograph and will include L1-S1 lumbar segments. T1 and T2 weighted images, acquired both in the axial and sagittal planes, will be obtained. Images will be used to assess the disc degeneration grade, presence and type of Modic changes and presence and severity of vertebral endplate defects. Disc degeneration grade will be estimated under the Pfirrmann classification from 1 to 5, with 1 corresponding to a normal disc and 5 corresponding to the most severe degeneration [42]. Modic changes will be evaluated for each endplate (from the lower L1 vertebra to the upper S1 vertebra) on sagittal scans using T1WI and T2WI [43]. Vertebral endplate defects will be ranged from 1 to 6 according to the Rajasekaran classification [44]. Then, total endplate scores (TEPS) will be calculated as the sum of the endplate defect scores of both upper and lower endplates in each L1-S1 spinal segment. Additionally, the height of the intervertebral discs and severity of osteophytes will be estimated based on Jarosz classification [45].

Clinical assessment will include demographic data (sex, age), self-reported ethnicity, height, weight, family status, physical activity, smoking, comorbidity, education level and job type.

All participants will complete the following clinical questionnaires: visual analogue scale (VAS) [46] back and VAS leg estimating the intensity of back and leg pain, correspondingly, and the Oswestry disability index (ODI) [47,48] questionnaire.

Sample collection, transportation and storage

Samples collected in NRITO will be frozen at -40C (whole blood and plasma) or at -80C (disc samples) and then transferred to ICG for long-time storage at -80C. All biosamples from Priorov CITO will be frozen and stored on site at -80C.

Blood and plasma sampling: from each patient, 13 ml of peripheral venous blood will be collected into two BD Vacutainer 4-ml and 9-ml K2 EDTA tubes (or three 4-ml K2 EDTA tubes depending on availability). The tubes will be labeled with a patient's unique code. During

the first 30 min after blood collection, the 4-ml vacutainer with whole blood will be placed into a test tube rack and transported to the freezer to be stored at -40C. The 9-ml blood tube will be used for plasma extraction according to the plasma extraction protocol (see [S2 File](#)). Once the plasma has been extracted, the tubes with plasma will be labeled and placed into the freezer at -40C.

Eventually, all the tubes from NRITO will be transported to ICG in thermoboxes with cooling agents and then placed into freezers at -80C. Samples from Priorov CITO will be stored locally at -80C at once.

Collected plasma samples could be used for cell-free DNA, glycomics, proteomics and metabolomics analyses.

Disc sampling

During the surgical procedure, the resected fragments of intervertebral discs will be placed into sterile falcon 50-ml tubes and stored in the operating room until the end of the intervention (for no longer than 40 min). The falcons will be labeled with unique codes and transported to the laboratory within 5 min where the discs' fragments will be put into liquid nitrogen.

Disc samples from NRITO will be transferred to ICG within thermoses with liquid nitrogen and then will be stored in freezers at -80C. Biospecimens obtained in Priorov CITO will be stored on site at -80C.

Genotyping

DNA will be extracted according to the standard protocol using the Qiagen DNeasy Blood & Tissue Kit. Genotyping for not less than 600,000 SNPs will be performed using whole-genome SNP-arrays of high coverage (tentatively, Illumina Infinium Global Screening Array). Each genotyping batch will be sampled using the block randomization method implemented in the OSAT R package [49] by Bioconductor (functions "setup.sample", "setup.container", "create.experiment.setup" and "get.experiment.setup" with case / control LDDD status, sex, age, availability of the disc tissue, place of birth, BMI, physical activity level, smoking status and date of blood sampling used as characteristics for optimization). The imputation procedure will be carried out using the Haplotype Reference Consortium [48] or a later reference panel.

Total plasma proteins N-glycosylation profiling

The plasma glycans profiling will be performed using the APTS glycan labelling kit (Genos) according to the standard protocol [49]. In short, plasma N-glycans are enzymatically released from proteins by PNGase F, fluorescently labeled with 2-aminobenzamide and cleaned up from the excess of reagents by hydrophilic interaction liquid chromatography solid phase extraction (HILIC-SPE) as described previously [50]. Fluorescently labeled and purified N-glycans are separated by HILIC on a Waters BEH Glycan chromatography column, 150 × 2.1 mm, 1.7 μm BEH particles, installed on an Acquity ultra-high-performance liquid chromatography (UHPLC) instrument (Waters, Milford, MA, USA) consisting of a quaternary solvent manager, a sample manager and a fluorescence detector set with excitation and emission wavelengths of 250 nm and 428 nm, respectively. Following chromatography conditions previously described in detail [50], glycan peaks (GPs)–quantitative measurements of glycan levels–are defined by manual integration of intensity peaks in the chromatograms. The abundance of N-glycans in each chromatographic peak is expressed as percentage area of the corresponding peak. Random allocation of the specimens will be carried out using the block randomization approach with the optimization characteristics similar to the genotyping procedure.

Total RNA profiling

The intraoperative material will be homogenized using TissueLyzer II homogenator (QUAGEN) and total RNA will be extracted and converted to cDNA using a kit for the isolation of total RNA and microRNA from cells and tissues (Biolabmix, Russia) and the M-MuLV-RH First Strand cDNA Synthesis Kit (Biolabmix, Russia), respectively. The amount of extracted RNA and its quality will be estimated using Bioanalyzer 2100 (Agilent).

Total RNA sequencing will be performed using Illumina-HiSeq 4000 under the PE-protocol. A read length of up to 100 bp and a sequencing coverage of 20M are expected. Specimens will be randomly selected for sequencing utilizing the block randomization approach in order to maintain a uniform distribution of optimization characteristics set for genotyping batches. In total, not less than 40 samples will be profiled.

Duration of the project

Specimen collection and processing will last for three years starting in May 2021, unless additional funding has been obtained to expand the biobank. Samples will be stored for fifteen years after collection according to the storage protocol described above.

Statistical data analysis

In this section, we will describe designs of further studies that we plan to conduct based on the omics data generated from the samples of the biobank.

Genome-wide association study of lumbar DD and *in silico* follow-up. We will perform a genome-wide association study (GWAS) using the GCTA software [51] using a mixed model linear regression model and quantitative scales of the assessed disc degeneration status. The GWAS results will be used for replication of genome-wide significant findings from the GWAS of DD conducted in 2013 by Williams et al. [35]. The associations reported by Williams et al. will be considered as replicated if the following two criteria are met: (1) the direction of the effect is the same in both studies; (2) the p-value of association in the replication sample is less than 0.05/4. Further, these two sets of GWAS will be meta-analyzed using the METAL software [52] to obtain the currently largest GWAS on lumbar DD for the European ancestry population. All loci passing the genome-wide significance threshold (p-value < 5e-08) for association will be identified using COJO [53]. The results of the meta-analysis will be used for *in silico* follow-up functional annotation and prioritization of genes involved in lumbar DD development.

Sample size calculation and power estimation. We do not set the upper limit for sample size, although we expect to recruit not less than 1,100 participants and perform a GWAS involving at least 384 patients. Thus, the total sample size of meta-analysis of the GWAS of lumbar DD conducted in 2013 by Williams et al. and in the current study will be about 5,000. This will allow us to detect the genomic loci with an odds ratio of 1.20 or higher with p-value < 5e-8 and 80% power [54].

Functional analysis and data integration. GWAS meta-analysis results will be annotated in order to predict the probable effects of replicated SNPs on gene expression and disease development. Furthermore, we will highlight the molecular pathways, cell and tissue types most likely to be involved in DD pathogenesis using tools like DEPICT [55] and FUMA [56]. This information will be used for gene prioritization alongside the results of the colocalization analysis [57] and Mendelian randomization [58,59]. This methodology will be applied for causal inference between DD and gene expression profiles from different tissues, also including the transcriptomic data from intervertebral discs obtained in this study. A gene network for LDDD regulation will be built and the key regulators will be revealed.

Glycomics data analysis. The association of plasma N-glycans levels with disc degeneration status will be studied. We are planning to analyze the differences in N-glycosylation patterns between patients with and without LDDD utilizing machine learning approaches including regression analysis. The results obtained could be helpful for the development of the dynamic glycan biomarkers for elucidating DD pathogenesis and for the development of a prognostic biomarker.

Transcriptomic data analysis. The patient group with available intraoperative disc material (see Fig 1) could be divided into “cases” or “degenerated disc” (grades 4–5 of disc degeneration according to the Pfirrmann classification) and “controls” or “healthy disc” (grades 1–3 of disc degeneration according to the Pfirrmann classification) [44,60].

This division does not influence the scheme of patient recruiting or sample collection but plays a role in the transcriptomic data analysis. The groups will include following ICD-10 codes: M51.1 “Thoracic, thoracolumbar and lumbosacral intervertebral disc disorders with radiculopathy” (cases); M51.3, M51.8 “Other thoracic, thoracolumbar and lumbosacral intervertebral disc degeneration” (controls).

Transcriptomic data obtained in this study will be used for detection of genes differentially expressed in discs between cases and controls. In short, the gene expression data will be mapped [61] onto the reference genome and the quality of the reads will be recalibrated (“Picard Toolkit.” 2019. Broad Institute, GitHub Repository. <http://broadinstitute.github.io/picard/>; Broad Institute). Subsequently, aiming to reveal the transcriptomic differences in intervertebral discs tissue between cases and controls, we will count the reads and identify differentially expressed genes (DEGs) using the edgeR package [62]. To assess the statistical significance of the results, we will apply a Benjamini-Hochberg correction for multiple testing [63]. DEGs will be filtered by the q -value < 0.05 and divided into up- and down-regulated genes. Finally, the revealed DEGs will be functionally annotated with Gene Ontology terms.

The data will also be used to identify expression quantitative trait loci (eQTLs). We will perform an eQTL analysis of gene expression levels in intervertebral discs using the methodology described in [64]. Resulting regional association summary statistics will be used for gene prioritization.

Data management

All data management and access procedures will be identical in both participating centers.

Each participant included in the study is assigned a unique code. Keys for these codes are saved in the locked storage of the internal hospital Electronic Data Capture (EDC) system, with access provided only for the curators of the study. All clinical data obtained from the patients are kept in the internal hospital EDC system accessible only by authorized researchers who are entering data into it. Information on who entered the data into the clinical database is available for viewing. During the study, internal monitoring will be conducted to maintain the quality of the study in accordance with the GCP principles. Participant and specimen codes are transferred to ICG along with the specimens from both clinical centers in an anonymous way.

Information on the physical location of specimens is kept with limited access only for curators of the study. Genotypes, transcriptomic data and data produced during their processing will be stored on a local server in ICG.

Data and sample access

All clinically relevant data, such as genotypic data, medical history, MRI scans, results of transcriptome and N-glycome profiling could not be deposited in open access repositories due to

legal restrictions (data contains sensitive information). Access to this data and accesses to the biological samples could be granted on a collaborative basis upon special formal request to the data access committee of the Institute of Cytology and Genetics SB RAS. For further information, please contact data access committee (statgenomicslab@gmail.com).

The data generated during the specific analyses in the project will be made available to the scientific community following accepted standards, e.g. controlled access via such platforms as the European Genome-phenome Archive (EGA) (<https://ega-archive.org/>) for personally identifiable data, and public access for less sensitive summary-level data (e.g. GWAS results) via organization's web-site and/or such archiving services as Zenodo (<https://zenodo.org/>).

The study was registered at clinicaltrials.gov, trial registration number NCT04600544.

Discussion

This study will be carried out in two different centers providing wider population coverage and more reliability to the sample storage, as data and specimens will be stored in two places. Participants will be recruited from unique medical centers (NRITO, Priorov CITO), which are the only centers in Russia specializing in LDDD. The distinctive features of our clinical centers are a large constant flow of LDDD patients, the opportunity to conduct spine MRI and a highly-qualified team [65,66]. In the same way, our multidisciplinary research group, who will generate and analyze omics data, has a broad experience in studying back pain (one of the main manifestations of LDDD) [67–69] and statistical and functional data analyses using integrative omics approaches [70,71].

Another notable advantage of the project is the availability of diverse and rare biological material that can be used for multiple profiling. Besides whole blood samples, plasma and the intervertebral disc tissue obtained during spine surgery will be collected. The presence of different biological specimens for the same study participants will allow conducting integrative multi-omics analyses. Thus, not only genomes, but also transcriptomes, glycomes and other '-omes' can be measured using these samples. To our knowledge, this is the first study with a multiple omics profiling of similar biological material. Current studies are limited by a single profiling of one or two tissues: one in [32,33]; two in [28,29,31,37]; or use omics datasets in open access, combining data from different cohorts [40]. The importance of the multi-omics approach is hard to overestimate, as it allows us to look at the disease from different points of view and improve our understanding of the pathology.

One more substantial strength of our study is the assessment of the disc degeneration grade by MRI scans, because this is the most accurate and precise method of DD diagnostics. In LDDD studies based on plain radiography [72] or CT [73], phenotype definition is rather subjective as these approaches provide only indirect evidence of disc degeneration such as disc height loss and osteophytes. Similarly, the use of self-reported questionnaires for LDDD identification [74] is not reliable enough. By contrast, the Pfirrmann grading system of disc degeneration used in the present study is based on MRI scans and estimates the main characteristics of DD by assessing the signal intensity and the height of the intervertebral disc: the more dehydrated the disc is, the more severe degenerative changes are in it [42].

This study has some limitations. First, the expected sample size is modest in comparison with national country-level biobanks or large prospective cohorts. The sample size of the present study is one-fourth as large as the one reported for meta-analysis of LDDD studies by Williams et al. [35]. This limitation is explained by funding, establishing the recruitment period of three years. Nonetheless, the expected number of disc tissue samples (~510) and their transcriptome profiles (~40) is comparable with other omics studies using disc specimens

(proteome [33]: 7 cases / 7 controls; metabolome [39]: 60 cases / 21 controls; transcriptome [28]: 39 annulus fibrosus (AF) and 21 nucleus pulposus (NP) samples; [29]: 24 AF and 24 NP samples).

The next limitation of our study is an imbalance between patients with “healthy disc” and “degenerated disc” in the group of patients subjected to surgery. Obviously, patients with “healthy discs” are subjected to spinal surgery under specific circumstances. However, according to our protocol, we can still expect some controls in this group (in total, the expected number of controls is 20). In the groups not subjected to surgery, we expect a ratio between cases and controls to be closer to that in the general population.

It should be noted that “healthy disc” is a conditional definition, not objective enough; however, the inter-observer agreement is quite high among different groups of researchers, therefore, the Pfirrmann classification is considered to be highly reliable [75]. The main point here is that the existence of adult patients with perfectly healthy lumbar discs is questionable. To find such patients, a large-scale exploratory study is needed. Nevertheless, according to a population-based study, there are no adults with lumbar discs containing no degenerative changes [13].

Results of the work will be of practical importance, as they will replenish the theoretical base necessary for the development of a model of LDDD prediction, for example, polygenic risk score models based on genetic markers. Moreover, the data obtained could potentially make a valuable contribution to the development of new minimally invasive methods for degenerative spine condition treatment. Nowadays, molecular genetics and cell technologies are actively applied in practice to treat degenerative disc disease. In particular, increasing attention is being given to the development of molecular and gene therapy directed towards slowing and reversing disc degeneration [16]. It determines the social significance of this project.

In conclusion, this protocol describes the recruitment and analysis of the RuDDS biobank. The main goal of the project is to collate various medical data and biological specimens for further multidisciplinary investigations of LDDD. The notable advantages of the study are the MRI-based phenotyping and accessibility of a wide range of biological materials (MRI scans; blood, plasma and disc samples with omics data generated based on their analysis). The information gained from the collected biobank data will significantly contribute to the development of novel therapeutic approaches for LDDD treatment and qualitatively improve our understanding of the mechanisms underlying lumbar DD.

Supporting information

S1 File. Informed consent form for patients.

(DOCX)

S2 File. Standard operating procedure for plasma collection.

(DOCX)

Author Contributions

Conceptualization: Olga N. Leonova, Elizaveta E. Elgaeva, Aleksandr V. Krutko, Yurii S. Aulchenko, Yakov A. Tsepilov.

Data curation: Elizaveta E. Elgaeva, Aleksandr V. Krutko, Yakov A. Tsepilov.

Formal analysis: Elizaveta E. Elgaeva.

Investigation: Yurii S. Aulchenko.

Methodology: Olga N. Leonova, Elizaveta E. Elgaeva, Tatiana S. Golubeva, Aleksandr V. Krutko, Yakov A. Tsepilov.

Project administration: Olga N. Leonova, Aleksandr V. Krutko.

Resources: Alexey V. Peleganchuk, Yurii S. Aulchenko, Yakov A. Tsepilov.

Writing – original draft: Olga N. Leonova, Elizaveta E. Elgaeva, Tatiana S. Golubeva, Alexey V. Peleganchuk, Aleksandr V. Krutko, Yakov A. Tsepilov.

Writing – review & editing: Olga N. Leonova, Elizaveta E. Elgaeva, Tatiana S. Golubeva, Alexey V. Peleganchuk, Aleksandr V. Krutko, Yurii S. Aulchenko, Yakov A. Tsepilov.

References

1. Berg AJ, Ahmadje U, Jayanna HH, Tréguët P, Sanville P, Kapoor V. The prevalence of lumbar disc degeneration in symptomatic younger patients: A study of MRI scans. *J Clin Orthop Trauma*. 2020; 11(5):932–6. <https://doi.org/10.1016/j.jcot.2020.07.021> PMID: 32879583
2. Brinjikji W, Luetmer PH, Comstock B, Bresnahan BW, Chen LE, Deyo RA, et al. Systematic Literature Review of Imaging Features of Spinal Degeneration in Asymptomatic Populations. *AJNR Am J Neuroradiol*. 2015; 36(4):811–6. <https://doi.org/10.3174/ajnr.A4173> PMID: 25430861
3. Kos N, Gradisnik L, Velnar T. A Brief Review of the Degenerative Intervertebral Disc Disease. *Med Arch (Sarajevo, Bosnia Herzegovina)*. 2019; 73(6):421–4. <https://doi.org/10.5455/medarh.2019.73.421-424> PMID: 32082013
4. Cheung KMC, Karppinen J, Chan D, Ho DWH, Song YQ, Sham P, et al. Prevalence and pattern of lumbar magnetic resonance imaging changes in a population study of one thousand forty-three individuals. *Spine (Phila Pa 1976)*. 2009; 34(9):934–40.
5. Matsumoto M, Okada E, Toyama Y, Fujiwara H, Momoshima S, Takahata T. Tandem age-related lumbar and cervical intervertebral disc changes in asymptomatic subjects. *Eur Spine J*. 2013; 22(4):708–13. <https://doi.org/10.1007/s00586-012-2500-z> PMID: 22990606
6. Teraguchi M, Yoshimura N, Hashizume H, Muraki S, Yamada H, Oka H, et al. The association of combination of disc degeneration, end plate signal change, and Schmorl node with low back pain in a large population study: The Wakayama Spine Study. *Spine J*. 2015; 15(4):622–8. <https://doi.org/10.1016/j.spinee.2014.11.012> PMID: 25433277
7. Jensen RK, Jensen TS, Koes B, Hartvigsen J. Prevalence of lumbar spinal stenosis in general and clinical populations: a systematic review and meta-analysis. *Eur Spine J*. 2020;(0123456789). <https://doi.org/10.1007/s00586-020-06339-1> PMID: 32095908
8. Brinjikji W, Diehn FE, Jarvik JG, Carr CM, Kallmes DF, Murad MH, et al. MRI findings of disc degeneration are more prevalent in adults with low back pain than in asymptomatic controls: A systematic review and meta-analysis. *Am J Neuroradiol*. 2015; 36(12):2394–9. <https://doi.org/10.3174/ajnr.A4498> PMID: 26359154
9. Oichi T, Taniguchi Y, Oshima Y, Tanaka S, Saito T. Pathomechanism of intervertebral disc degeneration. *JOR SPINE*. 2020; 3(1):1–9. <https://doi.org/10.1002/jsp2.1076> PMID: 32211588
10. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. *Lancet*. 2017; 389(10070):736–47. [https://doi.org/10.1016/S0140-6736\(16\)30970-9](https://doi.org/10.1016/S0140-6736(16)30970-9) PMID: 27745712
11. Middendorp M, Vogl TJ, Kollias K, Kafchitsas K, Khan MF, Maataoui A. Association between intervertebral disc degeneration and the Oswestry Disability Index. *J Back Musculoskelet Rehabil*. 2017; 30(4):819–23. <https://doi.org/10.3233/BMR-150516> PMID: 28372314
12. Teraguchi M, Yoshimura N, Hashizume H, Muraki S, Yamada H, Minamide A, et al. Prevalence and distribution of intervertebral disc degeneration over the entire spine in a population-based cohort: The Wakayama Spine Study. *Osteoarthritis Cartil*. 2014; 22(1):104–10. <https://doi.org/10.1016/j.joca.2013.10.019> PMID: 24239943
13. Oh CH, Yoon SH. Whole Spine Disc Degeneration Survey according to the Ages and Sex Using Pfirrmann Disc Degeneration Grades. *Korean J Spine*. 2018; 14(4):148–54.
14. Sääksjärvi S, Kerttula L, Luoma K, Paaanen H, Waris E. Disc Degeneration of Young Low Back Pain Patients: A Prospective 30-year Follow-up MRI Study. *Spine (Phila Pa 1976)*. 2020; 45(19):1341–7. <https://doi.org/10.1097/BRS.0000000000003548> PMID: 32453239

15. Teraguchi M, Yoshimura N, Hashizume H, Yamada H, Oka H, Minamide A, et al. Progression, incidence, and risk factors for intervertebral disc degeneration in a longitudinal population-based cohort: the Wakayama Spine Study [Internet]. Vol. 25, *Osteoarthritis and Cartilage*. Elsevier Ltd; 2017. 1122–1131 p. <https://doi.org/10.1016/j.joca.2017.01.001> PMID: 28089899
16. Kadow T, Sowa G, Vo N, Kang JD. Molecular Basis of Intervertebral Disc Degeneration and Herniations: What Are the Important Translational Questions? *Clin Orthop Relat Res*. 2015; 473(6):1903–12. <https://doi.org/10.1007/s11999-014-3774-8> PMID: 25024024
17. Williams FMK, Sambrook PN. Neck and back pain and intervertebral disc degeneration: Role of occupational factors. *Best Pract Res Clin Rheumatol*. 2011; 25(1):69–79. <https://doi.org/10.1016/j.berh.2011.01.007> PMID: 21663851
18. Xiao L, Ni C, Shi J, Wang Z, Wang S, Zhang J, et al. Analysis of correlation between vertebral endplate change and lumbar disc degeneration. *Med Sci Monit*. 2017; 23:4932–8. <https://doi.org/10.12659/msm.904315> PMID: 29032381
19. Jarman JP, Arpinar VE, Baruah D, Klein AP, Maiman DJ, Tugan Muftuler L. Intervertebral disc height loss demonstrates the threshold of major pathological changes during degeneration. *Eur Spine J*. 2015; 24(9):1944–50. <https://doi.org/10.1007/s00586-014-3564-8> PMID: 25212450
20. Desmoulin GT, Pradhan V, Milner TE. Mechanical Aspects of Intervertebral Disc Injury and Implications on Biomechanics. *Spine (Phila Pa 1976)*. 2020; 45(8):E457–64. <https://doi.org/10.1097/BRS.0000000000003291> PMID: 31651681
21. Otluoğlu GD, Konya D, Toktas ZO. The influence of mechanic factors idisc degeneration disease as a determinant for surgical indication. *Neurospine*. 2020; 17(1):215–20. <https://doi.org/10.14245/ns.2040044.022> PMID: 32252171
22. Dragsbæk L, Kjaer P, Hancock M, Jensen TS. An exploratory study of different definitions and thresholds for lumbar disc degeneration assessed by MRI and their associations with low back pain using data from a cohort study of a general population. *BMC Musculoskelet Disord*. 2020; 21(1):1–11. <https://doi.org/10.1186/s12891-020-03268-4> PMID: 32303267
23. Navone S, Marfia G, Giannoni A, Beretta M, Guarnaccia L, Gualtierotti R, et al. Inflammatory mediators and signalling pathways controlling intervertebral disc degeneration. *Histol Histopathol*. 2017; 32(6):523–42. <https://doi.org/10.14670/HH-11-846> PMID: 27848245
24. Feng C, Liu H, Yang M, Zhang Y, Huang B, Zhou Y. Disc cell senescence in intervertebral disc degeneration: Causes and molecular pathways. *Cell Cycle*. 2016; 15(13):1674–84. <https://doi.org/10.1080/15384101.2016.1152433> PMID: 27192096
25. Chen WK, Yu XH, Yang W, Wang C, He WS, Yan YG, et al. lncRNAs: novel players in intervertebral disc degeneration and osteoarthritis. *Cell Prolif*. 2017; 50(1):1–12. <https://doi.org/10.1111/cpr.12313> PMID: 27859817
26. Munir S, Rade M, Määtä JH, Freidin MB, Williams FMK. Intervertebral Disc Biology: Genetic Basis of Disc Degeneration. *Curr Mol Biol Reports*. 2018; 4(4):143–50. <https://doi.org/10.1007/s40610-018-0101-2> PMID: 30464887
27. Kawaguchi Y. Genetic background of degenerative disc disease in the lumbar spine. *Spine Surg Relat Res*. 2018; 2(2):98–112. <https://doi.org/10.22603/ssrr.2017-0007> PMID: 31440655
28. Riestler SM, Lin Y, Wang W, Cong L, Ali AM, Krauss W, et al. RNA sequencing identifies gene regulatory networks controlling extracellular matrix synthesis in intervertebral disk tissues. *J Orthop Res*. 2019; 36(5):1356–69.
29. Kazezian Z, Gawri R, Haglund L, Ouellet J, Mwale F, Tarrant F, et al. Gene Expression Profiling Identifies Interferon Signalling Molecules and IGFBP3 in Human Degenerative Annulus Fibrosus. *Sci Rep*. 2015; 5(September):1–13. <https://doi.org/10.1038/srep15662> PMID: 26489762
30. Sherafatian M, Abdollahpour HR, Ghaffarpasand F, Yaghmaei S, Azadegan M, Heidari M. MicroRNA Expression Profiles, Target Genes, and Pathways in Intervertebral Disk Degeneration: A Meta-Analysis of 3 Microarray Studies. *World Neurosurg*. 2019; 126:389–97. <https://doi.org/10.1016/j.wneu.2019.03.120> PMID: 30904808
31. Sarath Babu N, Krishnan S, Brahmendra Swamy C V., Venkata Subbaiah GP, Gurava Reddy A V., Idris MM. Quantitative proteomic analysis of normal and degenerated human intervertebral disc. *Spine J*. 2016; 16(8):989–1000. <https://doi.org/10.1016/j.spinee.2016.03.051> PMID: 27125197
32. Ye D, Liang W, Dai L, Zhou L, Yao Y, Zhong X, et al. Comparative and quantitative proteomic analysis of normal and degenerated human annulus fibrosus cells. *Clin Exp Pharmacol Physiol*. 2015; 42(5):530–6. <https://doi.org/10.1111/1440-1681.12386> PMID: 25739836
33. Yee A, Lam MPY, Tam V, Chan WCW, Chu IK, Cheah KSE, et al. Fibrotic-like changes in degenerate human intervertebral discs revealed by quantitative proteomic analysis. *Osteoarthr Cartil*. 2016; 24(3):503–13. <https://doi.org/10.1016/j.joca.2015.09.020> PMID: 26463451

34. Eskola PJ, Lemmelä S, Kjaer P, Solovieva S, Männikkö M, Tommerup N, et al. Genetic Association Studies in Lumbar Disc Degeneration: A Systematic Review. *PLoS One*. 2012; 7(11). <https://doi.org/10.1371/journal.pone.0049995> PMID: 23185509
35. Williams FMK, Bansal AT, Van Meurs JB, Bell JT, Meulenbelt I, Suri P, et al. Novel genetic variants associated with lumbar disc degeneration in northern Europeans: A meta-analysis of 4600 subjects. *Ann Rheum Dis*. 2013; 72(7):1141–8. <https://doi.org/10.1136/annrheumdis-2012-201551> PMID: 22993228
36. Song YQ, Karasugi T, Cheung KMC, Chiba K, Ho DWH, Miyake A, et al. Lumbar disc degeneration is linked to a carbohydrate sulfotransferase 3 variant. *J Clin Invest*. 2013; 123(11):4909–17. <https://doi.org/10.1172/JCI69277> PMID: 24216480
37. Zhang YG, Guo X, Sun Z, Jia G, Xu P, Wang S. Gene expression profiles of disc tissues and peripheral blood mononuclear cells from patients with degenerative discs. *J Bone Miner Metab*. 2010; 28(2):209–19. <https://doi.org/10.1007/s00774-009-0120-4> PMID: 19798547
38. Wang Y, Dai G, Wang L, Shang F, Jiang L, Li S, et al. Identification of Key Genes Potentially Related to Intervertebral Disk Degeneration by Microarray Analysis. *Genet Test Mol Biomarkers*. 2019; 23(9):610–7. <https://doi.org/10.1089/gtmb.2019.0043> PMID: 31368816
39. Pacholczyk-Sienicka B, Radek M, Radek A, Jankowski S. Characterization of metabolites determined by means of 1H HR MAS NMR in intervertebral disc degeneration. *Magn Reson Mater Physics, Biol Med*. 2015; 28(2):173–83. <https://doi.org/10.1007/s10334-014-0457-0> PMID: 25108703
40. Xu C, Luo S, Wei L, Wu H, Gu W, Zhou W, et al. Integrated transcriptome and proteome analyses identify novel regulatory network of nucleus pulposus cells in intervertebral disc degeneration. *BMC Med Genomics*. 2021; 14(1):1–12. <https://doi.org/10.1186/s12920-020-00855-1> PMID: 33397358
41. Battie MC, Levalahti E, Videman T, Burton K, Kaprio J. Heritability of lumbar flexibility and the role of disc degeneration and body weight. *J Appl Physiol* (1985). 2008; 104(2):379–85. <https://doi.org/10.1152/japplphysiol.01009.2007> PMID: 18048587
42. Pfirrmann CWA, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic Resonance Classification of Lumbar Intervertebral Disc Degeneration. *Spine (Phila Pa 1976)*. 2001; 26(17):1873–8. <https://doi.org/10.1097/00007632-200109010-00011> PMID: 11568697
43. Chen Y, Bao J, Yan Q, Wu C, Yang H, Zou J. Distribution of Modic changes in patients with low back pain and its related factors. *Eur J Med Res*. 2019; 24(1):1–9. <https://doi.org/10.1186/s40001-018-0359-0> PMID: 30606260
44. Rajasekaran S, Venkatadass K, Naresh Babu J, Ganesh K, Shetty AP. Pharmacological enhancement of disc diffusion and differentiation of healthy, ageing and degenerated discs: Results from in-vivo serial post-contrast MRI studies in 365 human lumbar discs. *Eur Spine J*. 2008; 17(5):626–43. <https://doi.org/10.1007/s00586-008-0645-6> PMID: 18357472
45. Sambrook PN, MacGregor AJ, Spector TD. Genetic influences on cervical and lumbar disc degeneration: A magnetic resonance imaging study in twins. *Arthritis Rheum*. 1999; 42(2):366–72. [https://doi.org/10.1002/1529-0131\(199902\)42:2<366::AID-ANR20>3.0.CO;2-6](https://doi.org/10.1002/1529-0131(199902)42:2<366::AID-ANR20>3.0.CO;2-6) PMID: 10025932
46. Collins S, Moore R, McQuay H. The visual analogue pain intensity scale: what is moderate pain in millimetres? *Pain*. 1997; 72(1–2):95–7. [https://doi.org/10.1016/s0304-3959\(97\)00005-5](https://doi.org/10.1016/s0304-3959(97)00005-5) PMID: 9272792
47. Fairbank J.C. Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)*. 2000; Nov 15; 25(22):2940–52. <https://doi.org/10.1097/00007632-200011150-00017> PMID: 11074683
48. Cherepanov EA. Russian version of the oswestry disability index: cross-cultural adaptation and validity. *Hir pozvonočnika (Spine Surgery)*. 2009; 0(3):093–8.
49. Yan L, Ma C, Wang D, Hu Q, Qin M, Conroy JM, et al. OSAT: A tool for sample-to-batch allocations in genomics experiments. *BMC Genomics*. 2012; 13(1):1. <https://doi.org/10.1186/1471-2164-13-689> PMID: 23228338
50. Trbojević Akmačić I, Ugrina I, Štambuk J, Gudelj I, Vučković F, Lauc G, et al. High-throughput glycomics: Optimization of sample preparation. *Biochem*. 2015; 80(7):934–42. <https://doi.org/10.1134/S0006297915070123> PMID: 26542006
51. Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, et al. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet*. 2019 Dec 25; 51(12):1749–55. <https://doi.org/10.1038/s41588-019-0530-8> PMID: 31768069
52. Willer CJ, Li Y, Abecasis GR, Overall P. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26(17):2190–1. <https://doi.org/10.1093/bioinformatics/btq340> PMID: 20616382
53. Yang J, Ferreira T, Morris A, Medland S, McCarthy M, Hirschhorn J, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012; 44(4):369–75. <https://doi.org/10.1038/ng.2213> PMID: 22426310

54. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. *Nat Rev Genet.* 2014 May; 15(5):335–46. <https://doi.org/10.1038/nrg3706> PMID: 24739678.
55. Pers TH, Karjalainen JM, Chan Y, Westra HJ, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun.* 2015. <https://doi.org/10.1038/ncomms6890> PMID: 25597830
56. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017 Dec; 8(1):1826. <https://doi.org/10.1038/s41467-017-01261-5> PMID: 29184056
57. Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charlotiaux B, et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat Commun.* 2018; 9(1):1–18. <https://doi.org/10.1038/s41467-017-02088-w> PMID: 29317637
58. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet.* 2020; 52(10):1122–31. <https://doi.org/10.1038/s41588-020-0682-6> PMID: 32895551
59. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016; 48(5):481–7. <https://doi.org/10.1038/ng.3538> PMID: 27019110
60. Ikuno A, Akeda K, Takebayashi SI, Shimaoka M, Okumura K, Sudo A. Genome-wide analysis of DNA methylation profile identifies differentially methylated loci associated with human intervertebral disc degeneration. *PLoS One.* 2019; 14(9):1–20. <https://doi.org/10.1371/journal.pone.0222188> PMID: 31513634
61. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
62. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2009; 26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616> PMID: 19910308
63. Benjamini, Yoav; Hochberg Y. Controlling the False Discovery Rate—a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B.* 1995; 57(1):289–300. <https://doi.org/10.2307/2346101>
64. Parisien M, Khoury S, Chabot-Doré AJ, Sotocinal SG, Slade GD, Smith SB, et al. Effect of Human Genetic Variability on Gene Expression in Dorsal Root Ganglia and Association with Pain Phenotypes. *Cell Rep.* 2017. <https://doi.org/10.1016/j.celrep.2017.05.018> PMID: 28564610
65. Krutko A V., Sanginov AJ, Baykov ES. Predictors of Treatment Success Following Limited Discectomy With Annular Closure for Lumbar Disc Herniation. *Int J Spine Surg.* 2020; 14(1):38–45. <https://doi.org/10.14444/7005> PMID: 32128301
66. Leonova O, Baykov E, Sanginov A, Krutko A. Cervical Disc Degeneration and Vertebral Endplate Defects After the Fused Operation. *Spine (Phila Pa 1976).* 2021 Sep 15; 46(18):1234–40. <https://doi.org/10.1097/BRS.0000000000004007> PMID: 33595261
67. Freidin MB, Tsepilov YA, Palmer M, Karssen LC, Suri P, Aulchenko YS, et al. Insight into the genetic architecture of back pain and its risk factors from a study of 509,000 individuals. Loos RJF, editor. *Pain.* 2019 Jun 6; 160(6):1361–73. <https://doi.org/10.1097/j.pain.0000000000001514> PMID: 30747904
68. Elgaeva EE, Tsepilov Y, Freidin MB, Williams FMK, Aulchenko Y, Suri P. ISSLS Prize in Clinical Science 2020. Examining causal effects of body mass index on back pain: a Mendelian randomization study. *Eur Spine J.* 2020 Apr 3; 29(4):686–91. <https://doi.org/10.1007/s00586-019-06224-6> PMID: 31797139
69. Freidin MB, Tsepilov YA, Stanaway IB, Meng W, Hayward C, Smith BH, et al. Sex- and age-specific genetic analysis of chronic back pain. *Pain.* 2021 Sep 30; 162(4):1176–87. <https://doi.org/10.1097/j.pain.0000000000002100> PMID: 33021770
70. Tsepilov YA, Sharapov SZ, Zaytseva OO, Krumsek J, Prehn C, Adamski J, et al. A network-based conditional genetic association analysis of the human metabolome. *Gigascience.* 2018; 7(12):1–11. <https://doi.org/10.1093/gigascience/giy137> PMID: 30496450
71. Sharapov SZ, Tsepilov YA, Klaric L, Mangino M, Thareja G, Shadrina AS, et al. Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Hum Mol Genet.* 2019; 28(12):2062–77. <https://doi.org/10.1093/hmg/ddz054> PMID: 31163085
72. Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio Le Graverand MP, Rosendaal FR, et al. Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: The GARP study. *Ann Rheum Dis.* 2005; 64(3):438–43. <https://doi.org/10.1136/ard.2004.024661> PMID: 15458958

73. Yau MS, Demissie S, Zhou Y, Anderson DE, Amanda L. Heritability of Thoracic Spine Curvature and Genetic Correlations With Other Spine Traits: The Framingham Study. *J Bone Min Res.* 2016; 31(12):2077–84. <https://doi.org/10.1002/jbmr.2925> PMID: 27455046
74. Allegri M, De Gregori M, Minella CE, Klersy C, Wang W, Sim M, et al. Omics' biomarkers associated with chronic low back pain: Protocol of a retrospective longitudinal study. *BMJ Open.* 2016; 6(10):1–8. <https://doi.org/10.1136/bmjopen-2016-012070> PMID: 27798002
75. Urrutia J, Besa P, Campos M, Cikutovic P, Cabezon M, Molina M, et al. The Pfirrmann classification of lumbar intervertebral disc degeneration: an independent inter- and intra-observer agreement assessment. *Eur Spine J.* 2016; 25(9):2728–33. <https://doi.org/10.1007/s00586-016-4438-z> PMID: 26879918