

1 **Protocol for disease-oriented Russian disc degeneration study (RuDDS)**
2 **biobank facilitating functional omics studies of lumbar disc degeneration**

3 Olga N. Leonova¹, Elizaveta E. Elgaeva^{2,3}, Tatiana S. Golubeva³, Alexey V. Peleganchuk¹, Aleksandr V.
4 Krutko⁴, Yurii S. Aulchenko^{2,3^}, Yakov A. Tsepilov^{2,3^}
5

6 1 Novosibirsk Research Institute of Traumatology and Orthopedics, Novosibirsk, Russia

7 2 Novosibirsk State University, Novosibirsk, Russia

8 3 Institute of Cytology and Genetics, Novosibirsk, Russia

9 4 Priorov National Medical Research Center of Traumatology and Orthopedics, Moscow, Russia
10

11 **Corresponding author:** Olga Leonova, MD, PhD, Neurosurgery Department, Novosibirsk Research
12 Institute of Traumatology and Orthopaedics (NRITO) n.a.Ya.L.Tsivyan, Frunze str, 17, Novosibirsk,
13 Russia, onleonova@gmail.com

14 **Keywords:** lumbar disc degeneration; biobank; lumbar spine; functional genomics; transcriptomics;
15 glycomics.

16 **Authors and Institutions:**

17 Olga Leonova, MD, PhD, Neurosurgery Department, Novosibirsk Research Institute of Traumatology
18 and Orthopaedics (NRITO) n.a.Ya.L.Tsivyan, Frunze str, 17, Novosibirsk, Russia, ORCID 0000-0002-
19 9916-3947;

20 Elizaveta E. Elgaeva, MSc, Novosibirsk State University, Novosibirsk, Russia, Institute of Cytology and
21 Genetics, Novosibirsk, Russia, ORCID 0000-0002-2484-553X

22 Tatiana S. Golubeva, PhD, Institute of Cytology and Genetics, Novosibirsk, Russia, ORCID 0000-0002-
23 4400-0665

24 Alexey V. Peleganchuk, MD, PhD, Neurosurgery Department, Novosibirsk Research Institute of
25 Traumatology and Orthopaedics (NRITO) n.a.Ya.L.Tsivyan, Frunze str, 17, Novosibirsk, Russia, ORCID
26 0000-0002-4588-428X

27 Aleksandr Krutko, MD, PhD, Neurosurgery Department, Priorov National Medical Research Center of
28 Traumatology and Orthopedics, Priorova str, 10, Moscow, Russia, ORCID 0000-0002-2570-3066

29 Yurii S. Aulchenko[^], Novosibirsk State University, Novosibirsk, Russia, Institute of Cytology and
30 Genetics, Novosibirsk, Russia

31 Yakov A. Tsepilov[^], Novosibirsk State University, Novosibirsk, Russia, Institute of Cytology and
32 Genetics, Novosibirsk, Russia

33 [^]contributed equally to this work

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

34 **ABSTRACT**

35 **Introduction**

36 Lumbar intervertebral disc degeneration (DD) disease is one of the main risk factors for low back pain.
37 The social and economic importance of low back pain is very high: back pain is among the leading
38 causes of absenteeism and the cost of treating back pain exceeds the cost of treatment of many other
39 serious diseases (cancer, in particular); however, therapy does not always provide the desired result.
40 Despite the variability of biological studies of lumbar DD, it is still not fully understood, partially due to
41 the fact that there are only few studies using systematic and integrative approaches. Hence, more
42 integrative omics studies are needed to link all pieces of knowledge together, build a complete picture
43 of biology of lumbar DD and obtain a deeper understanding of the processes underlying this pathology.

44 **Methods and analysis**

45 This disease-oriented biobank to study lumbar disc degeneration will be recruited from two clinical
46 centers. A total of 1100 participants with available lumbar MRI will be enrolled during the three-year
47 period. General information about a patient, medical history, lumbar MRI parameters and biological
48 material (whole blood and plasma) will be collected in the centers at baseline. Then, from those
49 patients, who will undergo a spine surgery during the treatment, disc tissue samples will be gained.
50 Eventually, postoperative clinical data will be collected from operated patients during the follow-up.

51 **Ethics and dissemination**

52 The study will be performed according to the Helsinki Declaration. The study protocol was approved
53 by the Local Ethical Committee of NRITO (№034/20 dated 02 Oct 2020) and by the Local Ethical
54 Committee of Priorov CITO (№1/21 dated 25 Feb 2021). Preliminary and final results will be presented
55 in peer-reviewed journals, at national and international congresses.

56 **Trial registration number** NCT04600544

57

58 **ARTICLE SUMMARY**

59 **STRENGTHS AND LIMITATIONS OF THIS STUDY**

- 60 • Two-center and multiple-discipline study: the study includes two centers (Moscow in the European
61 and Novosibirsk in the Asian part of Russia), with research teams specializing in following fields:
62 (1) clinical aspects of lumbar disc degeneration disease, (2) biology and genetics of pain, (3)
63 generation of omics data and (4) multi-omics data analysis
- 64 • Collection of variable biological material: whole blood, plasma and, for part of the participants,
65 intraoperative material of the lumbar disc (with different DD status) will be available for omics
66 studies
- 67 • Objective diagnosis: lumbar DD status is confirmed by MRI, not self-reported diagnosis
- 68 • Sample size is limited compared to nationwide biobanks
- 69 • Patients with severe lumbar DD status will be more prevalent in the biobank than these with mild
70 or no lumbar DD compared to the general population

71

72 INTRODUCTION

73 Intervertebral disc degeneration is a normal aging process, but in some cases it causes lumbar disc
74 degeneration disease (LDDD) [1,2]. Intervertebral disc degeneration (DD) often begins earlier than
75 degenerative changes in the ligaments, cartilages and other tissues of spinal segment [3]. DD is a major
76 contributor to low back pain [4], but also DD can proceed without back pain [5–8]. However, LDDD is
77 associated with lower health-related quality of life [9] and is the leading cause of population
78 absenteeism and disability [10,11].

79 The prevalence of lumbar DD in the general population is extremely high: up to 50% in people of 30-
80 39 years old [2,12]. There is not a single adult without even a small degree of degeneration in the
81 intervertebral discs [13].

82 List of well-known risk factors of LDDD includes female gender, advanced age [14], obesity, smoking
83 [15], absence of or extreme physical activity [16] and genetic risk factors [17]. However, even in the
84 absence of obvious risk factors, there are cases of LDDD among young people [1,14], as well as
85 progressive severe LDDD requiring a number of surgical interventions.

86 Biology of intervertebral DD is actively studied from different points of view. Besides various studies
87 of degenerative disc morphology [18–22] and molecular biology [16,23–25] there is an increasing
88 number of genetic [26,27], transcriptomic [28–30] and proteomic [31–33] studies of intervertebral DD.
89 Thus, nowadays based on candidate-gene [34] and genome-wide associated studies (GWAS) [35,36]
90 over 160 genes were considered as potentially involved in intervertebral disc degeneration, although
91 less than ten of them provide strong evidence for the association [34]. More than 500 genes were
92 highlighted as differentially expressed in degenerative disc tissue in comparison with healthy
93 intervertebral discs [37,38]. Moreover, proteomic [31–33] and metabolomics [39] changes were also
94 detected in degenerative intervertebral discs. However, despite a large number of studies of LDDD, it
95 is still not fully understood, partially due to the fact that there are only few studies using systematic
96 and integrative approaches [40]. Hence, more integrative omics studies are needed to link all pieces
97 of knowledge together, build a complete picture of biology of lumbar DD and obtain a deeper
98 understanding of the processes underlying this pathology.

99 Research aim and objectives

100 The main aim of this study is to establish disease-oriented biobank to facilitate research in biology of
101 the lumbar disc degeneration. Diverse biological samples (whole blood, plasma, disc tissue) along with
102 MRI imaging, clinical, socio-demographic and various omics data (e.g. genomic and transcriptomic) will
103 be available for researchers and clinicians for a variety of further multi-omics studies. It will lay the
104 groundwork for the development of early diagnostics of DDD and its personalized treatment.

105

106 METHODS AND ANALYSIS

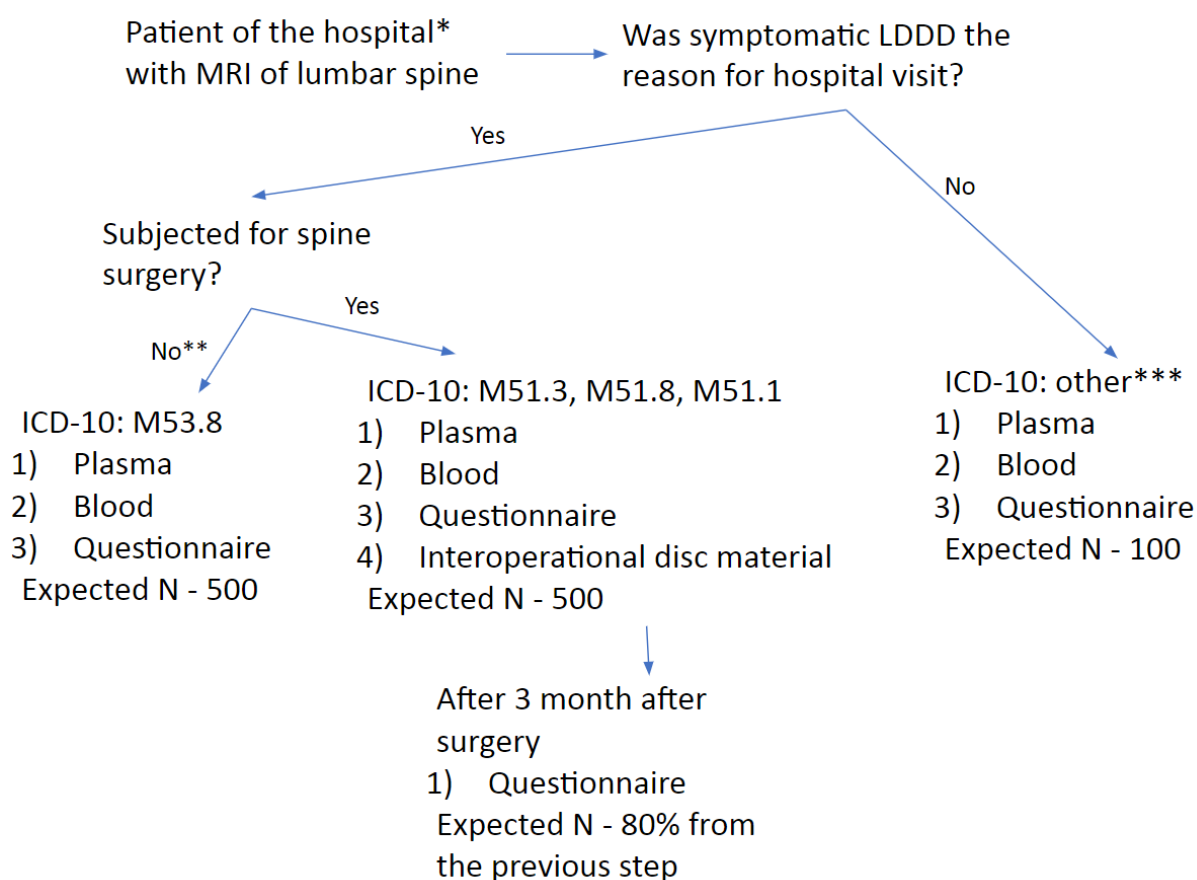
107 Patient and Public Involvement

108 Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination
109 plans of this research.

110 Study design and settings

111 This disease-oriented biobank to study lumbar disc degeneration will be recruited from two centers:
 112 Priorov National Medical Research Center of Traumatology and Orthopedics (Priorov CITO), Moscow,
 113 Russia and Novosibirsk Research Institute of Traumatology and Orthopedics (NRITO), Novosibirsk,
 114 Russia. It will include only patients with available MRI of the lumbar spine, who signed the informed
 115 consent. Clinical data and specimen collection will be performed in three visits as described in Table 1.
 116 At the first step general information about a patient, his medical history and MRI scans of lumbar spine
 117 will be obtained. Patient's biological material (whole blood and plasma) will be also sampled in the
 118 centers at baseline. Then, from those patients, who will undergo a spine surgery during the treatment,
 119 disc tissue samples will be gained. Eventually, postoperative clinical data will be collected from
 120 operated patients during the follow-up. During three years, we expect to collect information and
 121 biological material from a total of 1100. The full scheme of the study is presented at Figure 1.

122



123
 124 **Figure 1. Patient selection.** Sample size is depicted for 3 years timeline of project duration. Expected
 125 total sample size is 1100. *Novosibirsk Research Institute of Traumatology and Orthopedics (NRITO) /
 126 Priorov National Medical Research Center of Traumatology and Orthopedics (Priorov CITO).
 127 **Conservative treatment. ***ICD-10: I71 Aortic aneurysm and dissection, K81 Cholecystitis, K85
 128 Pancreatitis, K25, K26 Penetrating ulcer, M06.4 Inflammatory polyarthropathy, N00-N99 Diseases of
 129 the genitourinary system

130

131 **Table 1. Scheduled procedures for clinical data and samples 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	X		
Medical History	X		
Anamnesis Vitae	X		
Demographics	X		
Blood samples	X		
MRI	X		
ODI ³	X		X
VAS ⁴	X		X
Indications for surgery		X	
Surgical Procedure		X	
Intervertebral disc samples		X	
Adverse events / Serious adverse events	X	X	X

132 ¹only for patients subjected for surgery, ²optional, ³Oswestry Disability Index (ODI) scale, ⁴Visual
133 analogue scale (VAS) for back and leg pain.

134

135 All biological samples from NRITO will be then transported to the Institute of Cytology and Genetics
136 (ICG), Novosibirsk, Russia for long-term storage, further processing and omics profiling, including
137 genotyping of blood samples and RNA-Sequencing of disc tissue. Priorov CITO will store the specimens
138 on its own.

139 Patient selection

140 Participants aged over 18 with available MRI scans of lumbar spine, who will also sign an informed
141 consent and meet all the eligibility criteria (Table 2), will be included in the study. All participants will
142 be grouped by three groups based on their diagnosis and type of treatment (conservative or surgical).
143 All details and ICD-10 codes required for inclusion are presented in Figure 1. Restrictions on
144 participation in the study for patients with bloodborne pathogens is determined by requirements of
145 laboratory biosafety.

146 We plan to collect whole blood and plasma samples from two centers no less than 1100 patients (see
147 Figure 1). The SNP array based whole-genome genotyping and total plasma protein N-glycosylation
148 profiles will be measured for at least 384 participants. Expression profiles of approximately 40 disc
149 specimens will be measured.

150

151 **Table 2. Eligibility criteria.**

Inclusion criteria	Exclusion criteria
1. Age over 18; 2. Presence of lumbar MRI scans;	1. Any contraindication or inability to undergo baseline procedures;

3. Signed informed consent for voluntary participation is provided.	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.
---	---

152

153 **MRI imaging, clinical and socio-demographic data**

154 We require each patient from this study to have lumbar spine MRI images. MRI scanning will be
155 performed on a 1.5 Tesla (or higher) tomograph and will include L1-S1 lumbar segments. T1 and T2
156 weighted images, acquired both in the axial and sagittal planes, will be obtained. Images will be used
157 to assess the disc degeneration grade, presence and type of Modic changes and presence and severity
158 of vertebral endplate defects. Disc degeneration grade will be estimated under the Pfirrmann
159 classification from 1 to 5, with 1 corresponding to normal disc and 5 corresponding to the most severe
160 degeneration [41]. Modic changes will be evaluated for each endplate (from the lower L1 vertebra to
161 the upper S1 vertebra) on sagittal scans using T1WI and T2WI [42]. Vertebral endplate defects will be
162 ranged from 1 to 6 according to the Rajasekaran classification [43]. Then, total endplate scores (TEPS)
163 will be calculated as the sum of the endplate defect scores of both upper and lower endplates in each
164 L1-S1 spinal segment. Additionally, height of the intervertebral discs and severity of osteophytes will
165 be estimated based on Jarosz classification [44].

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

168 All participants will also complete the following clinical questionnaires: visual analogue scale (VAS) [45]
169 back and VAS leg estimating the intensity of back and leg pain correspondingly and the Oswestry
170 disability index (ODI) [46,47] questionnaire.

171 **Sample collection, transporting and storing**

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

175 Blood and plasma sampling: from each patient 13 ml of peripheral venous blood will be collected into
176 two BD Vacutainer 4 ml and 9 ml K2 EDTA tubes (or three 4 ml K2 EDTA tubes depending on
177 availability). Tubes will be labeled with a unique patient's code. During the first 30 min after blood
178 collection, the 4 ml vacutainer with whole blood will be placed into the test tube rack and transported
179 to the freezer to be stored at -40C. The other 9 ml blood tube will be used for plasma extraction
180 according to the plasma extraction protocol (see Supplementary Materials 1). When the plasma will
181 have been extracted, the tubes with plasma will also be labeled and placed into the freezer at -40C.

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

184 Plasma samples could be potentially used for cell-free DNA, glycomics, proteomics and metabolomics
185 analyses.

186 Disc sampling: during the surgical procedure the resected fragments of intervertebral discs will be
187 placed into sterile falcon 50 ml tubes that will be stored in the operating room till the end of

188 intervention (no longer than 40 min). The falcons will be labeled with unique codes and will be
189 transported to the laboratory in 5 min where the discs' fragments will be put into liquid nitrogen.

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

192 **Genotyping**

193 DNA will be extracted according to the standard protocol with the help of Qiagen DNeasy Blood &
194 Tissue Kit. Genotyping by at least 600,000 SNPs will be performed using whole-genome SNP-arrays of
195 high coverage (tentatively, Illumina Infinium Global Screening Array will be used). Imputation
196 procedure will be carried out using Haplotype Reference Consortium [48] or later reference panel.

197 **Total plasma proteins N-glycosylation profiling**

198 The plasma glycans profiling will be performed using APTS glycan labelling kit (Genos) according to
199 standard protocol [49].

200 **Total RNA profiling**

201 The intraoperative material will be homogenized using TissueLyzer II homogenator (QUAGEN) and
202 total RNA will be extracted and converted to cDNA using Kit for the isolation of total RNA and microRNA
203 from cells and tissues (Biolabmix, Russia) and M-MuLV–RH First Strand cDNA Synthesis Kit (Biolabmix,
204 Russia) respectively. The amount of extracted RNA and its quality will be estimated using Bioanalyzer
205 2100 (Agilent).

206 Total RNA sequencing will be performed using Illumina-HiSeq 4000 under PE-protocol. The read length
207 up to 100 bp and sequencing coverage of 20M are expected. In total not less than 20 samples will be
208 profiled.

209 **Duration of the project**

210 Specimens collection and processing will be lasting for three years, unless the additional funding is
211 obtained to expand the biobank. Samples will be stored for fifteen years after collection according to
212 the storage protocol described above.

213 **Statistical data analysis**

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

216 **Genome-wide association study of LDD and in-silico follow-up**

217 We will perform genome-wide association study (GWAS) using quantitative scales of assessed disc
218 degeneration status. The GWAS results will be used for replication of genome-wide significant findings
219 from GWAS of LDD conducted in 2013 by Williams et al. [35]. Further, these two sets of GWAS will be
220 meta-analysed to obtain the largest to date GWAS on LDD for European ancestry population. The
221 results of meta-analysis will be used for in-silico follow-up functional annotation and prioritization of
222 genes involved in LDD development.

223 **Functional analysis and data integration**

224 GWAS meta-analysis results will be annotated in order to predict the probable effects of replicated
225 SNPs on gene expression and disease development. Furthermore, we will highlight the molecular

226 pathways, cell and tissue types most likely to be involved in DD pathogenesis. All this information will
227 be used for gene prioritisation alongside with the results of analyses, based on colocalization analysis
228 [50] and Mendelian randomization [51,52]. This methodology will be applied for causal inference
229 between DD and gene expression profiles from different tissues, also including the transcriptomic data
230 from intervertebral discs obtained in this study. A gene network of DDD regulation will be built and the
231 key regulators will be revealed.

232 **Glycomics data analysis**

233 The association of plasma N-glycans levels with disc degeneration status will be studied. Any obtained
234 data may be helpful to use as dynamic glycan biomarkers for elucidating DD pathogenesis and may
235 ultimately provide prognostic information.

236 **Transcriptomic data analysis**

237 Patient group with available intraoperative disc material (see Figure 1) could be divided into “cases”
238 or “degenerated disc” (grades 4-5 of disc degeneration according to Pfirrmann classification) and
239 “controls” or “the healthy disc” (grades 1-3 of disc degeneration according to Pfirrmann classification)
240 [43,53].

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

245 Transcriptomic data obtained in this study will be used for detection of genes differentially expressed
246 in discs between cases and controls. In short, the gene expression data will be mapped [54] on the
247 reference genome and quality of the reads will be recalibrated (“Picard Toolkit.” 2019. Broad Institute,
248 GitHub Repository. <http://broadinstitute.github.io/picard/>; Broad Institute). Subsequently, aiming to
249 reveal the transcriptomic differences in intervertebral discs tissue between cases and controls, we will
250 count the reads and identify differentially expressed genes (DEGs) [55]. At last, the revealed DEGs will
251 be functionally annotated with Gene Ontology terms.

252 The data will also be used to identify expression quantitative trait loci (eQTLs). We will perform eQTL
253 analysis of gene expression levels in intervertebral discs using methodology described in [56]. Resulting
254 regional association summary statistics will be used for the gene prioritization.

255 **Data management and access**

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

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

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

268 To access the database, omics data and other relevant information, projects should be submitted to
269 the steering committee (contact the corresponding author) of the RuDDS.

270 **DISCUSSION**

271 This study will be carried out in two different centers providing wider population coverage and more
272 reliability to the sample storage, as data and specimens will be stored in two places. Participants will
273 be recruited at unique medical centers (NRITO, Priorov CITO), that are the only centers in Russia
274 specializing in studying DDD. The distinctive features of our clinical centers are: large constant flow of
275 DDD patients, opportunity to conduct spine MRI and a highly-qualified team [57,58]. In the same way,
276 our multidisciplinary research group, who will generate and analyze omics data, has a broad
277 experience in studying back pain (one of the main manifestations of DDD) [59–61] and statistical and
278 functional data analysis using integrative omics approaches [62,63].

279 Another notable advantage of the present project is availability of diverse and rare biological material
280 that can be used for multiple profiling. Besides the whole blood samples, plasma and the intervertebral
281 disc tissue obtained during the spine surgery will be collected. Presence of variable biological
282 specimens for the same study participants will allow conducting integrative multi-omics analyses.
283 Thus, not only genomes, but also transcriptomes, glycomes and other ‘-omes’ can be measured using
284 these samples. To our knowledge it is the first study with multiple omics profiling of similar biological
285 material. Current studies are limited by single profiling of one or two tissues [one - [32], [33]; two -
286 [37], [31], [28], [29] or use omics datasets in open access, combining data from different cohorts [40].
287 The importance of the multi-omics approach is hard to overestimate, as it allows us to look at the
288 disease from different points of view and level-up our understanding of the pathology.

289 One more substantial strength of our study is the assessment of disc degeneration grade by MRI scans
290 because it is the most accurate and precise method of DD diagnostics among the others. In LDD studies
291 based on plain radiography [64] or CT [65] phenotype definition is rather subjective as these
292 approaches provide only indirect evidence of disc degeneration such as disc height loss and
293 osteophytes. Similarly, the use of self-reported questionnaires for LDD identification [66] is not reliable
294 enough. On the contrary, Pfirrmann grading system of disc degeneration used in the present study is
295 based on MRI scans and estimates the main characteristics of DD by assessing the signal intensity and
296 height of the intervertebral disc - the more dehydrated the disc is, the more severe degenerative
297 changes are in it [41].

298 However, this study has some limitations. First, it should be noted that the expected sample size is
299 modest in comparison with national country-level biobanks or large prospective cohorts. Sample size
300 of the present study is four times less than one reported for meta-analysis of DDD studies reported by
301 Williams et al. [35]. This limitation is explained by funding, establishing the recruitment period of three
302 years. Nonetheless, the expected number of disc tissue samples (~510) and transcriptome profiles of
303 them (~40) is comparable with other omics studies using disc specimens [proteome [33] - 7 cases / 7
304 controls; metabolome [39] - 60 cases / 21 controls; transcriptome [28] - 39 annulus fibrosus (AF) and
305 21 nucleus pulposus (NP) samples, [29] - 24 AF and 24 NP samples].

306 The next limitation of our study is imbalance between patients with “healthy disc” and “degenerated
307 disc” in the group of patients subjected to surgery. Obviously, patients with “healthy discs” are

308 subjected to spinal surgery under specific circumstances. Although, according to our protocol we still
309 can expect some controls in this group (in total, the expected number of controls is 20). In groups not
310 subjected to surgery we expect the ratio between cases and controls closer to the general population.

311 It should be noted that “healthy disc” is a conditional definition, not objective enough; however, the
312 inter-observer agreement is quite high among different groups of researchers, therefore, Pfirrmann
313 classification is considered to be highly reliable [67]. The main point here is that the existence of adult
314 patients with perfectly healthy lumbar discs is questionable. To find such patients a huge large-scale
315 exploratory study is needed. Nevertheless, according to a population-based study there are no adults
316 with lumbar discs containing no degenerative changes [13].

317 **ETHICS AND DISSEMINATION**

318 The study will be performed according to the Helsinki Declaration; the study protocol was approved
319 by the Local Ethical Committee of NRITO (№034/20 dated 02 Oct 2020) and by the Local Ethical
320 Committee of Priorov CITO (№1/21 dated 25 Feb 2021). Preliminary and final results will be presented
321 in peer-reviewed journals, at national and international congresses.

322 **Acknowledgements**

323 **AUTHORS’ CONTRIBUTIONS**

324 ONL, EEE, AVK, YSA and YAT contributed to the study concept and design of the study protocol. ONL,
325 EEE, TSG, YAT contributed to the development of specimen collection, storage and profiling protocols.
326 ONL, AVP and AVK contributed to the setting of the logistics of working with hospitals and ethical
327 committee approval. ONL, EEE and YAT drafted the first version of the manuscript. All authors critically
328 reviewed and approved the final manuscript.

329 **FUNDING STATEMENT**

330 The work of YAT and EEE was supported by the Russian Foundation for Basic Research (project 19-015-
331 00151) and by the Ministry of Education and Science of the RF via the Institute of Cytology and Genetics
332 SB RAS (project number 0259-2021-0009 / AAAA-A17-117092070032-4). The work of ONL and YSA was
333 supported by the Russian Ministry of Science and Education under the 5-100 Excellence Programme.

334 **COMPETING INTERESTS STATEMENT**

335 YSA is a co-owner of Maatschap PolyOmica and PolyKnomics BV, private organizations, providing
336 services, research and development in the field of quantitative and computational genomics.

337

338 **Data statement section:** Prepublication history for this paper is available online. To view these files,
339 please visit the journal online.

340

341 **REFERENCES**

- 342 1. Berg AJ, Ahmadje U, Jayanna HH, Trégouët P, Sanville P, Kapoor V. The prevalence of lumbar
343 disc degeneration in symptomatic younger patients: A study of MRI scans. *J Clin Orthop Trauma*.
344 2020;11(5):932–6.
- 345 2. Brinjikji W, Luetmer PH, Comstock B, Bresnahan BW, Chen LE, Deyo RA, Halabi S, Turner JA,
346 Avins AL, James K, Wald JT, Kallmes DF JJ. Systematic Literature Review of Imaging Features of

- 347 Spinal Degeneration in Asymptomatic Populations. *AJNR Am J Neuroradiol.* 2015;36(4):811–6.
- 348 3. Kos N, Gradisnik L, Velnar T. A Brief Review of the Degenerative Intervertebral Disc Disease.
349 2019;73(6):421–4.
- 350 4. Cheung KMC, Karppinen J, Chan D, Ho DWH, Song YQ, Sham P, et al. Prevalence and pattern of
351 lumbar magnetic resonance imaging changes in a population study of one thousand forty-three
352 individuals. *Spine (Phila Pa 1976).* 2009;34(9):934–40.
- 353 5. Matsumoto M, Okada E, Toyama Y, Fujiwara H, Momoshima S, Takahata T. Tandem age-related
354 lumbar and cervical intervertebral disc changes in asymptomatic subjects. *Eur Spine J.*
355 2013;22(4):708–13.
- 356 6. Teraguchi M, Yoshimura N, Hashizume H, Muraki S, Yamada H, Oka H, et al. The association of
357 combination of disc degeneration, end plate signal change, and Schmorl node with low back
358 pain in a large population study: The Wakayama Spine Study. *Spine J [Internet].*
359 2015;15(4):622–8. Available from: <http://dx.doi.org/10.1016/j.spinee.2014.11.012>
- 360 7. Jensen RK, Jensen TS, Koes B, Hartvigsen J. Prevalence of lumbar spinal stenosis in general and
361 clinical populations: a systematic review and meta-analysis. *Eur Spine J [Internet].*
362 2020;(0123456789). Available from: <https://doi.org/10.1007/s00586-020-06339-1>
- 363 8. Brinjikji W, Diehn FE, Jarvik JG, Carr CM, Kallmes DF, Murad MH, et al. MRI findings of disc
364 degeneration are more prevalent in adults with low back pain than in asymptomatic controls:
365 A systematic review and meta-analysis. *Am J Neuroradiol.* 2015;36(12):2394–9.
- 366 9. Oichi T, Taniguchi Y, Oshima Y, Tanaka S, Saito T. Pathomechanism of intervertebral disc
367 degeneration. *JOR SPINE.* 2020;3(1):1–9.
- 368 10. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. *Lancet.*
369 2017;389(10070):736–47.
- 370 11. Middendorp M, Vogl TJ, Kollias K, Kafchitsas K, Khan MF, Maataoui A. Association between
371 intervertebral disc degeneration and the Oswestry Disability Index. *J Back Musculoskelet*
372 *Rehabil.* 2017;30(4):819–23.
- 373 12. Teraguchi M, Yoshimura N, Hashizume H, Muraki S, Yamada H, Minamide A, et al. Prevalence
374 and distribution of intervertebral disc degeneration over the entire spine in a population-based
375 cohort: The Wakayama Spine Study. *Osteoarthr Cartil [Internet].* 2014;22(1):104–10. Available
376 from: <http://dx.doi.org/10.1016/j.joca.2013.10.019>
- 377 13. Oh CH, Yoon SH. Whole Spine Disc Degeneration Survey according to the Ages and Sex Using
378 Pfirrmann Disc Degeneration Grades. *Korean J Spine.* 2018;14(4):148–54.
- 379 14. Sääksjärvi S, Kerttula L, Luoma K, Paajanen H, Waris E. Disc Degeneration of Young Low Back
380 Pain Patients: A Prospective 30-year Follow-up MRI Study. *Spine (Phila Pa 1976).*
381 2020;45(19):1341–7.
- 382 15. Teraguchi M, Yoshimura N, Hashizume H, Yamada H, Oka H, Minamide A, et al. Progression,
383 incidence, and risk factors for intervertebral disc degeneration in a longitudinal population-
384 based cohort: the Wakayama Spine Study [Internet]. Vol. 25, *Osteoarthritis and Cartilage.*
385 Elsevier Ltd; 2017. 1122–1131 p. Available from: <http://dx.doi.org/10.1016/j.joca.2017.01.001>
- 386 16. Kadow T, Sowa G, Vo N, Kang JD. Molecular Basis of Intervertebral Disc Degeneration and
387 Herniations: What Are the Important Translational Questions? *Clin Orthop Relat Res [Internet].*
388 2015;473(6):1903–12. Available from: <http://dx.doi.org/10.1007/s11999-014-3774-8>
- 389 17. Williams FMK, Sambrook PN. Neck and back pain and intervertebral disc degeneration: Role of
390 occupational factors. *Best Pract Res Clin Rheumatol [Internet].* 2011;25(1):69–79. Available

- 391 from: <http://dx.doi.org/10.1016/j.berh.2011.01.007>
- 392 18. Xiao L, Ni C, Shi J, Wang Z, Wang S, Zhang J, et al. Analysis of correlation between vertebral
393 endplate change and lumbar disc degeneration. *Med Sci Monit*. 2017;23:4932–8.
- 394 19. Jarman JP, Arpinar VE, Baruah D, Klein AP, Maiman DJ, Tugan Muftuler L. Intervertebral disc
395 height loss demonstrates the threshold of major pathological changes during degeneration. *Eur*
396 *Spine J*. 2015;24(9):1944–50.
- 397 20. Desmoulin GT, Pradhan V, Milner TE. Mechanical Aspects of Intervertebral Disc Injury and
398 Implications on Biomechanics. *Spine (Phila Pa 1976)*. 2020;45(8):E457–64.
- 399 21. Otluoğlu GD, Konya D, Toktas ZO. The influence of mechanic factors idisc degeneration disease
400 as a determinant for surgical indication. *Neurospine*. 2020;17(1):215–20.
- 401 22. Dragsbæk L, Kjaer P, Hancock M, Jensen TS. An exploratory study of different definitions and
402 thresholds for lumbar disc degeneration assessed by MRI and their associations with low back
403 pain using data from a cohort study of a general population. *BMC Musculoskelet Disord*.
404 2020;21(1):1–11.
- 405 23. Navone S, Marfia G, Giannoni A, Beretta M, Guarnaccia L, Gualtierotti R, et al. Inflammatory
406 mediators and signalling pathways controlling intervertebral disc degeneration. *Histol*
407 *Histopathol*. 2017;32(6):523–42.
- 408 24. Feng C, Liu H, Yang M, Zhang Y, Huang B, Zhou Y. Disc cell senescence in intervertebral disc
409 degeneration: Causes and molecular pathways. *Cell Cycle [Internet]*. 2016;15(13):1674–84.
410 Available from: <http://dx.doi.org/10.1080/15384101.2016.1152433>
- 411 25. Chen WK, Yu XH, Yang W, Wang C, He WS, Yan YG, et al. lncRNAs: novel players in intervertebral
412 disc degeneration and osteoarthritis. *Cell Prolif*. 2017;50(1):1–12.
- 413 26. Munir S, Rade M, Määttä JH, Freidin MB, Williams FMK. Intervertebral Disc Biology: Genetic
414 Basis of Disc Degeneration. *Curr Mol Biol Reports*. 2018;4(4):143–50.
- 415 27. Kawaguchi Y. Genetic background of degenerative disc disease in the lumbar spine. *Spine Surg*
416 *Relat Res*. 2018;2(2):98–112.
- 417 28. Riester SM, Lin Y, Wang W, Cong L, Ali AM, Krauss W, et al. RNA sequencing identifies gene
418 regulatory networks controlling extracellular matrix synthesis in intervertebral disk tissues. *J*
419 *Orthop Res*. 2019;36(5):1356–69.
- 420 29. Kazezian Z, Gawri R, Haglund L, Ouellet J, Mwale F, Tarrant F, et al. Gene Expression Profiling
421 Identifies Interferon Signalling Molecules and IGF1 in Human Degenerative Annulus Fibrosus.
422 *Sci Rep [Internet]*. 2015;5(September):1–13. Available from:
423 <http://dx.doi.org/10.1038/srep15662>
- 424 30. Sherafatian M, Abdollahpour HR, Ghaffarpasand F, Yaghmaei S, Azadegan M, Heidari M.
425 MicroRNA Expression Profiles, Target Genes, and Pathways in Intervertebral Disk Degeneration:
426 A Meta-Analysis of 3 Microarray Studies. *World Neurosurg [Internet]*. 2019;126:389–97.
427 Available from: <https://doi.org/10.1016/j.wneu.2019.03.120>
- 428 31. Sarath Babu N, Krishnan S, Brahmendra Swamy C V., Venkata Subbaiah GP, Gurava Reddy A V.,
429 Idris MM. Quantitative proteomic analysis of normal and degenerated human intervertebral
430 disc. *Spine J [Internet]*. 2016;16(8):989–1000. Available from:
431 <http://dx.doi.org/10.1016/j.spinee.2016.03.051>
- 432 32. Ye D, Liang W, Dai L, Zhou L, Yao Y, Zhong X, et al. Comparative and quantitative proteomic
433 analysis of normal and degenerated human annulus fibrosus cells. *Clin Exp Pharmacol Physiol*.
434 2015;42(5):530–6.

- 435 33. Yee A, Lam MPY, Tam V, Chan WCW, Chu IK, Cheah KSE, et al. Fibrotic-like changes in
436 degenerate human intervertebral discs revealed by quantitative proteomic analysis. *Osteoarthr*
437 *Cartil* [Internet]. 2016;24(3):503–13. Available from:
438 <http://dx.doi.org/10.1016/j.joca.2015.09.020>
- 439 34. Eskola PJ, Lemmelä S, Kjaer P, Solovieva S, Männikkö M, Tommerup N, et al. Genetic Association
440 Studies in Lumbar Disc Degeneration: A Systematic Review. *PLoS One*. 2012;7(11).
- 441 35. Williams FMK, Bansal AT, Van Meurs JB, Bell JT, Meulenbelt I, Suri P, et al. Novel genetic variants
442 associated with lumbar disc degeneration in northern Europeans: A meta-analysis of 4600
443 subjects. *Ann Rheum Dis*. 2013;72(7):1141–8.
- 444 36. Song YQ, Karasugi T, Cheung KMC, Chiba K, Ho DWH, Miyake A, et al. Lumbar disc degeneration
445 is linked to a carbohydrate sulfotransferase 3 variant. *J Clin Invest*. 2013;123(11):4909–17.
- 446 37. Zhang YG, Guo X, Sun Z, Jia G, Xu P, Wang S. Gene expression profiles of disc tissues and
447 peripheral blood mononuclear cells from patients with degenerative discs. *J Bone Miner Metab*.
448 2010;28(2):209–19.
- 449 38. Wang Y, Dai G, Wang L, Shang F, Jiang L, Li S, et al. Identification of Key Genes Potentially Related
450 to Intervertebral Disk Degeneration by Microarray Analysis. *Genet Test Mol Biomarkers*.
451 2019;23(9):610–7.
- 452 39. Pacholczyk-Sienicka B, Radek M, Radek A, Jankowski S. Characterization of metabolites
453 determined by means of 1H HR MAS NMR in intervertebral disc degeneration. *Magn Reson*
454 *Mater Physics, Biol Med*. 2015;28(2):173–83.
- 455 40. Xu C, Luo S, Wei L, Wu H, Gu W, Zhou W, et al. Integrated transcriptome and proteome analyses
456 identify novel regulatory network of nucleus pulposus cells in intervertebral disc degeneration.
457 *BMC Med Genomics*. 2021;14(1):1–12.
- 458 41. Pfirrmann CWA, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic Resonance Classification of
459 Lumbar Intervertebral Disc Degeneration. 2001;26(17):1873–8.
- 460 42. Chen Y, Bao J, Yan Q, Wu C, Yang H, Zou J. Distribution of Modic changes in patients with low
461 back pain and its related factors. *Eur J Med Res* [Internet]. 2019;24(1):1–9. Available from:
462 <https://doi.org/10.1186/s40001-019-0393-6>
- 463 43. Rajasekaran S, Venkatadass K, Naresh Babu J, Ganesh K, Shetty AP. Pharmacological
464 enhancement of disc diffusion and differentiation of healthy, ageing and degenerated discs:
465 Results from in-vivo serial post-contrast MRI studies in 365 human lumbar discs. *Eur Spine J*
466 [Internet]. 2008;17(5):626–43. Available from: doi: 10.1007/s00586-008-0645-6
- 467 44. Sambrook PN, MacGregor AJ, Spector TD. Genetic influences on cervical and lumbar disc
468 degeneration: A magnetic resonance imaging study in twins. *Arthritis Rheum*. 1999;42(2):366–
469 72.
- 470 45. Collins S, Moore R, McQuay H. The visual analogue pain intensity scale: what is moderate pain
471 in millimetres? *Pain*. 1997;72(1–2):95–7.
- 472 46. Fairbank, J.C. Pynsent PB. The Oswestry Disability Index. *Spine (Phila Pa 1976)*. 2000;Nov
473 15;25(22):2940–52.
- 474 47. Cherepanov EA. Russian version of the oswestry disability index: cross-cultural adaptation and
475 validity. *Hir pozvonočnika (Spine Surgery)*. 2009;0(3):093–8.
- 476 48. McCarthy S, Das S, Kretschmar W, Delaneau O, Wood A, Teumer A, et al. A reference panel of
477 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279–83.
- 478 49. Trbojević Akmačić I, Ugrina I, Štambuk J, Gudelj I, Vučković F, Lauc G, et al. High-throughput

- 479 glycomics: Optimization of sample preparation. *Biochem*. 2015;80(7):934–42.
- 480 50. Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charlotheaux B, et al. IBD risk
481 loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat*
482 *Commun*. 2018;9(1):1–18.
- 483 51. Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurler MR, et al. Phenome-wide Mendelian
484 randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet*
485 [Internet]. 2020;52(10):1122–31. Available from: [http://dx.doi.org/10.1038/s41588-020-0682-](http://dx.doi.org/10.1038/s41588-020-0682-6)
486 6
- 487 52. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from
488 GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481–7.
- 489 53. Ikuno A, Akeda K, Takebayashi SI, Shimaoka M, Okumura K, Sudo A. Genome-wide analysis of
490 DNA methylation profile identifies differentially methylated loci associated with human
491 intervertebral disc degeneration. *PLoS One*. 2019;14(9):1–20.
- 492 54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal
493 RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
- 494 55. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential
495 expression analysis of digital gene expression data. *Bioinformatics*. 2009;26(1):139–40.
- 496 56. Parisien M, Khoury S, Chabot-Doré A-J, Sotocinal SG, Slade GD, Smith SB, et al. Effect of Human
497 Genetic Variability on Gene Expression in Dorsal Root Ganglia and Association with Pain
498 Phenotypes. *Cell Rep* [Internet]. 2017 May;19(9):1940–52. Available from:
499 <https://linkinghub.elsevier.com/retrieve/pii/S2211124717306435>
- 500 57. Krutko A V., Sanginov AJ, Baykov ES. Predictors of Treatment Success Following Limited
501 Discectomy With Annular Closure for Lumbar Disc Herniation. *Int J Spine Surg*. 2020;14(1):38–
502 45.
- 503 58. Leonova O, Baykov E, Sanginov A, Krutko A. Cervical Disc Degeneration and Vertebral Endplate
504 Defects After the Fused Operation. *Spine (Phila Pa 1976)* [Internet]. 2021;(Feb 16):Ahead of
505 Print. Available from: doi: 10.1097/BRS.0000000000004007
- 506 59. Freidin MB, Tsepilov YA, Palmer M, Karssen LC, Suri P, Aulchenko YS, et al. Insight into the
507 genetic architecture of back pain and its risk factors from a study of 509,000 individuals. Loos
508 RJF, editor. *Pain* [Internet]. 2019 Jun 6;160(6):1361–73. Available from:
509 <https://dx.plos.org/10.1371/journal.pgen.1007601>
- 510 60. Elgaeva EE, Tsepilov Y, Freidin MB, Williams FMK, Aulchenko Y, Suri P. ISSLS Prize in Clinical
511 Science 2020. Examining causal effects of body mass index on back pain: a Mendelian
512 randomization study. *Eur Spine J* [Internet]. 2020 Apr 3;29(4):686–91. Available from:
513 <http://link.springer.com/10.1007/s00586-019-06224-6>
- 514 61. Freidin MB, Tsepilov YA, Stanaway IB, Meng W, Hayward C, Smith BH, et al. Sex- and age-specific
515 genetic analysis of chronic back pain. *Pain* [Internet]. 2020 Sep 30; Publish Ah. Available from:
516 <https://journals.lww.com/10.1097/j.pain.0000000000002100>
- 517 62. Tsepilov YA, Sharapov SZ, Zaytseva OO, Krumsek J, Prehn C, Adamski J, et al. A network-based
518 conditional genetic association analysis of the human metabolome. *Gigascience*. 2018;7(12):1–
519 11.
- 520 63. Sharapov SZ, Tsepilov YA, Klaric L, Mangino M, Thareja G, Shadrina AS, et al. Defining the genetic
521 control of human blood plasma N-glycome using genome-wide association study. *Hum Mol*
522 *Genet*. 2019;28(12):2062–77.

- 523 64. Riyazi N, Meulenbelt I, Kroon HM, Runday KH, Hellio Le Graverand MP, Rosendaal FR, et al.
524 Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings
525 with multiple joint involvement: The GARP study. *Ann Rheum Dis.* 2005;64(3):438–43.
- 526 65. Yau MS, Demissie S, Zhou Y, Anderson DE, Amanda L. Correlations With Other Spine Traits : The
527 Framingham Study. 2016;31(12):2077–84.
- 528 66. Allegri M, De Gregori M, Minella CE, Klersy C, Wang W, Sim M, et al. Omics’ biomarkers
529 associated with chronic low back pain: Protocol of a retrospective longitudinal study. *BMJ Open.*
530 2016;6(10):1–8.
- 531 67. Urrutia J, Besa P, Campos M, Cikutovic P, Cabezon M, Molina M, et al. The Pfirrmann
532 classification of lumbar intervertebral disc degeneration: an independent inter- and intra-
533 observer agreement assessment. *Eur Spine J.* 2016;25(9):2728–33.
- 534