scientific reports

Check for updates

OPEN

Proteomic analysis identifies distinct signaling pathways in lumbar intervertebral disc degeneration between type-2 diabetic and non-diabetic patients

F M Moinuddin[1,2,7], Jun Zhong[3,7], Maria D. Astudillo Potes[4], Asimina Dominari[1,6], Ashis Dhar[1,2], Akhilesh Pandey[3,5] & Mohamad Bydon[1,2,6]✉

Intervertebral disc degeneration (IDD) is a common condition associated with aging and degeneration of the disc. Type 2 diabetes mellitus (T2DM) is linked to various musculoskeletal disorders, including IDD, but its molecular effects on IDD remain understudied. To explore the molecular differences in IDD between T2DM and non-T2DM patients, we conducted a quantitative proteomic analysis of nucleus pulposus (NP) tissues from patients with T2DM-IDD or non-T2DM-IDD individuals. We identified 3,720 proteins, with 221 significantly upregulated and 233 downregulated in T2DM-IDD. Multiple mitochondrial proteins were upregulated (NDUFV2, NDUFB5, NDUFA13, MPC2) indicating a potential increase in oxidative stress in the NP tissues from T2DM-IDD. Conversely, downregulated proteins such as versican (VCAN), an extracellular matrix proteoglycan, IL17B, a pro-inflammatory cytokine, and vascular endothelial growth factor A (VEGFA) indicate possible alterations in ECM composition and structure that impact the mechanical properties and integrity of the IVD as well as a reduced inflammatory response in NP tissues from non-T2DM-IDD. Additionally, T2DM-IDD showed enrichment in fatty acid metabolism pathways, while non-T2DM-IDD samples were enriched in epithelial mesenchymal transition and hypoxia pathways. Altogether, these results provide novel insights into the molecular mechanisms of T2DM-IDD and may guide future therapies.

Keywords Intervertebral disc degeneration, Type 2 diabetes mellitus, Proteomics, Proline hydroxylation, Tandem mass tag, Nucleus pulposus

Intervertebral disc (IVD) degeneration (IDD) is quite common and affects a substantial portion of the global population[1]. IDD encompasses aging and degenerative changes to the integrity of IVD, a response to external insults such as mechanical injury or metabolic perturbation often marked by a range of molecular, biochemical, cellular and anatomical alterations of IVD[2]. The intricate network of proteins, especially the extracellular matrix (ECM) proteins, plays a critical role in maintaining the integrity of the IVD by regulating cell behavior, matrix synthesis, mechanical support, and tissue remodeling[3]. Disruptions in cellular, structural, and mechanical IVD integrity can culminate in a spectrum of degenerative spinal pathologies, including disc herniation and spondylolisthesis.

Type 2 diabetes mellitus (T2DM), a chronic, metabolic, and inflammatory condition, is characterized by insulin resistance, pancreatic beta-cell dysfunction, and hyperglycemia[4]. Its global incidence continues to soar, particularly among younger individuals, potentially accelerating the onset and progression of chronic conditions. Although several studies have suggested that patients with T2DM tends to develop more severe IDD compared with patients without T2DM[5,6], the biological effects of T2DM on IDD at a proteomic level have not been examined.

[1]Mayo Clinic Neuro-Informatics Laboratory, Mayo Clinic, Rochester, MN, USA. [2]Department of Neurological Surgery, University of Chicago, Chicago, IL, USA. [3]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. [4]Alix School of Medicine, Mayo Clinic, Rochester, MN, USA. [5]Manipal Academy of Higher Education, Manipal, Karnataka, India. [6]Department of Neurosurgery, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA. [7]These authors contributed equally to this work ✉email: mohamad.bydon@gmail.com

Mass spectrometry has been utilized to characterize the proteomes of aged and degenerative IVD. Sarath Babu et al. used iTRAQ-based quantitative proteomics to identify 759 and 692 proteins from the annulus fibrosus and nucleus pulposus tissues of IVD, respectively. Among these identified proteins, the differentially expressed proteins are involved in cell adhesion, cell migration and interleukin 13 signalling pathway[7]. Rajasekaran et al. documented identification of 409, 512, and 412 proteins in IVD from healthy young, healthy aged, and IDD, respectively. The analysis implicated inflammaging as the main differentiator between normal aging and IDD[8]. Later Rajasekaran et al. focused on proteomics analysis of collagens in IDD and showed altered collagen signature in IDD[9]. Wang et al. used label-free proteomics to identify 125 differentially expressed protein in degenerated nucleus pulposus (NP) tissues and found that inflammatory-sensitive chitinase-3-like protein 1 may play a protective role in IDD[10]. Qiu et al. used tandem mass tag (TMT)-based quantitative proteomics to quantify 2,691 of 3,259 proteins identified in NP tissues isolated from 5 fetal IVD and 5 geriatric IVD with Pfirrmann Grade IV ratings[11]. They also reported that proteins expressed highly in the geriatric IVD were enriched primarily in the complement and coagulation cascades pathway and the cytokine receptor interaction pathway and those expressed highly in fetal IVD were enriched mainly in the ribosome pathway. Fu et al. also used TMT-based quantitative proteomics to identify 2,042 proteins in 3 Pfirrmann Grade II and 3 Pfirrmann Grade IV NP tissues from adult patients[12]. Their KEGG analysis of differentially expressed proteins showed that Epstein–Barr virus infection, viral myocarditis, colorectal cancer, nonalcoholic fatty liver disease (NAFLD) and amyotrophic lateral sclerosis (ALS) are the main pathways involved in IDD. Yang et al. combined mass spectrometry-based proteomics with bulk and single-cell RNA sequencing to analyze 8 cases of Grade I-IV IDD[13]. Their analysis of multiomics data suggested that SERPINA1 can be used as a biomarker to regulate or predict the progress of disc degeneration. However, biological differences of IVD in T2DM-IDD patients and non-T2DM-IDD patients have not been investigated at a proteomics level.

Here, we employed a tandem mass tag (TMT)-based quantitative approach to analyze the proteomes of NP tissues isolated from lumbar IVD that were surgically removed from five T2DM-IDD and four non-T2DM-IDD patients. Our analysis identified 3,887 proteins, with 221 significantly upregulated and 233 significantly downregulated in T2DM-IDD compared to non-T2DM-IDD. Unsupervised hierarchical clustering demonstrated distinct proteomic profiles between T2DM-IDD and non-T2DM-IDD. Gene set enrichment analysis (GSEA) indicated that oxidative phosphorylation, MYC targets, and fatty acid metabolism were among the pathways enriched in T2DM-IDD, while epithelial mesenchymal transition and hypoxia pathways were among those enriched in non-T2DM-IDD. Protein-protein interaction network analysis revealed highly connected subnetworks that are highly upregulated in T2DM-IDD than non-T2DM-IDD. Furthermore, the analysis of proline hydroxylation in our proteomic dataset revealed a highly connected protein-protein interaction network in hydroxylated IVD proteins in IDD. Overall, the results shed light on the intricate proteomic landscape underlying IDD, with a particular focus on the substantial contribution of T2DM in patients with surgically treated LDH or spondylolisthesis. Our analysis of IVD proteomes from T2DM-diagnosed patients holds the potential to unveil novel therapeutic targets for managing IDD and the associated burden of chronic low back pain.

Results

Proteomics analysis of NP tissues in T2DM-IDD and non-T2DM-IDD patients

To understand the mechanisms underlying the IDD development in both T2DM and non-T2DM patients, we employed a TMT-based quantitative analysis to characterize proteomic differences of NP tissues in T2DM-IDD patients compared to non-T2DM-IDD patients (Fig. 1A). We identified 3,887 IVD proteins (Supplementary Table 1), representing a significant number of proteins for comparison with previous proteomics studies. [7–13]. The PANTHER Gene Ontology GO-Slim molecular function (MF) analysis of these IVD proteins revealed 1,318 proteins with binding activity, 1,112 with catalytic activity, 195 with molecular function regulator activity, 126 with structural molecule activity, 116 with transporter activity, 101 with ATP-dependent activity, 63 with molecular adaptor activity, 53 with molecular transducer activity, 48 with transcription regulator activity, 30 with antioxidant activity, 34 with translation regulator activity, and 25 cytoskeletal motor activity. Subcellular localization analysis revealed 500 proteins from extracellular region, 1,670 from cytoplasm, 451 from plasma membrane, and 570 from nucleus. Not surprisingly, by using the whole human proteome as a reference, the overrepresentation test revealed 2.2-fold enrichment for proteins in extracellular region and 1.6-fold enrichment for proteins in cytoplasm in our proteomics dataset (Fig. 1B), in consistent with that ECM proteins constitute the major components of IVD NP tissue.

Differentially expressed proteins in T2DM-IDD compared to non-T2DM-IDD

Among 3,887 identified proteins, we were able to quantify 3,720 proteins (Fig. 2A; Supplementary Table 1). We chose a cut-off of > 1.5-fold change for upregulated proteins, < 0.67-fold change for downregulated proteins, and p-value < 0.05 for significance. We identified 221 significantly upregulated IVD proteins and 233 significantly downregulated IVD proteins in T2DM-IDD as compared to non-T2DM-IDD. Unsupervised clustering of differentially expressed proteins shows the notable distinctions in IVD proteins between T2DM-IDD and non-T2DM-IDD (Fig. 2B, Supplementary Table 1, Tables 2 and 3). Gene set enrichment analysis for molecular function in the significantly upregulated IVD proteins in T2DM-IDD revealed 91 proteins with binding activity, 44 with catalytic activity, 29 with structural molecule activity, 4 with antioxidant activity, 10 with transporter activity, and 10 with molecular transducer activity. The GO cellular components (CC) analysis of the significantly upregulated IVD proteins in T2DM-IDD revealed 3 proteins from extracellular region, 122 from cytoplasm, 10 from plasma membrane, 54 from nucleus. The GO MF analysis of the significantly downregulated IVD proteins in T2DM-IDD revealed 68 proteins with binding activity, 76 with catalytic activity, 27 with molecular function

Figure 1

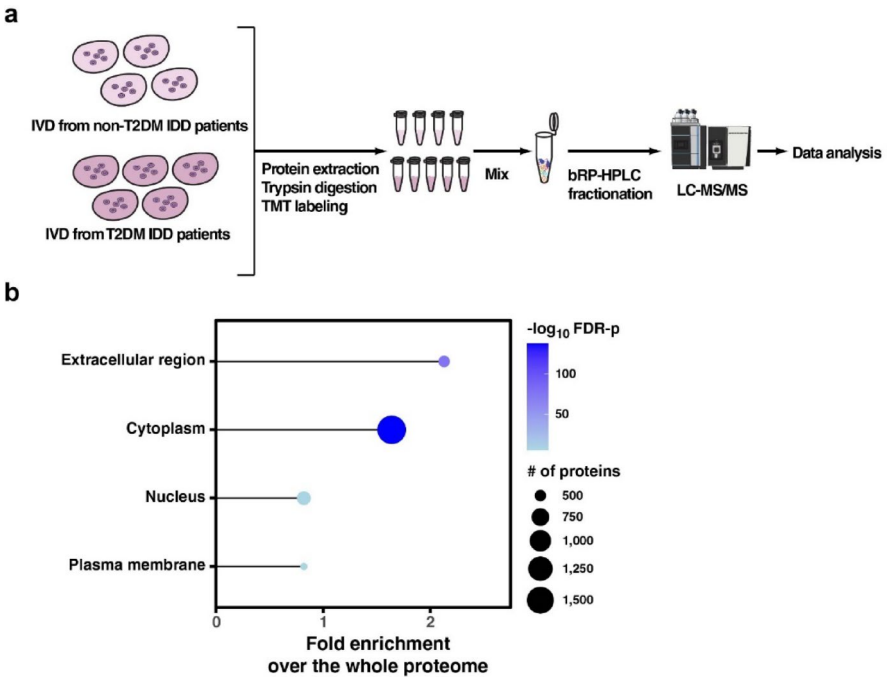

Fig. 1. Proteomic analysis of NP tissue from lumbar IVD in T2DM-IDD and non-T2DM-IDD patients. (**a**) The schematic illustration of TMT-based quantitative proteomics analysis. (**b**) The gene ontology cellular component enrichment analysis of identified NP proteins.

regulator activity, 7 with molecular transducer activity, 2 with antioxidant activity, and 87 without PANTHER GO MF term assignment. The GO cellular components (CC) analysis of the significantly downregulated IVD proteins in T2DM-IDD revealed 65 proteins from extracellular region, 65 from cytoplasm, 28 from plasma membrane, 9 from nucleus. Using all identified IVD proteins as a reference, the GO CC overrepresentation test revealed 2.15-fold enrichment with FDR-adjusted p-value of 1.24×10^{-7} for significantly downregulated IVD proteins in extracellular region in T2DM-IDD (i.e. significantly upregulated IVD proteins in non-T2DM-IDD) but no more than 2-fold enrichment for significantly upregulated IVD proteins for all four cellular components (Supplementary Table 2), suggesting more damage to extracellular compartment in T2MD IDD than non-T2DM-IDD.

Pathways differentially enriched in T2DM-IDD and non-T2DM-IDD

To understand the differences in the underlying pathways between T2DM-IDD and non-T2DM-IDD, we performed Gene Set Enrichment Analysis (GSEA) on our IVD proteomics dataset to identify pathways differentially enriched in T2DM-IDD and non-T2DM-IDD (Fig. 3A, Supplementary Table 3). In T2DM-IDD, oxidative phosphorylation is the most significantly enriched pathway, followed by six pathways - MYC targets, fatty acid metabolism, E2F targets, G2M checkpoint, mitotic spindle, and adipogenesis. Figure 3B shows 6 examples of significantly upregulated IVD proteins in T2DM-IDD. In non-T2DM-IDD, epithelial mesenchymal transition is the most enriched pathway followed by six pathways – hypoxia, glycolysis, angiogenesis, estrogen response early, KRAS signaling up, and coagulation. Figure 3C shows 6 examples of significantly upregulated proteins in non-T2DM-IDD. Our pathway analysis suggests that metabolic perturbation, such as increased oxidative phosphorylation, fatty acid metabolism, and adipogenesis, are evident in the NP tissues in T2DM-IDD and could contribute to the severity observed in T2DM-IDD.

Protein-protein interaction networks of differentially expressed proteins in T2DM-IDD and non-T2DM-IDD

To gain more insights into the functions of these differentially expressed proteins, we queried their interactions through the STRING database to construct the interaction network across 450 differentially expressed proteins in T2DM-IDD and non-T2DM-IDD (Fig. 4A). The network consists of 256 nodes and 1,234 edges. Of 256 nodes, 171 are upregulated IVD proteins and 85 downregulated ones. MCL clustering analysis revealed different clusters between T2DM-IDD and non-T2DM-IDD. In T2DM-IDD, the largest cluster is translation, followed by nucleosome core, spliceosome, aerobic electron transfer chain, mRNA processing, and ATP synthesis, while in non-T2DM-IDD, the clusters are desmosome, proteins from extracellular space, secreted, and collagen biosynthesis and modifying enzyme (Fig. 4B). The identification of the clusters of some transcription and translation machineries in T2DM-IDD along with the enrichment of MYC targets V1, E2F targets, G2M

Figure 2

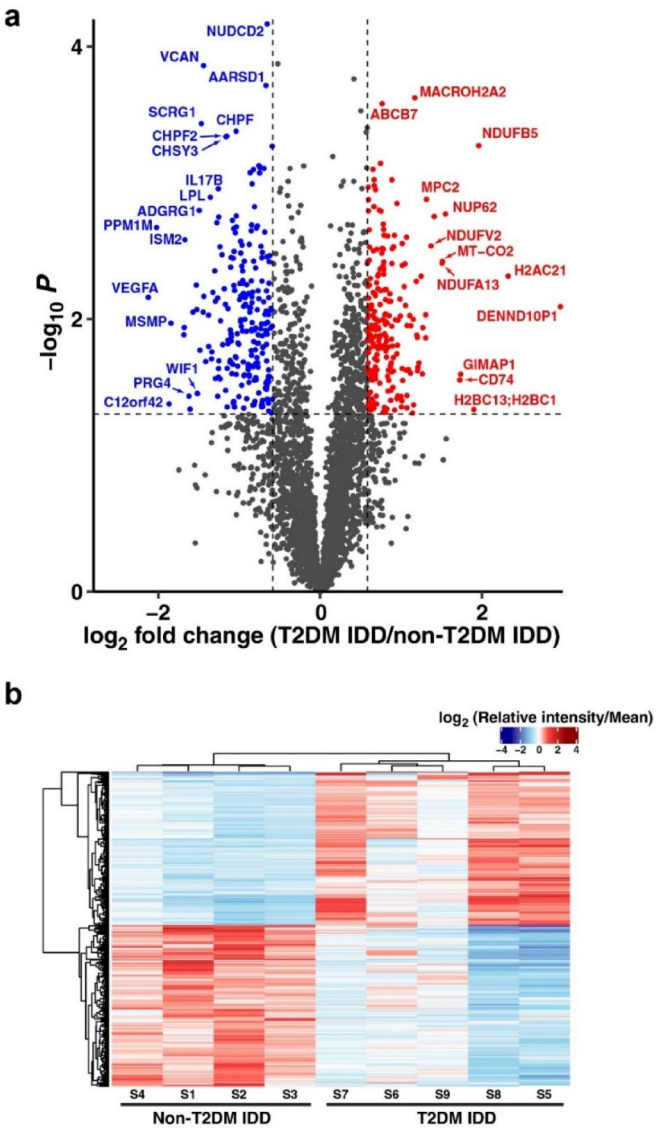

Fig. 2. Differentially expressed NP proteins in T2DM-IDD patients compared to non-T2DM-IDD patients. (**a**) Volcano map depicting log2 (fold change) vs. –log10 (p-value) calculated for each protein/protein group. (**b**) Heatmap showing unsupervised clustering of differentially expressed proteins.

checkpoint, and mitotic spindle suggest that aberrant cell growthy activities in T2DM-IDD, which requires further investigation. The identification of aerobic electron transfer chain and ATP synthesis is also consistent with the enrichment of oxidative phosphorylation in T2DM-IDD. It is worth noting that the identification of the clusters of desmosomes, proteins from extracellular space, secreted, and collagen biosynthesis and modifying enzyme in non-T2DM-IDD further support that extracellular matrix and cell-cell adhesion are damaged more in T2DM-IDD.

Differential proline hydroxylation of IVD proteomes in T2DM-IDD and non-T2DM-IDD

Given that proline hydroxylation is one of the most abundant post-translational modifications in the ECM, [14] we examined proline hydroxylation in our dataset. Interestingly, we identified 2,278 proline hydroxylation sites on 230 proteins (Fig. 5A). We chose a cut-off of > 1.5-fold change for upregulated proline hydroxylation, < 0.67-fold change for downregulated proline hydroxylation, and p-value < 0.05 for significance. We identified 72 significantly upregulated proline hydroxylation sites on 19 NP proteins and 36 significantly downregulated sites on 19 NP proteins in T2DM-IDD compared to non-T2DM-IDD. The GO cellular component (CC) analysis of the proline-hydroxylated proteins revealed 9 collagens, 47 proteins from collagen-containing extracellular matrix, 105 from extracellular region, 75 from cell periphery (Fig. 5B). As expected, we observed 16.47-fold, 7.96-fold, and 3.46-fold enrichment for collagen, collagen-containing extracellular matrix, and extracellular region, respectively,

Figure 3

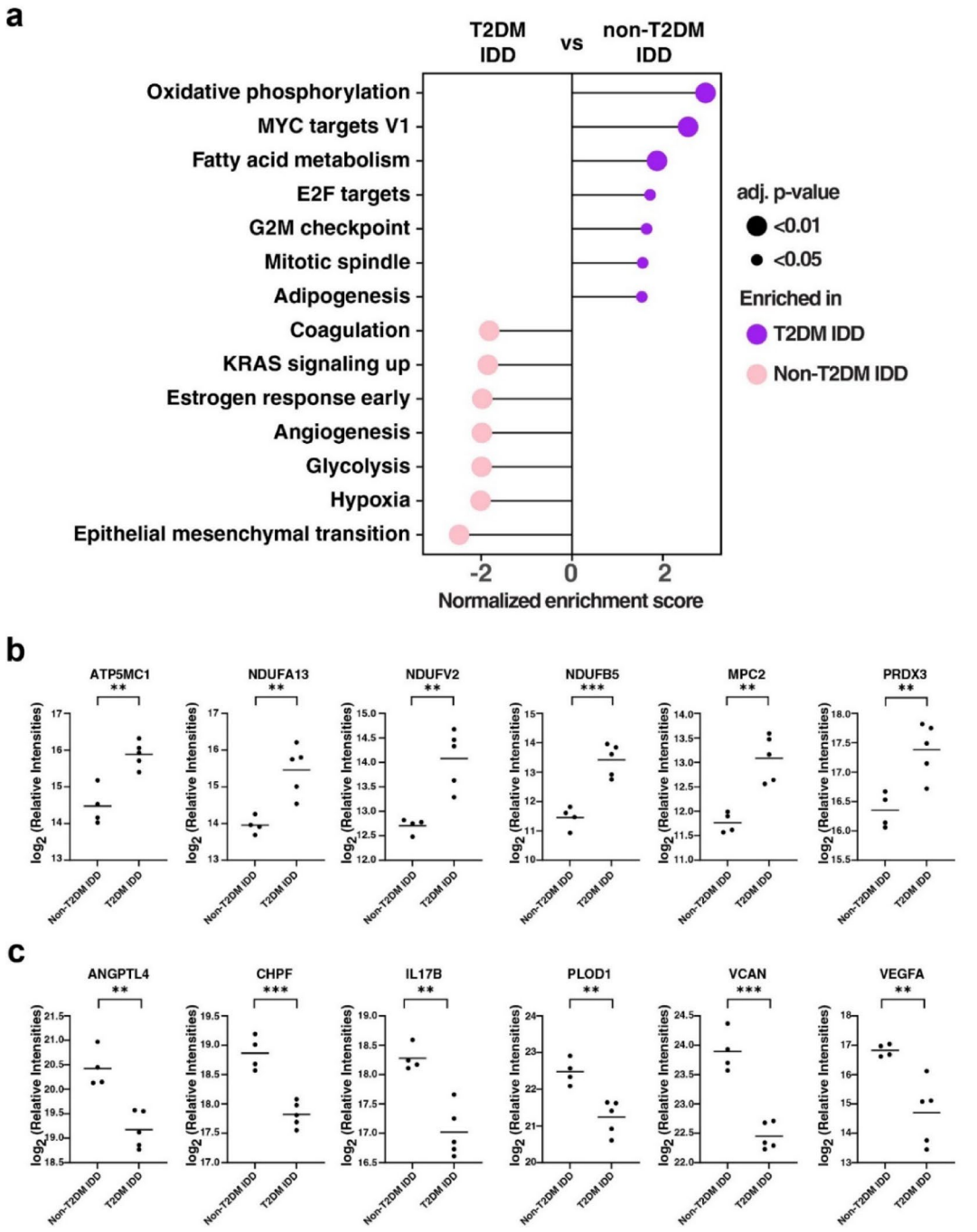

Fig. 3. Pathways differentially enriched in T2DM-IDD and non-T2DM-IDD revealed by Gene Set Enrichment Analysis (GSEA). Lollipop plot showing pathway enrichment (**a**) and box plots showing 6 significantly upregulated (**b**) and 6 significantly downregulated (**c**) NP proteins in T2DM-IDD compared to non-T2DM IDD.

with great significance. Furthermore, the protein-protein network analysis using the STRING (physical) database revealed a highly connected network among proteins with proline hydroxylation (Fig. 5C). In the network, 12 collagens have significantly upregulated proline hydroxylation and 2 have significantly downregulated in T2DM-IDD compared to non-T2DM-IDD, suggesting oxidative stress environment in T2DM-IDD. The higher proline hydroxylation on collagens in T2DM-IDD indicates oxidative stress.

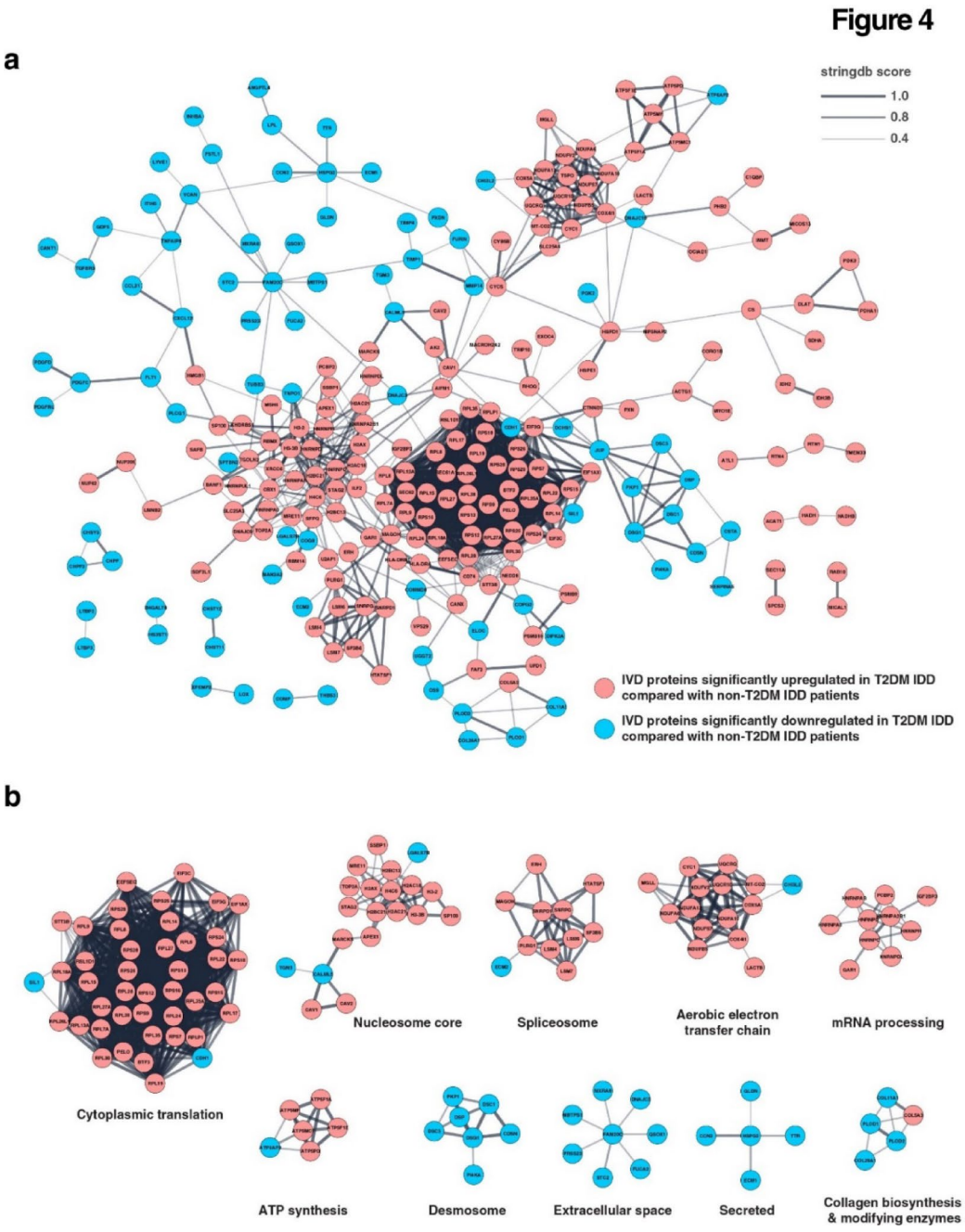

Fig. 4. Protein-protein interaction network (**a**) and clusters (**b**) of differentially expressed NP proteins in T2DM-IDD and non-T2DM-IDD using STRING database (http://www.string-db.org).

Discussion

Low back pain (LBP) is one of the most common disabilities, accounting for 149 million missed workdays annually, with an estimated loss of $90 billion in the United States[15]. More than 50% of LBD is associated with IDD[16]. Although IDD is often associated with aging, it also affects a younger population with multiple risk factors, including injury, genetics, chronic diseases, and environmental factors such as smoking or a combination thereof[17]. T2DM, the most common metabolic disorder, also increases the risk of developing IDD[18]. The association between T2DM and IDD has been suggested in some cross-sectional and retrospective studies. For example, a case–control study that enrolled 160,911 patients with IVDD and 315,225 controls in a group of military members also suggested that diabetes was a risk factor for developing IDD[19]. In a more recent study using a cohort from the European population, it was found that patients with T2DM have a 6.9% higher risk of IDD than those without T2DM. They also used multivariable Mendelian randomization (MR) analysis to examine the direct effect of T2DM on IDD adjusting for BMI[20]. The studies consistently suggest that T2DM is a risk factor for IDD, independent of BMI.

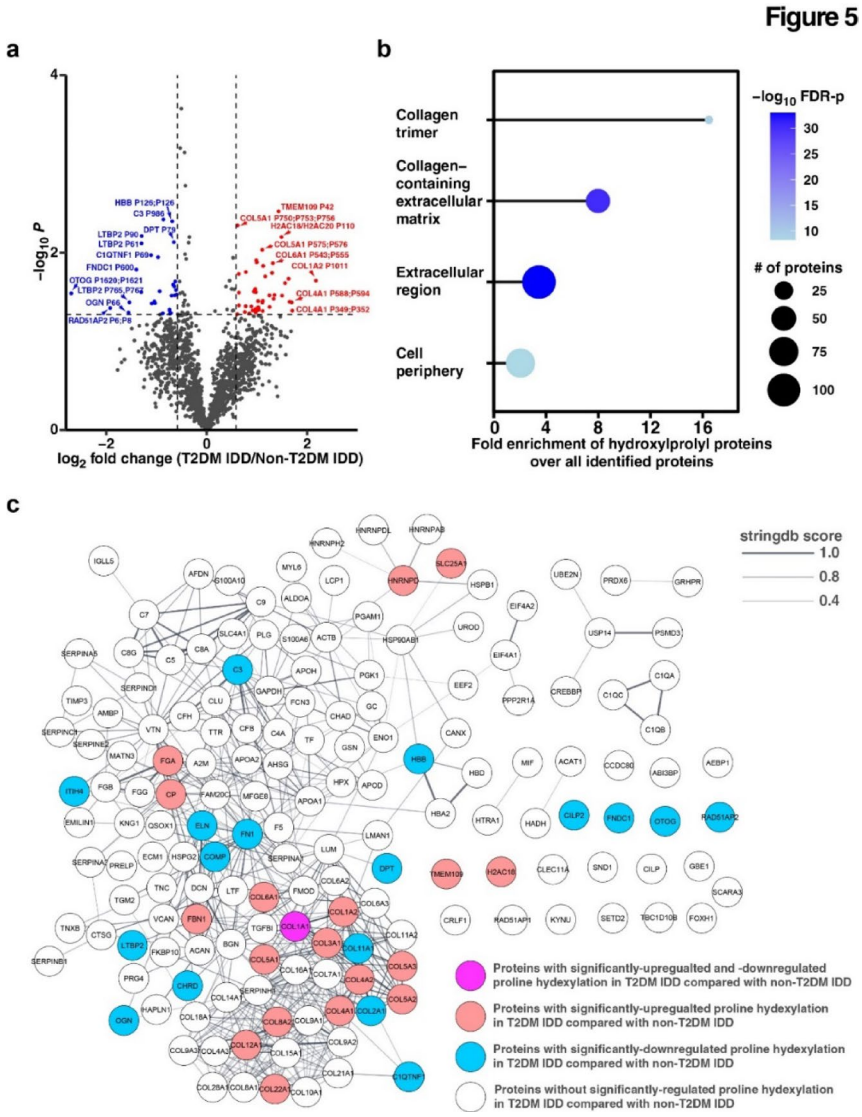

Fig. 5. Protein proline hydroxylation in IVD. (**a**) Volcano plot showing the differential proline hydroxylation in T2DM-IDD and non-T2DM-IDD. (**b**) Lollipop plot showing the GO CC enrichment analysis of protein with proline hydroxylation. (**c**) Protein-protein interaction network of proteins with proline hydroxylation using STRING database (http://www.string-db.org).

To understand this interrelation, several preclinical studies have demonstrated different aspects of the overall pathogenesis. Specifically, the effect of high level of advanced glycation end-products (AGEs) due to high glucose concentrations was responsible for IVD stiffening, altering the biomechanical properties of the IVD[21]. In addition, IVD cell stress under high glucose conditions is also characterized by the increased release of inflammatory cytokines, such as Tumor Necrosis Factor (TNF) α, interleukin (IL)-1, and IL-6. These cytokines lead to the ECM degradation by activating the matrix-metalloproteinases (MMPs) and A disintegrin and metalloproteinase with thrombospondin motifs (ADAM-TS)[22]. Similarly, high glucose condition induces cell apoptosis and senescence, which further weakens the IVD ECM and, therefore, accelerates its degeneration[23]. However, systemic examination of signaling pathways behind IDD in T2DM patients has not been carried out in an omics level.

Previously, we employed RNA sequencing technologies to identify the key regulatory players in degenerative discs[24]. Several groups have used quantitative proteomics approaches to study differential protein expression in healthy control discs and various types of degenerative discs[7–13]. Here, we used a mass spectrometry-based quantitative analytic pipeline to quantify changes in the NP proteomes from both T2DM-IDD and non-T2DM-IDD. Our analysis revealed 221 significantly upregulated NP proteins and 233 significantly downregulated NP proteins in T2DM-IDD compared to non-T2DM-IDD. Key NP proteins upregulated in T2DM-IDD include mitochondrial proteins such as NDUFV2, NDUFB5, NDUFA13, and MPC2, all involved in oxidative phosphorylation. In fact, mitochondrial dynamic is related to ROS production under hyperglycemic condition[25]. Therefore, the upregulation of these mitochondrial proteins indicates increased oxidative phosphorylation

and mitochondrial activity in diabetic IVD tissues, suggesting higher oxidative stress levels, which can lead to cellular damage, apoptosis, and further degeneration of the disc tissue. Interestingly, Liang Kang et al. suggested anti-oxidant MitoQ protect against IDD by ameliorating mitochondrial dysfunction and redox imbalance[26]. Additionally, CD74, another protein upregulated in diabetic tissues, is consistent with FGSEA analysis and previous findings by Xiong et al. [27], emphasizing its role in inflammatory responses and its association with various diseases, including IDD. Nucleoporin 62 (NUP62), upregulated in diabetic IVD tissues, indicates potential disruptions in nuclear-cytoplasmic transport, affecting cellular function and viability. Overall, the mechanical function of the IVD depends on the integrity of its tissue structure. In degenerative IVD, ECM catabolism increases, while anabolism decreases due to cellular death and abnormal function. Oxidative stress is found to involve in abnormal ECM metabolism, thereby promoting IDD progression[28]. Specifically, in type 2 diabetes (T2D), elevated glucose levels can trigger or worsen oxidative stress, accelerating the process of IDD.

The hallmark of IVD degeneration is the extracellular matrix decomposition. Rajasekaran et al. reported remarkable decrease of the structural and matrix proteins like BGN, COL2A1, DCN and VCAN in degenerated discs[8]. Our analysis revealed an even lower expression of DCN and VCAN in T2DM-IDD, suggesting even worse matrix disassembly in T2DM-IDD. IL17B, a pro-inflammatory cytokine, was also downregulated, suggesting a reduced inflammatory response in diabetic IVD tissues. Interestingly, vascular endothelial growth factor A (VEGFA), known for its roles in influencing microinflammation and abnormal neural elongation[29], also showed intriguing results. We found that VEGFA expression was lower in NP tissue of T2DM-IDD samples than in non-T2DM-IDD samples. While VEGF expression was found to be increased in an injured IVD rat model, suggesting that it may be clinically important in discogenic pain[30], VEGF vascularization pathway in human IVDs does not change during degeneration process[31]. These contradictory results suggest a new hypothesis that T2DM-associated IDD may trigger an alternative immune response to protect against disc degeneration, which requires further investigation.

Pathway enrichment analysis uncovered multiple pathophysiological factors such as oxidative phosphorylation, MYC targets, and fatty acid metabolism. The upregulation of oxidative phosphorylation and fatty acid metabolism pathways in diabetic IVD tissues indicates heightened metabolic activity and oxidative stress. The enrichment of MYC targets, E2F targets, G2M checkpoint and mitotic spindle not only suggests alterations in cell growth and proliferation in degenerated discs but also increased transcription and translation activities in cells, both of which may contribute to the accelerated IVD degeneration observed in T2DM patients.

Proline hydroxylation, a common post-translational modification in the ECM, was examined in our study. We identified 2,278 proline hydroxylation sites on 230 proteins, with significant enrichment in collagen and collagen-containing ECM proteins. Protein-protein interaction network analysis revealed a highly connected network among these proteins, particularly involving collagen, emphasizing the critical role of ECM modifications in diabetic IVD degeneration. Our findings align with previous research, such as studies by Sarath et al. [7]and Rajasekaran et al. [8], which identified distinct proteomic signatures in normal and degenerated human IVDs. These studies highlight the significance of inflammation and metabolic dysregulation as seen by increased expression of proteins related to oxidative stress, apoptosis, and immune defense mechanisms, indicating active inflammation and tissue degradation in disc degeneration. The identification of hypoxia and glycolysis pathways enriched in T2DM-IDD not only indicate nutrient deprivation in degenerated discs but also support the observation of oxidative stress which may contribute to higher hydroxylation in collagens in T2DM IDD. Similarly, Fu et al. identified critical proteins involved in IDD pathogenesis using TMT-based proteomic analysis, proposing potential biomarkers and therapeutic targets. Along with our findings, these studies suggest a shift in clinical practice from relying solely on MRI to incorporating proteomic analysis for a more accurate understanding of disc health and disease. This could lead to the development of molecular targets for preventive therapies, potentially reducing the need for surgical interventions.

Overall, our study produced point-reference datasets of T2DM-IDD and non-T2DM-IDD. The differences in protein composition hold potential for informing future research and aiding in the discovery of therapeutic targets for managing IDD in T2DM patients and the associated burden of chronic low back pain. Further investigation is needed to unravel the complex interplay between T2DM and IDD at the molecular level, paving the way for potential treatments and interventions.

Our study provides a detailed proteomic profile of NP tissue from diabetic and non-diabetic IDD patients, revealing significant differences in protein expression that underscore the complex interplay between T2DM and IVD degeneration. The identified proteins and pathways highlight potential mechanisms by which diabetes exacerbates IDD, including increased oxidative stress, altered ECM composition, disrupted cell signaling, and heightened inflammation. These insights pave the way for future research aimed at developing targeted therapies to mitigate the impact of T2DM on IVD degeneration and improve outcomes for patients suffering from chronic low back pain. Further investigation using complementary approaches such as immunohistochemistry and Western blotting is needed to validate these findings and explore their therapeutic potential.

Previous studies have shown that the development of IDD has been associated with multiple factors, such as age[32,33], sex[34], body mass index (BMI)[35,36], smoking[37–41], and other lifestyle factors[42]. Machine learning (ML) and deep learning (DL) have emerged as powerful tools for identification of biomarker for clinical applications; however, low sample sizes in most proteomics studies limit the usage of ML/DL in the proteomics data analysis[43]. Our future study will expand the collection of the IDD samples to enable further investigations of these additional confounding factors in the development of IDD in T2DM patients using ML/AL and identification of diagnostic and prognostic IDD biomarkers for clinical applications.

Limitations and Future Directions.

This study has several limitations. First, the sample size was modest ($n = 9$), which may limit statistical power and the ability to fully control for potential confounders such as degeneration severity, age, BMI, and other metabolic variables. Second, this investigation was based on discovery-driven quantitative proteomics

Patient sample	Age (year)/Sex	Group	Discectomy levels	Pfirrman Grade per level
1	28/M	non-T2DM-IDD	L5-S1	Grade IV
2	58/M	non-T2DM-IDD	L5-S1	Grade V
3	66/M	non-T2DM-IDD	L4-5	Grade V
4	45/M	non-T2DM-IDD	L5-S1	Grade V
5	88/F	T2DM-IDD	L4-5	Grade IV
6	57/M	T2DM-IDD	L5-S1	Grade V
7	72/M	T2DM-IDD	L4-5	Grade IV
8	56/M	T2DM-IDD	L4-5	Grade V
9	74/F	T2DM-IDD	L5-S1	Grade IV

Table 1. Clinical and imaging characteristics of patients with IDD from whom the disc samples were obtained.

without orthogonal validation (e.g., Western blotting or qRT-PCR) due to the unavailability of remaining tissue specimens following institutional transition. Therefore, the identified protein and pathway alterations should be interpreted as hypothesis-generating rather than definitive mechanistic conclusions. Third, functional assays were not performed to directly assess oxidative phosphorylation activity, oxidative stress levels, or the mechanistic basis of altered proline hydroxylation.

Future studies will involve prospective collection of a larger, clinically matched cohort of T2DM-IDD and non-T2DM-IDD specimens to enable targeted validation of prioritized proteins and pathways. Orthogonal validation using qRT-PCR, Western blotting, and targeted proteomics will be performed, along with mechanistic studies examining oxidative stress, mitochondrial function, and collagen-modifying enzymes to clarify causal relationships. Such investigations will strengthen the translational relevance of these proteomic findings and may identify actionable therapeutic targets for diabetic intervertebral disc degeneration.

Materials and methods

Surgical tissue collection

All samples were obtained from surgically treated degenerative pathology under a single tissue acquisition protocol. A total of 9 intervertebral disc specimens were collected for research use from 9 adult IDD patients undergoing lumbar discectomy between November 2020 and May 2022. Out of the 9 patients, 5 were identified as known cases of T2DM and referred to T2DM-IDD group, while the remaining 4 patients were non-T2DM and referred to as the non-T2DM-IDD group (Table 1). The study is limited by small sample size and the absence of a prospectively designed matching scheme for degeneration grade; therefore, residual confounding by severity cannot be excluded. The indication of discectomy for patients who consented and enrolled in this study was to treat symptomatic degenerative disc disease presenting with lumbar disc herniation or with spondylolisthesis. Trauma and patients presenting with infection or neoplastic process to the spine were excluded. Following a previously established protocol for tissue collection[24], the NP tissue was carefully dissected by a Mayo Clinic staff surgeon and obtained from the surgical waste material containing disc or endplate tissue that would be otherwise discarded during discectomy in the operating room. The NP tissue specimens were thoroughly irrigated with sterile saline to minimize blood contamination during sample preparation, snap-frozen in liquid nitrogen, and stored at − 80 °C until proteomics analysis. Collection of the specimens used in this study was approved by Mayo Clinic institutional review board minimal risk protocol (IRB 17-003293). Because surgical waste product was used, the need for informed consent for NP tissue was waived by the ethics committee. All patients provided written informed consent for their surgical procedure under standard clinical practice. We confirm that all methods were carried out in accordance with relevant guidelines and regulations in all the experiments.

Protein extraction and trypsin digestion

Frozen NP specimens were dried in a lyophilizer, and 50 mg of dried tissue were used for subsequent extraction with a mortar and pestle on dry ice. 1 ml of 5% SDS in 50 mM TEAB (triethylammonium bicarbonate, pH 8.5) was added to the tissue powder and probe sonicated at 60% power for 20 s three times. The extract was centrifuged at 4000 g for 30 min, and a BCA protein assay was performed. 200 µg of protein from each sample was taken for protein digest. 23 µl of 10% SDS in 100 mM TEAB was added. Samples were reduced using 2 µl of 500 mM TCEP (tris(2-carboxyethyl) phosphine) for 18 min in a 60 °C incubator, then alkylated with 2 µl of 500 mM iodoacetamide for 10 min in the dark. Samples were acidified with 5 µl of 55% phosphoric acid and loaded onto S-Trap mini cartridges (Protifi, Huntington, NY) that had been pre-equilibrated with wash buffer (90% methanol in 50 mM TEAB). The solution was spun in the S-Trap cartridge at 4000 g for 30 s and the flow through was discarded. The wash buffer was spun through the S-trap 5 more times. Then trypsin was added in a 1:10 trypsin: protein ratio and pulse spun to draw it into the cartridge. The samples were then digested overnight at 37 °C. The resulting peptides were eluted with washes of 50 mM TEAB, 0.2% formic acid, and 60% acetonitrile with 0.1% formic acid. Peptides were frozen, lyophilized, and reconstituted in 400 µl 100 mM TEAB for TMT labeling. After checking for 99% TMT labeling efficiency with a preliminary LC-MS run, peptides from each sample were mixed so that TMT reporter ion channels gave the same signal.

Basic pH reversed phase high performance liquid chromatography (bRP-HPLC)

The mixed peptides were acidified and desalted using a Sep-Pak Plus Light C_{18} cartridge and then fractionated by bRP-HPLC on a reversed phase C_{18} column (4.6 × 100 mm column) using an Ultimate 3000 UHPLC System. 96 fractions collected for a total run time of 80 min were concatenated into 24 fractions. These concatenated 24 fractions were dried down in a speed vacuum system and stored at -80 °C before LC/MS-MS analysis.

Liquid chromatography tandem mass spectometry (LC-MS/MS)

Each fraction of peptides was analyzed by LC-MS/MS on a high-resolution Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific). Briefly, each fraction of peptides were resuspended in 0.1% trifluoroacetic acid, 0.2% formic acid, and 0.002% zwittergent 3–16, loaded onto an Halo EXP 2.7 µm C_{18} trap column (100 mm × 2 cm, Fisher Healthcare) at a flow rate of 8 µl/min and separated on a PepSep RSCL C_{18} column (1.5 µm, 100 µm × 40 cm, Bruker) heated at 50 °C using a 150 min gradient by an Ultimate 3000 system (Thermo Fisher Scientific) coupled with a high-resolution Fourier transform Orbitrap Exploris 480 mass spectrometer (Thermo Scientific). Ionization of eluting peptides was performed using an EASY-Spray source kept at an electric potential of 2.2 kV. All mass spectra were acquired in a data-dependent acquisition (DDA) mode set as: MS1 survey scan data from m/z 340-1,600 at 120,000 resolution (at m/z 120), 300% $AGC_{,}$; MS/MS scan from m/z 110 at 45,000 resolution, a minimum precursor intensity of 50,000, isolation width of 0.7 m/z, 200% AGC target, max fill time of 100 ms, HCD collision energy of 32%, for precursor charge states of 2–4.

Mass spectrometric data analysis

The mass spectra were searched against a UniProt human protein database (2024_01 version) by Andromeda algorithm on the MaxQuant (ver. 2.2.0.0) proteomics analysis platform. The search parameters were set: carbamidomethylation on cysteine residues, TMTpro-16plex modification on N-terminal and lysine residues as fixed modifications; protein N-terminal acetylation, oxidation on methionine residues, hydroxylation on proline residues; a maximum of two missed cleavages. The data were searched against the target decoy database and the false discovery rate was set to 1% at the peptide level. The reported ion intensities were used to quantitate the changes of identified proteins and protein proline hydroxylation sites.

Bioinformatics analysis

The TMT reporter intensities were log_2-transformed, normalized and then used to calculate the fold-changes of each NP protein and proline hydroxylation site in T2DM-IDD patients compared with non-T2DM-IDD patients. Statistical significance of protein expression differences between groups was assessed using a Student's t-test, and the resulting p-values were adjusted for multiple hypothesis testing using the Benjamini–Hochberg false discovery rate (FDR) correction. Proteins were considered differentially expressed based on the FDR-adjusted p-values together with fold-change criteria. The gene ontology (GO) analysis of the identified proteins and proline-hydroxylated proteins were performed using PANTHER[44,45]. Heatmap of the differentially expressed proteins was created using the R package ComplexHeatmap[46]. Gene set enrichment analysis (GSEA) of NP proteins was conducted using the "fgsea' R package[47]. The analysis used genes ranked by log-transformed fold change and FDR-adjusted *p*-value obtained from protein expression comparison between T2DM-IDD and non-T2DM-IDD. The HALLMARK gene set collection from MSigDB 2023.2 database served as input files. The resulting GSEA outputs were visualized in R using the ggplot2 package[48]. Protein-protein interaction network among differentially expressed NP proteins in T2DM-IDD compared to non-T2DM-IDD and proline hydroxylated proteins were generated using stringApp (version 2.1.1) in Cytoscape (version 3.10.02)[49] against the STRING (physical) Database version 11.5 [50]. The protein-protein interaction clusters in differentially expressed NP proteins in T2DM-IDD compared to non-T2DM-IDD were generated using MCL Clustering (granularity parameter: 4) in the stringApp (version 2.1.1) in Cytoscape (version 3.10.02)[49].

Data availability

Data is provided within the manuscript or supplementary information files.

Received: 14 October 2024; Accepted: 16 March 2026
Published online: 26 March 2026

References

1. Global national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1789–1858. https://doi.org/10.1016/s0140-6736(18)32279-7 (2018).
2. Kushchayev, S. V. et al. ABCs of the degenerative spine. *Insights Imaging.* **9**, 253–274. https://doi.org/10.1007/s13244-017-0584-z (2018).
3. Oichi, T., Taniguchi, Y., Oshima, Y., Tanaka, S. & Saito, T. Pathomechanism of intervertebral disc degeneration. *JOR Spine*. **3**, e1076. https://doi.org/10.1002/jsp2.1076 (2020).
4. Hameed, I. et al. Type 2 diabetes mellitus: From a metabolic disorder to an inflammatory condition. *World J. Diabetes*. **6**, 598–612. https://doi.org/10.4239/wjd.v6.i4.598 (2015).
5. Liu, X., Pan, F., Ba, Z., Wang, S. & Wu, D. The potential effect of type 2 diabetes mellitus on lumbar disc degeneration: a retrospective single-center study. *J. Orthop. Surg. Res.* **13**, 52. https://doi.org/10.1186/s13018-018-0755-8 (2018).
6. Kakadiya, G., Gandbhir, V., Soni, Y., Gohil, K. & Shakya, A. Diabetes Mellitus-A Risk Factor for the Development of Lumbar Disc Degeneration: A Retrospective Study of an Indian Population. *Global Spine J.* **12**, 215–220. https://doi.org/10.1177/21925682209480 35 (2022).
7. Sarath Babu, N. et al. Quantitative proteomic analysis of normal and degenerated human intervertebral disc. *Spine J.* **16**, 989–1000. https://doi.org/10.1016/j.spinee.2016.03.051 (2016).

8. Rajasekaran, S. et al. Inflammaging determines health and disease in lumbar discs-evidence from differing proteomic signatures of healthy, aging, and degenerating discs. *Spine J.* **20**, 48–59. https://doi.org/10.1016/j.spinee.2019.04.023 (2020).
9. Rajasekaran, S. et al. Novel Biomarkers of Health and Degeneration in Human Intervertebral Discs: In-depth Proteomic Analysis of Collagen Framework of Fetal, Healthy, Scoliotic, Degenerate, and Herniated Discs. *Asian Spine J.* **17**, 17–29. https://doi.org/10.31616/asj.2021.0535 (2023).
10. Wang, R. et al. Inflammatory-sensitive CHI3L1 protects nucleus pulposus via AKT3 signaling during intervertebral disc degeneration. *FASEB J.* **34**, 3554–3569. https://doi.org/10.1096/fj.201902096R (2020).
11. Qiu, C. et al. Differential proteomic analysis of fetal and geriatric lumbar nucleus pulposus: immunoinflammation and age-related intervertebral disc degeneration. *BMC Musculoskelet. Disord.* **21**, 339. https://doi.org/10.1186/s12891-020-03329-8 (2020).
12. Fu, Y., Huang, X. Q., Qu, H. B., Ge, Y. Z. & Ru, X. L. Tandem Mass Tag-Based Proteomic Analysis of Normal and Degenerated Human Intervertebral Discs. *J. Pain Res.* **17**, 1313–1326. https://doi.org/10.2147/JPR.S449044 (2024).
13. Yang, X., Lu, Y., Zhou, H., Jiang, H. T. & Chu, L. Integrated proteome sequencing, bulk RNA sequencing and single-cell RNA sequencing to identify potential biomarkers in different grades of intervertebral disc degeneration. *Front. Cell. Dev. Biol.* **11**, 1136777. https://doi.org/10.3389/fcell.2023.1136777 (2023).
14. Rappu, P., Salo, A. M., Myllyharju, J. & Heino, J. Role of prolyl hydroxylation in the molecular interactions of collagens. *Essays Biochem.* **63**, 325–335. https://doi.org/10.1042/ebc20180053 (2019).
15. The Department of Economic and Social Affairs of the United Nations Secretariat. World Population Prospects The 2015 Revision.
16. Yelin, E., Weinstein, S. & King, T. The burden of musculoskeletal diseases in the United States. *Semin Arthritis Rheum.* **46**, 259–260. https://doi.org/10.1016/j.semarthrit.2016.07.013 (2016).
17. Rider, S. M., Mizuno, S. & Kang, J. D. Molecular Mechanisms of Intervertebral Disc Degeneration. *Spine Surg. Relat. Res.* **3**, 1–11. https://doi.org/10.22603/ssrr.2017-0095 (2019).
18. Mahmoud, M., Kokozidou, M., Auffarth, A. & Schulze-Tanzil, G. The Relationship between Diabetes Mellitus Type II and Intervertebral Disc Degeneration in Diabetic Rodent Models: A Systematic and Comprehensive Review. *Cells* **9** https://doi.org/10.3390/cells9102208 (2020).
19. Steelman, T. et al. Population-based Risk Factors for the Development of Degenerative Disk Disease. *Clin. Spine Surg.* **31**, E409–e412. https://doi.org/10.1097/bsd.0000000000000682 (2018).
20. Jin, P. et al. Diabetes and intervertebral disc degeneration: A Mendelian randomization study. *Front. Endocrinol. (Lausanne).* **14**, 1100874. https://doi.org/10.3389/fendo.2023.1100874 (2023).
21. Yang, F. et al. Role of Advanced Glycation End Products in Intervertebral Disc Degeneration: Mechanism and Therapeutic Potential. *Oxid Med Cell Longev* 7299005 (2022). (2022). https://doi.org/10.1155/2022/7299005
22. Illien-Junger, S. et al. Combined anti-inflammatory and anti-AGE drug treatments have a protective effect on intervertebral discs in mice with diabetes. *PLoS One.* **8**, e64302. https://doi.org/10.1371/journal.pone.0064302 (2013).
23. Li, S. et al. The Potential Role of Cytokines in Diabetic Intervertebral Disc Degeneration. *Aging Dis.* **13**, 1323–1335. https://doi.org/10.14336/ad.2022.0129 (2022).
24. Bydon, M. et al. Lumbar intervertebral disc mRNA sequencing identifies the regulatory pathway in patients with disc herniation and spondylolisthesis. *Gene* **750**, 144634. https://doi.org/10.1016/j.gene.2020.144634 (2020).
25. Rovira-Llopis, S. et al. Mitochondrial dynamics in type 2 diabetes: Pathophysiological implications. *Redox Biol.* **11**, 637–645. https://doi.org/10.1016/j.redox.2017.01.013 (2017).
26. Kang, L. et al. The mitochondria-targeted anti-oxidant MitoQ protects against intervertebral disc degeneration by ameliorating mitochondrial dysfunction and redox imbalance. *Cell. Prolif.* **53**, e12779. https://doi.org/10.1111/cpr.12779 (2020).
27. Xiong, C., Huang, B., Cun, Y., Aghdasi, B. G. & Zhou, Y. Migration inhibitory factor enhances inflammation via CD74 in cartilage end plates with Modic type 1 changes on MRI. *Clin. Orthop. Relat. Res.* **472**, 1943–1954. https://doi.org/10.1007/s11999-014-3508-y (2014).
28. Wang, Y. et al. Oxidative stress in intervertebral disc degeneration: Molecular mechanisms, pathogenesis and treatment. *Cell. Prolif.* **56**, e13448. https://doi.org/10.1111/cpr.13448 (2023).
29. Shibuya, M. Differential roles of vascular endothelial growth factor receptor-1 and receptor-2 in angiogenesis. *J. Biochem. Mol. Biol.* **39**, 469–478. https://doi.org/10.5483/bmbrep.2006.39.5.469 (2006).
30. Sato, J. et al. Vascular endothelial growth factor in degenerating intervertebral discs of rat caudal vertebrae. *Spine Surg. Relat. Res.* **2**, 42–47. https://doi.org/10.22603/ssrr.2017-0026 (2018).
31. Capossela, S. et al. VEGF vascularization pathway in human intervertebral disc does not change during the disc degeneration process. *BMC Res. Notes.* **11**, 333. https://doi.org/10.1186/s13104-018-3441-3 (2018).
32. Wen, P. et al. The role of ageing and oxidative stress in intervertebral disc degeneration. *Front. Mol. Biosci.* **9**, 1052878. https://doi.org/10.3389/fmolb.2022.1052878 (2022).
33. Vo, N. V. et al. Molecular mechanisms of biological aging in intervertebral discs. *J. Orthop. Res.* **34**, 1289–1306. https://doi.org/10.1002/jor.23195 (2016).
34. Asai, T. et al. Sex- and Age-related Differences in Spinal Degeneration: An Anatomical and Magnetic Resonance Imaging Study of the Human Spine. *Prog Rehabil Med.* **7**, 20220011. https://doi.org/10.2490/prm.20220011 (2022).
35. Coppock, J. A. et al. Increasing BMI increases lumbar intervertebral disc deformation following a treadmill walking stress test. *J. Biomech.* **121**, 110392. https://doi.org/10.1016/j.jbiomech.2021.110392 (2021).
36. Al-Zoubi, M. S. et al. Elevated BMI is considerably associated with IDD rather than polymorphic variations in interleukin-1 and vitamin D receptor genes: A case-control study. *J. Med. Biochem.* **40**, 129–137. https://doi.org/10.5937/jomb0-26367 (2021).
37. Battié, M. C. et al. Volvo Award in clinical sciences. Smoking and lumbar intervertebral disc degeneration: an MRI study of identical twins. *Spine (Phila Pa 1976)* 16, 1015–1021 (1991). (1991).
38. Akmal, M. et al. Effect of nicotine on spinal disc cells: a cellular mechanism for disc degeneration. *Spine (Phila Pa. 1976).* **29**, 568–575. https://doi.org/10.1097/01.brs.0000101422.36419.d8 (2004).
39. Mattila, V. M., Saarni, L., Parkkari, J., Koivusilta, L. & Rimpelä, A. Early risk factors for lumbar discectomy: an 11-year follow-up of 57,408 adolescents. *Eur. Spine J.* **17**, 1317–1323. https://doi.org/10.1007/s00586-008-0738-2 (2008).
40. Hanley, E. N. Jr. & Shapiro, D. E. The development of low-back pain after excision of a lumbar disc. *J. Bone Joint Surg. Am.* **71**, 719–721 (1989).
41. Glassman, S. D. et al. The effect of cigarette smoking and smoking cessation on spinal fusion. *Spine (Phila Pa. 1976).* **25**, 2608–2615. https://doi.org/10.1097/00007632-200010150-00011 (2000).
42. Pye, S. R., Reid, D. M., Adams, J. E., Silman, A. J. & O'Neill, T. W. Influence of weight, body mass index and lifestyle factors on radiographic features of lumbar disc degeneration. *Ann. Rheum. Dis.* **66**, 426–427. https://doi.org/10.1136/ard.2006.057166 (2007).
43. Mann, M., Kumar, C., Zeng, W. F. & Strauss, M. T. Artificial intelligence for proteomics and biomarker discovery. *Cell. Syst.* **12**, 759–770. https://doi.org/10.1016/j.cels.2021.06.006 (2021).
44. Thomas, P. D. et al. Making genome-scale phylogenetics accessible to all. *Protein Sci.* **31**, 8–22. https://doi.org/10.1002/pro.4218 (2022).
45. Mi, H. et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* **14**, 703–721. https://doi.org/10.1038/s41596-019-0128-8 (2019).
46. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313 (2016).

47. Korotkevich, G. et al. Fast gene set enrichment analysis. *bioRxiv* 060012. https://doi.org/10.1101/060012 (2021).
48. Villanueva, R. A. M. & Chen, Z. J. ggplot2: Elegant Graphics for Data Analysis, 2nd edition. *Meas-Interdiscip Res* 17, 160–167 (2019). https://doi.org/10.1080/15366367.2019.1565254
49. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. https://doi.org/10.1101/gr.1239303 (2003).
13. /11/2498 [pii].
50. Szklarczyk, D. et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368. https://doi.org/10.1093/nar/gkw937 (2017).

Author contributions

F M Moinuddin: Study design, data collection, data analysis and manuscript writing; Jun Zhong: Data collection, data analysis and manuscript writing; Maria D. Astudillo Potes: Study design, manuscript writing; Akhilesh Pandey: data analysis and manuscript revision; Mohamad Bydon: manuscript revision and supervision. All authors read and approved the final manuscript.

Funding

No funding was received for this study.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-44970-y.

Correspondence and requests for materials should be addressed to M.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.