INTEGRATIVE METHDOS FOR LINKING GENOTYPE TO PHENOTYPE: AN APPLICATION TO INTERVERTEBRAL DISC DEGENERATION

FRANCESCO GUALDI

TESI DOCTORAL UPF / 2024

Thesis supervisors

Dr. Baldomero Oliva Miguel

Dr. Janet Piñero

Department of Medicine and Life Sciences

Acknowledgements

I would like to start to thank my family, my mum, that always supported me every decision, no matter how difficult it was and my dad, despite not being physically here, still helps me during the challenging times.

A special thank goes to Elena, that shaped my character and was always by my side during most of this journey.

I would like to express my deepest gratitude to my advisors, Janet Piñero and Baldo Oliva for their continuous support, guidance, and encouragement throughout the course of this project. They let me free of doing what I liked yet providing priceless advices that guided my research.

I am also grateful to the members of my thesis committee, Laura Furlong, Adria Fernandez, and Joan Planas, for their valuable suggestions, time and careful evaluations.

I extend my thanks to my lifetime friends, Manu, Edo, Zinna, Ale, Giammi, Albi and Marco with which I shared education, experience and they are part of my past present and future. They've been always there for me during challenging times and I will for them.

I feel very lucky I spent these 3 years in Barcelona surrounded by amazing people, you really made this experience special.

My GRIB colleagues, Lili, Karo, Patri, David, Juanma, Paula, Marta, Jose Carlos, Cova, Miguel, Alejandro, Marta, Josema, Paula, Adrian, Xabi, Xoel and all of the 4th floor. I thoroughly appreciated every single lunch, beach volley, brunch and salida we had together.

The Disc4all fellows, Roger Mapo Dimitris Terence, Katherine, Paola, Andrea, Estefano, Sai, Francis, Nash, Ahmad, Sofia, and Ex. I will forever cherish our activities across Europe and US, collaborations and cultural exchanges.

The CRG people, Alessio, Grisha, Leo, Dami, Francesca, Zoe, Zach, Giuli, Mico, Luisa, Helene, Javi, Oz, Julia, Gregor and Carlotta. I enjoyed all the football games, bamboo sessions and all the laughter together.

Abstract

Intervertebral disc degeneration (IDD) is a complex condition influenced by genetics, environment, and lifestyle. Despite advancements in sequencing technologies, challenges remain in understanding the biology of complex diseases. The increasing use of artificial intelligence has led to innovative approaches in studying these conditions, benefiting translational medicine, drug discovery, and personalized treatment. This thesis introduces novel computational frameworks for studying the genetics of complex disorders, including a tool based on knowledge graph embeddings and Genopyc, a Python library for variant analysis. We exploited these tools in the study of biological underpinnings of IDD. Additionally, we investigated the autoimmune basis of Modic change, a comorbidity of IDD, using TwinsUK, the largest twin cohort worldwide. Our findings demonstrate the successful utility of these computational tools in advancing the understanding of IDD biology, emphasizing the importance of integrating diverse data sources to achieve a comprehensive understanding of this intricate condition.

Resum

La degeneració del disc intervertebral (IDD) és una condició complexa influïda per la genètica, l'entorn i l'estil de vida. Malgrat els avanços en tecnologies de seqüenciació, persisteixen els reptes en comprendre la biologia de malalties complexes. L'augment de l'ús de la intel·ligència artificial (IA) ha portat a enfocaments innovadors en l'estudi d'aquestes condicions, beneficiant la medicina translacional, la descoberta de medicaments i el tractament personalitzat. Aquesta tesi presenta nous marcs computacionals per estudiar la genètica de trastorns complexes, incloent-hi una eina basada en "knowledge-graph embeddings" (KGE) i Genopyc, una llibreria de Python per a l'anàlisi de variants. Hem aprofitat aquestes eines en l'estudi de les bases biològiques de l'IDD i hem investigat la base autoimmunitària del canvi de Modic, una comorbiditat de l'IDD, utilitzant TwinsUK, la cohort de bessons més gran del món. Els nostres resultats demostren la implementació reeixida d'aquestes eines computacionals en l'avenç de la comprensió de la biologia de l'IDD, posant èmfasi en la importància d'integrar fonts de dades diverses per assolir una comprensió completa d'aquesta condició complexa.

Preface

I've always been intrigued by getting to the heart of things because I believed that some hidden truths were there for us to be learned and to enrich our understanding of nature and ourselves. But uncovering these truths takes effort. It's only through learning and doing that we can grow as people and make our world a bit better by adding to our knowledge.

This thought motivated me when I undertook the career path to become a scientist. At that time, I had the choice to stay where I was, study a “safer” subject, and aim for a job at a company nearby my hometown. But at that time, I knew I wanted to understand the world better both by doing science, finding “the reason why”, and by travelling to gather diverse experiences. I was fascinated by the intricate mechanisms of life to study and I wanted to study them with a scientific approach. Years of university have been great, I studied some subjects that I liked and some I didn’t but surely everything was fulfilling, and I was happy no matter what, I found joy in learning and growth. Yet, in biology, what I found challenging yet fascinating was the inherent uncertainty, the unpredictability of findings, attributed to the complex and ever-changing nature of biological systems.

For this reason, when I discovered bioinformatics, I was compelled about the application of mathematical algorithms to biological data, for me it was a way to find the “reason why” of life, enclosing complex biological systems into well-defined boxes that we understand and of which we can predict the behaviour. The truth is that things are way more complicated than that and even if we made huge advancements in knowing the mechanisms of life we are still at the beginning of the puzzle of life.

However, this doesn’t mean the puzzle can’t be completed. It’s being completed every day in front of our eyes when we experience reality, it’s only a challenge in finding the right pieces to advance. While some might feel discouraged from the challenge deriving from the incomplete picture, I must admit I'm incredibly enthusiastic about the multitude of life's mysteries awaiting discovery. Just as we can't fully explain an imaginary world from a science fiction movie, reality remains equally inexplicable. The imaginary world can’t be experienced and explained, reality can be experienced but not explained.

And regarding experience, I am glad that I had the opportunity to live the bittersweet years of the PhD trying to find the “reason why” both during my research, by applying mathematical algorithms to biological systems, and outside the lab, where algorithms really don’t work most of the times. Everything surrounded by fellow researchers and humans, close to the sea in the vibrant city of Barcelona.

Contents

1. Introduction

1.1 Intervertebral disc degeneration

Intervertebral disc degeneration is a multifactorial condition that significantly impacts the health of the spine. The intervertebral disc, composed of a gel-like nucleus pulposus (NP) surrounded by a tough annulus fibrosus (AF), play a crucial role in providing flexibility, shock absorption, and stability to the spinal column. However, with aging and various contributing factors, the disc undergoes degenerative changes that compromise its biomechanical properties. These changes include alterations in the composition and structure of the extracellular matrix, such as a decrease in proteoglycan content and disorganization of collagen fibers, leading to reduced hydration, diminished disc height, and osteophytes formation. The progressive loss of disc integrity can lead to the development of pathological conditions like disc herniation, spinal stenosis, and facet joint osteoarthritis, leading to pain and impaired spinal function [1].

Several factors contribute to IDD, including genetic predisposition, biomechanical loading, lifestyle factors, and environmental influences [2]. Genetic predisposition plays a significant role in determining an individual's susceptibility to disc degeneration, with certain gene polymorphisms associated with an increased risk of developing degenerative disc disease. Additionally, repetitive mechanical loading and trauma, as well as poor posture and sedentary lifestyle habits, can accelerate disc degeneration by inducing microstructural damage and promoting inflammatory responses within the disc tissue. Furthermore, lifestyle factors such as obesity and smoking can impair disc metabolism, further incentivizing degenerative changes [3].

The clinical manifestations of intervertebral disc degeneration are very broad and can range from asymptomatic to debilitating, depending on the severity and location of the degenerative changes. Common symptoms include chronic low back pain (LBP), most of the time due to spinal stenosis. In fact, the loss of function of the disc leads to a degenerative process affecting the surrounding anatomical areas such as joints, muscles and ligaments resulting in the narrowing of the spinal canal and compression of the nerve tissue [4]. Another frequent manifestation strongly correlated with the LBP and IDD are Modic Change (MC). MC are bone marrow signal intensity changes classified in 3 types depending on how they appear in magnetic resonance imaging (MRI). The etiology of this condition is poorly understood, it has been theorized a multifactorial model in which genetics, trauma, inflammation and autoimmunity contribute to the development of the clinical symptoms [5]. The relationship between MC and LBP has been extensively reported in literature. For example, it was shown that especially MC type I, characterized by signal reduction in T1-weighted (T1w) MRI and a signal increase in T2-weighted (T2w) MRI are strongly correlated with more severe LBP outcomes [6]. MC are most common at L4-S1 spinal level and often adjacent to degenerated or herniated discs [7].

The current approaches for the treatment of IDD comprise conservative, interventional, and surgical approaches. Conservative treatment focuses on alleviating LBP and improving quality of life through methods like physical therapy, medication, and lifestyle adjustments. In cases of severe symptoms interventional treatments involve procedures such as Intra-Discal Electrothermal Therapy (IDET), radiofrequency myeloplasty, and ozone therapy to modify disc mechanics and manage pain. Surgical options, including intervertebral disc fusion, aim to provide pain relief and functional improvement by removing damaged discs, inserting support cages, and fixing vertebrae with pedicle screws. The choice

of treatment depends on the severity of symptoms and individual patient factors, with the goal of addressing IDD symptoms effectively and improving patient outcomes [8].

In recent years, advancements in regenerative medicine, tissue engineering, and biological therapies have provided promising avenues for the treatment of intervertebral disc degeneration. Strategies such as mesenchymal stem cell therapy, tissue engineering and gene therapy aim to enhance matrix synthesis, treat disc injuries and inhibit inflammatory processes within the degenerated discs. Despite these innovative approaches holding great promise for revolutionizing the treatment of degenerative disc disease, management of IDD remains a great challenge [9].
The treatments of IDD have been supported from a deeper insight into the biological landscape of this complex condition. In fact, current technologies have revolutionized our understanding of complex diseases by allowing the study of biological molecules on a large scale.

1.2 The genetic landscape of complex diseases

What is complex about complex diseases is the intricate interplay between genetics, environmental and lifestyle factors that contribute to the onset of these disorders. It was shown that these disorders don't follow typical patterns of Mendelian inheritance [10]. This is the result of various phenomena such as the already mentioned polygenicity, the gene-gene interaction (epistasis), gene-environment interaction, penetrance and phenotype definition. If attention is solely focused on the genetic component of complex disorders, the situation remains intricate. Indeed, many unanswered questions persist regarding how genetics contributes to the etiology of this group of disorders.

In recent years, the advent of omics technologies has significantly advanced our comprehension of complex diseases by offering a holistic perspective on the molecular terrain that underlies these conditions. Omics data can be integrated enabling the identification of disease-associated biomarkers, the characterization of cell types within disease-relevant tissues, patient stratification based on molecular profiles elucidating how the genetics is reflected on downstream biological pathways [11]. This led to the creation of extensive repositories to collect and harmonize information regarding genes associated to complex and Mendelian disorders such as DisGeNET [12].

The increasing amount of clinical data has facilitated the creation of extensive biological repositories which house genetic, phenotypic, and health data. For example, TwinsUK is one of the largest repositories of twins worldwide [13]. It focuses on the genetic basis of healthy aging and complex diseases such as cardiovascular, metabolic, musculoskeletal, and ophthalmologic disorders. It offers comprehensive 'omics' data, including genome-wide scans, next-generation sequencing, exome sequencing, epigenetic markers, gene expression arrays, RNA sequencing, telomere length measures, metabolomic profiles, and gut microbiomics. The integration of big data analytics with this repository enhances the discovery of new complex biological insights supporting the advancement in the healthcare sector.

1.3 A brief introduction to omics studies

With the term "omics" we consider disciplines such as genomics, proteomics, metabolomics, and transcriptomics. The study of "omics" represents a transformative approach in biology and medicine, offering comprehensive insights into the complex biological landscape of living organisms. At its core,

omics' studies aim to characterize and quantify the entirety of biological molecules within cells, tissues, or organisms, providing a holistic view of their composition, interactions, and dynamics [14].

Genomics, the study of an organism's entire genome, elucidates the genetic blueprint underlying its traits, diseases, and evolutionary history. Proteomics, on the other hand, explores the vast array of proteins present in biological systems, deciphering their structures, functions, modifications, and interactions to uncover the intricacies of cellular processes and signalling pathways. Metabolomics complements these efforts by profiling the diverse set of small molecules, or metabolites, produced by cellular metabolism, thereby offering insights into metabolic pathways, disease biomarkers, and environmental responses. Finally, transcriptomics focuses on the transcriptome, capturing the full spectrum of RNA transcripts expressed in a cell or tissue at a given time, unveiling gene expression patterns, alternative splicing events, and regulatory mechanisms.

By integrating data from these omics disciplines, researchers can understand the complexities of biological systems, from molecular mechanisms underlying diseases to the discovery of novel drug targets. This integrative approach holds immense promise for advancing personalized medicine, drug discovery, agriculture, environmental science, and beyond, ultimately driving innovations that benefit human health [15].

1.4 Genome Wide Association Studies

Genome-wide association studies (GWAS) stand as a testament to the remarkable progress made in genetics research over the last two decades. They trace their roots back to the early 20th century when geneticists first began exploring the inheritance patterns of traits in plants and animals. However, it wasn't until the completion of the Human Genome Project in 2003 [16] that the foundation for large-scale genome-wide studies was laid. This monumental effort provided researchers with a reference map of the entire human genome, enabling them to embark on the ambitious task of identifying genetic variants associated with complex traits and diseases.

In the years following the Human Genome Project, technological advancements in DNA sequencing and genotyping paved the way for the emergence of GWAS as a powerful tool in genetics research. The landmark study by Klein et al. in 2005 [17], which identified a genetic variant associated with age-related macular degeneration, marked the beginning of a new era in the study of human genetics. This pioneer study not only provided insights into the genetic basis of macular degeneration but also demonstrated the potential of GWAS to unravel the genetic architecture of complex diseases. Since then, GWAS have rapidly expanded in scope and scale, with researchers conducting studies involving hundreds of thousands or even millions of individuals. Subsequent studies have uncovered genetic associations for a myriad of conditions, ranging from cardiovascular disease and Alzheimer's disease to psychiatric disorders and autoimmune diseases.

Despite their success, GWAS have faced criticism and challenges. Critics have pointed out that the variants identified through GWAS often explain only a small fraction of the heritability of complex traits and diseases the so called "missing heritability" [18]. Additionally, issues such as population

stratification, sample size, and the functional interpretation of identified variants have posed challenges to researchers [19].

1.5 Beyond GWAS: from association to function

GWAS unravelled thousands of genetic loci offering unprecedented insights into the genetic basis of human traits and diseases. Nevertheless, the interpretation of the mechanistic contribution of genetic variants to the development of complex diseases is still a challenge. Many theories have been postulated to interpret why the variants that have emerged from GWAS don't explain the overall variance of a given trait. Due to the phenomena called linkage disequilibrium (LD) that causes the correlation of SNPs in the human genome, the signals coming from GWAS could be only the tip of the iceberg. Then, many other SNPs associated to the ones detected from GWAS but that cannot be detected from the study due to absence of power, could explain the variance of the trait. Another theory could be that GWAS could be related to rare, high penetrance SNPs that segregate in the population and that play a role in the complex trait or just a phenotype that contributes to the global disease [20].

A great challenge in studying the genetic underlying complex disease is accurately defining a phenotype. In fact, complex traits are usually a set of clinical manifestations that occur jointly. For example, IDD can be conceived as a mixture of disc narrowing, osteophytes presence, and imaging signal that could have different biological underpinnings. A possible approach could be to break down complex traits into endotypes that are closer to the biology underlying the trait and are more useful to detect variants that are exclusively related to those functions [21].
Missing heritability and accurately defining a suitable phenotype are not the sole hurdles faced in GWAS. Another significant challenge lies in interpreting the outcomes and comprehending how the identified variants impact downstream biological pathways, as well as understanding the functional consequences of genetic variations. It's noteworthy that, in many instances, SNPs prioritized via GWAS are located in non-coding regions, indicating their probable regulatory role.

In the past, a prevalent method to link the SNPs to their functional consequence involved designating the nearest upstream or downstream gene as the target gene. However, in recent years, there has been an emergence of a more thorough analysis regarding the functional consequences of specific variants. In fact, relying on physical proximity between the variant and the gene considered affected by the variant can be misleading as SNPs can affect gene expression across extensive genomic distances [22]. Research utilizing expression quantitative trait loci (eQTL) data indicates that approximately two-thirds of the genes causally linked to GWAS loci are not the closest ones [23]. Consequently, we find ourselves transitioning into what could be termed as the post-GWAS era [24].

Pinpointing the true causal variant, targeting the gene whose function is affected from a particular SNP and understanding the dysregulation of biological pathways that eventually lead to a condition are vital for understanding disease etiology, and for identifying potential therapeutic targets in an optic of personalized medicine. For these reasons, a number of heterogeneous datasets and analyses must be employed to understand the effect of the tagged GWAS SNPs. [25]

The investigation of non-coding SNPs requires a deep understanding of the genomic landscape. The regulation of gene expression is a complex tissue-specific mechanism that is regulated by many

elements such as enhancers, suppressors, microRNAs (miRNAs), transcription factors and features such as chromatin conformation and accessibility [26].

With the advancements of sequencing technology many repositories that collect data relative to the regulatory genome have been created. Examples of the more popular repositories are listed below.

- Ensembl [27]: Provides comprehensive and up-to-date genomic information for a wide range of species, including humans and other vertebrates, as well as model organisms and non-model organisms.

- ENCODE (Encyclopedia of DNA Elements) [28] : A project aimed at identifying all functional elements in the human genome, including regions involved in gene regulation, chromatin structure, and transcription.

- Roadmap Epigenomics Project [29]: Provides epigenomic data from various human tissues and cell types, including DNA methylation, histone modifications, and chromatin accessibility.

- GTEx (Genotype-Tissue Expression) [30]: Characterizes gene expression patterns and regulation across different human tissues, providing insights into tissue-specific regulatory mechanisms.

- FANTOM (Functional Annotation of the Mammalian Genome) [31] : Identifies and annotates enhancers, promoters, and other regulatory elements across the human genome, with a focus on transcriptional regulation.

- RegulomeDB [32]: Integrates data on regulatory elements, such as transcription factor binding sites and DNA motifs, with genomic annotations and functional annotations.

- TRANSFAC [33]: A database of transcription factors, their binding sites, and regulatory elements, along with information on their interactions and functional annotations.

- JASPAR [34]: A database of curated, non-redundant transcription factor binding profiles derived from experimental data, facilitating the study of transcriptional regulation.

- 3CDB (Chromatin Conformation Capture DataBase) [35]: A repository for storing and sharing data generated from chromatin conformation capture experiments (3C). These experiments involve the capture of spatial interactions between genomic regions within the three-dimensional structure of the DNA.

Additionally, a series of techniques/approaches have been developed that exploit this wealth of data. Some of the key methods include:

SNP enrichment analysis helps to understand whether specific categories of genetic variants are more prevalent among SNPs associated with the trait of interest compared to what would be expected by

chance. SNP enrichment can detect if the variants from a GWAS are enriched in a specific cell type, present specific consequences or affect genes with a determined function [36].

Fine mapping aims to prioritize variants within each genetic locus pinpointed by GWAS, with a focus on those with a higher likelihood of being directly linked to the target phenotype. It utilizes patterns of linkage disequilibrium and association statistics. Fine mapping methods may include various techniques such as SNP selection based on p-value or LD thresholds, regression analysis, or Bayesian statistics [37].

Colocalization analysis: Is the statistical process used to investigate whether two or more traits or diseases share a common genetic basis at a particular genetic locus. Many different methods have been produced to carry out this analysis based on different statistic tests [38]. When applied to the interpretation of the functional effects of the variants, colocalization between a variant associated to a trait and an eQTL of the variant is a powerful method to detect likely target genes [39]

Variant Annotation [40], [41]*:* Is the process in which variants are investigated in relation to functional genomics aspects. Several characteristics of the variants can be investigated: often prediction or assessments of the functional impact of a variant on gene structure or function is studied.

- **Variant effect:** This may involve predicting variants as synonymous (not affecting the encoded amino acid), missense (changing a single amino acid), nonsense (creating a premature stop codon), frameshift (altering the reading frame), or splice site disrupting (affecting mRNA splicing).
- **Conservation scores** indicate the degree of evolutionary conservation of a genomic region across different species. Highly conserved regions are more likely to be functionally important, and variants occurring in these regions may have a greater likelihood of affecting biological function.
- **Functional genomics**: Annotations may include information on the functional genomic context of a variant, such as its location within regulatory elements (e.g., enhancers, promoters) or its potential effects on transcription factor binding sites.

Many tools that rely on different principles have been created to predict the effect of genomic variants such as variant effect predictor (VEP) [42] CADD [43] SIFT [44] Polyphen [45] and UNET [46].

The abundance of data and analysis pipelines has led to the creation of numerous tools for conducting post-GWAS analysis. These tools aim to integrate different analyses and data in order to reach a greater understanding on how non-coding variants affect the downstream biological pathways.

- **FUMA** (Functional Mapping and Annotation of GWAS) [47]: : a web-based platform that performs functional annotation and mapping of GWAS results. It integrates data on gene expression, functional annotations, protein-protein interactions, and pathway analysis to prioritize candidate genes and biological pathways implicated by GWAS signals.

- **MAGMA** (Multi-marker Analysis of GenoMic Annotation) [48]: : a tool for gene-based analysis of GWAS results. It aggregates SNP-level association statistics to identify genes or gene sets that are significantly associated with a phenotype, accounting for linkage disequilibrium and gene size.

- **GCTA** (Genome-wide Complex Trait Analysis) [49]: a software package for analyzing complex traits using GWAS data. It provides tools for estimating heritability, conducting genetic correlation analysis, and performing genome-wide association analysis with mixed linear models.

- **Finemap** [50]: is a tool used to identify and prioritize candidate causal variants within a genomic region associated with a complex trait or disease. The building blocks of this method are a likelihood function, priors, efficient likelihood evaluation and efficient search algorithm.

- **Coloc** (Colocalization Analysis for GWAS) [51]: is a tool used to assess the likelihood that two or more traits or diseases share a common genetic signal at a particular genomic locus. Specifically, Coloc evaluates the probability that the same genetic variant(s) influence both traits (diseases and eQTLs), indicating potential genetic overlap or colocalization.

1.6 In-silico investigation of complex diseases

In recent years, alongside the advancement of GWAS, computational methodologies have emerged to explore and elucidate the impact of DNA sequences on gene expression and regulatory mechanisms. These methods analyze large-scale biological data in order to identify possible biomarkers associated with disease, physiological states or response to treatment [52].

Through omics technologies we assess cellular features challenging to interpret. However, the complexity of analyzing this data is multiple; not only the volume of data is huge, but also identifying a mathematical framework that can efficiently represent this complexity in a scalable manner that can be interpreted computationally, presents a significant obstacle. This collaborative effort, involving contributions from biology, mathematics, and computer science, aims to personalize treatments and enhance health outcomes but poses formidable challenges [53].

In biology and nature in general, entities do not exist in isolation but are integral components of larger, interconnected systems. Numerous examples illustrate this concept: people engage in daily interactions, planets and stars interact through gravitational forces, and atoms form chemical bonds. Therefore, examining entities within the context of broader interacting systems is essential for comprehending their functions.

As far as it concerns IDD, in the last decade many advancements have been made in the study of risk factors, phenotype definition and genetic influences [54]. However molecular in – silico investigation of the disease relies solely on the implementation of network analysis mainly on protein – protein interaction networks [55], [56]. Thus, the connection between genetic variability, biological pathways and phenotypic manifestations is far from being reached.

1.7 A brief introduction to graph theory

Complex systems wherein entities interact through specific relationships can be represented as networks. A network or graph (G) can be defined as a pair $G = (V,E)$, where V represents a set of elements known as nodes (or vertices), and E denotes a set of paired nodes, with its elements referred to as edges (or links). Thus, entities can be depicted as nodes, and relationships as edges connecting these nodes. Edges in a network are considered directed when interactions possess a specific direction, moving from a source to a target. Differently, edges in the network are undirected when interactions lack a specified direction.

Every node in a specific network has a degree that refers to the number of edges incident to a particular node, indicating its level of connectivity within the network. Nodes with a high degree are often considered important hubs or central points of connectivity. Conversely, nodes with a low degree may serve as peripheral or isolated components. Additionally, the neighbourhood of a node encompasses its immediate vicinity within the network, consisting of neighbouring nodes directly connected to it.

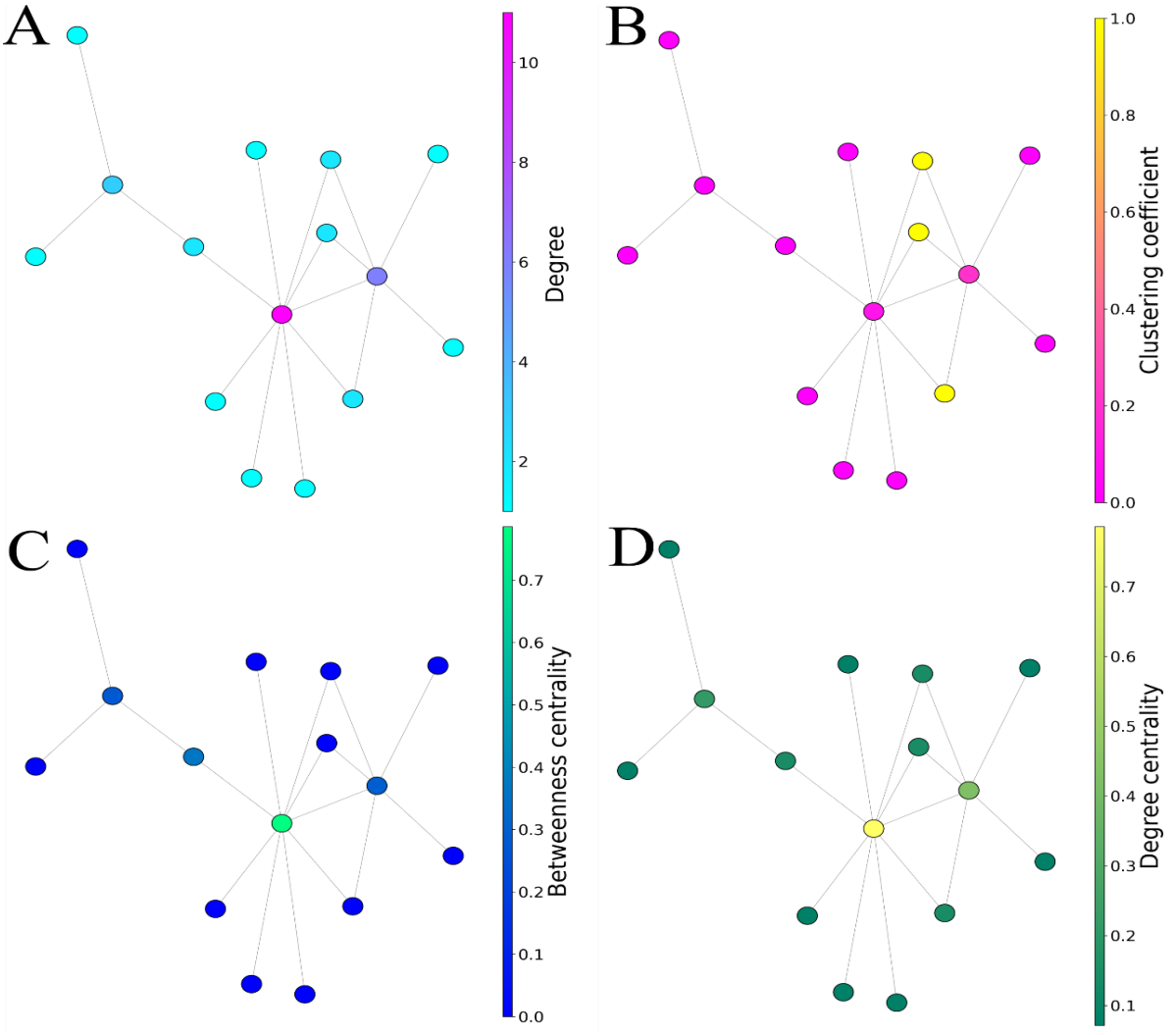


Figure 1 Characteristics of the networks; A node degree, B clustering coefficient, C Betweenness centrality and D degree centrality

Figure 1 summarizes some of the important network properties such as the degree (Panel A), clustering coefficient, which quantifies the degree to which nodes in the network tend to cluster together and form local communities or clusters (Panel B), betweenness centrality (Panel C) and degree centrality, which identify important nodes based on their position and interconnectivity in the network (Panel D). Other important metrics are modularity, which quantifies the degree to which a network can be divided into distinct, densely connected communities or modules and path length, which measures the average distance between pairs of nodes in the network and indicates its overall connectivity. An important feature of networks is degree distribution, which describes the distribution of node degrees within the network and provides information about the connectivity of the nodes in the network. Often biological networks follow a scale-free degree distribution in which node degree follows a power-law distribution [57]. In such networks, most nodes have relatively few connections (low degree), while a small number of nodes, known as hubs, have a disproportionately high number of connections. This results in a highly skewed distribution (Figure 2).

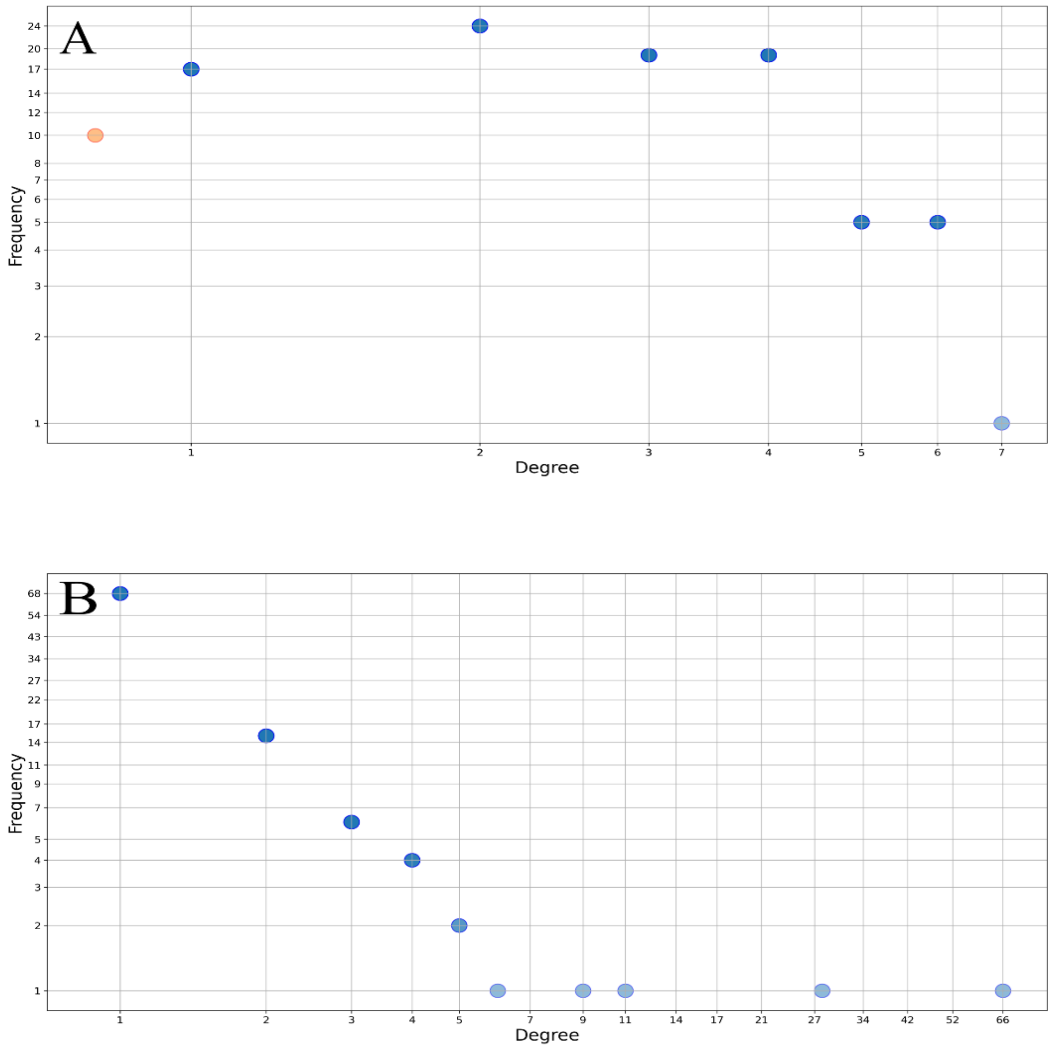


Figure 2 Degree Distribution of 2 different networks with the same number of nodes and edges. In the panel A is depicted the degree distribution of a random network while in panel B the degree distribution of a scale free network

Finally, a subgraph in network theory refers to a subset of nodes and edges that are derived from a larger, original graph. By extracting a subgraph from a larger network, researchers can focus on specific regions or components of interest, allowing for more detailed examination and analysis. Subgraphs can represent various aspects of network topology, providing insights into local interactions and functional relationships within the network.

1.8 Investigating Human Diseases through biological networks

Analysis of biomedical data can help to understand a*etiology* and causes of complex conditions. The interplay between biological entities can be represented as biological networks where different entities interact through different types of relationships. In the biomedical field multiple types of networks can be constructed each of one having nodes and edges reflecting biological entities and relationships (refer to Table 1 for a summary of the primary types of biological networks).

Table 1: Types of biological networks

Network Type	Description	Nodes	Possible edges
Gene Regulatory Networks (GRNs)	Model the regulatory interactions between genes, transcription factors, and other regulatory molecules.	Proteins, genes, RNAs	Activation, suppression, silencing
Protein-Protein Interaction Networks (PPIs)	Represent physical interactions between proteins within a cell.	Proteins	interaction
Disease Networks	Represents relationships between associated diseases (comorbidities) and different phenotype associated to a specific condition	Diseases	Comorbidity, phenotype, correlation
Drug-Target Interaction Networks	Model the interactions between drugs and their target molecules (e.g., proteins).	Drugs and proteins	Inhibitor, antagonist, blocker
Biological Knowledge Graphs	Integrate multiple types of biological data to infer functional relationships between genes, proteins, or other biological entities.	Multiple biological entities	Heterogeneous biological interactions
Co-expression Networks	Constructed based on the correlation patterns of gene expression across different experimental conditions or samples.	Genes	Correlations in the expression levels
Signalling Networks	Model the flow of molecular signals within cells, including pathways such as cell signalling cascades, signal transduction, and cellular communication.	Proteins	Activation, inhibitions, phosphorylation

With the advancements in high throughput technologies, the Human Protein Reference Database (HPRD) [58], a pioneering effort to systematically catalog and annotate PPI in humans was established. By providing a comprehensive resource of experimentally validated protein interactions, HPRD

facilitated deeper insights into the complex networks that govern cellular processes. This laid the basis for subsequent studies in systems biology, leading to a more holistic understanding of biological systems and their dysregulation in disease states.

With the increasing number of repositories that were created there was the necessity of the integration of the different resources. Moved from this necessity databases such as BIANA a comprehensive repository integrating multiple biological entities and relations in the form of network enabling the possibility of inferring new biomolecular relationships exploiting similar features of the biological entities [59].

Subsequent studies demonstrate that proteins linked to similar diseases tend to directly interact and form clusters within the same regions of the interactome. An early study in 2007 [60] showed that proteins encoded by genes associated with similar diseases exhibit a propensity to interact with one another, suggesting the presence of distinct functional modules within the interactome associated to the diseases. Expanding on this insight, they pioneered the construction of the inaugural human disease network, termed the human diseasome. This network was established by linking diseases with shared genetic components, with disease-gene associations extracted from the Online Mendelian Inheritance in Man (OMIM) database [61].

The approaches of identifying diseases modules can be roughly classified into 2 classes:

prior knowledge identification: This category includes methods utilizing prior knowledge about disease-associated genes, often referred to as seed genes. These methods focus on identifying the vicinity of proteins that exhibit closer topological relationships with those encoded by the seed genes.

“Ab-initio” identification: This category comprises methods that identify modules "ab-initio" by employing community structure detection algorithms. These approaches, relying on the network's topology, pinpoint neighborhoods of proteins characterized by high within-edge density of connections.

Different algorithms can be implemented to investigate disease modules (diffusion state distance, kernel clustering, modularity optimization and random walk-based approaches). In [62] different algorithms were applied to a network constructed by compiling a panel of diverse human molecular networks. These networks were extracted from various databases and sources to provide a benchmark for the comparison of the algorithms. The identified disease modules were shown to belong to biological pathways relevant to the disease which comprised therapeutic targets. This work benchmarked the application of network medicine to complex disease and showed that the integration of heterogeneous networks leads to the discovery of complementary types of modules and is thus beneficial.

Network based approaches were successfully implemented to discover functional modules for many complex diseases. In [63] researchers utilized advanced Network Medicine methodologies, specifically DIseAse MOdule Detection (DIAMOnD) and SWItch Miner (SWIM), to identify a novel gene

signature associated with Alzheimer's disease. By starting with a set of 99 known Alzheimer's disease-associated genes (seed proteins), DIAMOnD predicted 238 novel putative disease genes by analyzing their interactions in the human interactome. Through the integration of SWIM, 14 additional genes were identified that had significant interactions with the seed proteins, forming a statistically significant disease module. This novel gene signature, comprising the 99 seed proteins and the 14 additional genes, provides valuable insights into the molecular determinants of Alzheimer's disease and may offer new diagnostic biomarkers and therapeutic targets for this neurodegenerative condition.

The algorithms can be embedded into webservers to facilitate the user experience, allowing also non – expert coders to benefit from their use. An example of this is Guildify [64] a web – server with the scope of prioritizing disease associated genes through network – based algorithms. The output of the model is a score reflecting the relevance to the phenotype of interest and can be used to short-list the set of candidate genes.

In [65] through the analysis of differentially expressed genes in metastatic cancer and the identification of their subnetwork in the interactome, the study detected a set of novel hub genes that are significantly associated with liver metastasis in gastric cancer patients. Subsequent analysis revealed the involvement of these hub genes in key biological pathways implicated in cancer progression and metastasis. The findings from this study have important clinical implications, as they provide potential biomarkers for predicting metastatic risk and offer new targets for therapeutic intervention tailored to gastric cancer patients with liver metastasis.

Finally, in [66] by calculating the distance between diseases modules in the interactome the authors investigated the molecular relationship between complex diseases. Through this analysis the study revealed that diseases with overlapping modules have shared molecular basis and helped in identifying novel disease-disease relationships and uncover potential therapeutic targets shared across seemingly unrelated disorders. Thus, network models can capture the complexity of molecular interactions that bridge comorbidities providing an explanation on how common pathways are dysregulated and contribute to the etiology of certain conditions.

Earlier studies like the ones cited before relied mainly on protein-protein interaction data. However, remaining bounded to a single omics source offers only a small portion of the global picture on the biology underlying complex disorders and methods to integrate different types of data holds promise for unravelling complex disease mechanisms in an optic of advancing precision medicine. Precision medicine tailors healthcare to individual patients by exploiting all clinical and molecular data available to optimize treatment effectiveness and minimize adverse effects [67]. A careful analysis and integration of different sources of biological data presents a challenge due to differences in format, scale, and quality. Harmonizing these data requires advanced computational methods and interdisciplinary expertise. Additionally, the dynamic nature of biological systems necessitates integrating temporal and spatial information, thus addressing technical and methodological hurdles is crucial to harness the full potential of biological networks in studying complex diseases.

Such integrative analysis was made possible thanks to projects such as the Cancer Genome Atlas (TCGA) who provided a comprehensive multi - omics landscape of various cancer types. Initiated by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) in 2006, TCGA aimed to characterize the molecular alterations underlying cancer development and progression. Through extensive multi-platform genomic analyses, including whole-genome sequencing, whole-exome sequencing, DNA methylation profiling, RNA sequencing, and proteomics, TCGA has generated vast amounts of data from thousands of tumour samples across numerous cancer types. These data have facilitated the identification of key genetic mutations, gene expression patterns, epigenetic alterations, and molecular subtypes associated with different cancers, shedding light on the heterogeneity and complexity of the disease.

An example of heterogeneous data integration is provided in [68] where the authors researched disease candidate metabolites using a network method that integrates multi-omics information. The developed method (MetPriCNet) is based on a random walker with restart to prioritize candidate metabolites based on their global distance similarity with seed nodes in a composite network, which integrates multi-omics information from the genome, phenome, metabolome, and interactome. The method was tested on 87 phenotypes with a total of 602 metabolites and achieved a high AUC value of up to 0.918. It was also tested on 18 disease classes, with 4 classes achieving an AUC value over 0.95. MetPriCNet demonstrated effectiveness for diseases without known disease metabolites and was able to predict new high-risk metabolites for certain disease.

Network-based approaches exhibit promising outcomes across numerous biomedical domains, including gene prioritization and biomarker discovery in complex diseases. Furthermore, the rapid advancements in artificial intelligence (AI) have led to exponential growth in computational methodologies aimed at exploring how genetics influence the development of these conditions. These AI-driven computational approaches offer enhanced capabilities to detect complex relationships within omics data, thereby increasing our understanding of disease mechanisms and facilitating the discovery of personalized medicine strategies.

1.9 AI in genomics

Artificial intelligence (AI) has revolutionized the field of genomics by offering powerful computational tools and algorithms for analyzing vast amounts of genomic data [69]. With the rapid advancements in sequencing technologies, genomics has entered the era of big data, presenting both challenges and opportunities for researchers AI techniques, such as machine learning (ML), deep learning (DL), have been increasingly applied to genomics to extract meaningful insights from complex datasets [70]. These AI-driven approaches enable the identification of genetic variants, regulatory elements, and disease-associated genes, facilitating the understanding of genetic mechanisms underlying complex conditions [71], [72], [73].

AI methods can be broadly classified into unsupervised and supervised methods. Supervised learning involves training a model on a labelled dataset, where each input is associated with a corresponding target output. The model learns to map inputs to outputs by minimizing the discrepancy between its predictions and the true targets. This type of learning is used for tasks like classification and regression, where the goal is to predict discrete labels or continuous values[74].

In contrast, unsupervised learning operates on unlabelled data, seeking to find inherent structures or patterns within the data without explicit guidance. Algorithms in unsupervised learning aim to discover relationships, clusters, or distributions in the data [75].

Machine learning models can be *generative* or *discriminative*. If the primary focus of the application is to understand which variable led the model to decide for a specific outcome i.e. the interpretability of the model, *generative* model is preferred. These models learn a joint distribution between features and labels, learning the underlying structure of the data and can generate new samples that closely resemble the training data distribution. On the other hand, *discriminative* model in which the only aim is to predict new labels generating a conditional distribution [76].

Machine learning in genomics can be implemented in a multitude of different tasks. For instance, to acquire the ability to identify cis-regulatory elements such as transcription start sites (TSSs) within a genome sequence. Similarly, algorithms can be instructed to detect splice sites, promoters, enhancers, or positioned nucleosomes. Generally, if there exists a compilation of sequence elements of a specific type, it's likely that a machine learning approach can be trained to recognize those elements [77]. Moreover, by combining models that recognize individual types of genomic elements with learned logic regarding their relative positions, it becomes possible to construct machine learning systems capable of annotating genes across entire eukaryotic chromosomes [78].

Deep Learning DL is a subset of machine learning focused on neural networks with many layers, enabling automatic learning and improvement from large amounts of raw data. These layers are composed of neurons using activation functions like ReLU, sigmoid or thanh to process the information. Neural network have achieved substantial improvements for tasks such as image recognition, natural language processing and speech recognition [79], [80].

The initial successful applications of neural networks in regulatory genomics involved substituting a traditional machine learning approach with a deep model while keeping the input features unchanged. For instance, Xiong et al. [81] utilized a fully connected feedforward neural network to forecast the splicing activity of individual exons. They trained the model using over 1,000 predefined features extracted from the candidate exon and adjacent introns. Despite the relatively limited training data of 10,700 samples combined with the model's complexity, this method significantly enhanced the accuracy of splicing activity prediction compared to simpler approaches. Moreover, it effectively pinpointed rare mutations associated with splicing dysregulation.

1.10 AI on multi-omics

Building predictive models that work on heterogeneous biological data poses several challenges that can limit the potential of this approach. One major challenge is the non-uniform missing data across different omics datasets, which can complicate the integration process and lead to biased results. Additionally, the diverse signal-to-noise ratios in various omics data types can affect the quality and reliability of integrated analyses.

Poor biological interpretation of integrated results is another challenge, as the complexity of multi-omics data may hinder the extraction of meaningful biological insights. Inconsistent sample annotation

and lack of standardized protocols for multi-omics experiments further complicate data integration efforts.

Addressing these challenges requires the development of integration-aware methods for data imputation, efficient computation, and robust biological interpretation, as well as the establishment of community standards for multi-omics data collection and analysis.
An important aspect of AI models applied to heterogeneous data is the integration strategy of the different types of omics, there are roughly 3 approaches in which data can be integrated [82]:

Vectorizing each type of data: for example, in [83] the researchers after collecting gene expression, sequence patterns, PPI data etc., represented them using different matrices where the rows represented genes and columns had a different meaning depending on the types of data.
Kernel transformations: data can be transformed before being fed to the model by the implementation of kernel functions. The output of this transformation is a fixed length vector for each data type and, depending on the function, prior knowledge can be encoded in the transformation e.g. the kernel function could include the correlation between data types.
Probability-based approaches: In this model, diverse data types are depicted within a probabilistic framework inherent to the model itself. An instance of this utilization could be seen in the work by Troyanskatya et. al [84] with the implementation of Bayesian networks that merges data from gene interactions, transcription factors, and gene expression patterns, the network can make more accurate predictions about how genes work together by generating posterior probabilities of genes being involved in the same biological process. This approach helps improve the accuracy of gene function predictions by considering information from multiple sources in a systematic way.

These methods play a crucial role in extracting valuable insights from multi-omics data and advancing our understanding of complex biological systems.

1.11 AI applied to networks

As already mentioned in this thesis, complex systems can be represented as networks that offer a simple and direct way to integrate different sources of data. AI applications to network biology spans to understand disease biology to drug repurposing and could help to investigate alterations in the relationships among various biomolecules in a disease state versus healthy state. This will lead to infer new knowledge allowing the discovery of new biomarkers, treatments, and patient stratification [85].

Various categories of algorithms can be employed for graph analysis, each based on distinct frameworks. Table 2 presents the primary categories of these algorithms along with some illustrative examples. Despite their differences, these graph AI algorithms share common characteristics that persist across various algorithmic classes:

- The generation of node representations in the latent space, namely node embeddings
- The necessity of a task definition for training the model. The tasks can be general and applicable to many different algorithms, such as node classification or link prediction, or they can be very algorithmic specific such as the minimization of an ad-hoc loss function.

Table 2: Classes of graph learning algorithms

Algorithm classes	**Description of the Class**	**Algorithm Examples**
Random Walk-based Methods	Generate embeddings by simulating random walks on the graph and learning embeddings based on the sequences of visited nodes.	Node2Vec[86], DeepWalk [87], Metapath2Vec[88]
Tensor decomposition	Decompose the graph adjacency matrix or related matrices into lower-dimensional representations, which serve as node embeddings.	ComplEx [89], DistMult[90]
Deep Learning Methods	Utilize neural network architectures to learn embeddings by optimizing objective functions that capture structural properties of the graph.	Graph Convolutional Networks (GCNs) [91], Relational Graph Convolutional Networks (RGCN) [92]
Translational Models	Represent relationships between nodes in a graph as translations in a latent space, modeling connections through addition or subtraction operations.	TransE [93], RotatE [94]

AI can be applied to complex systems in order to reach a greater understanding of the interplay of different biological entities and infer new possible interactions at different levels, from gene regulations to species interactions.

The main tasks of graph learning models related can be classified into 3 categories [95]:

- **Node level:** the most common task at node level is node classification. In this task the model is trained to predict whether the node belongs to a specific class. This task can be performed in a supervised, unsupervised and semi-supervised fashion.

- **Edge level:** predicting edges in a network is applicable to many different biological fields. Interaction, association, activation, inhibition are different types of interactions in complex biological systems that are of interest in bioinformatic research. An important aspect to consider when predicting a connection is the directionality and quality of the predictions. Conceiving models that can handle these aspects is important for a better understanding of biological networks in which a single interaction can be of many different types.

- **Graph level:** this task is mainly related to the generation of graphs. This is particularly important in the context of drug development in which the interest is to generate chemical compounds (that can be represented as graphs) that have similar properties of a given molecule.

These algorithms have been successfully applied to biological networks; in a work by Jha et al [96] a graph neural network was implemented in order to predict PPI. The study proposes two graph-based architectures, GCN-based and GAT-based, to learn features from protein representations by integrating spatial structure and sequence features. The methodology consists of three modules: protein graph construction, feature extraction, and a classifier to predict interactions. The analysis includes the construction of sequence embeddings generated using a language model (LSTM-based) as node features. The experiments were conducted using PPI datasets from human and yeast , showing promising results in predicting protein interactions.

In another study, Gao et al [97] proposed a KG based disease-gene prediction system called GenePredict-KG. They constructed a comprehensive knowledge graph with over 2 million associations between various entities related to genotypic and phenotypic information. By developing a knowledge graph embedding model, they were able to learn low-dimensional representations of entities and relations, which were then used to predict new disease-gene interactions. The architecture of the model was based on 2 modules, an encoder and a decoder. The encoder utilized a composition-based multi-relational graph convolutional network to acquire representations of entities and relationships. In the decoder module, the InteractE model was deployed to assess unseen interactions, with those receiving higher scores being more likely to be accurate. GenePredict-KG outperformed several state-of-the-art models in terms of performance metrics such as AUROC, AUPR, and MRR, showcasing its effectiveness in inferring disease-gene associations. The study highlights the importance of leveraging semantic relationships from diverse biological databases to enhance disease-gene prediction and offers valuable insights for understanding disease mechanisms and identifying potential therapeutic targets.

Machine learning applied to biological networks offers a pathway towards understanding complex biological systems with unprecedented depth and precision. By integrating large datasets and sophisticated algorithms, it allows the identification of patterns allowing new discoveries in different fields such as disease biology, drug discovery and personalized medicine.

2. Objectives

This thesis aims to fulfill the following objectives:

1. To develop a framework for gene-disease association prioritization tool based on knowledge graph embeddings.
    a. To generate a biomedical KG to be exploited for predicting genes associated to human diseases
    b. To compare different algorithms for KGE production in supervised and unsupervised tasks.
    c. To prioritize genes associated with IDD by exploiting the created framework

2. To develop Genopyc, a Python library designed to investigate the effects of genetic variants identified through GWAS.
    a. To exploit the Genopyc to investigate GWAS variants related to IDD

3. To investigate if Modic change, a comorbidity of IDD, present an autoimmune etiology.

Section 3.1 covers the initial three objectives. First, I created a knowledge graph containing information about proteins, drugs, diseases and biological processes. Furthermore, I assessed various algorithms for generating KGE, and tested their performance in unsupervised (biological clustering) and supervised (GDA prediction) tasks. The most effective model was then chosen to develop a tool for prioritizing genes linked to complex disorders. Subsequently, I applied this tool to prioritize genes associated with IDD, demonstrating that the highest-ranking genes were already documented in literature as being linked with this condition. Additionally, the enriched functions of the gene set were involved in ECM turnover.

Objective 2 is covered in section 3.2, where I created a Python library that combines various analyses to investigate SNPs in their genomic context. This library facilitates the retrieval of information concerning functional elements at specific genomic sites, assesses linkage disequilibrium, and prioritizes causal genes at GWAS loci through the integration of diverse analyses and data sources. To illustrate its utility, I applied the library to interpret GWAS variants associated with IDD. The results highlighted transcription factors previously linked to the condition in literature.

Finally, objective 3 is addressed in section 3.3 by exploiting TwinsUK data. I performed a cross sectional study to assess if autoimmune positive individuals presented a higher prevalence or severity of (MC), a comorbidity of IDD. Our results showed that no significant differences were found in autoimmune positive participants and thus that having an autoimmune phenotype doesn't contribute to a more severe inflammation of the vertebral body.

In the appendix are listed 2 additional studies indirectly related with this thesis in which I have participated and contributed. In appendix A is included a review on IDD with a focus on the crosstalk between the IVD, immune system, and shifted metabolism during the degeneration of the disc. Appendix B is a review focusing on the cartilaginous endplate and its role in degenerated and healthy disc.

3. Results

3.1 Predicting gene disease associations with knowledge graph embeddings for diseases with curtailed information

This chapter is based on:

Predicting gene disease associations with knowledge graph embeddings for diseases with curtailed information.

Francesco Gualdi[1,2], Baldomero Oliva[2], Janet Piñero[1,3,*]

[1]*Integrative Biomedical Informatics, Research Programme on Biomedical Informatics (IBI-GRIB), Hospital del Mar Medical Research Institute (IMIM), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, C/Dr Aiguader88, E-08003 Barcelona, Spain.*

[2]*Structural Bioinformatics Lab, Research Programme on Biomedical Informatics (SBI-GRIB), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, C/Dr Aiguader88, E-08003 Barcelona, Spain, Barcelona, Spain*

[3]*Medbioinformatics Solutions SL, Barcelona, Spain*

** To whom correspondence should be addressed. Email: janet.pinero@upf.edu*

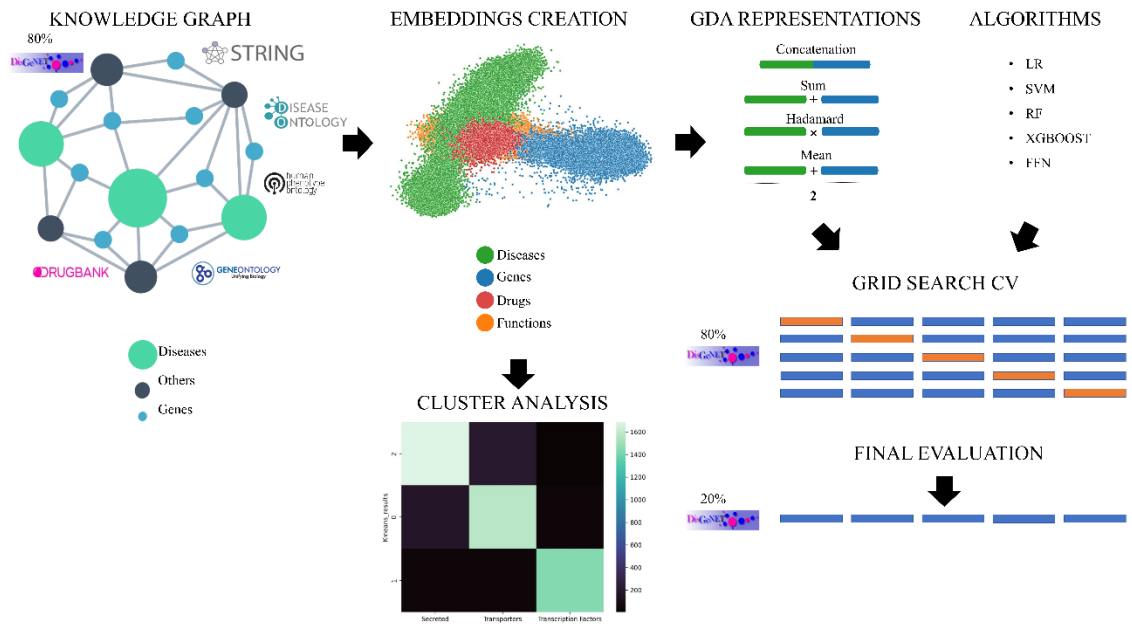
Published in: Nucleic Acid Research Genomics and Bioinformatics

DOI: 10.1093/nargab/lqae049

Summary of the work

In this chapter we developed a methodology based on knowledge graph embeddings (KGE) to prioritize genes associated with complex diseases. After building a knowledge graph (KG) merging multiple heterogeneous biological repositories, we carried out a thorough comparison of different algorithms for the creation of node embeddings. We evaluated the performance of the embeddings in an unsupervised task by unsupervised clustering and in a supervised task consisting of gene disease association predictions (GDA). Once the model with highest predictive capabilities was determined, we built a GDA tool that is freely available for the scientific community at https://github.com/freh-g/KGE. With this part of the work, we satisfied objectives 1.a and 1.b. Then, to accomplish objective 1.c, we implemented the tool for prioritizing genes in intervertebral disc degeneration (IDD) a complex condition and one of the major causes of low back pain (LBP). Of the 20,951 tested, the model predicted 445 genes associated with the condition. The top 10 genes prioritized from the framework were largely reported in literature to be associated to disc degeneration in in vitro and in vivo models. To better assess the nature of the prioritized genes, we performed a function enrichment analysis on the 93 genes predicted to be associated to IDD with probability greater than 0.95. The functions enriched in this set of genes were related to extracellular matrix turnover and homeostasis; functions relevant to the disease. We showed that through the implementation of KGE algorithms on a KG composed by heterogeneous biological data sources we were able to prioritize genes involved in the homeostasis of the extracellular matrix and already reported in literature to be associated with the pathology. Moreover, predicted genes such as TGFβ - 1 and SMAD3 are regulated from AP2α a transcription factor which importance was highlighted from the model we developed in chapter 2. Taken together these results show that we successfully developed a KGE framework for prioritizing genes in complex diseases and that there is a concordance in the output of models built in different chapters of the thesis.

GRAPHICAL ABSTRACT



ABSTRACT

Knowledge graph embeddings (KGE) are a powerful technique used in the biomedical domain to represent biological knowledge in a low dimensional space. However, a deep understanding of these methods is still missing, and, in particular, regarding their applications to prioritize genes associated with complex diseases with reduced genetic information. In this contribution, we built a knowledge graph (KG) by integrating heterogeneous biomedical data and generated KGE by implementing state-of-the-art methods, and two novel algorithms: Dlemb and BioKG2vec. Extensive testing of the embeddings with unsupervised clustering and supervised methods showed that KGE can be successfully implemented to predict genes associated with diseases and that our novel approaches outperform most existing algorithms in both scenarios. Our findings underscore the significance of data quality, preprocessing, and integration in achieving accurate predictions. Additionally, we applied KGE to predict genes linked to Intervertebral Disc Degeneration (IDD) and illustrated that functions pertinent to the disease are enriched within the prioritized gene set.

INTRODUCTION

Predicting genes associated with diseases is a challenging task. Recent advancements in genomic technologies have contributed to reach a deeper understanding of the genetics underlying complex diseases. However, the difficulties related to costs and time of these technologies have prompted the development of *in silico* methods to perform this task [98].

In this regard, network approaches have emerged as valuable tools for building meaningful models allowing the integration of heterogeneous biological knowledge from numerous sources [99].

These heterogeneous networks, defined as knowledge graphs (KGs), contain structured depictions of biological systems wherein different biological entities interact through complex relationships. Elucidating these intricate relations is crucial to better interpret complex biological data and thus the plausible causes of diseases.

KGs are increasingly implemented in the biomedical field due to their potential in representing and analyzing complex biomedical data. Recent research has highlighted their importance in enabling intelligent applications such as recommendation systems, semantic search, and logical reasoning. Automated schemes have been shown to significantly reduce the cost of building knowledge graphs [100]. Current research is addressing challenges such as knowledge graph completion and extraction methods for unstructured data. There is also a growing emphasis on constructing KGs from natural language text, with a focus on named entity recognition and relation extraction [101]. The field is still facing technical challenges, but the ongoing research aims to enhance the quality and reliability of knowledge graphs through novel techniques, models, and frameworks.

One commonly used approach to infer new interactions between biological entities involves expressing the entities within KGs as low-dimensional vectors using vectorial representations that preserve the graph's local structure known as knowledge graph embeddings (KGE). This method outperforms other approaches in terms of accuracy and scalability of their prediction [102].

Numerous methods have been developed to generate embeddings from KGs, and they can be broadly categorized into five main families: translational models, matrix factorization, semantic matching, random walks-based models, and deep neural networks. Refer to [103], [104] for a comprehensive overview of these methods. Recently, new techniques that combine these existing methods have emerged [105]. For example translational methods or PageRank [106] are merged with graph attention networks (GAT) to improve predictive powers of the embeddings [107]. Several studies have been conducted to explore the potential of KGE for predicting gene-disease associations (GDAs). For instance, *Nunes et al* investigated the impact of employing rich semantic representations based on more than one ontology to predict GDAs by testing different embedding creation models and machine learning algorithms [108]. Other works have focused on the heterogeneous integration of knowledge bases with the development of a single deep learning framework for predicting GDAs starting from a KG [97], [109].

In the biomedical domain KGE have been implemented for a wide range of downstream machine learning tasks, such as drug – target prediction [110], protein – protein interaction prediction [111] and therapeutic indications [112]. Also, KGE has demonstrated the ability to achieve prediction capabilities similar to those of raw data, while also offering the advantage of reduced dimensionality compared to the original dataset. [113].While previous studies have made progress in implementing KGE methods in GDA research, we lack a proper benchmark of available methods. Existing works in this field are limited to evaluating the proposed method [97] or the comparison of different algorithms [108] without providing a deeper insight into the generated embeddings or validating a particular use case. In this work we conducted a comparison of different methods of KGE creation with unsupervised and supervised machine learning tasks. We first generated KGE from multiple ontologies and biological knowledge bases, and we implemented four state-of-the-art methods, and two novel algorithms. Subsequently, we analyzed

the generated embeddings using unsupervised clustering algorithms. Furthermore, we evaluated the performance of the embeddings in a GDAs prediction task. Finally, we used the best performing model to predict potential genes associated to intervertebral disc degeneration (IDD).

METHODS

Data sources

To build the KG, we mined different types of biological data from publicly available repositories: *Protein – protein interactions:* We partially integrated data from multiscale interactome (downloaded 29/06/2022) [112]. Specifically, the data were integrated from:

The biological general repository for interaction dataset (BioGRID) [114]. This is a repository of manually curated both physical and genetic interactions between proteins from 71,713 high – throughput and low – throughput publications.

The database of interacting proteins (DIP)[115] in which only physical protein – protein interactions are reported with experimental and curated evidence.

Four protein-protein interaction networks from the human reference protein interactome mapping project [116]): (HI-I-05: 2,611 interactions between 1,522 proteins; HI-II-14 13,426 interactions between 4,228 proteins, Venkatesan-09: 233 interactions between 229 proteins; Yu-11 1,126 interactions between 1,126 proteins). In addition, we integrated the last version of the Human reference interactome (HI-III-20) [116].

Physical protein-protein interaction from Menche et al. [66]). This repository integrates different resources of physical protein – protein interaction data from experimental evidence. It integrates regulatory interactions from TRANSFAC [117] database, binary interactions from yeast-two-hybrid datasets and curated interactions from IntAct [118], BioGRID and HPRD [119]. It integrates also metabolic-enzyme interactions from KEGG [120] and BIGG [121], protein complex interactions from CORUM [122], kinase-substrate interactions from PhosphositePlus [123] and signalling interactions from Vinayagam et al. [124]

Only human proteins for which existed direct experimental evidence of a physical interaction were considered.

Ontologies: Ontologies are computational structures that aim to describe and classify the entities belonging to a certain domain in a structured and machine-readable format in order to be implemented in a broad range of applications. The main components of the ontology are classes that represent specific entities and usually are associated with an identifier. These classes are arranged in a hierarchical way from general to more specific and are connected to each other through relations. Finally, ontologies feature metadata, formats and axioms [125] For our purpose we integrated the following types of ontologies:

Gene Ontology (GO) [126], (downloaded 18/07/2022) is a knowledge base that aims to computationally describe biological systems ranging from molecules to organisms, as of 2023 it comprises 43,248 terms, 7,503,460 annotations across 5,267 species.

Disease Ontology (DO) [127], (downloaded 02/08/2022) is an ontological structure of standardized disease descriptors across multiple resources. The aim of the project is to provide a computable structure of integrated biomedical data in order to improve the knowledge on human diseases.

Human Phenotype Ontology (HPO) [128], (downloaded 22/08/2022) is a comprehensive logical structure that describes phenotypic abnormalities found in human diseases. This enables computational inference and interoperability in digital medicine.

We integrated HPO and DO and mapped the common codes to UMLS CUIS [129]

Gene product annotations to biological processes Proteins in the KG were mapped to their specific biological process through GO. GO annotations are statements about the function of a particular gene product, in this way, it is possible to obtain a snapshot of the current biological knowledge. We included gene annotations from the gene ontology association file (downloaded 29/06/2022).

Gene products annotations to phenotypes We integrated data of genes associated to phenotypes from 2 sources:

DisGeNET [130] is one of the largest publicly available collections of genes and variants associated with human diseases, it integrates GDAs data from curated resources with data automatically mined from the scientific literature using text-mining approaches. For our purposes we exploited DisGeNET curated (version 7.0) that integrates expert curated human gene disease associations from different data sources.To create a dataset, we used curated data from DisGeNET, comprising a total of 84,037 associations (hereafter considered as positives). We generated the same number of gene-disease non-associations (i.e. negatives) by considering that such associations were not reported in the text – mining version of DisGeNET, hence taking randomly any gene-disease pair not reported as positive.

HPO gene annotations to phenotypes: HPO (downloaded 02/08/2022) provides a file that links between genes and HPO terms. If variants in a specific gene are associated with a disease, then all the phenotypes related to that specific disease are assigned to that gene.

Phenotypes annotated to diseases We integrated annotations of phenotypes to disease from the phenotype.hpoa file from HPO ontology (downloaded 15/12/2022).

Drug-disease associations We integrated data of drug-disease pairs from the multiscale interactome [112]. This dataset is integrated by a collection of FDA approved treatments for diseases including different sources:

The drug repurposing database [131] is a database of gold-standard drug-disease pairs extracted from DrugCentral [132] and ClinicalTrials.gov

The drug repurposing hub [133] is a collection of drug-disease including 4,707 compounds. The database contains information mined from publicly and proprietary datasets that undergo manual curation.

The drug indication database [134] integrates data from 12 openly available, commercially available and proprietary information sources.

The dataset was filtered by keeping only human proteins resulting in a total number of drug – disease pairs of 5,926.

Drug – target interaction We obtained a dataset of drugs and their mode of actions on target proteins by integrating DrugBank [135] and the drug repurposing hub. Proteins that were not included in the protein – protein interaction network were removed.

KGE generation algorithms

We tested four state-of-the-art algorithms based on different principles and we implemented two novel methods to generate embeddings, referred to as BioKG2vec and Dlemb. For all experiments, the embeddings vector dimension was 100 and we set the number of epochs to 15.

RotatE RotatE [94] is a KGE generation algorithm that maps relations and entities to the complex vector space. The relations are considered as rotations from the source entity to the target entity. The principle lays on the assumption that given the triple "(h,r,t)", where h is the head, r is the relation, t is the tail e.g. "(protein1, interacts with, protein2)"the embeddings are obtained by the relation t = h ◦ r where ◦ denotes the Hadamard operation between the h and r vectors.

TransE TransE [136] is an algorithm that relies on a translational – based model. It represents relationships as translations in the embedding space. The principle lays on the assumption that given the triple *"(h,r,t)"*, where h is the head, r is the relation, t is the tail e.g. *"(protein1, interacts with, protein2)"*, the embedding of the tail should be similar to the head embedding plus the relationship embedding.

Relational graph convolutional networks (R-GCN) [92] R-GCN is an architecture for calculating the forward pass of relational graphs with multiple edge types. The propagation model is calculated as follows:

$$h_i^{(h+1)} = \sigma\left(\sum_{r\epsilon R}\sum_{j\epsilon N_i^r}\frac{1}{c_{i,r}}W_r^{(l)}h_j^{(l)} + W_0^{(l)}h_i^{(l)}\right)$$

Where N_i^r is the set of neighbours of node *I* under the relation $r \in R$ and $c_{i,r}$ is a problem-specific normalization constant that is chosen beforehand. $h_j^{(l)}$ is the node vector of neighbour *j* on which applies weight matrix $W_r^{(l)}$ of relation r in the iteration *l*. $W_0^{(l)}h_i^{(l)}$ is the representation of node *I* at layer *l* i.e. a self-representation at antecedent iteration.

Metapath2Vec [88] is an extension of the Node2Vec model [86] well suited for heterogeneous networks. The algorithm relies on meta-path-based random walks that capture both semantic and structural correlations between different types of nodes.

BioKG2vec BioKG2vec relies on a biased random-walk approach in which the user can prioritize specific connections by assigning a weight to edges. In the KG defined in this work we used 4 different node-types: drug, protein, function and disease. Then, the probability of visiting a specific neighbour at every step is given by the equation:

$$P(n_i) = \frac{\left(n_i \left(1 + \frac{w_i}{n_i} \right) \right)}{W}$$

where $P(n_i)$ is the probability for the random walker to visit a specific node type, n_i is the number of paths leading to the node (of the same type), w_i is the assigned weight (also specific for the type) and W equals to the node degree plus the sum of all weights (i.e. $\sum_i w_i$). To detect the optimal weights for the prediction of GDAs we performed a grid search assigning weights prioritizing drug -> protein -> function -> disease. Moreover, the walker stores the information of the visited edge type, and this information is used as input for Word2Vec algorithm in the embedding generation step. Thus, the algorithm handles different edges and nodes behaving differently for each node type being visited and storing the edge type of information too. BioKG2vec is available at https://zenodo.org/badge/latestdoi/624339823.

Dlemb Dlemb is a shallow neural network (NN) that consists of 3 layers: the input layer, embedding layer and output layer. The input layer takes as input KG entities as numbers and outputs them to the embedding layer. In the dot layer the scalar product of the vector is computed and normalized so the result is a number that ranges between -1 and 1. A false relation yields -1 while true relations produce +1. Then, the RMSE is calculated between the dot product and the expected value. Finally, the ADAM optimizer is used to adjust the embeddings layer directly since these are parameters of the neural network so that the model can be fitted to the data.

Dlemb is available at https://zenodo.org/badge/latestdoi/635382680.

Methods to combine embeddings

We used 4 strategies to combine gene and disease embeddings to obtain GDAs representations: 1) Sum, which consisted of the addition of both vectors; 2) Average, in which we averaged them; 3) concatenation, in which the result is a vector in a larger dimension, representing a pair gene-disease by concatenating both vectors; 4) Hadamard product (i.e. each element is produced by the product of the elements of the two vectors). For this work we produced embeddings of fixed dimension (i.e. 100) in the space of reals (i.e.$\mathbb{R}^{100}$).

Unsupervised analysis of the embeddings

We assessed the quality of the embeddings performing k-means unsupervised clustering. Specifically, we used function and compartment-based classification to group gene products in 16 different categories from human protein atlas (HPA) [137]. For diseases, we used annotations from UMLS to ICD-9 [138], that classify diseases into macro classes. We then used various evaluation scores for the comparison, such as the silhouette score, defined as:

$$\frac{b - a}{max(a, b)}$$

where b is the mean distance between a sample and all other points in the nearest cluster (nearest – cluster distance) and a is the mean distance between a sample and all other points in the same class (inter – cluster distance). We calculated this score for different cluster sizes ranging from 10 to 20 for genes (the gold standard number of clusters is 16) and from 10 to 20 for diseases (the gold standard number of clusters is 16).

Finally, we evaluate the homogeneity score, defined as:

$$1 - \frac{H(Y_{true}|Y_{pred})}{H(Y_{true})}$$

That is a measure that quantifies the similarity of samples in each cluster. Where the Y_{true} is the number of classes, Y_{pred} is the number of clusters and $H(Y_{true}|Y_{pred})$ represents the ratio between the number of classes Y_{true} in cluster Y_{pred} and the total number of samples in cluster Y_{pred}. When all the entities in the cluster belong to a class the homogeneity score equals 1.

Then, for visualization purposes, we performed UMAP dimensionality reduction on the embeddings and plotted the first 2 UMAP embeddings of gene and disease embeddings. Only 3 classes of genes and diseases are plotted.

Grid Search to select the best predictive model.

We performed a grid search cross-validation to find the best combination of embedding creation algorithm, GDAs representation and predictive machine learning (ML) and deep learning (DL) algorithms implemented in Scikit-learn [139] and Pytorch [140] respectively. In the grid-search experiment we created a KG in which we integrated all the biological data and 80% of curated GDAs from DisGeNET. We tested the predictions in the remaining 20% of GDAs that weren't used in the embeddings creation step. To avoid data leakage, we excluded diseases with over 20 associated genes, of which more than 90% were shared with another disease. Additionally, we made sure that in the validation dataset there were no GDAs included in the HPO data. For each algorithm, we fitted a grid of parameters (Table 1) maximizing the area under the receiver

operating-characteristic curve (ROCAUC). With this, we tested a total of 120 combinations for the grid search (Supplementary Table 1). Then, the best parameter combination was evaluated on the test set by assessing additional evaluation metrics, such as:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$recall = \frac{TP}{TP + FN}$$

$$precision = \frac{TP}{TP + FP}$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

$$FPR = \frac{FP}{FP + TN}$$

We also report the area under the precision recall curve (AUPRC).

Table 1: Search spaces of the algorithms tested during the grid search cross validation.

ALGORITHM	**PARAMETERS**	**VALUES**
LR	C	0.001, 0.01, 1, 5, 10, 25
	PENALTY	L1, L2
RANDOM FOREST	MAX DEPTH	2, 4, 6, None
	N. OF ESTIMATORS	20, 50, 100
XGBOOST	COLSAMPLE BY TREE	0.3, 0.7
	GAMMA	0, 0.5
	LEARNING RATE	0.03, 0.3
	MAX DEPTH	2, 6
	N. OF ESTIMATORS	100, 150
	SUBSAMPLE	0.4, 0.6
SVM	C	0.1, 1, 10

	GAMMA	0.001, 0.01, 0.1
	KERNEL	rbf, poly
FFN	N. OF LAYERS	2, 3
	N. OF NODES FIRST LAYER	50, 100, 150
	N. OF NODES SECOND LAYER	20, 50
	ACTIVATION FUNCTION	sigmoid, tanh, relu
	LOSS FUNCTION	Binary cross-entropy, hinge
	BATCH SIZE	30, 100
	EPOCHS	20, 60

Ontology preprocessing and heterogeneous data integration

Once we selected the model with the highest predictive power, we investigated the influence of integrating heterogeneous biological data in the KG on the GDAs predictions. For this experiment we only used ontological data. Ontologies are complex, standardized data structures composed of classes, relations, axioms and metadata all of which are included in the raw ontology. Moreover, we tested the effect of implementing a pre-processing step in the ontology in which only classes and relations were maintained as a graph structure (axioms and metadata were excluded). We studied the following combinations of data sources:

HPO + HPO annotations raw

HPO + HPO annotations pre-processed

HPO + HPO annotations + GO + GO annotations (all) pre-processed

We used a comparison based on two metrics. For this experiment, we created embeddings with the Mtapath2vec algorithm, using concatenation for GDAs representation, and SVM as classification algorithm. For the processing of the ontologies nxontology and pronto [141] Python libraries were used.

Influence of GDAs in the KG for GDA-predictions.

We tested the influence of adding increasing GDAs proportions in the KG. For this experiment, we used 20% 50% 80% and 100% of DisGeNET and we included it in the KG. Then we generated

embeddings from the KGs with Metapath2Vec and we trained a SVM on 80% of DisGeNET. We tested the model on the 20% of remaining associations and calculated ROCAUC and AUPRC as evaluation metrics.

Comparison with randomly generated embeddings

To show that the information is efficiently translated from the KG to the vectorial space, we compared the performance of Metapath2Vec generated embeddings and random embeddings of the same size. We aimed to assess the effectiveness of translating the information encoded in the KG into embeddings by comparing KGE with a null model. To conduct this evaluation, we created 100-dimensional random embeddings for each gene and disease, represented GDAs through concatenation, and tested their predictive capabilities. The number of associations is a latent variable that can be learned by ML to produce good predictions. This can be considered a potential bias. Therefore, we further tested the effect of removing the number of associations stratifying DisGeNET diseases by the number of associated genes. We divided the data into 23 groups in which the number of associations for every disease has a maximum difference of 20. Then we selected a disease belonging to every class, generated negative associations and performed a five-fold cross validation on the data with the best performing algorithm. We evaluated accuracy, precision, recall, f1 score and ROCAUC across every fold.

Generalizability of the model.

The predictive model selected was tested to predict associations for diseases not used in the training set. The rationale behind this experiment was to understand the capabilities of the model to predict gene-disease associations of new diseases, proving that the biological information encoded in the embeddings was generalizable.

To assess this, we trained the model on GDAs belonging to diseases of a specific ICD-9 disease class and then we tested the model on all other classes.

Performance of the algorithms

We compared the performance of the algorithm with the top predictive power i.e. Metapath2Vec, BioKG2vec and Dlemb. We performed n = 10 experiments by randomly selecting 1000 nodes from the knowledge graph, creating the subnetwork and producing the embeddings. We calculated the difference of the running time (in seconds) as percentage with the following formula:

$$\frac{T_1 - T_2}{T_1} \times 100$$

Being T1 the running time of Metapath2Vec and T2 the running time of either BioKG2vec or Dlemb. The experiment was conducted on an 8-core intel i7 machine.The experiment was conducted on an 8-core intel i7 machine.

Intervertebral Disc Degeneration Biomarker Prediction

We tested the model to predict genes associated with IDD. We used the model selected through grid search cross validation with concatenation of the embeddings for the GDAs representation. Lastly, we performed a function enrichment analysis using g:Profiler [142] on the set of prioritized genes with a probability greater than 0.95 to be associated to IDD

RESULTS

Data integration and KG structure

We integrated multiple sources of data in the form of KG for a total of 95952 nodes and 2,183,603 edges. The KG contains 4 types of nodes: drugs (n = 2,991), phenotypes (n = 28,374), proteins (n =21,019) and functions (n = 43,568). These entities are connected by 81 different types of relationships represented as edges. The relationships are obtained through different data sources, 18,282 proteins interacting among each other (87, 1356 edges), 19,409 proteins annotated to 18,813 biological functions (303,404 edges) and 8, 053 proteins annotated to 13,525 phenotypes (246,006 edges). Moreover, drug information was included: 1,551 drugs annotated to 828 phenotypes for a total of 5,744 edges and 2,887 connected to 2,074 proteins they target for a total of 14,491 edges. The degree distribution of the graph follows a scale free law (Supplementary Figure 1) [143].

Unsupervised clustering of the embeddings reflects the biological classification

From the KG, we generated embeddings using six algorithms. Figure 1 shows the first 2 UMAP embeddings of genes and diseases. The embeddings tend to differentiate among gene products belonging to different groups: secreted, transcription factors, and transporters (Figure 1 A to G). Metapath2Vec, BioKG2vec, and Dlemb from a visual perspective achieve the best clustering of genes. In Figure 1, G to L only 3 categories of diseases are represented, corresponding to the ICD chapters disease of blood and blood-forming organs, diseases of the musculoskeletal system and connective tissue and mental disorders. As above, algorithms Metapath2Vec, BioKG2vec and Dlemb visually distinguished disease classes better than others.

The algorithm producing the best clustering of disease classes and gene products was Metapath2Vec, which has a higher homogeneity score for both genes and diseases. (Table 2). For

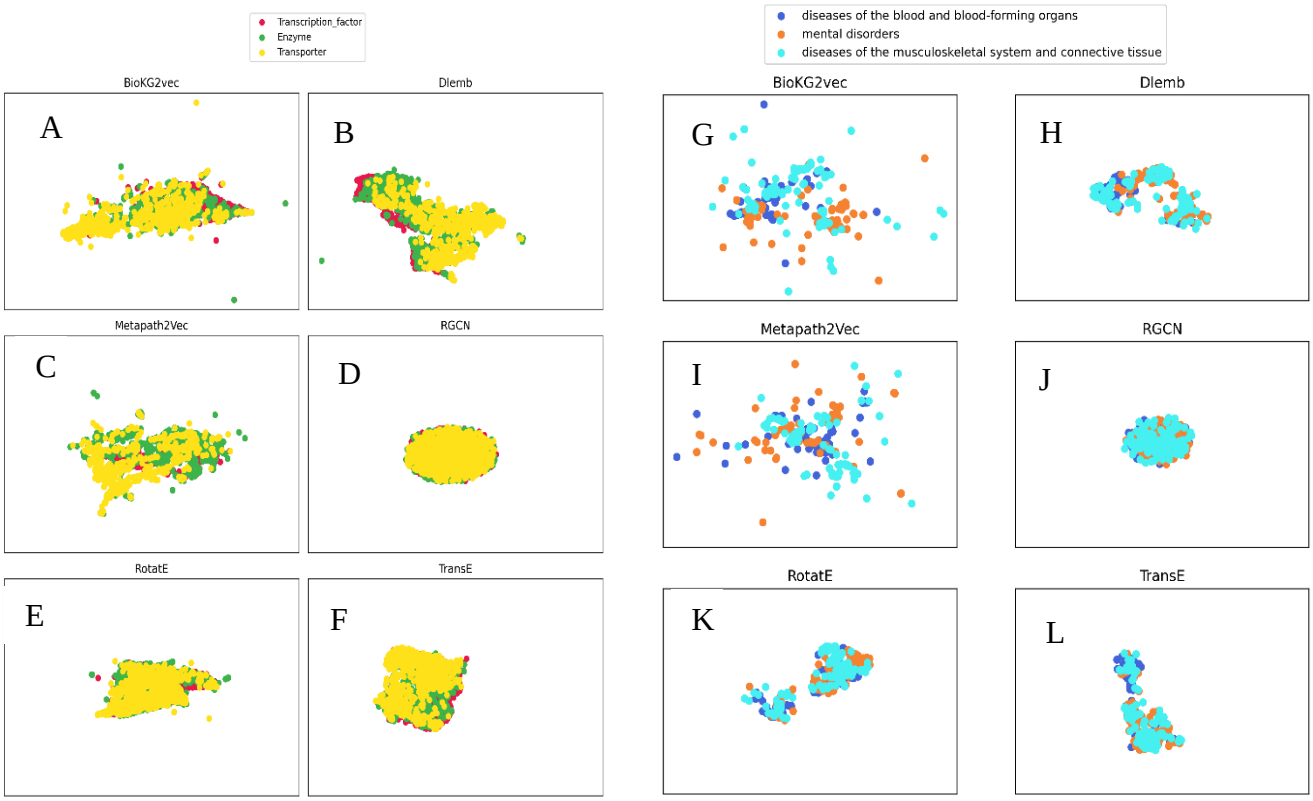


Figure 1 UMAP of gene-embeddings (panels A–F) and disease-embeddings (panels G–L) generated with BioKG2vec (A, G), Dlemb (B, H), Metapath2Vec (C, I), RGCN (D, J), RotatE (E,K) and TransE (F, L). Genes classes were obtained from human protein atlas while diseases classes are ICD – 9 classification.

the case of diseases, the silhouette score of the embeddings produced with any algorithm couldn't match the gold standard number of clusters (Supplementary Figures 2 and 3).

	Homogeneity score	
	Genes	**Diseases**
Metapath2Vec	**0.49**	**0.28**
Dlemb	0.35	0.17
RotatE	0.35	0.15
Trans-E	0.29	0.09
BioKG2vec	0.2	0.20
RGCN	0.008	0.02

Table 2: Homogeneity score of K – means algorithm calculated for genes (number of clusters = 16) and diseases (number of clusters = 16). True labels are classification from ICD-9 and HPA for diseases and genes respectively

The embeddings of gene products generated with Metapath2Vec produced more homogeneous clusters. Assigning every different gene and disease correctly to their category is a very complex task because of the high granularity of genes and disease classes (Supplementary Figures 4 and 5).

Model selection through grid search cross-validation

The best performing combination for GDA prediction was Metapath2Vec. Metapath2Vec coupled with concatenation of the gene and disease embedding as association representation and SVM with parameters C = 10 and kernel = rbf as classification algorithm. The whole output of the experiment is available in Supplementary Table 1. The following experiments were run using this combination.

Heterogeneous data integration and preprocessing

Pre-processing the ontologies leads to better ROCAUC and AUPRC compared to using embeddings generated with raw data. Nevertheless, adding heterogeneous data in the KG did not significantly affect the predictions of GDAs (Table 3). Integrating more data leads to similar performances which can be appreciated when comparing the results of generating the KG using HPO data with HPO and GO data. We must note the different results on the use of Dlemb algorithm (Supplementary Table 2). While the predictive power of Metapath2Vec is not affected by the preprocessing of the ontologies, Dlemb significantly improves the AUPRC and ROCAUC after preprocessing.

Table 3: ROCAUC and AUPRC of different experiments of GDAs predictions using Human Phenotype Ontology (HPO) + annotations (A), HPO ontology processed (B) and HPO + Gene Ontology (GO) + GO annotations. The embeddings were generated with Metapath2Vec and we used SVM as predictive algorithm, and operator concatenation for combining the embeddings.

Experiment	ROCAUC	AUPRC
HPO + HPO annotations raw	**0.95**	**0.98**
HPO + HPO annotations processed	0.93	0.97
HPO + HPO annotations + GO + GO annotations processed	0.93	0.97

The amount of training GDAs in the KG affects the prediction of GDAs

We tested the effect on the predictions caused by the increase of GDAs in the KG. We expect that increasing the amount of GDAs in the KG will increase the quality of the predictions. Figure 2 shows that the increase in the number of GDAs used for training the knowledge graph embeddings increases the values of ROCAUC and AUPRC.

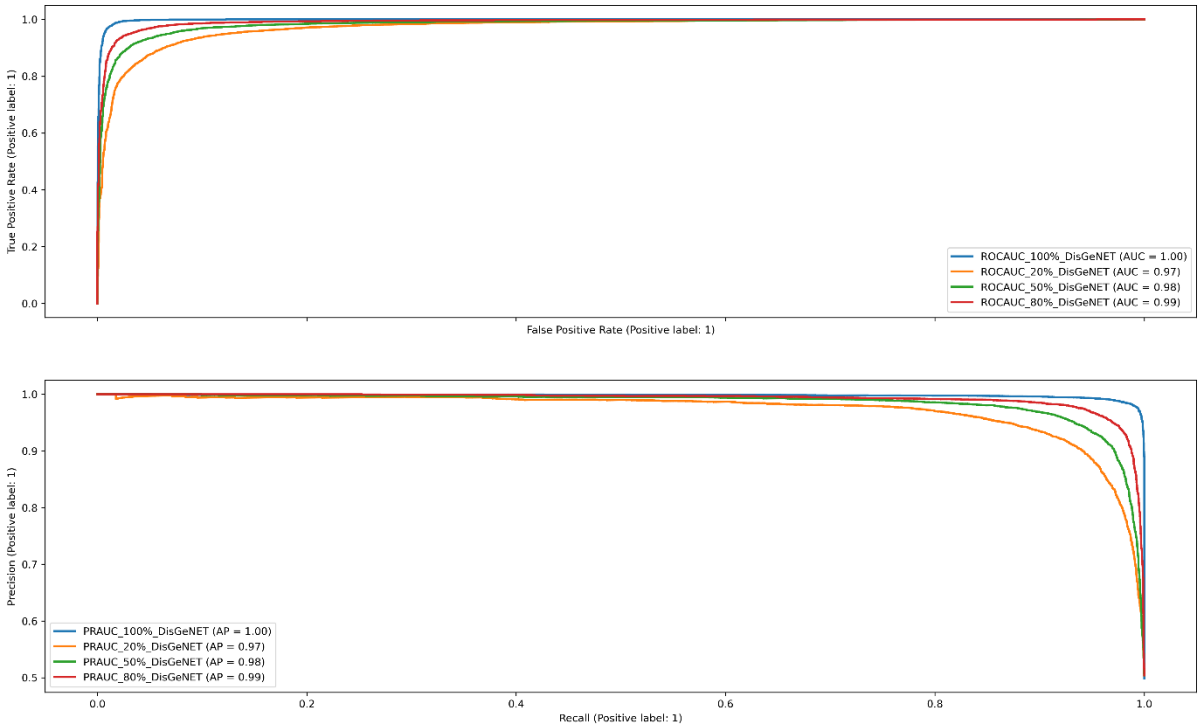


Figure 2: ROCAUC and PRAUC of the prediction of GDAs. Several KG embeddings are obtained using increasing percentages of known GDAs from 20% to 100%. Note: embeddings were generated with Metapath2Vec, using concatenation for combining embeddings and SVM for the classification/prediction algorithm.

Comparison with randomly generated embeddings

Supplementary Table 3 presents the outcomes of the experiment contrasting embeddings produced by Metapath2Vec with those generated randomly. Metapath2Vec embeddings reach an average ROCAUC of 0.93 while random generated embeddings have random metrics. These results are due to the biological information intrinsic to the embeddings since the effect of the number of GDAs was prevented by selecting associations of one disease only. In fact, the number of associations is a latent variable that is learned by the model.

Model generalization across different disease classes

Figure 3 shows the performance of the model trained on a specific ICD9 disease class and tested on all the others. Training and testing in diseases belonging to the same class leads to accurate predictions. However, embeddings generated with Metapath2Vec have poor prediction capabilities across different ICD-9 classes. Similar results were observed with randomly generated embeddings. Biological information encoded in Dlemb generated embeddings is translated across disease classes and we can see that some pairs of disease classes achieved a noteworthy prediction (e.g. the model trained for neoplasms predicts genes associated with diseases of circulatory system with an ROCAUC > 0.7) (Supplementary Figure 6). As expected, randomly generated embeddings show ROCAUC in the heatmap with random values.

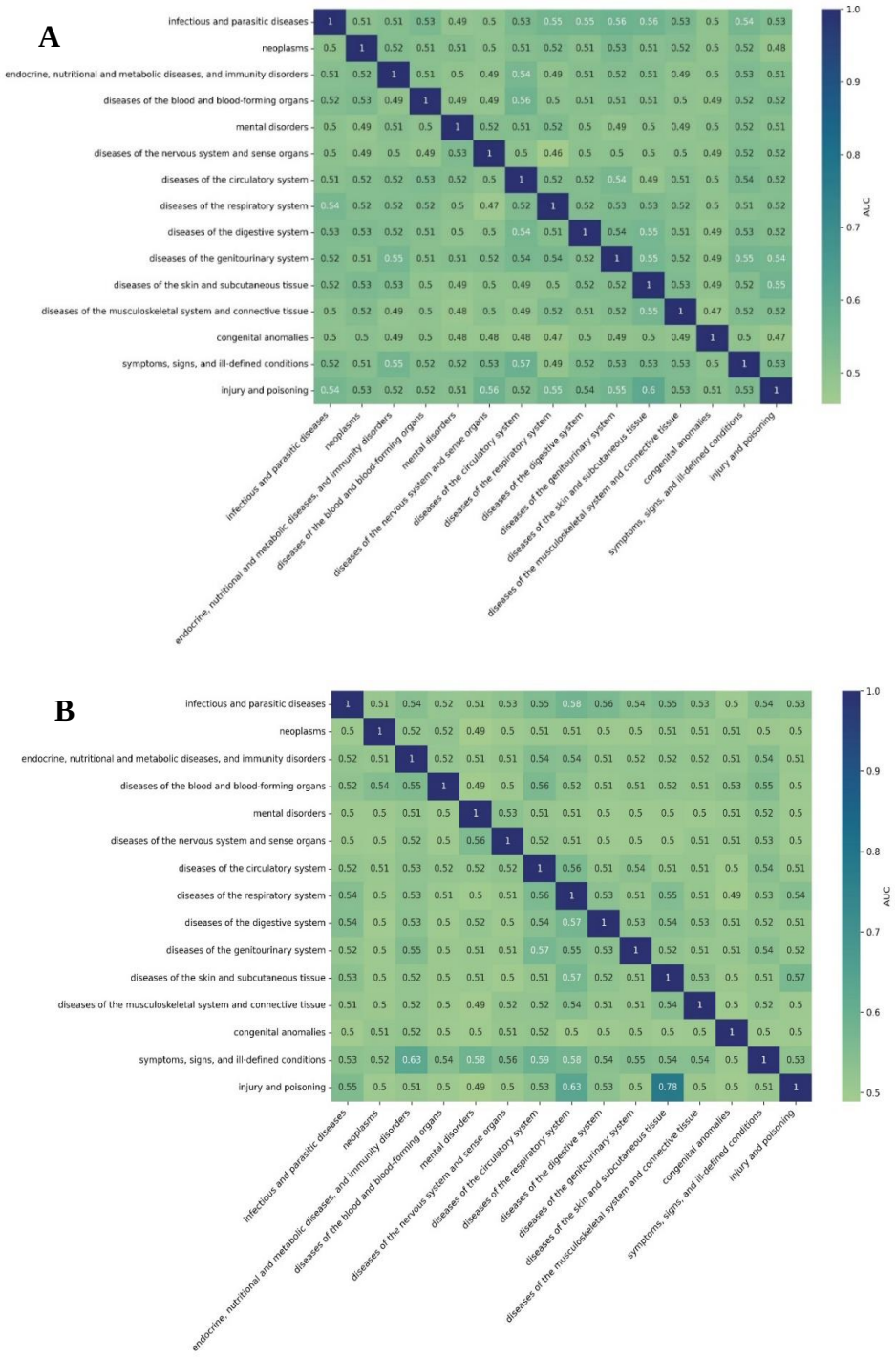


Figure 3: ROCAUC of the best performing combination for the prediction on different disease classes. A) Results for randomly generated embeddings, the ROCAUC shows random values. B) The embeddings were generated with Metapath2Vec , the GDAs representation was concatenation, and the algorithm was SVM with parameters C = 10 and kernel = rbf. We show the results of training a model on a specific ICD-9 disease class (rows), and then testing on the others (columns).

Computational Performance of the best algorithms for KGE

We compared the performance of the three algorithms that reached the highest ROCAUC during the grid search cross validation in terms of running time. In the supplementary figure 7 are reported the running times of 10 experiments. BioKG2vec and Dlemb are respectively ~100% and 360% faster than Metapath2Vec.

KGE successfully predicts genes associated to IDD

Finally, we used the selected prediction model with the best parametrization to predict GDAs for IDD. IDD is one of the main causes of low back pain, the largest cause of morbidity worldwide affecting 80% of people from Western countries during their lifetime [144]. IDD consists of the gradual deterioration of the intervertebral disc (IVD) in which the content of collagen and glycosaminoglycan decreases, and it becomes more dehydrated and fibrotic. Due to this, its anatomical areas nucleus pulposus (NP) and anulus fibrosus (AF) becomes less distinguishable [145]. Also, during IDD there is a catabolic shift in the biochemical processes of the disc environment with an increased expression of matrix degrading enzymes promoted by catabolic cytokines and vascularization of the tissues [2]. According to DisGeNET (curated sources), IDD is associated to TGFβ-1, HTRA1 and SPARC. We ran predictions for 20,951 genes, of those 445 were predicted to be associated to the disease and 93 with a probability > 0.95. The results of the top 10 prioritized genes are shown in Table 4.

The predictive analysis identifies the TGFβ-1 gene as the most promising candidate associated with Intervertebral Disc Degeneration (IDD), with isoforms TGFβ-2 and TGFβ-3 also receiving prioritization. Notably, TGFβ-1 emerges as the highest-scoring gene in DisGeNET's curated dataset related to disc degeneration. TGFβ plays a multifaceted role in various pathways associated with the homeostasis and turnover of the extracellular matrix in IDD [146]. Additionally, SMAD3 and SMAD2, integral genes in disc homeostasis, participate in the TGF-β pathway. [147]. Matrix metalloproteinase 9 (MMP9) and matrix metalloproteinase 2 (MMP2) enzymes contribute significantly to IDD by participating in matrix degradation, targeting proteins expressed in the intervertebral disc like collagens and aggrecan. [148]. Moreover, LOX, crucial for cartilage homeostasis, presents a potential strategy for cartilage regeneration[149] , with studies indicating its anti-apoptotic effects in TNF-α treated rat NP-cells. [150]. These genes were shown to have a role in IDD and could be further investigated to elucidate the mechanisms that lead to the degeneration of the disc.

To further explore the biological functions of these candidate genes, we performed a function enrichment analysis (Figure 4). The top prioritized genes are enriched in processes related to the extracellular matrix organization, pathways related to collagen formation, and extracellular matrix degradation, all of them related to IDD.

Table 4: Top 10 genes prioritized from the model with highest predictive capabilities

Gene ID	Gene Symbol	Probability
7040	TGFB1	1
4088	SMAD3	1
4318	MMP9	1
4015	LOX	1
7043	TGFB3	1
7046	TGFBR1	1
7042	TGFB2	1
1277	COL1A1	1
4313	MMP2	1
4087	SMAD2	1

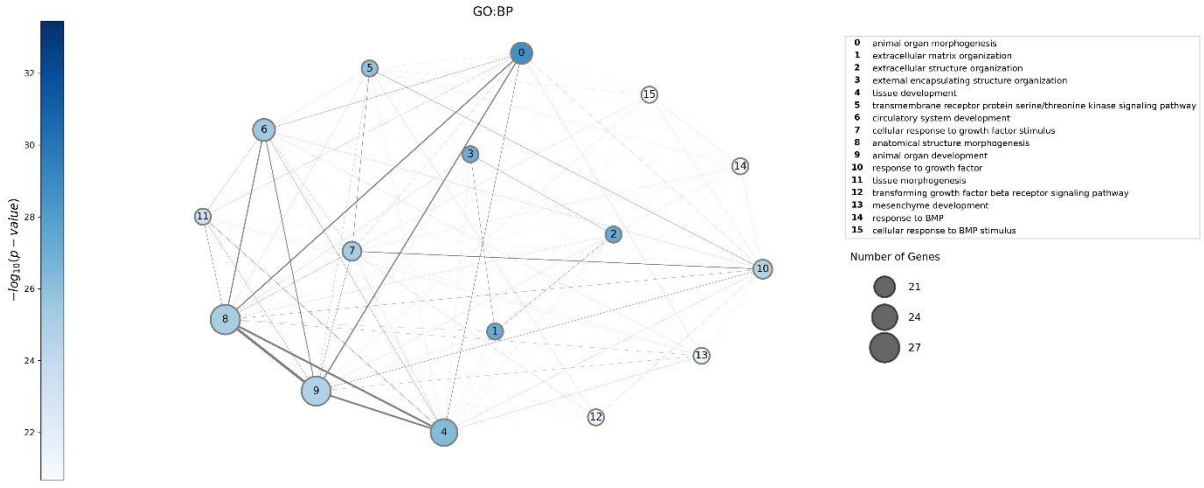


Figure 4: Gene ontology biological processes (GO:BP) function enrichment analysis on the genes with probability higher than 0.95 to be associated to C0158266 (n=93). To run the functional enrichment, we used g:Profiler. The nodes correspond to the pathway enriched in the gene set, their size is proportional to the number of genes belonging to that specific pathway and the colour is related to the significance of the enrichment in the gene set (calculated through hypergeometric distribution). An edge exists between 2

nodes if there are genes shared between the two pathways and the width of the edge is proportional to the number of the genes shared.

DISCUSSION

In this work we investigated how KGE perform to predict gene-disease associations. First, we generated a KG by implementing heterogeneous biological information such as protein-protein interactions, gene-disease associations, drug-disease associations and drug-protein interactions, and ontologies The integration of multiple knowledge-based datasets prevented us from using syntactic-based approaches for embedding-creation such as OPA2VEC [151]. Syntactic approaches rely in the set of axioms only for obtaining the embeddings without the intermediate graph-based representation [57], so the input of the algorithm must be in Web-Ontology Language (OWL) format. Moreover, the integration of different ontologies is a challenging task and an active research topic [153].

In this study, we systematically assessed diverse methodologies for KGE construction and introduced two novel algorithms, namely BioKG2vec and Dlemb. Our comprehensive evaluation reveals that these algorithms exhibit superior performance compared to most existing methods. Notably, the parallelized implementation of both BioKG2vec and Dlemb results in substantially reduced running times in comparison to Metapath2Vec. This enhanced scalability facilitates the effective utilization of computational resources.

We conducted an extensive analysis of embeddings utilizing unsupervised machine learning techniques. Our investigation encompassed the integration of diverse data types and the comparison of GDA predictions using random features. Our findings revealed that augmenting the proportion of GDA within the KG enhances model performance. This observation suggests that task-specific embeddings implementation could enhance predictions, potentially leveraging the learning of pertinent features, as indicated elsewhere [113]. Furthermore, we applied KGE to prioritize new genes associated with IDD, illustrating their utility in inferring disease biomarkers even in scenarios with limited genetic data. Notably, our model, trained on a DisGeNET curated dataset containing merely 3 associations, prioritized 445 genes, which effectively reflected the underlying biology of IDD. In fact, the polygenic nature and epistatic interactions characteristic of non-communicable diseases pose challenges to comprehending the intricate biology underlying the development of complex conditions [154].

Finally, we emphasize the significance of scrutinizing the data quality employed in embedding creation, as predictive models can glean numerous latent features, potentially introducing bias to the outcomes.

CONCLUSIONS

In this work we carried out an extensive investigation on KGE from the generation and evaluation of the produced embeddings to the development of two new models for KGE generation and the utilization of the created embedding in a GDA prediction task. We showed that embeddings can effectively be implemented in the biomedical field to infer new knowledge over a certain domain. Nevertheless, many challenges remain open that require interdisciplinary collaboration to reach better outcomes in the healthcare sector.

DATA AVAILABILITY

Tool with embeddings generated with the top 3 best performing algorithms (Metapath2Vec, Dlemb and BioKG2vec) for GDAs association are available at:

https://github.com/freh-g/KGE

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

FUNDING

This project was supported by the Marie Sklodowska-Curie International Training Network "disc4all" under grant agreement #955735.

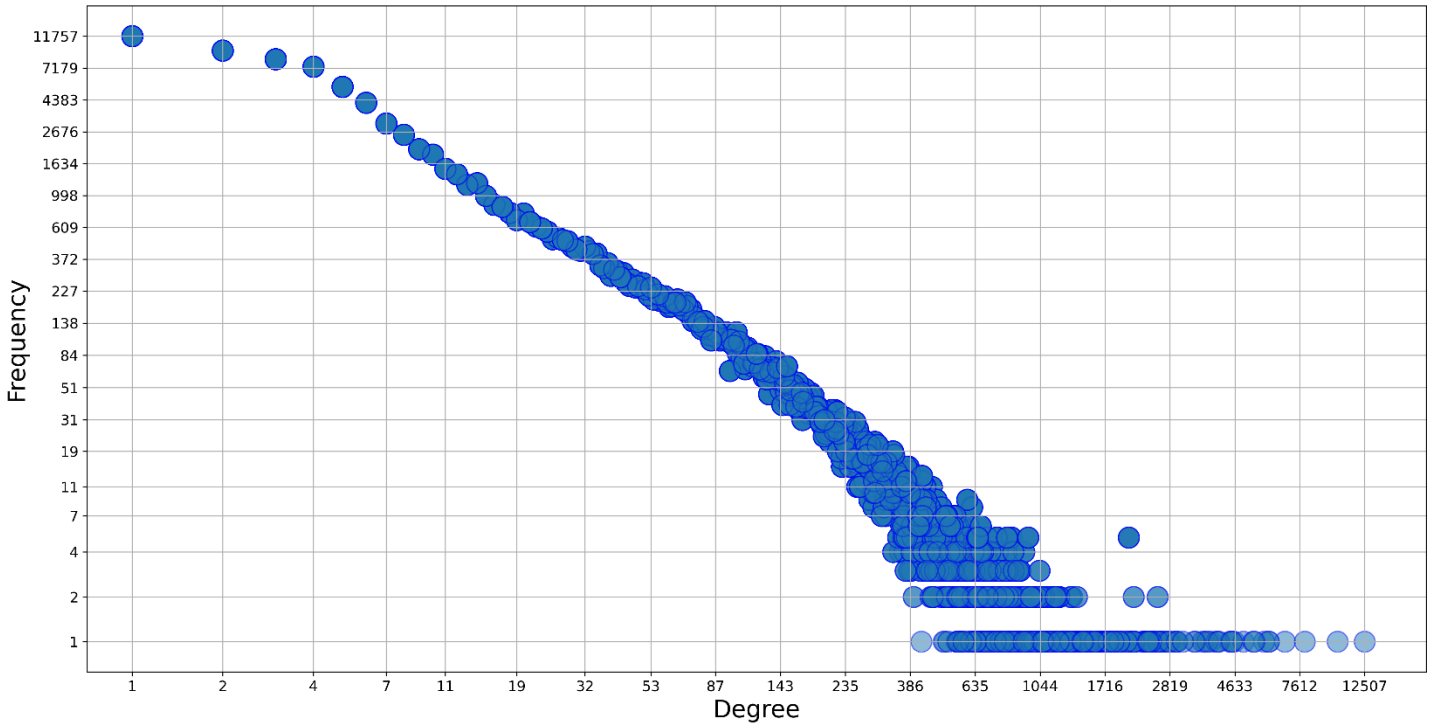
ACKNOWLEDGEMENTS

BO acknowledges support from MCIN and the AEI (DOI: 10.13039/501100011033) by grants PID2020-113203RB-I00 and "Unidad de Excelencia María de Maeztu" (ref: CEX2018-000792-M).

CONFLICT OF INTERESTS

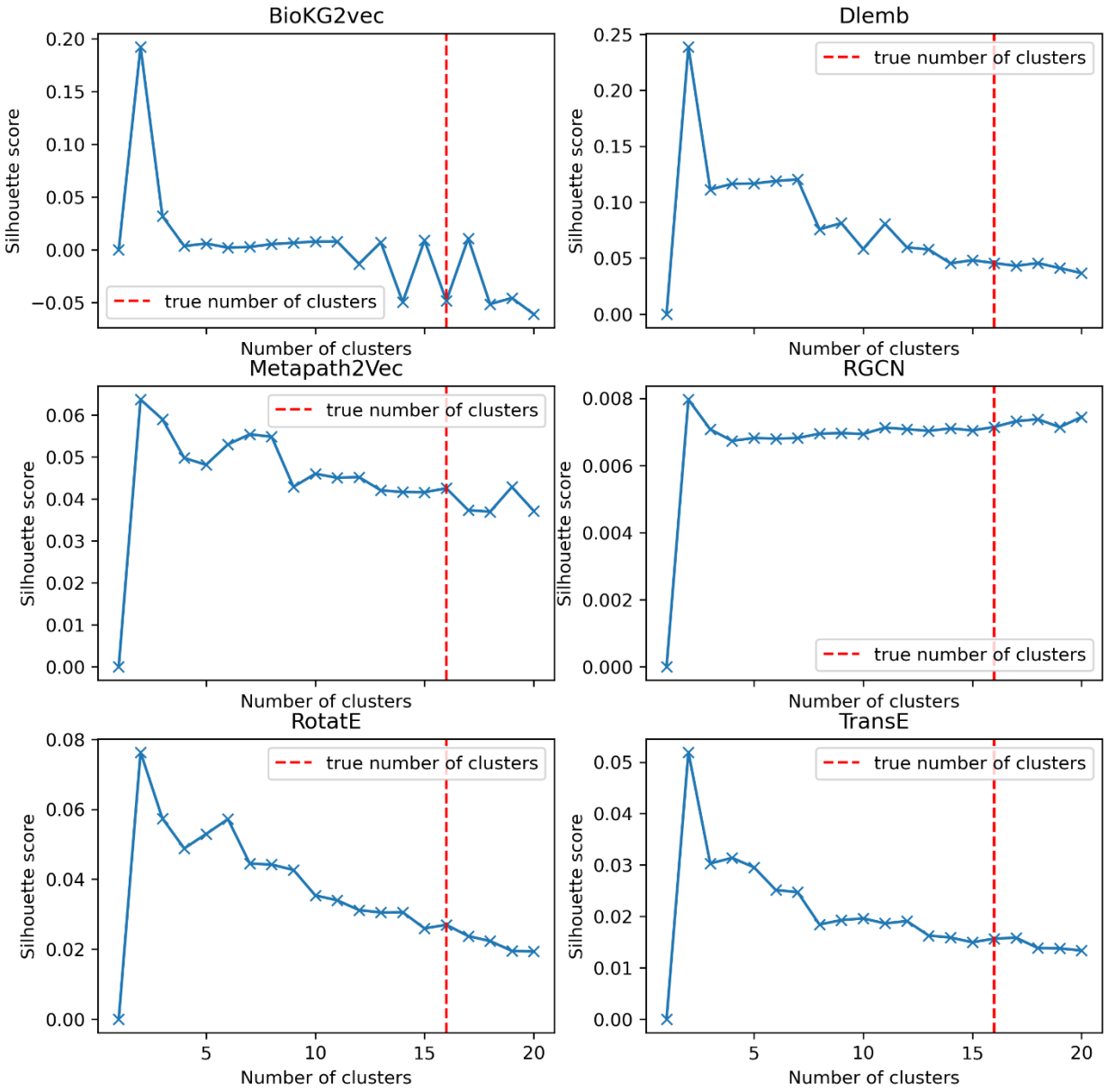
JP is an employee of Medbioinformatics Solutions SL. JP is co-founder and holds shares of Medbioinformatics Solutions SL.

SUPPLEMENTARY MATERIAL

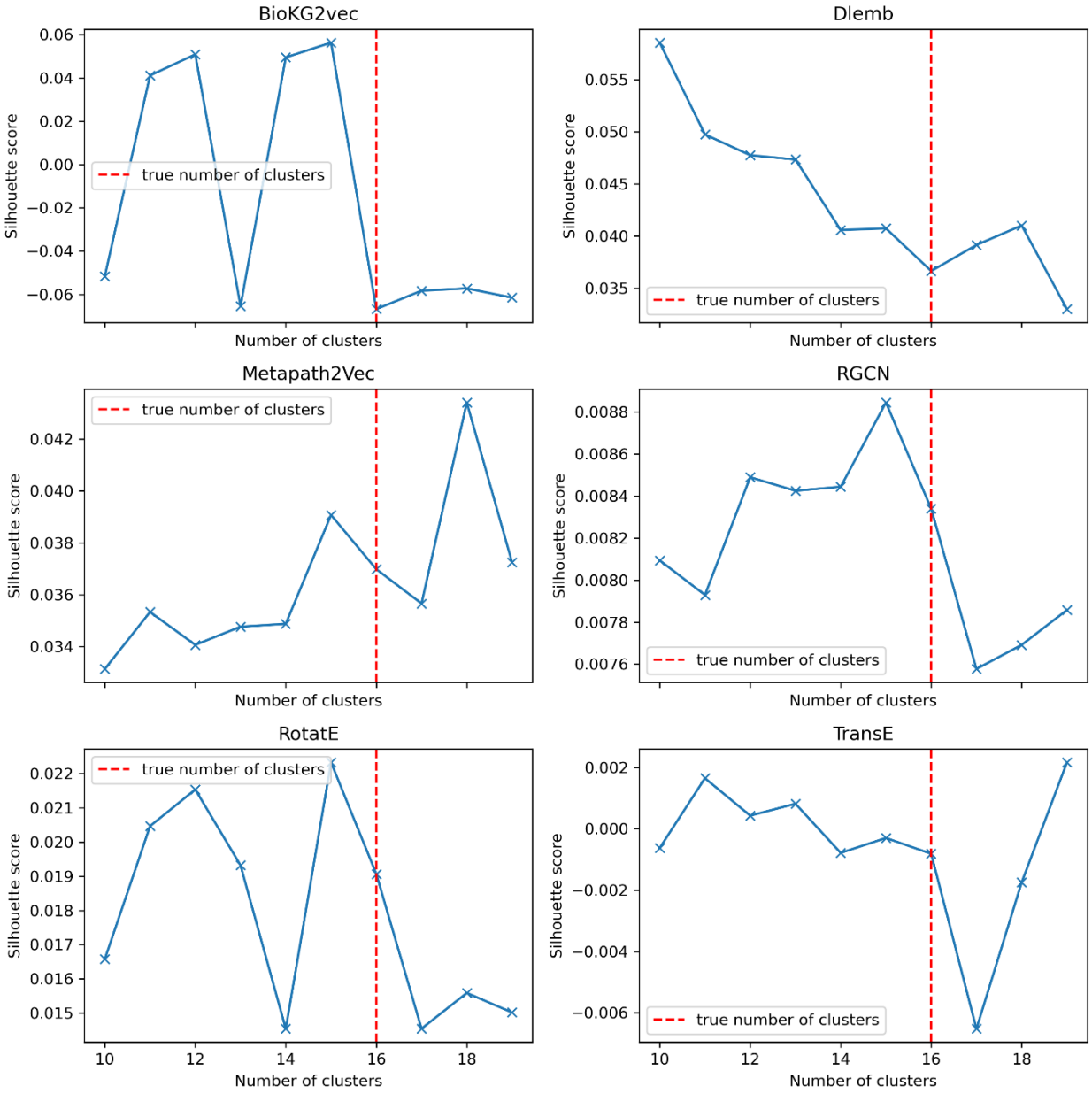


Supplementary Figure 1: Degree distribution of the knowledge graph, the x-axis represents the degree i.e. number of edges adjacent to a specific node and the y-axis is the frequency of nodes with that specific degree in the graph. The nodes in the KG follows a scale free degree distribution

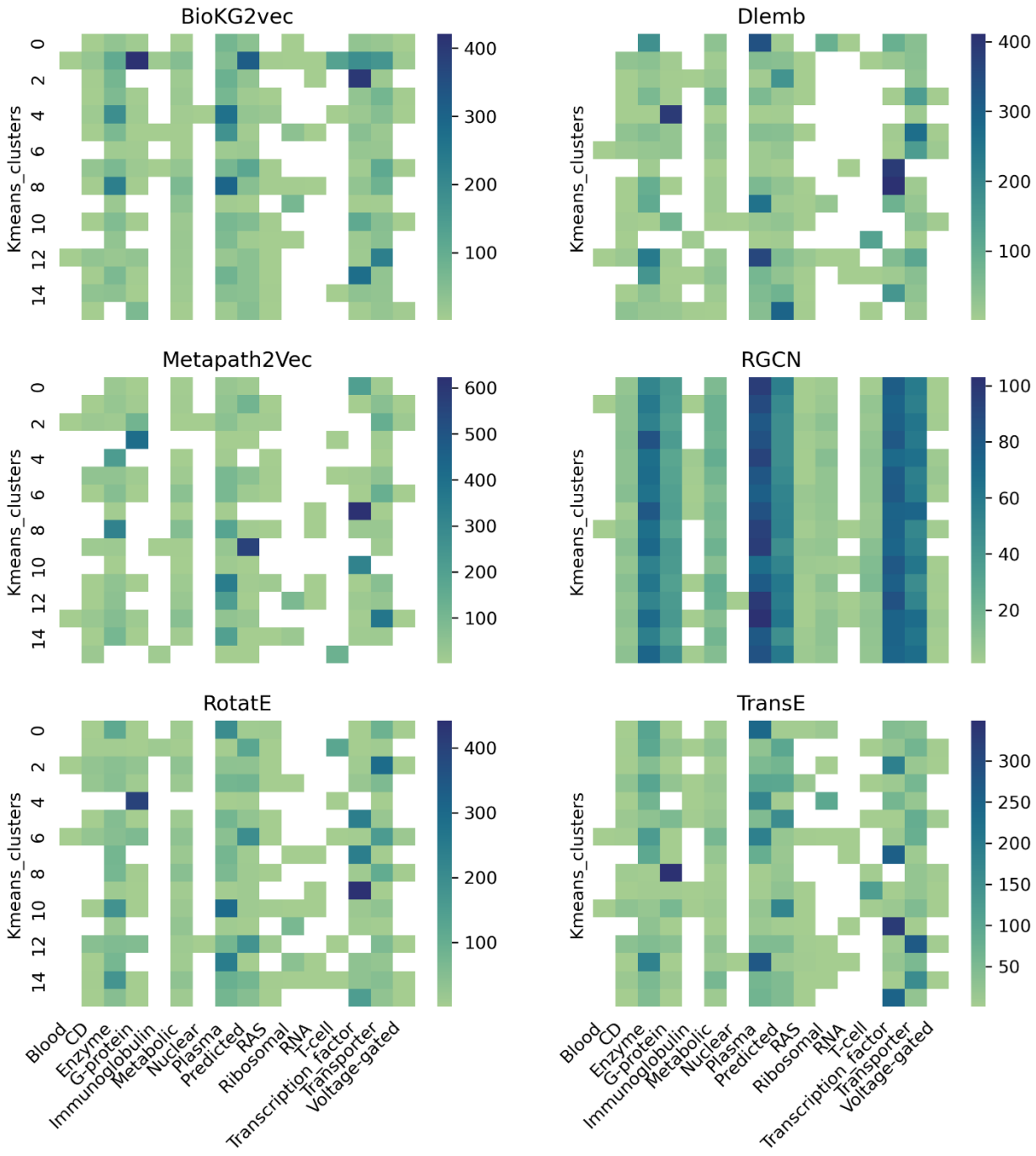
.



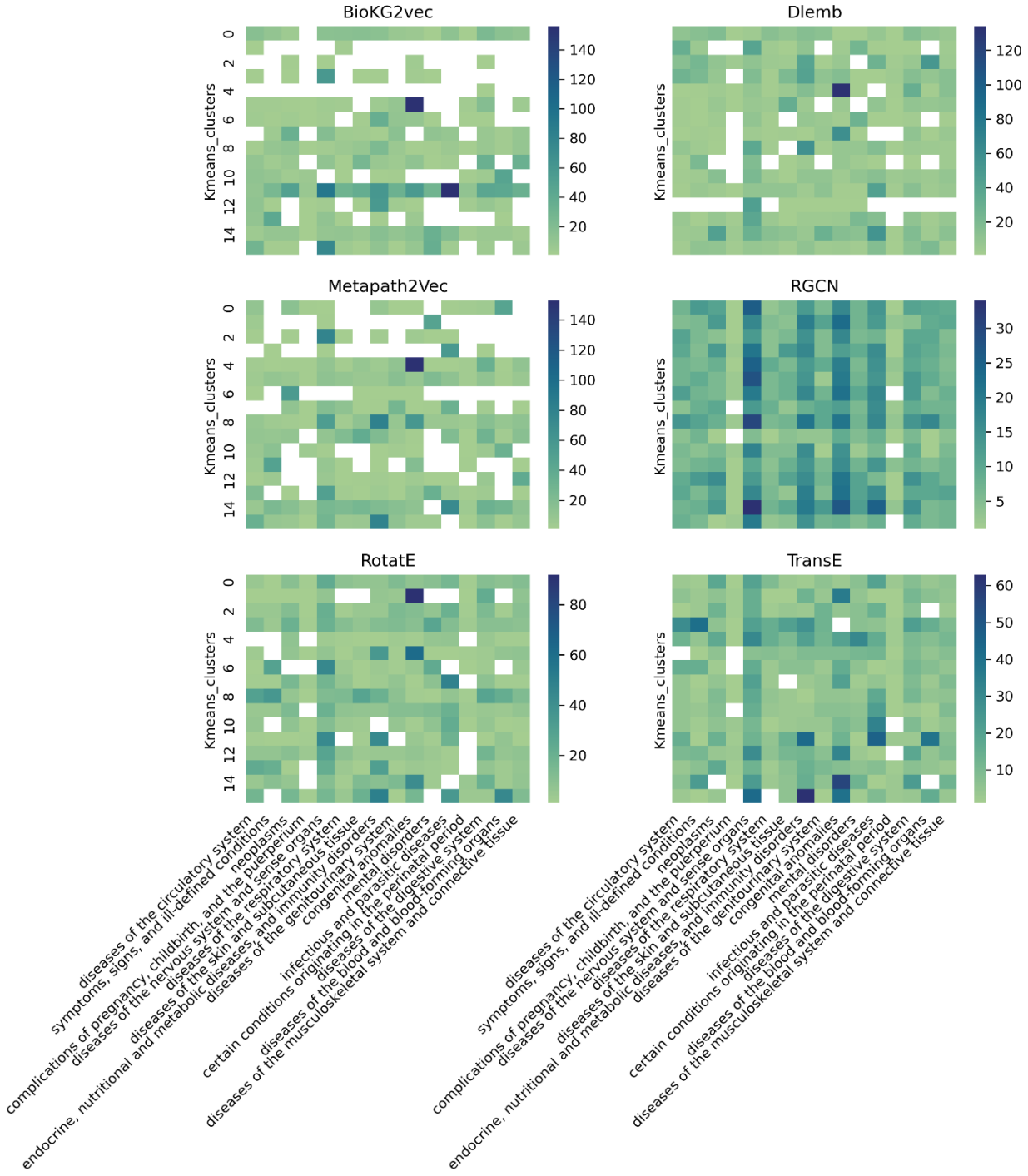
Supplementary Figure 2: Silhouette scores calculated for different numbers of K-means clusters for gene embeddings. The red line represents n = 16 i.e. the actual number of gene classes from Human Protein Atlas.



Supplementary Figure 3: Silhouette score calculated for different numbers of K-means clusters of diseases embeddings. The red line represents n = 16 i.e. the actual number of disease classes from ICD-9.



Supplementary Figure 4: K-means clusters on gene product embeddings separated by Human Protein Atlas protein categories. On the y axis are the 16 clusters produced from the algorithm and on the x axis the protein classes. The color indicates the number of gene products in each cluster.



Supplementary Figure 5: K-means clusters on disease embeddings separated by ICD-9 disease codes. On the y axis are the 16 clusters produced from the algorithm and on the x axis the disease classes. The color indicates the number of diseases in each cluster. BioKG2vec

Supplementary table 1 : Results of grid search cross-validation. The results are ordered by ROCAUC. DLemb Metapath2vec and BioKG2vec with Concatenation GDA representations are the best performing algorithms. Svm is support vector machine, lr logistic regression, xgb xgboost, rf random forest, and ffn feedforward neural network

Combination	***F1***	***PRECISION***	***RECALL***	***ACCURACY***	***ROCAUC***	***PRAUC***
Metapath2vec_Concatenation_svm	*0.88*	*0.97*	*0.79*	*0.88*	*0.88*	*0.87*
Metapath2vec_Sum_svm	*0.89*	*0.94*	*0.83*	*0.89*	*0.89*	*0.86*
BioKG2vec_Concatenation_svm	*0.88*	*0.94*	*0.8*	*0.88*	*0.88*	*0.85*
Metapath2vec_Concatenation_ffn	*0.87*	*0.95*	*0.77*	*0.87*	*0.87*	*0.85*
Metapath2vec_Concatenation_xgb	*0.87*	*0.94*	*0.79*	*0.87*	*0.87*	*0.85*
BioKG2vec_Concatenation_ffn	*0.88*	*0.92*	*0.83*	*0.88*	*0.88*	*0.84*
Metapath2vec_Average_svm	*0.86*	*0.95*	*0.75*	*0.86*	*0.86*	*0.84*
Metapath2vec_Average_ffn	*0.86*	*0.93*	*0.77*	*0.86*	*0.86*	*0.83*
Metapath2vec_Sum_ffn	*0.86*	*0.92*	*0.79*	*0.86*	*0.86*	*0.83*
DLemb_Concatenation_svm	*0.87*	*0.89*	*0.84*	*0.87*	*0.87*	*0.83*
DLemb_Concatenation_xgb	*0.87*	*0.9*	*0.83*	*0.87*	*0.87*	*0.83*
BioKG2vec_Average_svm	*0.86*	*0.91*	*0.8*	*0.86*	*0.86*	*0.82*
BioKG2vec_Concatenation_xgb	*0.86*	*0.9*	*0.81*	*0.86*	*0.86*	*0.82*
BioKG2vec_Sum_svm	*0.86*	*0.87*	*0.85*	*0.86*	*0.86*	*0.81*
RotatE_Concatenation_svm	*0.85*	*0.89*	*0.79*	*0.85*	*0.85*	*0.81*
DLemb_Concatenation_rf	*0.85*	*0.88*	*0.82*	*0.85*	*0.85*	*0.81*
Metapath2vec_Hadmard_svm	*0.82*	*0.94*	*0.7*	*0.83*	*0.83*	*0.8*
BioKG2vec_Average_ffn	*0.84*	*0.89*	*0.78*	*0.84*	*0.84*	*0.8*
DLemb_Hadmard_xgb	*0.84*	*0.88*	*0.79*	*0.84*	*0.84*	*0.8*
BioKG2vec_Sum_ffn	*0.84*	*0.89*	*0.77*	*0.84*	*0.84*	*0.8*
DLemb_Hadmard_rf	*0.84*	*0.88*	*0.77*	*0.84*	*0.84*	*0.8*
Metapath2vec_Hadmard_xgb	*0.82*	*0.93*	*0.7*	*0.82*	*0.82*	*0.8*
Metapath2vec_Sum_xgb	*0.83*	*0.89*	*0.75*	*0.83*	*0.83*	*0.79*
Metapath2vec_Average_xgb	*0.83*	*0.89*	*0.75*	*0.83*	*0.83*	*0.79*
DLemb_Concatenation_ffn	*0.84*	*0.85*	*0.83*	*0.84*	*0.84*	*0.79*

Metapath2vec_Hadmard_ffn	*0.81*	*0.93*	*0.68*	*0.82*	*0.81*	*0.79*
Metapath2vec_Concatenation_rf	*0.83*	*0.87*	*0.78*	*0.83*	*0.83*	*0.79*
DLemb_Sum_svm	*0.83*	*0.87*	*0.78*	*0.84*	*0.83*	*0.79*
RotatE_Sum_svm	*0.83*	*0.88*	*0.77*	*0.83*	*0.83*	*0.79*
DLemb_Average_svm	*0.83*	*0.87*	*0.78*	*0.83*	*0.83*	*0.78*
DLemb_Hadmard_ffn	*0.83*	*0.87*	*0.77*	*0.83*	*0.83*	*0.78*
DLemb_Hadmard_lr	*0.83*	*0.86*	*0.77*	*0.83*	*0.83*	*0.78*
DLemb_Hadmard_svm	*0.83*	*0.86*	*0.78*	*0.83*	*0.83*	*0.78*
BioKG2vec_Hadmard_ffn	*0.82*	*0.87*	*0.75*	*0.82*	*0.82*	*0.78*
Metapath2vec_Hadmard_lr	*0.8*	*0.91*	*0.67*	*0.81*	*0.81*	*0.78*
BioKG2vec_Hadmard_svm	*0.81*	*0.88*	*0.72*	*0.81*	*0.81*	*0.78*
DLemb_Average_rf	*0.83*	*0.85*	*0.8*	*0.83*	*0.83*	*0.78*
DLemb_Sum_rf	*0.83*	*0.85*	*0.8*	*0.83*	*0.83*	*0.78*
DLemb_Average_xgb	*0.82*	*0.86*	*0.78*	*0.82*	*0.82*	*0.78*
DLemb_Sum_xgb	*0.82*	*0.86*	*0.78*	*0.82*	*0.82*	*0.78*
RotatE_Average_svm	*0.82*	*0.87*	*0.75*	*0.82*	*0.82*	*0.77*
RotatE_Concatenation_xgb	*0.82*	*0.86*	*0.77*	*0.82*	*0.82*	*0.77*
Metapath2vec_Hadmard_rf	*0.8*	*0.89*	*0.69*	*0.8*	*0.8*	*0.77*
BioKG2vec_Hadmard_xgb	*0.81*	*0.86*	*0.74*	*0.81*	*0.81*	*0.77*
BioKG2vec_Concatenation_rf	*0.82*	*0.84*	*0.8*	*0.82*	*0.82*	*0.77*
BioKG2vec_Average_xgb	*0.82*	*0.85*	*0.77*	*0.82*	*0.82*	*0.77*
BioKG2vec_Sum_xgb	*0.82*	*0.85*	*0.77*	*0.82*	*0.82*	*0.77*
RotatE_Concatenation_ffn	*0.82*	*0.83*	*0.8*	*0.82*	*0.82*	*0.76*
RotatE_Concatenation_rf	*0.81*	*0.85*	*0.75*	*0.81*	*0.81*	*0.76*
Metapath2vec_Average_rf	*0.81*	*0.83*	*0.79*	*0.81*	*0.81*	*0.76*
Metapath2vec_Sum_rf	*0.81*	*0.83*	*0.79*	*0.81*	*0.81*	*0.76*
RotatE_Sum_ffn	*0.8*	*0.83*	*0.75*	*0.8*	*0.8*	*0.75*
BioKG2vec_Hadmard_lr	*0.78*	*0.86*	*0.68*	*0.79*	*0.79*	*0.74*
TransE_Concatenation_svm	*0.8*	*0.82*	*0.75*	*0.8*	*0.8*	*0.74*
RotatE_Average_xgb	*0.79*	*0.84*	*0.71*	*0.79*	*0.79*	*0.74*
RotatE_Sum_xgb	*0.79*	*0.84*	*0.71*	*0.79*	*0.79*	*0.74*

DLemb_Sum_ffn	*0.79*	*0.8*	*0.78*	*0.8*	*0.79*	*0.73*
BioKG2vec_Sum_rf	*0.8*	*0.79*	*0.8*	*0.8*	*0.8*	*0.73*
RotatE_Sum_rf	*0.78*	*0.84*	*0.69*	*0.78*	*0.78*	*0.73*
RGCN_Concatenation_svm	*0.78*	*0.84*	*0.69*	*0.78*	*0.78*	*0.73*
RotatE_Average_rf	*0.78*	*0.83*	*0.69*	*0.78*	*0.78*	*0.73*
BioKG2vec_Hadmard_rf	*0.79*	*0.8*	*0.78*	*0.79*	*0.79*	*0.73*
BioKG2vec_Average_rf	*0.79*	*0.79*	*0.8*	*0.79*	*0.79*	*0.73*
RotatE_Hadmard_xgb	*0.77*	*0.84*	*0.68*	*0.77*	*0.77*	*0.73*
RotatE_Hadmard_rf	*0.77*	*0.83*	*0.68*	*0.77*	*0.77*	*0.73*
TransE_Concatenation_rf	*0.78*	*0.81*	*0.72*	*0.78*	*0.78*	*0.72*
DLemb_Average_ffn	*0.78*	*0.79*	*0.77*	*0.78*	*0.78*	*0.72*
RotatE_Hadmard_svm	*0.77*	*0.83*	*0.68*	*0.77*	*0.77*	*0.72*
TransE_Concatenation_xgb	*0.78*	*0.8*	*0.75*	*0.78*	*0.78*	*0.72*
RotatE_Average_ffn	*0.78*	*0.78*	*0.77*	*0.78*	*0.78*	*0.72*
RGCN_Average_svm	*0.75*	*0.83*	*0.63*	*0.75*	*0.75*	*0.71*
RGCN_Concatenation_xgb	*0.75*	*0.82*	*0.65*	*0.75*	*0.75*	*0.7*
RotatE_Hadmard_ffn	*0.75*	*0.8*	*0.68*	*0.76*	*0.76*	*0.7*
RotatE_Hadmard_lr	*0.74*	*0.81*	*0.64*	*0.75*	*0.75*	*0.7*
RGCN_Concatenation_rf	*0.75*	*0.79*	*0.67*	*0.75*	*0.75*	*0.69*
TransE_Concatenation_ffn	*0.75*	*0.76*	*0.74*	*0.75*	*0.75*	*0.69*
DLemb_Concatenation_lr	*0.75*	*0.76*	*0.72*	*0.75*	*0.75*	*0.69*
TransE_Sum_svm	*0.74*	*0.76*	*0.7*	*0.74*	*0.74*	*0.68*
RotatE_Concatenation_lr	*0.74*	*0.75*	*0.72*	*0.74*	*0.74*	*0.68*
Metapath2vec_Concatenation_lr	*0.73*	*0.75*	*0.7*	*0.73*	*0.73*	*0.67*
RGCN_Sum_svm	*0.73*	*0.74*	*0.7*	*0.73*	*0.73*	*0.67*
BioKG2vec_Concatenation_lr	*0.73*	*0.74*	*0.7*	*0.73*	*0.73*	*0.67*
TransE_Concatenation_lr	*0.73*	*0.74*	*0.71*	*0.73*	*0.73*	*0.67*
TransE_Average_svm	*0.71*	*0.74*	*0.65*	*0.71*	*0.71*	*0.66*
TransE_Hadmard_xgb	*0.71*	*0.74*	*0.65*	*0.71*	*0.71*	*0.66*
Metapath2vec_Sum_lr	*0.72*	*0.73*	*0.69*	*0.72*	*0.72*	*0.66*
Metapath2vec_Average_lr	*0.72*	*0.73*	*0.69*	*0.72*	*0.72*	*0.66*

TransE_Hadmard_rf	*0.71*	*0.75*	*0.63*	*0.71*	*0.71*	*0.65*
BioKG2vec_Sum_lr	*0.71*	*0.72*	*0.69*	*0.71*	*0.71*	*0.65*
BioKG2vec_Average_lr	*0.71*	*0.72*	*0.69*	*0.71*	*0.71*	*0.65*
DLemb_Average_lr	*0.71*	*0.71*	*0.7*	*0.71*	*0.71*	*0.65*
DLemb_Sum_lr	*0.71*	*0.71*	*0.7*	*0.71*	*0.71*	*0.65*
TransE_Average_rf	*0.71*	*0.73*	*0.66*	*0.71*	*0.71*	*0.65*
TransE_Sum_rf	*0.7*	*0.73*	*0.65*	*0.71*	*0.71*	*0.65*
TransE_Sum_xgb	*0.71*	*0.72*	*0.68*	*0.71*	*0.71*	*0.65*
TransE_Average_xgb	*0.71*	*0.72*	*0.68*	*0.71*	*0.71*	*0.65*
RGCN_Concatenation_ffn	*0.7*	*0.72*	*0.65*	*0.7*	*0.7*	*0.65*
TransE_Sum_ffn	*0.71*	*0.71*	*0.7*	*0.71*	*0.71*	*0.65*
RotatE_Sum_lr	*0.7*	*0.71*	*0.69*	*0.7*	*0.7*	*0.64*
RotatE_Average_lr	*0.7*	*0.71*	*0.69*	*0.7*	*0.7*	*0.64*
TransE_Hadmard_svm	*0.69*	*0.72*	*0.62*	*0.69*	*0.69*	*0.64*
TransE_Hadmard_lr	*0.69*	*0.72*	*0.62*	*0.69*	*0.69*	*0.63*
TransE_Hadmard_ffn	*0.69*	*0.7*	*0.65*	*0.69*	*0.69*	*0.63*
TransE_Average_ffn	*0.68*	*0.71*	*0.6*	*0.68*	*0.68*	*0.63*
TransE_Sum_lr	*0.67*	*0.66*	*0.68*	*0.67*	*0.67*	*0.61*
TransE_Average_lr	*0.67*	*0.66*	*0.67*	*0.67*	*0.67*	*0.61*
RGCN_Average_ffn	*0.65*	*0.67*	*0.59*	*0.65*	*0.65*	*0.6*
RGCN_Sum_ffn	*0.65*	*0.65*	*0.64*	*0.65*	*0.65*	*0.6*
RGCN_Average_rf	*0.6*	*0.6*	*0.61*	*0.6*	*0.6*	*0.56*
RGCN_Sum_rf	*0.59*	*0.59*	*0.59*	*0.59*	*0.59*	*0.55*
RGCN_Average_xgb	*0.58*	*0.58*	*0.6*	*0.58*	*0.58*	*0.55*
RGCN_Sum_xgb	*0.58*	*0.58*	*0.6*	*0.58*	*0.58*	*0.55*
RGCN_Concatenation_lr	*0.56*	*0.56*	*0.59*	*0.56*	*0.56*	*0.53*
RGCN_Hadmard_xgb	*0.54*	*0.53*	*0.63*	*0.54*	*0.54*	*0.52*
RGCN_Average_lr	*0.54*	*0.54*	*0.59*	*0.54*	*0.54*	*0.52*
RGCN_Sum_lr	*0.54*	*0.54*	*0.59*	*0.54*	*0.54*	*0.52*
RGCN_Hadmard_rf	*0.48*	*0.51*	*0.81*	*0.53*	*0.53*	*0.51*
RGCN_Hadmard_svm	*0.51*	*0.51*	*0.64*	*0.51*	*0.51*	*0.5*

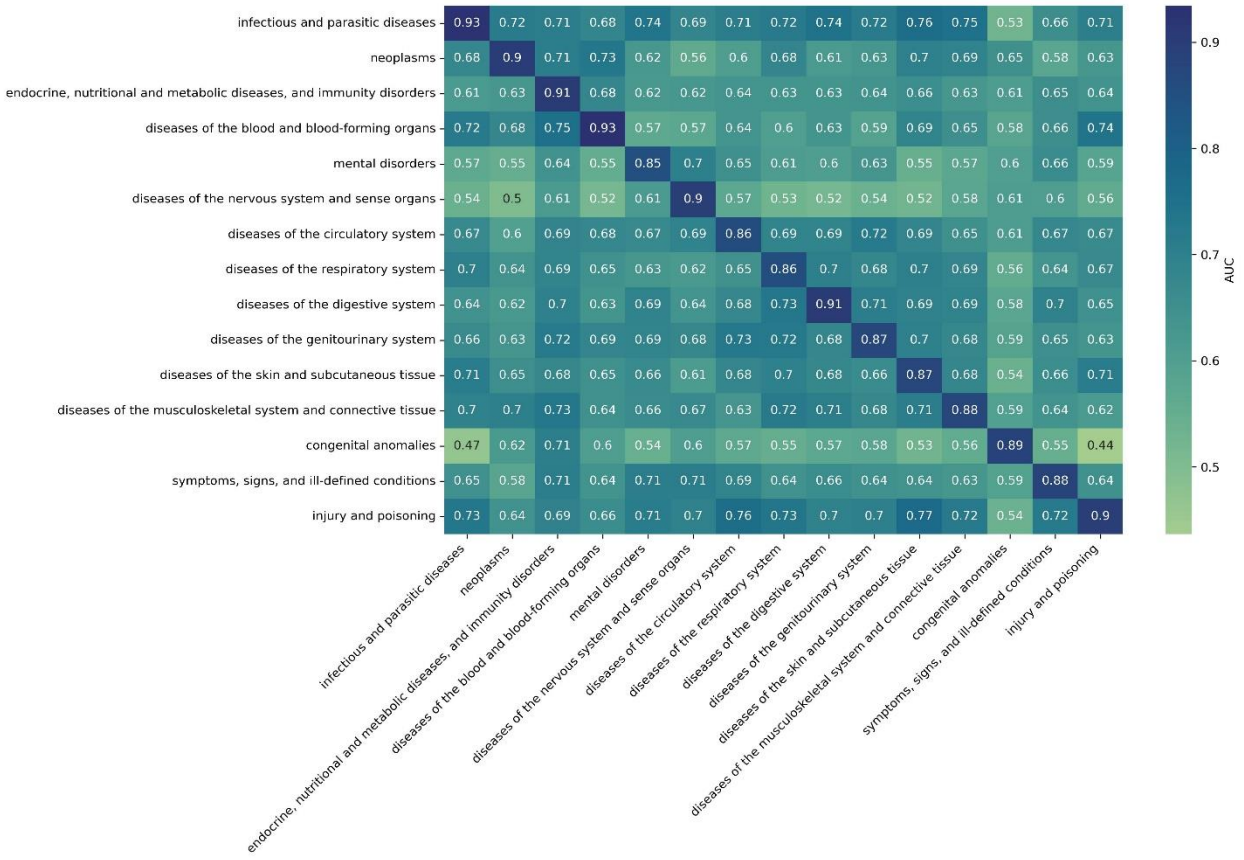
RGCN_Hadmard_ffn	*0.51*	*0.51*	*0.6*	*0.51*	*0.51*	*0.5*
RGCN_Hadmard_lr	*0.49*	*0.5*	*0.65*	*0.5*	*0.5*	*0.5*

Supplementary Table 2 effect of data preprocessing and integration on ROCAUC and PRAUC using DLemb algorithm. Preprocessing of the data significantly improves predictive performance of the model.

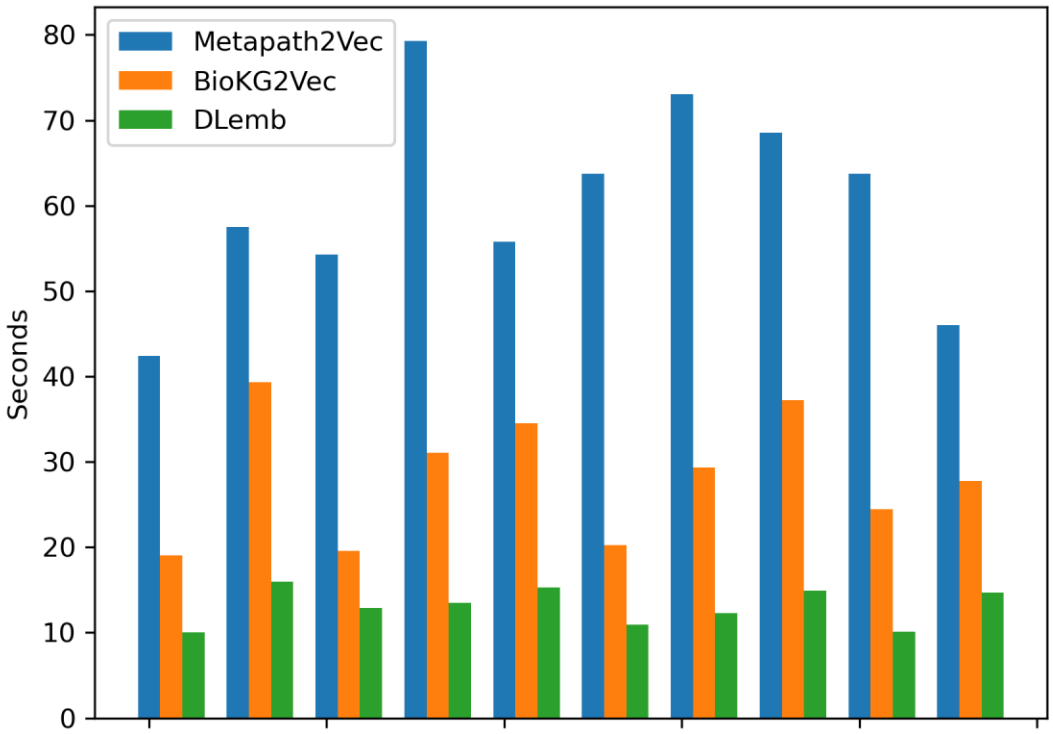
Experiment	**ROCAUC**	**AUPRC**
HPO + HPO annotations raw	0.76	0.87
HPO + HPO annotations processed	0.92	0.96
HPO + HPO annotations + GO + GO annotations processed	0.92	0.93

Supplementary Table 3: Results of 5-fold cross-validation on randomly selected diseases belonging to different number of GDA stratifications (Embedding creation algorithm is Metapath2vec, classification algorithm is support vector machine and GDA representation is concatenation).

model		**Metapath2vec**					**Random**				
CUI	**number of associations**	**test_accuracy**	**test_precision**	**test_recall**	**test_f1**	**test_roc_auc**	**test_accuracy**	**test_precision**	**test_recall**	**test_f1**	**test_roc_auc**
C0026764	42	0.88	0.92	0.85	0.88	0.95	0.58	0.57	0.6	0.56	0.66
C0020517	64	0.92	0.94	0.91	0.92	0.99	0.51	0.51	0.57	0.53	0.48
C0013421	86	0.93	0.93	0.95	0.94	0.98	0.48	0.48	0.51	0.49	0.46
C0023890	103	0.89	0.89	0.9	0.89	0.96	0.46	0.45	0.45	0.45	0.43
C0007134	128	0.84	0.85	0.84	0.84	0.92	0.52	0.52	0.53	0.52	0.53
C0032460	144	0.78	0.79	0.76	0.77	0.84	0.45	0.45	0.49	0.47	0.44
C0014175	161	0.85	0.85	0.84	0.84	0.92	0.49	0.49	0.48	0.49	0.46
C0151744	176	0.85	0.84	0.85	0.85	0.92	0.48	0.48	0.45	0.46	0.46
C0015397	212	0.95	0.94	0.96	0.95	0.98	0.45	0.45	0.46	0.46	0.41
C0004352	261	0.86	0.85	0.87	0.86	0.93	0.45	0.45	0.45	0.45	0.42



Supplementary Figure 6: Generalization capabilities of DLemb algorithm, the KGE obtained with the implementation of this algorithm can predict GDAs across different disease classes.



Supplementary Figure 7: Comparison of algorithms performance. On the x - axis are reported the ten experiments of creating embeddings of subnetworks generated by random sampling 10000 nodes from the knowledge graph, on the y-axis, the time in seconds.

3.2 Genopyc: a Python library for investigating the functional effects of genomic variants associated to complex diseases

This chapter is based on:

Genopyc: a Python library for investigating the genomic basis of complex diseases

Francesco Gualdi[1,2], Baldomero Oliva[2] and Janet Piñero[1,3 *]

[1]*Integrative biomedical informatics, Research program on Biomedical informatics (IBI-GRIB), Hospital del mar medical research institute IMIM Department of experimental and health sciences, C/Dr Aiguader 88, E-08003 Barcelona, Spain*

[2]*Structural bioinformatics lab, Research program on Biomedical informatics (SBI-GRIB), Hospital del mar medical research institute IMIM Department of Experimental and health sciences, C/Dr Aiguader 88, E-08003 Barcelona, Spain*

[3] *Medbioinformatics Solutions SL, Barcelona, Spain*

Published in: Oxford bioinformatics

DOI: *doi.org/10.1093/bioinformatics/btae379*

Summary of the work

In this chapter, I will accomplish objective 2 and 2.a. Firstly I developed a Python library to investigate functional consequences of genetic variants associated with complex diseases through Genome wide association studies (GWAS). Subsequently, I implemented it to interpret a set of single nucleotide polymorphisms (SNPs) associated with intervertebral disc degeneration (IDD). GWAS have successfully associated genetic loci to complex diseases, however, interpreting how the variants affect genes is a complex task that requires the integration of heterogeneous data and the implementation of different types of analysis. Often this is accomplished by accessing multiple repositories and with the implementation of many tools resulting in a time-consuming process. Thus, we developed Genopyc, a Python library to investigate genomic underpinnings of variants associated with complex traits. The library allows to perform multiple tasks on genomic data such as obtaining variants associated with a condition, retrieving linkage disequilibrium (LD) data, expression quantitative trait loci (eQTL), mapping between different variants and gene IDs and investigate functional effects of variants through the implementation of different tools such as variant effect predictor (VEP) and Locus to Gene pipeline from Open targets platform. Moreover, the tool provides different visualization features to better understand the results obtained. The library is freely available at https://pypi.org/project/genopyc/ and can be integrated in any Python pipeline. To showcase the library, we analized a set of SNPs associated with IDD (See more details at https://github.com/freh-g/genopyc/blob/main/tutorials/Genopyc_tutorial_notebook.ipynb). By performing a functional enrichment analysis in the set of genes obtained through the implementation of the library, we were able to prioritize transcriptional pathways that are largely reported in literature to be associated with the diseases in *in vitro* and *in vivo* models. These results are in concordance with chapter 3.1 where we prioritized genes that were reported to act in synergy with genes prioritized from Genopyc and with chapter 3.3 where we investigated the autoimmune basis of Modic change, a comorbidity of IDD. In fact, it was theorized that IDD could have autoimmune basis also due to the fact that many SNPs related to the conditions lay on chromosome 6 in the major histocompatibility complex (MHC). However we showed that by a careful data integration the attention is shifted from autoimmune pathways to pathways already associated with the condition such as AP-2α , HIF-1 and SP1. Taken together these results suggest that IDD does not have an autoimmune etiology and that Genopyc can be successfully implemented to better understand functional consequences of variants associated with complex traits.

Abstract

Motivation: Understanding the genetic basis of complex diseases is a paramount challenge in modern genomics. However, current tools often lack the versatility to efficiently analyze the intricate relationships between genetic variations and disease outcomes. To address this, we introduce Genopyc, a novel Python library designed for comprehensive investigation of the genetics underlying complex diseases. Genopyc offers an extensive suite of functions for heterogeneous data mining and visualization, enabling researchers to delve into and integrate biological information from large-scale genomic datasets with ease.

Results: In this study, we present the Genopyc library through application to real-world genome wide association studies variants. Using Genopyc to investigate variants associated to intervertebral disc degeneration (IDD) enabled a deeper understanding of the potential dysregulated pathways involved in the disease, which can be explored and visualized by exploiting the functionalities featured in the package. Genopyc emerges as a powerful asset for researchers, fostering advancements in the understanding of complex diseases and thus paving the way for more targeted therapeutic interventions.

Availability: Genopyc is available on pip https://pypi.org/project/genopyc/.The source code of Genopyc is available at https://github.com/freh-g/genopyc. A tutorial notebook is available at https://github.com/freh-g/genopyc/blob/main/tutorials/Genopyc_tutorial_notebook.ipynb

Finally, a detailed documentation is available at: https://genopyc.readthedocs.io/en/latest/

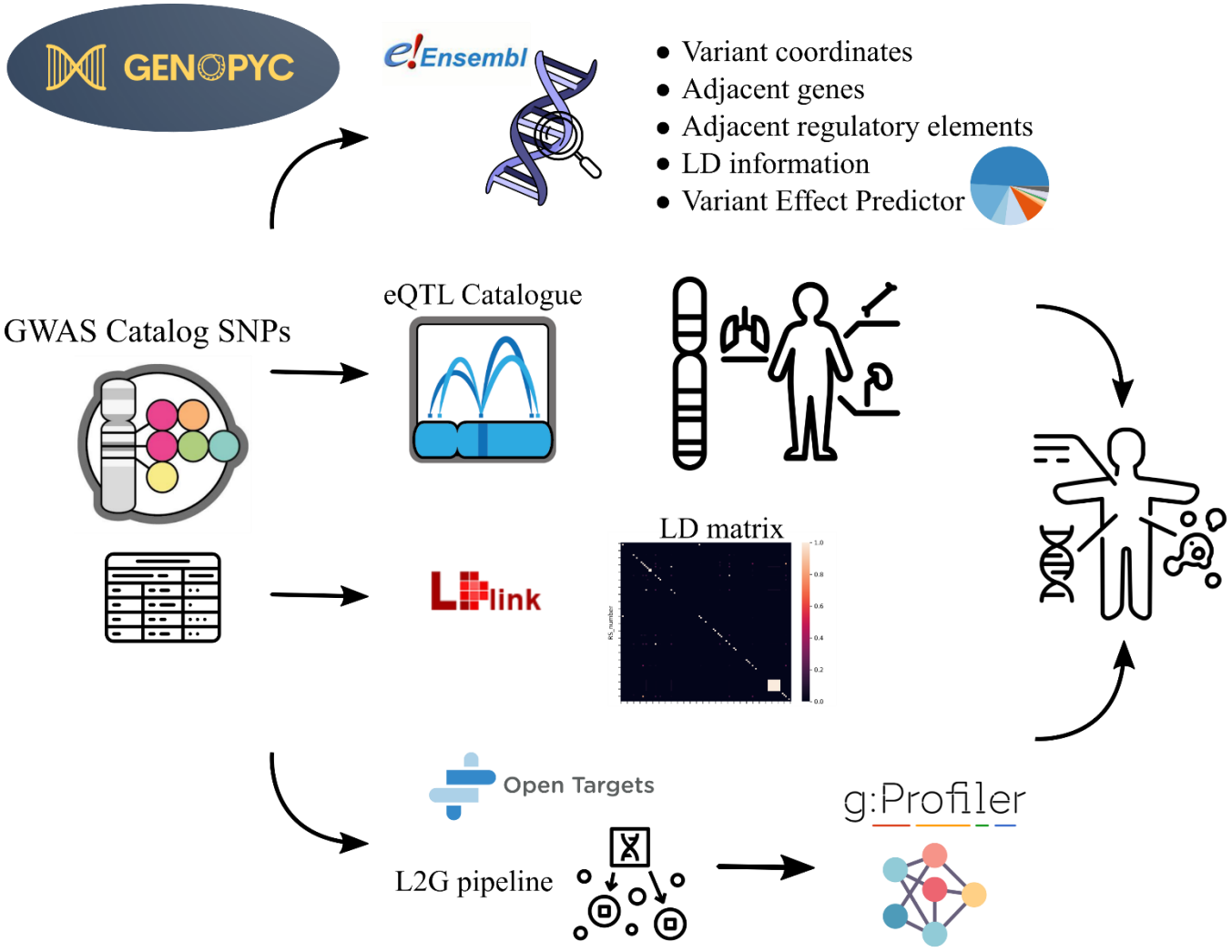


Figure 1 *The main Genopyc features and knowledge bases accessed schematically represented. Variants associated with a specific trait are initially obtained from GWAS catalog and then subjected to various analyses, including examination of genomic context, LD features, eQTLs, VEP and Locus2gene pipeline. Subsequently, as the variants are linked to genes through these analyses, the functions enriched within the gene set can be explored to identify potential dysregulated pathways relevant to the disease.*

Introduction

The onset of complex disorders is influenced by a multitude of components that include lifestyle, diet, environmental and genetic factors. In the last decades genome wide association studies (GWAS) have emerged as a powerful tool to investigate the genetic architecture underlying complex diseases [155]. However, now that thousands of genetic risk factors for numerous phenotypes have been discovered, we are facing another challenge: the interpretation of these associations in the biological context, we are thus entering in the so called post-GWAS Era [24]. Understanding how genetic variants are translated into biological pathways remains a challenging

task [156] and led to the development of numerous approaches to interpret GWAS results (see [25] for a comprehensive review of the type of analysis and tools).

Ascertaining the precise functional implications of the genetic variants discovered through GWAS has proven to be a formidable challenge due to the extensive amount of data required to perform these studies [156]. In response, a plethora of novel methodological approaches has emerged to address this knowledge gap [98]. These techniques rely on the large-scale omics datasets and repositories available to researchers such as Gene Expression Omnibus [157], the genotype – tissue expression project [158] and the Encode project [28]. The enormous amount of data regarding genes and variants associated to diseases is collected in knowledge bases such as the GWAS Catalog [159], containing a curated collection of GWAS, and DisGeNET [130] that offers a standardized integration from different sources. However, 90% of the genetic variation associated to complex diseases are non-coding type and a benchmark of methods to interpret how they alter genes, perturb biological pathways and ultimately lead to disease is still missing [160]. Moreover, the application and integration of different tools to analyze GWAS data lead to discordant results, thus an unbiased assessment of the methods available is still required [161]. Finally, the tools and data repositories useful for analyzing this type of data are scattered, which makes understanding GWAS results a time-consuming process.

An advancement in associating genes to non-coding variants has been made by the Open Target Genetics platform, which implemented a pipeline consisting of a machine learning model that uses heterogeneous features such as distance from variant to the gene, expression quantitative trait loci (eQTL), chromatin conformation and variant effect predictions. This method outperformed the naïve distance-based methods in the prioritization of causal genes related to complex diseases loci [162]. In this context we present Genopyc, a Python library for investigating the genetic basis of complex diseases.

Genopyc users to programmatically access multiple sources with the aim of understanding how non-coding variants impact the biological pathways and thus infer the mechanisms underlying the development of complex diseases (Figure 1). Moreover, being fully integrated in Python, all the downstream statistical analysis can be carried out in the same environment.

Implementation and features

Genopyc is a Python package integrating information from several knowledgebases. The tool can receive as an input a trait, coded with Experimental Factor Ontology identifiers [163], or the results of a GWA study. If an EFO code is used as an input, the variants associated to the trait are retrieved from the GWAS Catalog. Information such as the reported β coefficient that quantifies the association between a genetic variant and a trait or phenotype of interest, standard error (relevant for understanding the dispersion of the estimated beta coefficient), risk allele frequency (the frequency of the allele in the reference population) and the mapped genes closest to the SNPs are also retrieved.

Additionally, other features such as genomic coordinates, linkage disequilibrium (LD) correlated SNPs and neighboring functional elements can be obtained by querying Ensembl Genome Browser [27]. Genopyc also integrates the variant effect predictor (VEP) to obtain the consequences of the

SNPs on the transcript and its effect on neighboring genes and functional elements [42]. Often SNPs associated to complex phenotypes fall in non-coding regions of the genome and are more likely to have regulatory effects [164]. Therefore, it is possible to retrieve the expression eQTLs related to variants through the eQTL Catalogue [165]. Finally, Genopyc integrates the locus to gene (L2G) pipeline from Open Target Genetics to uncover the target gene or genes of variants located in non-coding regions.

Once a variant is associated to a gene or genes, the significantly enriched pathways are retrieved through G:profiler [142]. In this way the user can elucidate the functions whose perturbation could ultimately lead to the disease. Genopyc package also offers a functionality to visualize the results of the functional enrichment as an interactive network. In this network, genes of interest are mapped to a protein-protein interaction network derived from the HIPPIE database [166] in which nodes represent the gene products and edges correspond to the physical interactions between proteins. A dropdown menu allows the user to select the function enriched in the gene set and, when a function is selected, the gene-products belonging to that function are highlighted.

Genopyc can also be implemented to retrieve a linkage-disequilibrium (LD) matrix for a set of SNPs by using LDlink [167], convert genome coordinates between genome versions and retrieve genes coordinates in the genome. LDlink calculates the LD matrix through the population-specific 1000 genomes haplotype panels [168]. Retrieving genome coordinates and mapping between genome builds are made possible by accessing Ensembl genome browser.

Comparison with state-of-the-art algorithm

A comparison between the main functionalities of Genopyc and other tools for post-GWAS analysis is shown in Table 1. Genopyc is the only library that integrates multiple analysis to connect variants to genes (conditional, colocalization, fine mapping), gather functional information to annotate variants (eQTLs, HI-C, linkage disequilibrium, VEP, functional genomic elements), maps between different vocabularies of gene and variant identifiers and perform functional enrichment to detect possible pathways perturbed by genetic variations.

Table 1 *Comparison between Genopyc and the main tools for post GWAS analysis. Genopyc integrates diverse functionalities allowing a more flexible investigation of variants related to diseases*

Tool	Mapping Ids	Retrieve trait associated variants	Conditional analysis	Fine mapping	eQTL Co-localization	HI-C	Functional Enrichment	Linkage Disequilibrium	Genomic features	Variant annotation
Genopyc	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Coloc	×	×	×	×	✓	×	×	×	×	✓
FUMA	×	×	✓	×	✓	✓	×	✓	×	×
Finemap	×	×	×	✓	×	×	×	✓	×	×
Ensemble API	✓	×	×	×	×	×	×	✓	✓	✓
Open Targets: Genetics	✓	✓	✓	✓	✓	✓	×	✓	×	✓

To compare a feature of Genopyc capabilities of associating variants to genes we will compare we will compare the set of genes prioritized from the L2G pipeline included in Genopyc with the SNP2GENE pipeline of FUMA [169]. To perform this comparison, we will use variants from a GWAS on Chron's disease (CD) by Frank *et. al* [170]. By the implementation of L2G through Genopyc we were able to prioritize 216 genes while FUMA a total of 255 genes were associated with the studied variants, of this set there was an overlapping of 142 genes, thus the majority of the genes belonging to each set is in common between the 2 methods (Figure 2).

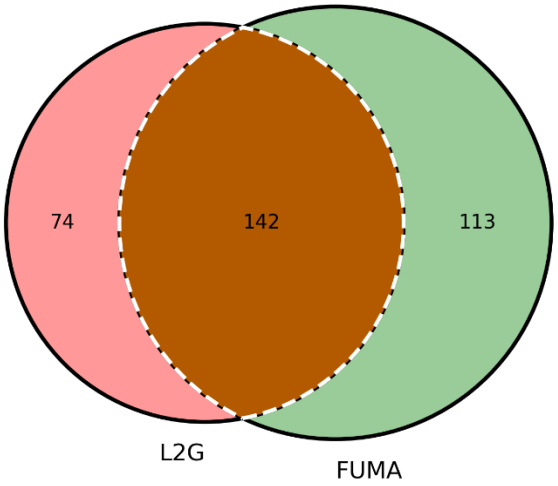


Figure 2 *Venn plot showing the prioritized genes through the implementation of L2G pipeline (pink circle) and FUMA (green circle).*

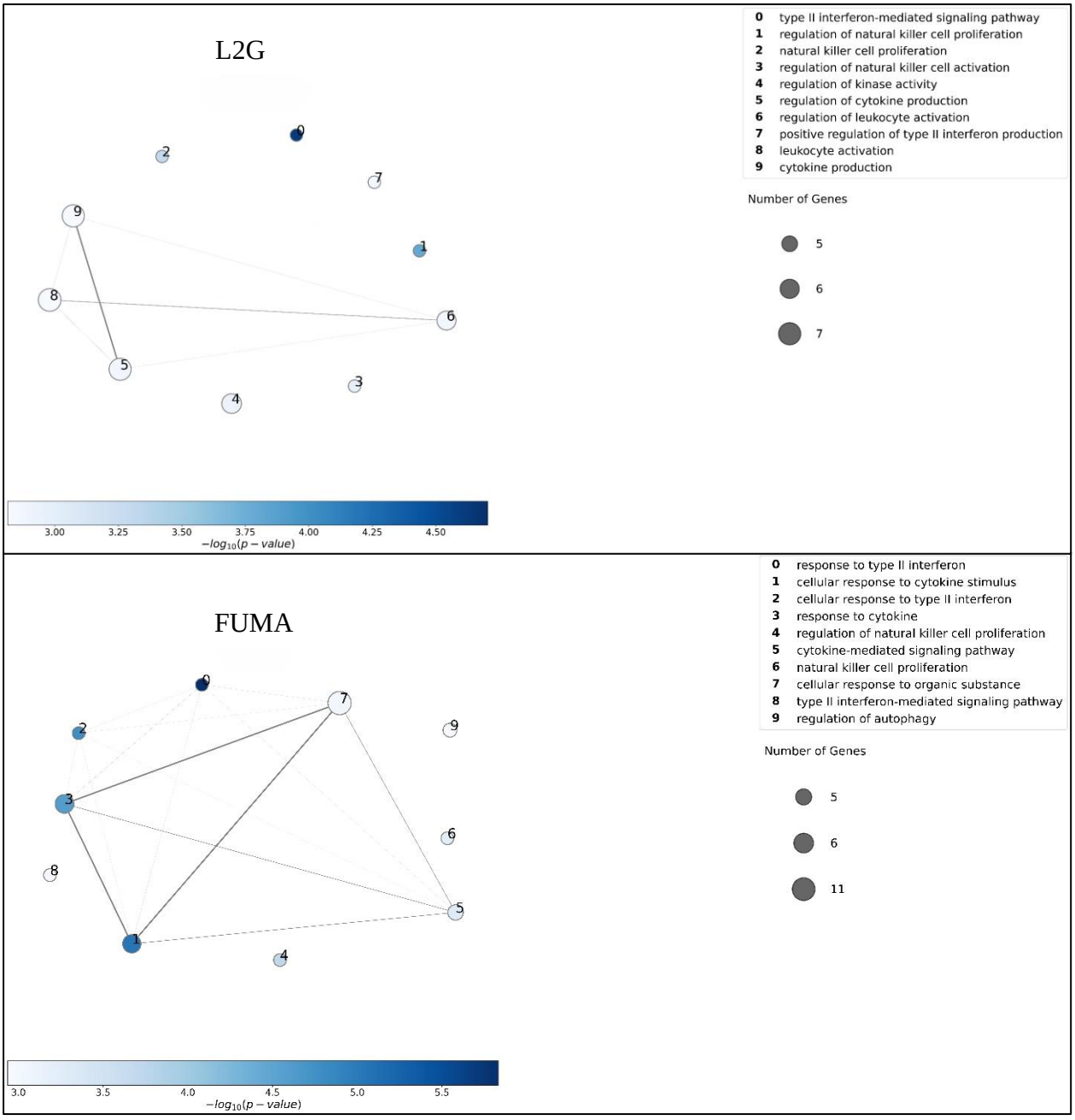


Figure 3 *Function enrichment analysis on gene set prioritized by the implementation of L2G and FUMA pipeline. Nodes in the network correspond to functions. Node size reflects the number of genes associated to that specific function while node color is related to the significance of the function enrichment calculated through hypergeometric distribution and expressed as -log₁₀ p-value. An edge exists between 2 nodes if the function shares genes; the number of shared genes is proportional to the edge thickness. Both gene sets present IFN type II pathway as the most represented function.*

By performing a function enrichment analysis on the prioritized gene sets we obtained similar results (Figure 3). In fact, both results reported that the interferon type II signalling pathway is the most represented function among the gene. This finding is further supported in literature in which it was shown that Chron's disease patients intestinal CD4+ cells produce an increased amount of

IFN – γ, in fact during CD there is an accumulation of T helper 1 cells that mediates the production of type II interferon [171]. This show that the genes obtained by the implementation of L2G pipeline through Genopyc brings to comparable results to state – of – the – art algorithms. Moreover, it must be underlined that L2G is only one of the many features that Genopyc offers, in fact the library includes many more functionalities, some of which will be presented in the next paragraph.

Use Case: Intervertebral disc degeneration

To illustrate the utility of Genopyc, we applied it to the variants associated to IDD that are available in the GWAS catalog (EFO:0004994). IDD is a complex multifactorial condition for which the molecular mechanisms are poorly understood [172]. In Table 1 are displayed all the variants associated to IDD that we obtained by querying GWAS catalog through Genopyc.

Table 2 *Variant associated to intervertebral disc degeneration from GWAS catalog. The table reports the risk allele (RA), the p-value of the association (PV), the risk allele frequency (RAF) the beta reported from the GWAS association (B) the confidence interval of the association (CI) and the mapped gene from the authors of the study (MG). A ? is reported in the RA section if the author of the study didn't specify the risk allele related to the SNP.*

RA	PV	RAF	B	CI	MG
rs17034687-C	2E-09	0.09	0.23 unit increase	[0.16-0.30]	CRBN,SUMF1
rs2187689-C	3E-08	0.08	0.23 unit increase	[0.15-0.31]	HLA-Z,PPP1R2P1
rs926849-C	3E-08	0.31	0.13 unit decrease	[0.083-0.177]	PRKN
rs7744666-C	6E-08	0.10	0.2 unit increase	[0.13-0.27]	PPP1R2P1,HLA-Z
rs11969002-A	6E-08	0.10	0.2 unit increase	[0.13-0.27]	HLA-Z,PPP1R2P1
rs4802666-A	0.000005	0.27	0.13 unit decrease	[0.073-0.187]	MYH14
rs7896691-C	0.000002	0.10	0.17 unit increase	[0.10-0.24]	PFKP
rs10998466-A	0.000004	0.01	0.53 unit decrease	[0.31-0.75]	STOX1
rs1981483-A	0.000004	0.42	0.11 unit increase	[0.065-0.155]	PIGQ
rs1154053-C	0.000004	0.17	0.13 unit decrease	[0.075-0.185]	CSMD1
rs2484990-C	0.000004	0.01	0.68 unit increase	[0.39-0.97]	LINC02664,ZEB1-AS1
rs1250307-A	0.000004	0.01	0.68 unit increase	[0.39-0.97]	LINC02664,ZEB1-AS1

rs7204439-C	0.000004	0.42	0.11 unit increase	[0.065-0.155]	RAB40C
rs2484992-C	0.000005	0.01	0.68 unit increase	[0.39-0.97]	ZEB1-AS1,LINC02664
rs9488238-A	0.000005	0.04	0.28 unit decrease	[0.16-0.40]	LINC02541,MARCKS
rs1205863-G	0.000006	0.06	0.21 unit increase	[0.12-0.30]	AMD1P4,HIVEP1
rs11918654-C	0.000007	0.27	0.11 unit decrease	[0.063-0.157]	ARL8B
rs2657195-A	0.000008	0.22	0.13 unit decrease	[0.075-0.185]	SLC26A7,RN7SKP231
rs11754641-C	0.000008	0.03	0.29 unit increase	[0.16-0.42]	EYS
rs12805875-A	0.000009	0.42	0.09 unit increase	[0.049-0.131]	MIR4693,DYNC2H1
rs980238-A	0.000009	0.30	0.1 unit decrease	[0.055-0.145]	CSMD1
rs7103004-C	0.000009	0.42	0.09 unit increase	[0.049-0.131]	DYNC2H1,MIR4693
rs4554859-G	0.000009	0.42	0.09 unit increase	[0.049-0.131]	DYNC2H1,MIR4693
rs7118412-A	0.000009	0.42	0.09 unit increase	[0.049-0.131]	DYNC2H1,MIR4693
rs2017567-C	0.000009	0.42	0.1 unit increase	[0.059-0.141]	RAB40C,PIGQ
rs6457690-A	9E-08	0.10	0.19 unit increase	[0.12-0.26]	PPP1R2P1,HLA-Z
rs1029296-C	9E-08	0.10	0.19 unit increase	[0.12-0.26]	PPP1R2P1,HLA-Z
rs6936004-C	1E-07	0.10	0.19 unit increase	[0.12-0.26]	PPP1R2P1,HLA-Z
rs3749982-A	1E-07	0.10	0.19 unit increase	[0.12-0.26]	PPP1R2P1,HLA-Z
rs9469300-A	1E-07	0.10	0.19 unit increase	[0.12-0.26]	PPP1R2P1,HLA-Z
rs10214886-A	2E-07	0.10	0.19 unit increase	[0.12-0.26]	HLA-Z,PPP1R2P1
rs10046257-A	3E-07	0.10	0.19 unit increase	[0.12-0.26]	HLA-Z,PPP1R2P1

rs4875102-A	4E-07	0.26	0.12 unit decrease	[0.073-0.167]	CSMD1
rs1029295-C	5E-07	0.10	0.19 unit increase	[0.12-0.26]	HLA-Z,PPP1R2P1
rs9301951-C	9E-07	0.04	0.26 unit decrease	[0.15-0.37]	GPC6
rs1245582-?	4E-08	0.44	-	[1.13-1.29]	CHST3,SPOCK2
rs1245582-?	0.000001	0.44	-	-	CHST3,SPOCK2
rs7907616-?	2E-08	NR	-	[1.05-1.11]	CHST3
rs4148933-?	3E-08	NR	-	[1.05-1.11]	CHST3
rs4284332-?	3E-08	NR	-	[1.05-1.11]	CHST3
rs7163797-?	4E-10	NR	-	[0.89-0.94]	SMAD3

The data from retrieved from GWAS catalog report the variant associated with the trait of interest, the p-value that express the significance of the association (significant variants usually have a pvalue < $5x10^{-8}$), the beta coefficient that quantifies the influence of the SNP on the trait of interest, the coefficient interval referred to the beta coefficient and the mapped gene reported from the authors of the study that is solely based on the distance to the detected SNP https://www.ebi.ac.uk/gwas/docs/fileheaders.

Subsequently, in order to provide a more comprehensive view of the genomic context in which the variant are located Genopyc allow the user of retrieving variant genomic locations (Table 3) and the neighboring functional elements. In Table 4 are shown an example of 6 transcription factor binding sites (TFBS) motifs in a 500kb window centered on the SNP rs1981483.

Table3 *Genomic locations of variants associated to IDD. The columns correspond to the rsid of the variant, the chromosome the position expressed as base pair.*

rsid	**chromosome**	**position**
rs17034687	3	3638168
rs2187689	6	32884870
rs926849	6	161740587
rs7744666	6	32891935

rs11969002	6	32891971
rs4802666	19	50217817
rs7896691	10	3112981
rs10998466	10	68866673
rs1981483	16	580665
rs1154053	8	4427868
rs2484990	10	31226203
rs1250307	10	31207045
rs7204439	16	611335
rs2484992	10	31223169
rs9488238	6	113695931
rs1205863	6	11943293
rs11918654	3	5146561
rs2657195	8	91547687
rs11754641	6	64926030
rs12805875	11	103658904
rs980238	8	4425096
rs7103004	11	103655296
rs4554859	11	103659638
rs7118412	11	103655026
rs2017567	16	587212
rs6457690	6	32887940
rs1029296	6	32888604
rs6936004	6	32889157
rs3749982	6	32894830
rs9469300	6	32892975
rs10214886	6	32889642
rs10046257	6	32886920
rs4875102	8	4427170
rs1029295	6	32888705
rs9301951	13	94300578

rs1245582	10	72018509
rs1245582	10	72018509
rs7907616	10	72000418
rs4148933	10	72000132
rs4284332	10	71974194
rs7163797	15	67072574

Table4 *TFBS overlapping rs1981483. The table reports the ensembl ID of the sequence (binding_matrix_stable_id) the ending coordinates, the feature type, the effect of the SNP on the motif score (score) that is a quantification of the binding affinity between the genomic sequence and the TF, the chromosome (seq_region_name), the ensembl id, the starting bp coordinates (start), the strand (1 if forward and -1 if reverse) and the transcription factor complex that binds that specific*

binding_matrix_stable_id	end	feature_type	score	seq_region_name	stable_id	start	strand	transcription_factor_complex
ENSPFM0550	580571	motif	0.167248	16	ENSM00536901928	580547	-1	TEAD4::FOXI1
ENSPFM0455	580579	motif	-0.61226	16	ENSM00607925648	580550	1	POU2F1::EOMES
ENSPFM0547	580578	motif	-0.35399	16	ENSM00621307960	580551	-1	TEAD4::FOXI1
ENSPFM0548	580578	motif	2.04745	16	ENSM00555128830	580552	1	TEAD4::FOXI1
ENSPFM0467	580568	motif	0.571308	16	ENSM00625515756	580554	1	PRDM1
ENSPFM0330	580572	motif	3.997266	16	ENSM00364342102	580556	1	HOXB2::SOX15

The genomic location of a SNP is a really important feature, but it doesn't bring information relative to the actual effect of the variant. To determine the functional consequences of SNPs, we should investigate their impact on gene transcription. Through the implementation of Genopyc we can retrieve all the genes that are eQTLs for that specific variant in different tissues. This is done by querying eQTL catalogue a repository of harmonized eQTL studies. Moreover, the library gives the functionality of plotting the results as a network to have a more comprehensive view on how the set of variants investigated is affecting the transcription of the genes. Since the eQTL catalogue does not include relevant tissue for IDD, the results will be plotted for musculoskeletal tissue. (Figure 4).

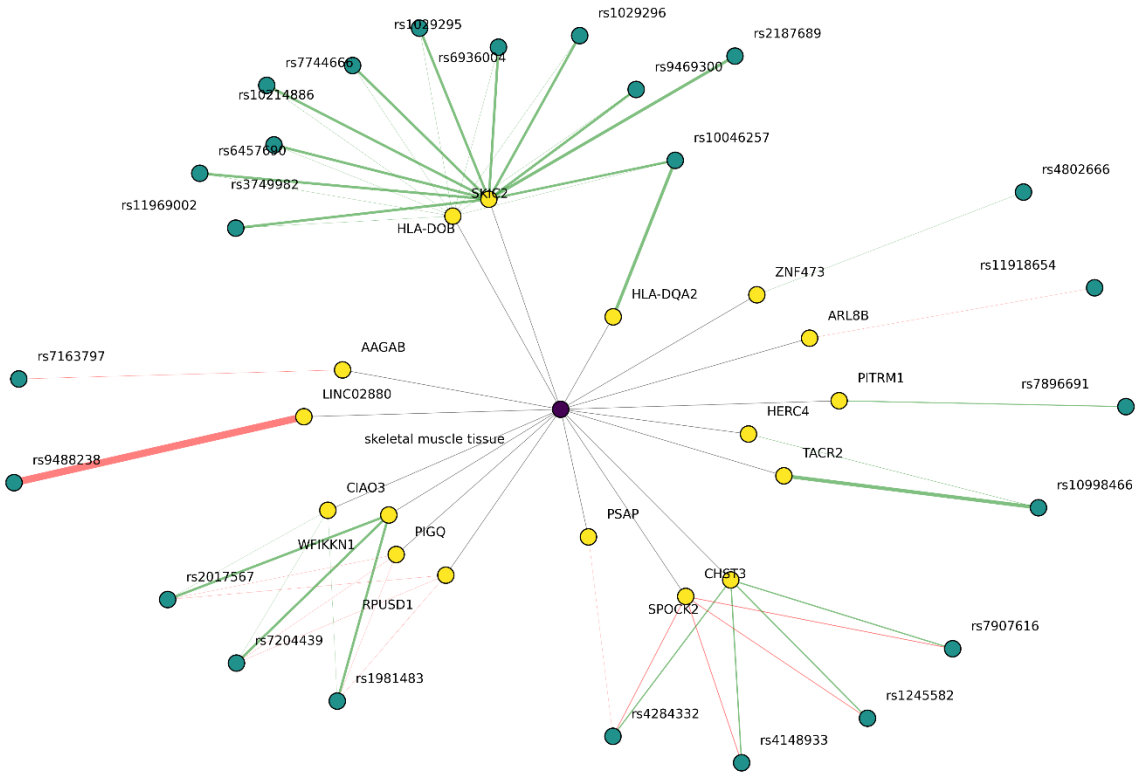


Figure 4 *eQTL network for musculoskeletal tissue. In this plot, genes are represented by yellow nodes, variants by green nodes, and tissues by black nodes. The edges connecting the genes and variants are of two types: red edges indicate that the variant decreases the expression of the gene, while green edges indicate that it increases expression. The thickness of the edges signifies the magnitude of the increase or decrease in expression.*

To precisely investigate the consequences of the set of variants associated with IDD Genopyc gives the possibility of running VEP a tool used to predict the impact of genetic variants on genes, transcripts. The output of Genopyc is a pie chart that show the proportion of different variant effects in the set of variants analyzed (Figure 5).

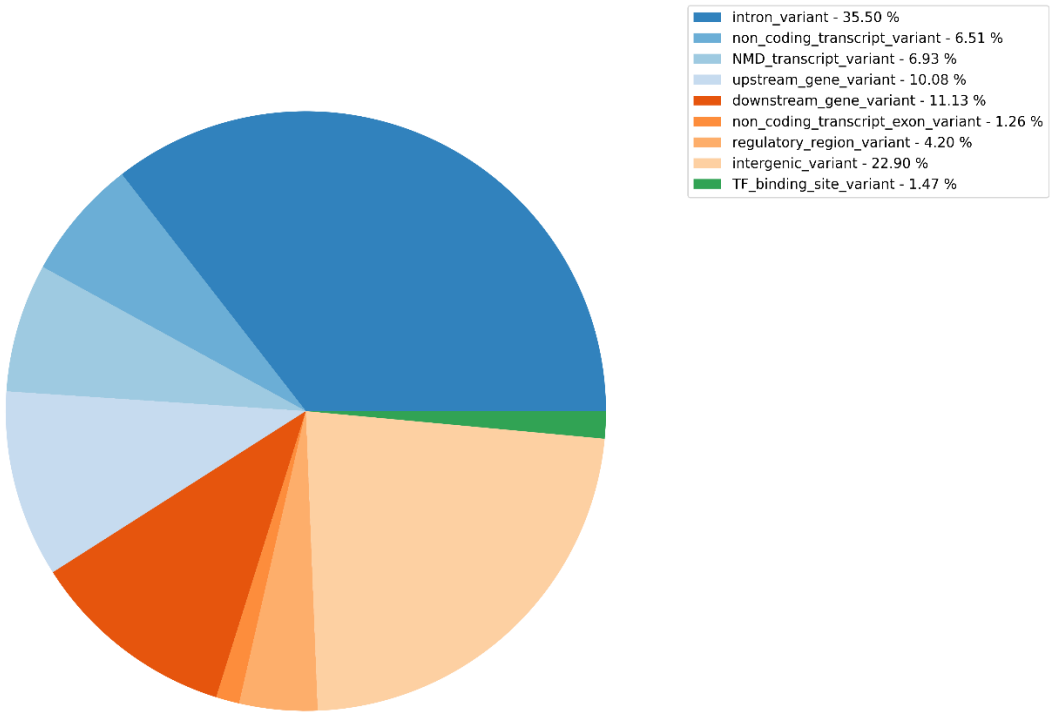


Figure 5 *VEP consequences on the set of variants associated with IDD.*

The majority of variants associated with the pathology (35%) fall in introns and thus is non - coding type, other effects include intergenic variants, TFBS variants and upstream and downstream gene variants. As expected, severe effects such as missense variants, frameshift variants or stop gained are not included in the set. This is a common characteristic of variants linked to complex conditions. SNPs related to common complex disorders usually have a mild impact and are considered susceptibility loci, which only affect the likelihood of developing the condition. [173].

Thus, it is reasonable to think that the variants have mainly an influence on the regulation of the gene transcription. Variants are typically assigned to the nearest gene element as their target functional element. However, recent advancements have emphasized the importance of investigating more significantly the functional characterization of risk loci [174]. Genopyc includes the Locus to Gene (L2G) pipeline of Open Target Genetics to assign target genes to variants associated with complex conditions. This pipeline, through the integration of transcriptomic, proteomic and epigenomic data and statistical analysis, assign genes to GWAS loci. This method outperformed the naive distance – based method that was adopted for assigning genes to variants [175]. We thus applied the pipeline obtaining a total of 30 genes associated to the SNPs of IDD.

To investigate which functions where significantly enriched in the set of gene prioritized from L2G pipeline, Genopyc includes the possibility of performing function enrichment analysis through G:profiler. This is a computational method used to identify and interpret the biological significance of a list of genes or proteins. The goal of this analysis is to determine whether certain biological functions, pathways, or processes are overrepresented (enriched) in the list

compared to what would be expected by chance. Moreover, Genopyc allows the visualization of the results through network – based plot (Figure 6).

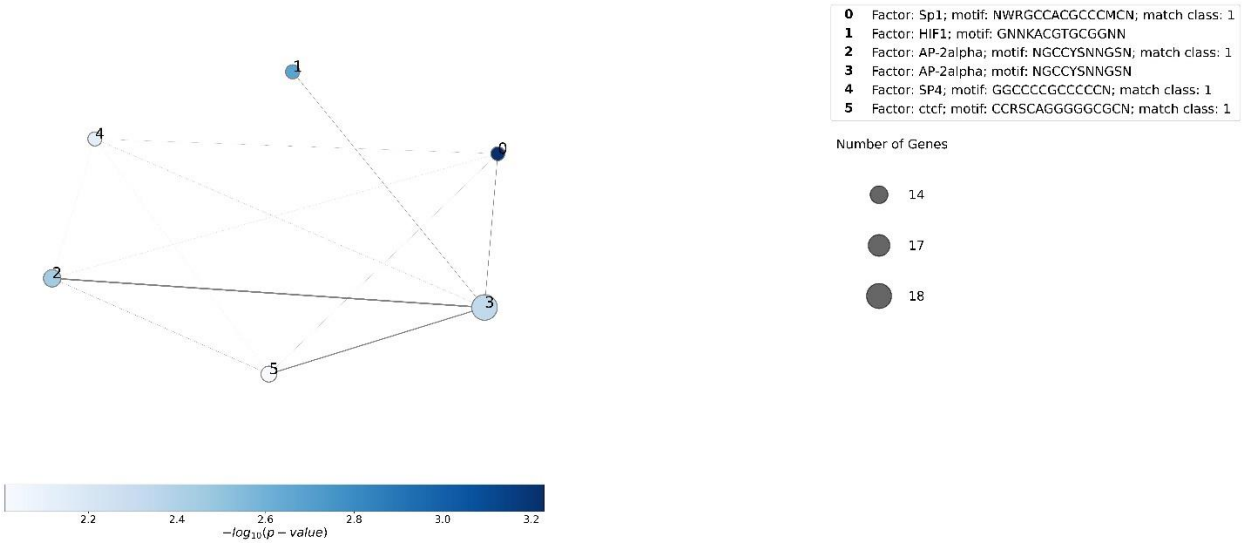


Figure 6 *Function enrichment analysis on the gene set prioritized by the implementation of L2G pipeline. The functions obtained shows the activity of pathways related to studied transcription factors in the context of IDD such as HIF1-α SP1 and AP- 2α and less known factors such as SP4.*

The function enrichment analysis reveals that among the genes prioritized through L2G pipelines transcription factors such as SP1, HIF1-α and AP-2α are significantly affected highlighting their role in the pathogenesis of IDD. Conversely, the functional enrichment on the genes reported from GWAS catalog didn't bring any result or valuable information on the pathways that could be dysregulated in the disease.

From a literature perspective these proteins have already been investigated to be related to the condition in different animal and in vitro model. SP1 was shown to be induced from TNF-α and interleukin-1β, moreover inhibitors of SP1 - DNA binding reduced the expression of pro catabolic enzymes such as for MMP3, ADAMTS4, and ADAMTS5. This suggest that SP1 is an effective target for mitigating extracellular matrix degradation during IDD [176].

Similarly, AP-2α ςασ upregulated along with TGF-β1 in NP tissues of patients and rats with IDD. Moreover, Silencing of AP – 2α was shown to diminish levels of MMP-2, MMP-9 and Smad3 expression (catabolic markers) and increase Aggrecan and Col-2 expression (anabolic marker) in NP tissues of rats with IDD thus improving pathological changes [177]

Finally, HIF – 1α plays a crucial role in maintaining the homeostasis of the IVD by regulating anaerobic glycolysis. This is particularly important given that the IVD is an anatomical structure characterized by low oxygen tension. It was demonstrated that in constitutively active HIF – 1α NP cells there was an upregulation of the expression of Glut-1, Glut-3, aggrecan, type II collagen,

and Sox9. Moreover, cells with constitutively active HIF – 1α showed reduced apoptosis through Fas ligand ligation [178] [179].

We weren't able to find any literature supporting the involvement of SP4 in IDD but further investigation should be carried out in order to clarify its possible role in the molecular biology underlying the pathology.

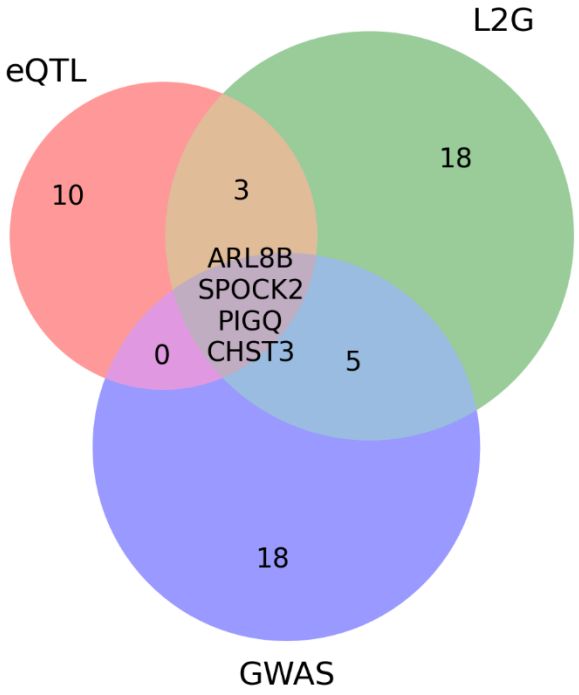


Figure 7 *Overlapping between the genes prioritized using 3 different Genopyc features. GWAS genes, eQTL genes and L2G genes. 4 genes were prioritized by all the methods: ARL8B SPOCK2 PIGQ and CHST3*

Additionally, visualizing the results can provide greater insight into the functional elements potentially involved in the etiology of the disease. To investigate even further the possible molecular underpinnings and the mechanisms of action of the variants related to IDD we overlapped the genes that were prioritized from 3 different methods: i) the genes prioritized from GWAS ii) the genes associated to variants through eQTL and iii) the genes prioritized through the implementation of the L2G pipeline. Through this method we detected 4 genes that were prioritized from all the methods (Figure 7). Finally, to understand how the overlapping genes relate to the transcription factors detected through functional enrichment analysis, we utilized the ENCODE transcription factor targets dataset. This repository collects gene–TF associations by examining the binding of transcription factors near the transcription start site of genes to investigate TF–gene relationships.

We were able to detect a functional cluster comprising SP1 SP4 HIF1-α AP2-α ARL8-β and CHST3 (Figure 8).

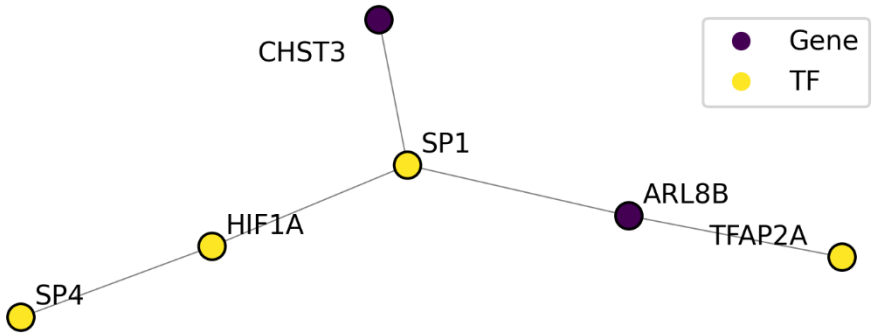


Figure 8 *Functional cluster of genes prioritized from 3 different sources using Genopyc (ARL8-β and CHST3) and factors coming from the functional enrichment analysis performed on the set of genes prioritized implementing L2G pipeline SP1 SP4 HIF1-α and AP2-α. An edge exist between the genes if the transcription factor acts on the regulates the activity of the target gene. Regarding the SP4 SP1 HIF1-α interaction, the activity of hypoxia factor is regulated by SP1 and SP4.*

This section showed the capabilities of Genopyc in investigating different aspects of variants associated with complex diseases. It allows to obtain information on the genomic context of the variants by listing regulatory elements in the neighboring genomic location, retrieve the genes that are eQTLs in a specific tissue, investigate the effect of the set of variants on the transcript and associate the genes that are affected through the implementation of L2G pipeline from open targets.

We showcased how Genopyc can be implemented to investigate how variants may dysregulate molecular pathways and contribute to disease. While our results require further validation using in-vivo and in-vitro models, we have developed an all-in-one tool for in-silico investigations. This data-driven approach allows us to explore potential molecular interactions contributing to disease.

The possible outcomes of using Genopyc in disease research and treatment include the identification of novel drug targets, facilitating the development of personalized treatment plans, and aiding in the discovery of predictive biomarkers.

The visualization capabilities of the library help the user to directly unveil biological associations and can be fully exploited in an interactive computational environment such as jupyter notebook. In summary we provide an all-in-one tool to retrieve and interpret the effect of genomic variants on the development of complex diseases. Genopyc is easily installable via pip and can be integrated into Python environments being built upon main Python libraries.

Availability

The library can be installed via pip: https://pypi.org/project/genopyc/

The source code is available at: https://github.com/freh-g/genopyc

The notebook with the use case is available at: https://github.com/freh-g/genopyc/blob/main/tutorials/Genopyc_tutorial_notebook.ipynb

The documentation of the package is available at: https://genopyc.readthedocs.io/en/latest/

Conflict of interests

JP is an employee of Medbioinformatics Solutions SL. JP is co-founder and holds shares of Medbioinformatics Solutions SL.

Funding

This project was supported by the Marie Sklodowska-Curie International Training Network "disc4all" under grant agreement #955735.

BO acknowledges support from MCIN and the AEI (DOI: 10.13039/501100011033) by grants PID2020-113203RB-I00 and "Unidad de Excelencia María de Maeztu" (ref: CEX2018-000792-M)

Conflict of Interest: None declared.

3.3 Modic change is associated with increased BMI but not autoimmune diseases in TwinsUK

This chapter is based on:

Modic change is associated with increased BMI but not autoimmune diseases in TwinsUK

Francesco Gualdi[1,2,3], Isabelle Granville Smith[3], Roger Compte[3], Frances MK Williams[3]

[1]*Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Pompeu Fabra University (UPF), Barcelona, Spain*

[2]*Structural Bioinformatics Lab (GRIB-IMIM), Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Barcelona 08005, Catalonia, Spain*

[3]*Department Twin Research and Genetic Epidemiology, King's College London, St Thomas Hospital, Westminster Bridge Road, London SE1 7EH, UK*

Published in: European Spine Journal

DOI: 10.1007/s00586-023-07870-7

Summary of the work

This part of the thesis was carried during a stage as part of the ETN in one of my trainings in which I worked in the department of Genetic Epidemiology at King's College in London supervised by Professor Frances Williams. For this study we investigated, through a cross sectional study, the autoimmune basis of Modic change (MC), an inflammation of the vertebral body strictly correlated to intervertebral disc degeneration (IDD) and low back pain (LBP). It is reported in literature that MC could have an autoimmune etiology, this is supported also from the fact that many genetic variants associated with IDD are found on chromosome 6 in the major histocompatibility process. With the implementation of Genopyc in chapter 3.2 we showed that by performing a function enrichment analysis on the genes in the surroundings of the variants associated with IDD many autoimmune pathways were enriched, but, by performing a more careful mapping of the genes, we were able to prioritize different pathways already reported in literature to be associated with the condition. For this study we exploited the large cohort of TwinsUK, we selected participants with autoimmune diseases diagnosis (sample size = 764) and we investigated MC size, prevalence and severity in this group. We weren't able to detect an association between autoimmunity and MC, further supporting the absence of an autoimmune etiology in this condition as we also showed in chapter 3.2 from the genetic perspective. Interestingly we found an association between Body mass index (BMI) and the size of the MC lesion. In conclusion, results seem to suggest that IDD does not have an autoimmune etiology, this is supported from the outcomes of chapter 3.2 from a genetic level and from the cross sectional study performed in this chapter**.**

Abstract

Purpose

Low back pain (LBP) is one of the largest causes of morbidity worldwide. The etiology of LBP is complex and many factors contribute to the onset. Bone marrow lesions within the vertebra adjacent to an intervertebral degenerate disc (IDD) named Modic change (MC) have been suggested as a diagnostic subgroup of LBP. Autoimmune response has been proposed to be one of the causes that promotes the development of MC. The aim of the current investigation is to assess prevalence and severity of MC and LBP in participants with an autoimmune disease diagnosis in a well-documented cohort of adult twin volunteers.

Methods

Multivariate generalized mixed linear models (GLMM) were implemented in order to calculate the association between having an autoimmune disorder and MC prevalence, width and severe and disabling LBP. The model was corrected for family structure as well as for covariates such as age, BMI and smoking.

Results

No association was found between diagnosis of autoimmune disorder and MC. Interestingly, BMI was independently associated with MC width but not to MC prevalence. These results help to shed light on the relationship between MC and autoimmunity as well as the role of BMI in the development of the lesions.

Conclusion

This study is the first to examine autoimmune disorders and MC prevalence in a large, population-based female cohort. The study was well powered to detect a small effect. No association was found between having a diagnosis of one or more autoimmune conditions and MC prevalence, width or LBP.

Key words: intervertebral disc degeneration, low back pain, modic change, autoimmune disease.

Introduction

Low back pain (LBP) is one of the largest causes of disability worldwide. It is a leading cause of work absenteeism and medical consultation [180]. This is reflected by a huge medical and economic social burden. The prevalence of LBP is higher in females and a 2019 estimation of 568.4 million LBP cases overall makes it the leading morbidity worldwide [181]. One of the main causes of LBP is spine degeneration, in particular the age-related changes of the intervertebral discs (IVD) [182]. The IVD is the largest avascular organ in the body and it composed of three anatomical regions: nucleus pulposus (NP) a gelatinous proteoglycan rich central structure, annulus fibrosus (AF): a ring of ligamental fibers surrounding the NP and cartilaginous endplates (CEP) that enclose the disc and separate it from vertebral bodies. Vertebral CEPs play an important

role in the homeostasis of IVDs serving as an interface between discs and bones and providing nutrients to IVD cells. Another important function of CEP is to prevent contact between host immune cells and the IVD acting as a physical barrier [144]. With age IVD faces biochemical and morphological changes and starts to degenerate. Intervertebral disc degeneration (IDD) is a pathological process that leads to the loss of function of this anatomical structure favoring the onset of a plethora of disorders that eventually lead to LBP[183].

One of the IDD is a bone marrow lesion within the vertebra and change to adjacent bony endplates detectable with magnetic resonance imaging (MRI) named Modic change (MC) [184]. Evidence suggests these manifestations are part of the same pathological process influencing the onset of LBP [185]. MC were first reported by Ross in 1987 [186] and subsequently described by Modic *et al*. Risk factors for MC are similar to those for IDD and include increased age, high body mass index (BMI) and smoking, however the condition also has a heritable component estimated at 30% [187].

Despite known risks and heritability, it is not clear how the onset of MC is triggered, and why this only occurs in a subset of people affected by IDD. One possible cause could be the subjective ability of bone marrow to respond to inflammation coupled with the damage of CEP (ED) [188]. When bone marrow and protruded disc tissue get in contact, the immune-privileged disc tissue could trigger an autoimmune response which could enhance the onset of MC [189], [190]. One characteristic of IDD is a compromised endplate, which, when damaged, along with the ingrowth of blood vessels leads to the infiltration of activated immunocytes and inflammatory cytokines in the disc space.

If MC have an autoimmune component, then people with autoimmune conditions would be expected to manifest higher prevalence and severity of MC. An autoimmune diagnosis may be associated with altered immunoreactivity [191], including a more severe response if NP tissue is exposed to immune cells. The aim of the current investigation was to assess prevalence and severity of MC and LBP in participants with an autoimmune disease diagnosis in a well-documented cohort of adult twin volunteers having prospectively gathered data over many years

Methods

This cross-sectional retrospective study determined whether adults diagnosed with an autoimmune condition demonstrate a higher prevalence or severity of MC. The cohort is composed of twins enrolled in the TwinsUK adult twin registry based at King's College London [192]. Currently, the registry comprises over 15,000 twins, mostly female, aged 18 - 88 making it one of the largest twin registries in the world. Participants are sent regular questionnaires that include questions regarding lifetime diagnosis of an autoimmune disease. All twins have signed informed consent forms for research approved by St. Thomas' Hospital Ethics Committee and Liverpool East Research Ethics Committee (REC reference 19/NW/0187), IRAS ID 258513. The TwinsUK spine study was started in 1996.

T2 weighted images (T2WI) sagittal spine MRIs were obtained using a Siemens scanner (Munich, Germany) with 1.0-T superconducting magnet and supporting data were collected from twins who

were included in the study. Variables such as age, sex and BMI were obtained, as well as lifestyle, and clinical information.

Only participants with availability of MRI and autoimmune questionnaires were included in the study. Data were collected from self-reported questionnaires and health assessment visits of twins previously invited to participate in various studies examining a wide range of traits and common medical conditions. Questionnaires included doctor's diagnosis of inflammatory bowel disease, rheumatoid arthritis, coeliac disease, vitiligo, autoimmune thyroid disease (hypothyroidism, hyperthyroidism), lupus, multiple sclerosis and type I diabetes were included as autoimmune conditions of interest (Figure 1). Circulating anti-thyroid peroxidase antibodies (TPOAb) were measured merging the results coming from two different assays: Roche assay (TPOAb titre of >34 kU/l considered positive) and Abbott assay (TPOAb titre of >6 kU/l considered positive). Patients with missing data were assumed to be negatives for that specific condition.

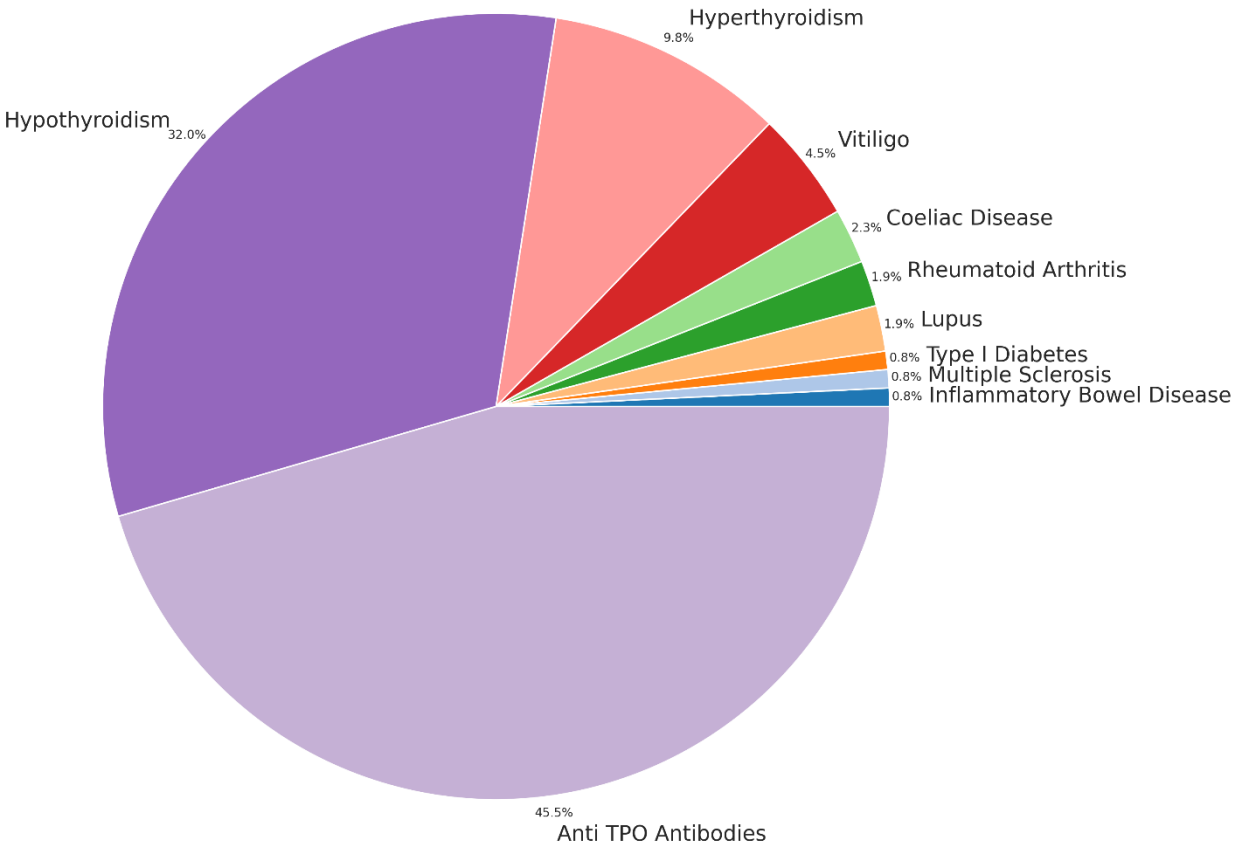


Figure 1 *The pie chart shows the prevalence in the sample (n = 764) of the autoimmune conditions included in the study.*

An individual was considered autoimmune positive if they had ever received a diagnosis of at least one of the conditions above or had circulating antibodies over the positivity threshold. Coding of the largest autoimmune sub-group, thyroid related was cross-referenced and integrated with prescription data that listed medications to treat autoimmune disorders affecting the thyroid.

Coding MRI scans

Lumbar MC was identified on MRI by presence of altered imaging signal at vertebral body levels as previously [187]. MC was coded as a positive binary variable if MC was present at any one of the lumbar segments [193]. A quantitative variable was derived from the size of both superior and inferior bone marrow lesions evaluated on a scale ranging from 0-5 as previously described [194], and values were summed to create a MC lesion size score for each lumbar level [195].

IDD had been coded as the sum of disc bulge, disc imaging signal intensity, disc height and osteophytes formation at each spine level previously [196]. Briefly, each measure was assigned a score (0-3), grading the severity of the phenotype and a total degeneration score, summing scores for each of the 5 lumbar discs was assigned for each participant (total IDD range = 0-60). Back pain was evaluated through several methods. First, the Medical Research Council Back and Neck Pain Questionnaire was administered at the MRI scan visit. In addition, self-reported episodes of disabling LBP lasting more than one month at any period during lifetime were collected on subsequent questionnaires [197].

Statistical analysis

Statistical parametric and non-parametric tests were used for comparing the 2 groups. T-test was used for comparing age and BMI, Chi-squared for MC prevalence and Mann Whitney U for MC width. Multivariable generalized mixed linear models (GLMM) were fitted to the data to calculate odds ratios for risk of developing MC and LBP in participants with an autoimmune disorder.

A general formulation of the model conditionally on random effects b_i:

$$f(y_{ij}|b_i) = \exp\left[\frac{y_{ij}n_{ij} - a(n_{ij})}{\emptyset} + c(y_{ij}, \emptyset)\right]$$

with mean $E(y_{ij}|b_i) = a'(n_{ij}) = \mu_{ij}(b_i)$ and variance $\mathrm{Var}(y_{ij}|b_i) = \emptyset a''(n_{ij})$, GLMM permit the incorporation of hierarchical random effects in multiple levels thus allowing the flexibility of correcting for effect that act among outcomes[198], [199] .

Moreover, it was investigated associations between diagnosis of an autoimmune disorder and the size of MC lesion measured in a normalized scale from 1 to 5. Multivariable analysis was adjusted for covariates included as dummy variables such as sex, smoking and episodes of disabling LBP, continuous as age and BMI and categorical such as family structure i.e., twin relatedness and number of autoimmune disorders. An individual was considered as smoker if reported to smoke more than 10 cigarette packets per year. Data processing and analysis were planned and executed

in Python version 3.10.5 using “statsmodels” and “scipy” packages and R Studio version 2022.07.1 using lme4 library. A post-hoc power analysis was carried out using G*Power software [200].

Results

Data were obtained for 764 twins having MRI coded for MC and completed autoimmune self-report questionnaires (Table 1). Twins were predominantly female (n=737 (96%)) with mean age of 54 years (range 34-73 years) and mean BMI = 25 kg/m^2 (range 16.23-51.40 kg/m^2). Participants reported smoking more than 10 packets a year or more (n = 213 (28%)), and 164 (22%) participants had had an episode of disabling LBP over their lifetime lasting more than one month (Table 1). BMI was significantly different between participants diagnosed with an autoimmune disease and those with no autoimmune diagnosis (t-test p-value = 0.001). This is because included autoimmune disorders were affecting thyroid and is well grounded that these conditions influence BMI [201].

Table 1 *Baseline and outcome characteristics of the participants divided into autoimmune positive and autoimmune negative groups*

	Autoimmune Negatives				**Autoimmune Positives**		
		n	**%**	**Mean(±SD)**	**n**	**%**	**Mean(±SD)**
Age		561		53(8)	203		54(7)
Sex							
	Female	535	95		202	99	
	Male	26	5		1	1	
BMI		561		25(4.3)	258		26(5)
Smoking							
	No	400	71		151	74	
	Yes	161	29		52	26	
Modic Change							
	No	376	67		140	69	
	Yes	185	33		63	31	
Modic Change Size		166		5.9(3.5)	176		7(5)
LBP							
	No	441			159	78	
	Yes	120			44	22	
IDD		508		12(8)		180	11.9(8)
N of Autoimmune conditions							
	1				137	68	
	2				56	27	
	3				9	4.5	
	4				1	0.5	

Modic change

MC was defined as any individual having at least one MC affected endplate. A prevalence of 31% and 33% was detected in the autoimmune positive and negative participants respectively, there was no significant difference between the two proportions (χ^2 p-value = 1). Both incidence and size of MC were higher in L4-5 and L5-S1 than other lumbar levels (Figure 2).

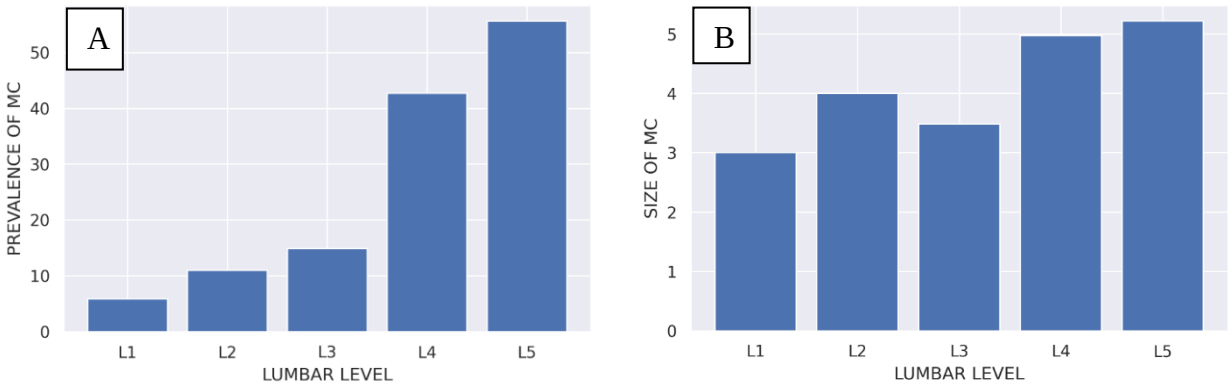


Figure 2 *Prevalence of MC (Panel A) is measured as total prevalence of MC at specific level among MC positive participants. Size of MC (Panel B) is measured as mean MC width at specific level among all MC positive participants.*

Table 2 summarizes the results of generalized mixed logistic regression model corrected for age, BMI, smoking and family structure. Neither having a diagnosis of autoimmune disease nor the number of morbidities were associated with the presence of MC (p-value > 0.05). MC was found to be associated independently to IDD (OR 1.22 CI 1.16 – 1.30, p-value = 6.7×10^{-13}). MC size and autoimmune diagnosis ($r^2 = 0.4$) were not associated, results are summarized in Table 3. No association between autoimmune diagnosis and MC size was detected. Interestingly, BMI was associated with MC size even after correcting for IDD, providing evidence of an independent association between BMI and MC size.

Table 2 *Risk factors for MC. Odd ratios were calculated through multivariate regression analysis. Significant results are highlighted in grey.*

	$R^2 = 0.48$			
Coefficient	OR	CI	P-VALUE	Z-VALUE
AGE	1.00	0.96 – 1.03	0.72	-0.35
BMI	1.04	0.98 – 1.09	0.15	1.44
SMOKING	1.13	0.70 – 1.84	0.61	1.51
AUTOIMMUNE	0.76	0.46 – 1.27	0.30	-1.03
IDD	1.22	1.16 – 1.30	1.46e-12 ***	7.08

Table 3 *Risk factors for MC width. Betas were calculated through multivariate regression analysis. Significant results are highlighted in grey.*

Coefficient	$R^2 = 0.40$			
	β	CI	P-VALUE	T-VALUE
AGE	-0.01	-0.06 – 0.03	0.48	-0.70
BMI	0.07	0.005 – 0.135	0.03 *	2.13
SMOKING	0.18	-0.40 – 0.77	0.54	0.61
AUTOIMMUNE	-0.23	-1.20 – 0.61	0.52	-0.64
IDD	0.24	0.20 – 0.29	2e-16 ***	10.41

Table 4 *Risk factors for LBP. Odd ratios were calculated through multivariate regression analysis. Significant results are highlighted in grey.*

Coefficient	$R^2 = 0.40$			
	OR	CI	P-VALUE	Z-VALUE
AGE	0.91	0.92 – 1.02	0.24	-1.18
BMI	1.06	0.99 – 1.13	0.10	1.65
SMOKING	1.83	0.97 – 3.44	0.06	1.88
AUTOIMMUNE	1.03	0.40 – 2.70	0.95	0.06
MC WIDTH	1.13	1.02 – 1.25	0.01 *	2.48
IDD	1.08	1.02 – 1.14	0.007 **	2.70

Finally, we examined the association between autoimmune disorders and episodes of LBP in the twin participants while correcting for BMI, age and smoking (r^2=0.4, Table 4). No association was found between autoimmune phenotype and LBP (p-value > 0.05).

Discussion

This is the first study using a large population sample having MRI scans coded for MC and recorded autoimmune diagnosis. Autoimmune diagnosis does not appear to influence the development of MC suggesting that the development of MC is not autoimmune in etiology. We explored MC both as a binary variable and as continuous measure of size. None of the models showed autoimmune diagnosis was correlated to MC size, width, or measures of back pain but our study does reveal that MC is related to raised BMI.

Related to our hypothesis that more severe autoimmune diathesis following an autoimmune diagnosis could result in larger MC lesions, we investigated the correlation between autoimmune diagnosis and MC width but found no association. BMI and MC were associated when correcting for disc degeneration which suggests an independent association, as we [202] and others have reported [161]. Clarity around how BMI influences MC prevalence or size is needed [203]. Whether high BMI places a physiological burden upon endplates, subjecting them to increased microtrauma or whether systemic inflammation by lipid-induced endo- or paracrine responses promotes MC is not clear. Our findings that BMI is independently associated with size but not prevalence of MC could suggest the association is driven by adipose-derived inflammation. In support to this, Teichtahl *et al*. showed that increased spinal adiposity is correlated with MC [204]. Conversely, high BMI increasing spine workload and influencing MC prevalence was not supported by our findings.

Autoimmunity has been suggested to play a role in pain perception, a dysregulated interplay between nervous and immune systems, especially the interaction of nociceptor and immune cells [205]. Pathological changes promoted by altered metabolite transport due to ED and the subsequent recruitment of inflammatory cytokines such as interleukin-1 (IL-1), tumor necrosis factor alpha (TNF-α) and IL-6 have been demonstrated to stimulate pain receptors in tissues and play a causal role in LBP [206]. For this reason, we examined whether diagnosis of an autoimmune disorder increased the reporting of back pain. This question fell within our general investigation that autoimmune diagnosis and MC may correlate, thus MC symptomology, for example LBP may also correlate with autoimmune diagnosis. We found no evidence to support either of these relationships. We found previously in TwinsUK that smoking and high BMI was associated to LBP [207]. High BMI and smoking likely have several influences on back pain from reducing blood flow and disc nutrition, to socio-economic or physical workload factors [168][208]. The endocrine responses of adipose tissue chronically elevate inflammatory markers [209] with may increase pain susceptibility. Even if autoimmunity has a preponderant component in chronic pain perception, LBP is directly influenced by IDD, smoking, and BMI.

The etiology of autoimmune disease depends, in part, from genetic susceptibilities, along with several other risk factors; lifestyle, environment and epigenetics have been proposed and demonstrated to contribute [210]. Autoimmunity is a diverse phenomenon classified into “systemic” and “organ specific” which share risk alleles at HLA locus on chromosome 6 [211]. In

contrast, alleles associated with cartilage production and vertebral health have been associated with MC. Both candidate and genome wide gene association studies fail to provide evidence of a shared genetic risk for autoimmune diagnosis and MC [212], [213], [214]. Autoimmune disorders are highly heterogeneous and future investigations would likely benefit from disease specific foci when examining links with other conditions [215]. MC does not appear to be an autoimmune disorder, the autoimmune-like response of cells may not relate to autoimmune diagnosis.

Finally, alternative hypotheses cannot be discounted; mounting evidence suggest a role for occult infection in the IVD [216], [217], [218], [219]. The resulting low-level or occult infection could promote inflammatory processes, leading to MC [220]. This theory is supported by studies investigating the presence of bacteria in degenerate and MC adjacent discs[221], Propionibacterium acnes (among others) has been cultured, and whole bacterial genome sequencing studies have reported an array of findings in disc tissue [222], [223] as well its injection in animal models disc caused the development of MC [224], [225]. Moreover, it was shown that patients with LBP and type I MC treated with antibiotics improved pain symptoms and MRI Modic grading [226]. There has however been relatively little scale, scientifically robust research investigating the presence of bacterial infection in the disc especially in relation to chronic LBP and MC [227] and bacterial contaminants are often posited to explain findings of disc microbes [228]. Whether dysregulated bacteria and occult infection or pain-generating cytokines are the impetus for an inflammatory stimulus, leading to MC is not clear.

A study may represent a false negative or lack of detectable association may arise where studies are of insufficient size. According to the power estimation of our study, we had 80% power to detect an association between the occurrence of MC and autoimmune disease of effect size of 0.1. Despite the high power we did not detect any increased MC prevalence. This suggests if MC does occur more frequently in participants with autoimmune diagnoses, it must be at a rate less than 10%. Our best conclusion is there is no relationship between autoimmune diagnosis and MC development. Exposure of normally immune privileged NP cells to the external circulus may be necessary for this response, however we did not find evidence this reaction is promoted by carrying an autoimmune disorder but instead could depend on a specific adaptive immunity alteration. We can't therefore exclude that an autoimmune response is ongoing during MC onset, but this is not related to a susceptibility to autoimmunity proper of autoimmune disorders.

We acknowledge several limitations to this study. We considered autoimmunity as a single phenotype obtained by merging very different diseases. For this reason, further investigations should focus on studying relationships between autoimmunity and MC in specific autoimmune disorders by including biological samples and genetic data in order obtain a deeper understanding of the complex mechanisms underlying. Self-report was used to classify diagnosis of autoimmune conditions and reliance on participant self-report may bias results, although the hypothesis being tested was obscure to participants. The cohort is predominantly female, which may have favored our study as autoimmune diseases are twice as prevalent in females [229], allowing us to include conditions rare in the general population. It does however prevent us drawing conclusions about autoimmune diagnosis and MC risk in males. Lastly, our MC categorization was based on T2WI images only. To distinguish MC types (I, II or III) both T1WI and T2WI are required – but not available in this cohort for funding reasons.

Conclusions

This study is the first to examine autoimmune disorders and MC prevalence in a large, population-based cohort. The study was well powered to detect a small effect. No association was found between having a diagnosis of one or more autoimmune conditions and the prevalence, width or severity of MC. Interestingly, MC extension was associated to increased BMI after correcting for IDD, sign of an independent association. These results are applicable to females since the results presented were replicated excluding male participants

4. Discussion

Recent advancements in omics technologies have boosted biomedical research into a new era of discovery and understanding. With breakthroughs in genomics, transcriptomics, proteomics, metabolomics, scientists can now explore biological systems at an unprecedented level of detail. The exponential increase in biological data volume alongside the growth of computational resources has prompted the beginning of the application of computational models to interpret complex biological systems. These innovative techniques allow researchers to analyze vast amounts of data, uncovering intricate molecular pathways, identifying biomarkers for diseases, and revealing the underlying mechanisms of various biological processes. Moreover, the integration of multiple omics datasets has enabled a holistic approach to studying complex biological phenomena, providing insights into the interplay between genes, proteins, metabolites, and their regulatory networks. These recent developments in omics offer great potential for enhancing our comprehension of the origins of complex disorders, with a focus on personalized medicine, drug discovery, and precision healthcare.

In this thesis, we applied innovative methods to heterogeneous biological data in order to shed light on the mechanisms leading to the development of IDD and its comorbidities. The developed method can be applied to other pathologies and are publicly available to support researchers in the investigation of the biology of complex disorders. We approached the problem of understanding how the genotype is influencing the phenotype of IDD from multiple directions. In Chapter 1, we developed a new method based on KGE that allows the integration of heterogeneous biological data in a graph-based structure. We then compared the capabilities of different embedding generation algorithms for translating the knowledge encoded in the graph into low-dimensional vectorial space, using supervised and unsupervised methods. The best performing algorithm was then selected to build an AI-based GDA prediction tool that is freely available for the research community.

Chapter 2 introduces a tool that integrates multiple datasets to explore the genetics of complex diseases. In this section, we describe Genopyc, a Python library to explore how genetic variations linked with complex traits impact gene transcription and regulation. The lack of a consensus approach to reach the understanding of the functional effects of the SNPs on the downstream biological pathways hinders the discovery of effective therapeutical strategies. Moreover, a Python library for the investigation of variants detected from GWAS is still missing and to carry out the necessary analysis the user must rely on different sources resulting in a time-consuming approach. We expect this tool to be useful to the scientific community.

Finally, in Chapter 3 we investigated IDD and its comorbidities from an epidemiological perspective. This study was conducted in collaboration with the department of epidemiology at King's College in London. The cross-sectional study took advantage of a large cohort of twins (TwinsUK) to investigate the connection between autoimmune positive participants and Modic change, an inflammation of the vertebral body strictly related to IDD.

4.1 Innovative KGE GDA framework

In this thesis we have developed a new framework to prioritize genes that could be involved in complex diseases. The technology utilized allows the integration of heterogeneous data sources

into a graph that reflects the state-of-the-art biological knowledge, that can be easily updated and that can accommodate a variety of biomedical data types. Then, this knowledge is transformed into a vectorial space and was utilized to train an AI model to prioritize genes associated with

complex conditions. We compared different algorithms to create embeddings relying on different methodologies and we created 2 novel algorithms, namely DLemb and BioKG2vec, to produce KGE and that outperformed most existing methods. The results of this part of the thesis showed that Metapath2Vec surpasses other algorithms both in supervised and unsupervised tasks and was the selected methodology to build the GDA framework.

Our results showed that heterogeneous data integration in the KG leads to better GDA predictions. In fact, the complex interplay of genes, epigenetic, regulatory and epistatic interactions captured from multiple data sources can depict a more meaningful biological scenario compared to single sources taken singularly.

In contrast, we showed the consequences of data preprocessing: random – walk based approaches such as Metapath2Vec seem to benefit from integrating the whole ontological data, comprising classes, properties, restrictions and metadata. In comparison, Dlemb prediction capabilities increase after ontology preprocessing in which only the graph hierarchical structure of the ontology was kept, leaving out the restrictions and metadata. This difference could be driven by the natural language processing capabilities of models such as Metapath2Vec or Node2vec that are based on language models such as Word2Vec.

In summary, we have developed a framework for GDA predictions, which is accessible for free at https://github.com/freh-g/EmBioMark. This tool serves to facilitate the exploration of genes associated with complex diseases, offering an in-silico solution for better interpretation of GWAS data, that require access to restricted genetic data, which can be intricate to handle.

4.2 Elucidating the consequences of genetic variation

Despite GWAS prioritizing many different genomic loci associated to complex conditions, understanding how the genotype influences the development of complex conditions is still not fully understood. Multiple genes contribute to disease susceptibility in conjunction with environmental influences. Moreover, many disease-associated genetic variants prioritized from GWAS have modest effects, making it challenging to pinpoint their specific roles in disease pathogenesis. Integrative approaches, combining genetic data with functional genomics, epigenetics, and other omics technologies, are increasingly being used to untangle the complexities of disease genetics.

However, despite these approaches holding promise for uncovering novel disease mechanisms and identifying potential therapeutic targets, a benchmark pipeline to understand the involvement of the variants in the etiology of complex condition is still to be achieved. We are entering the era in which we can mine multi layered data repositories to reach a greater understanding of the biological underpinnings of complex disorders. A service that unifies these heterogeneous repositories and carries out multiple analyses is still to be determined also because of lack of a gold-standard pipeline to understand the mechanistical effect of genome variants on the development of the diseases.

Several post-GWAS analysis tools are accessible for scrutinizing GWAS discoveries across various objectives. Despite the existence of different tools aimed at similar outcomes (e.g., fine-mapping or colocalization), a big variation in their data prerequisites, assumptions, and outcomes is present. Currently, there is no established protocol or standard procedure for determining the optimal tool for each scenario, and perhaps a combination of tools would be the most effective approach [161] although very time consuming. While some methodological comparisons have been conducted, they have predominantly remained at the mathematical level without digging into the biological basis of their findings. Moreover, there is a lack of a benchmark dataset to aid in evaluating the performance of post-GWAS tools. Although attempts to create a collection of manually curated genes with moderate to high confidence in their functional significance has been proposed to assist in prioritizing causal genes at GWAS loci [230], [231], the use of high-quality, gold standard GWAS datasets that encompass a broad spectrum of molecular mechanisms and genetic architectures is essential to mitigate potential bias in the interpretation of the effects of the genetic variants.

For these reasons, we developed Genopyc, a Python library enabling users to integrate multiple repositories and analyses. Genopyc serves as a versatile toolkit for navigating the genomic landscape, facilitating a deeper comprehension of variant effects on nearby functional genomic elements through the integration of diverse data and analyses. The package's code is freely accessible at https://github.com/freh-g/genopyc and can be easily installed via the Python Package Index (PyPI). While the package currently offers features like data integration, variant-gene binding, genetic identifier mapping, and result visualization, its open-source nature allows for seamless integration of additional functionalities. Designed as a comprehensive Python library for genetic data management, Genopyc holds potential for further enhancements and extensions.

4.3 Biological implications of the results

The aim of this thesis extended beyond the creation of a methodological approach to prioritize genes linked to the etiology of IDD. It sought to explore and comprehend the influence of genetics on disease development and progression, as well as its comorbidities. Consequently, we implemented tools into the study of IDD's genetic foundations. Our findings revealed that the prioritized genes had previously been identified in the literature as playing a significant role in the disease's etiology. Additionally, through functional enrichment analysis, we observed that these genes are involved in pathways crucial for the homeostasis and turnover of IVD tissue.

Analyzing biological systems is a challenging task due to their complexity and heterogeneity. Biological data exhibit intricate hierarchical organization and non-linear interactions, making it difficult to understand the underlying regulation mechanisms especially due to the absence of a robust statistical technique for accurate analysis. We modelled this complexity by the implementation of network approaches and applied learning techniques capable of extrapolating it to build predictive models. When applying our KGE-based tool, we obtained TGF-β as the most likely associated gene to the pathology confirming what was already reported in literature and in expert curated databases. In fact, TGF-β is connected with the homeostasis of the disc and plays a key double-edged role in the IVD turnover stimulating the proliferation of the cells even if was showed to be deleterious in high concentrations [232]. Interestingly, other top scoring genes prioritized from the model are part of TGF-β pathway, meaning that the detection of the disease

module, often composed that genes that have a role in the same functional niche is synthesized into the embeddings. Among the top 10 genes prioritized from the algorithm there are typical components of the disc such as COL1A1 [233] and MMP-2 a key enzyme that degrades the extracellular matrix and whose activity is positively correlated with the degeneration of the disc [234]. These results suggest that the genes prioritized from the model are involved in pathways relevant to the disc health and that are expressed in the tissue.

Our tool could be used to shortlist candidate genes that should be validated in vitro and in vivo with knock-down models or in a data – driven method through gene candidate approaches. However, to reach a greater understanding of complex conditions and how specifically the genomic sequence affects the expression and the function of certain proteins, genomic variants should be included in computational models to understand the effect of the haplotypes on disease susceptibility. For this reason, we developed Genopyc a Python library to help the investigation of genomic data and interpretation of functional effects of the variants. Interpreting the variants associated with complex diseases is a difficult task that involves the mining of multiple repositories and the implementation of different statistical analysis. To support this, we included in Genopyc the possibility of querying genetic and expression data such as eQTL to understand if a variant is associated with the differential expression of a gene, LD to discover the correlation among variants.

To show the potential of the library in interpreting the functional targets of the SNPs associated with a condition we applied it to the results from a GWAS on IDD. By applying the locus to gene pipeline from Open Target genetics included in Genopyc we prioritized a different gene set compared to the variant - gene distance - based methods of GWAS. The original study highlighted PARK2 as the most significantly associated gene to the condition. It is encoded in chromosome six and contributes to the development of Parkinson disease, being involved in the targeting of unwanted proteins and general functions such as cellular activity and growth. The function of this gene on the etiology and progression of IDD was not clear and the authors stated that further studies should be carried out to unveil its role [235].

Through a function enrichment analysis of the genes performed by the authors, no significant pathways were enriched in the gene set, thus not giving hint on the possible functional implications of the variants. On the other hand, after performing the same analysis on the gene set prioritized from Genopyc, we found out that the target genes of the SNPs play a role in regulating the activity of transcription factors involved in pathways related to IDD and already reported in literature as important component of the disease progression such as HIF1-α, SP1 and AP-2α.

Furthermore, the results obtained from the application of the 2 tools also show concordance, Genopyc highlighted the importance of AP-2α, HIF-1α and SP-1. These transcription factors are important regulators of proteins that were prioritized with the KGE framework. In fact, AP-2α was shown to modulate TGFβ-1 and SMAD-3 in rat IDD models. TGFβ-1 acts as a double-edged sword in IDD having normally a protective effect but, if excessively activated, contributes to the degeneration of the disc [236]. Another concordance is shown in the HIF – 1α pathway prioritized from Genopyc. This transcription factor is very important in the homeostasis of the IVD [237] . It was shown that constitutively activated HIF – 1α leads to upregulation of proteins such as SOX9, ACAN and COL2A1 [237], chondrogenic proteins prioritized with the KGE framework. Having

prioritized the genes under the control of these transcription factors laid the basis of further studies. Finally, SP – 1 was shown to act in synergy with_TNF-α and interleukin-1β, proteins that are largely reported in literature associated with IDD [238] and that were predicted to be positively associated with the condition.

These findings shifted the attention from autoimmune – related genes to genes that could be involved in the disease; the majority of SNPs prioritized from the GWAS were located on chromosome six close the human leukocyte antigen (HLA) locus aligning with the theory that autoimmunity plays a role in the development of the degenerative disk disease [239].

Motivated by this, we investigated the autoimmune bases of Modic change, the inflammation of the vertebral body closely correlated with IDD and LBP. The rationale behind the study was to understand if autoimmune positive participants showed significantly increased prevalence of MC compared to non-autoimmune participants. Thus, we exploited the large cohort of TwinsUK to perform a cross-sectional study to associate MC prevalence, size and LBP to the diagnosis of an autoimmune disorder. The results showed the absence of association between being affected by an autoimmune condition and increased prevalence, size and pain of MC. Interestingly, the increased BMI was significantly associated with the size of the lesion but not the prevalence. This could mean that increased BMI leads to more severe lesions in the vertebral body probably due to the increased widespread inflammation typically associated with higher BMI [240]. This study showed that most likely MC don't have an autoimmune nature and that the autoimmune processes going on during IDD are a restricted phenomenon following the exposure to NP tissue to the blood circulation and that isn't more severe in persons that have an autoimmune phenotype.

4.4 Data curation and interpretability

The KG GDA tool created in this thesis allows us to capture multiple levels of biological information, the integration of the data into a network depicts the multi-layered nature of biological systems in which their proteins are expressed in a given tissue and are related to certain pathways interacting with each other and being associated with particular phenotypes.

The KGE produced in this thesis were implemented in a GDA prediction task. However, the breadth of biological information included in the KG holds potential to explore different biological associations. By capturing intricate patterns and associations within biological, ontological and biomedical data, these embeddings can be effectively implemented to predict similar diseases or therapeutic indications. For instance, embeddings trained on disease – phenotype relationships can identify diseases with shared pathways and symptoms, thus clarifying how diseases are related. In a work from Biswas *et al*. [241] a biological KG composed of ontological and PPI data was created and by applying a R-GCN they were able to predict new links between diseases - phenotypes and drug – phenotypes. This study highlighted the capabilities of KGE to be implemented in diverse biological tasks.

However, despite the usefulness of this method, we showed that the quality of the data is very relevant to obtain quality predictions and not bias the results. During the validation process we discovered latent features that could lead to biased prediction capabilities of the model. In fact, due to both the nature of polygenic diseases that are associated with multiple genes and the intrinsic

bias in literature for which there are some diseases and genes that are significantly more studied compared to others. This is an outcome of the biological characteristics of key genes involved in many pathways and polygenic diseases, that must be acknowledged when a predictive model is framed to be aware of this potential bias.

Graph learning models applied to heterogeneous data face potential challenges due to the integration of diverse biological knowledge bases, including the risk of data leakage. This risk arises when incorporating multiple vocabularies and data sources, as discrepancies in disease definitions across ontological frameworks may occur. These discrepancies can lead to the mapping of identical entities to the same identifiers, resulting in redundant entries and inflated predictions. Additionally, biases may emerge from recurrent associations among distinct knowledge bases. For example, when training a predictive model to infer new links using a specific knowledge base with a conventional training-test split, embeddings created during the process may inadvertently encode some links from the test set. Consequently, this may lead to inflated predictions during model validation.

A critical challenge in biology is data privacy due to the sensitive nature of genomic, medical, and other biological data. Personal information contained within biological datasets, including genetic predispositions, health records, and potentially identifiable traits, necessitate stringent privacy measures to safeguard individuals' confidentiality and autonomy. This led to difficult and time-consuming processes to access the sources of information slowing the process of discovering new biological insight that could improve healthcare leading to the discovery of new therapeutic strategies. As biology increasingly intersects with fields like machine learning and genomics, balancing the need for keeping data private with moving science forward is of primary importance. Reaching a balance between facilitating research and protecting individuals' privacy rights requires robust regulatory frameworks, ethical guidelines, and technological safeguards to handle data responsibly while exploring new scientific discoveries.

In this thesis we explored different aspects of biological studies using ontological data, publicly available biomedical repositories and large-scale patient specific databases. We implemented different techniques, from non-interpretable deep learning frameworks to more understandable regressions. The complexity of biological landscape causes a disconnection between structured data such as ontologies and patient specific cohort data that often are populated with missing values, unharmonized entries and unstructured data. Consortia that aim to organize discoveries coming from cohorts of patients into publicly available datasets faces an enormous amount of challenge coming from data protections, poor description of the conditions under which the study is carried out and lacking of a unified vocabulary for biological entities. For this reason, future efforts in data collection and distribution should be aimed at improving data quality and interpretability to reach more precise outcomes of *in-silico* models.

Another layer of difficulty comes from the interpretability of AI models applied to networks. While these models often demonstrate impressive performance in tasks like node classification, link prediction, and graph embedding, understanding how they arrive at their decisions is equally important as data quality for real-world applications. The inherent complexity of network data poses unique challenges for interpretability, as relationships and interactions between nodes can be intricate and heterogeneous. Reaching a greater interpretability in AI models would be

accompanied by an easier understanding of the results and consequently the inference of causality. However, discerning causality from correlation remains a challenge itself. Causality comes into play when we aim to understand not just what happens, but why it happens. This involves identifying causal relationships between variables and determining how changes in one variable influence another. While deep learning models can implicitly capture some causal relationships through their learned representations, explicitly modeling causality often requires additional techniques [242]. Efforts to enhance interpretability include developing explainable AI techniques tailored to network data, such as attention mechanisms, node importance measures, and visualization [243].

5. Conclusions

1. We developed a knowledge graph embedding based tool able to prioritize genes associated to human diseases.
2. The results of applying these tools to IVD highlighted genes previously associated to this condition, and functional enrichment of these genes yielded processes related to extracellular matrix turnover and structure.
3. We showed that latent features such as number of associations, literature bias, data preprocessing are latent features that needs to be considered when applying AI in the biomedical field.
4. From the integration of heterogeneous data, the functions enriched in the set of genes associated to intervertebral disc degeneration variants are shifted from autoimmunity to transcription factors such as SP1 AP2α and HIF1-α.
5. SP1 and AP2-α are transcription-factors prioritized by the application of Genopyc. They contribute to the homeostasis of the catabolic/anabolic environment of the disc through downregulation of MMPs, ADAMTS4, Cox2 and through the TGF-β/SMAD3 signaling pathway respectively. MMPs, ADAMTs, TGF-β and SMAD3 were also prioritized by the implementation of the knowledge graph tool, yielding concordant outcomes from the different tools.
6. From a cross-sectional study performed on TwinsUK, we found that Body mass index is correlated with Modic change size, but not with its prevalence.
7. The study also shows that there is no correlation between an autoimmune diagnosis and severity, size and prevalence of MC.

6. Appendix

A. Immuno modulatory effects of intervertebral disc cells

This chapter is based on:

Immuno modulatory effects of intervertebral disc cells

Paola Bermudez-Lekerika [1,2†‡], Katherine B. Crump [1,2†‡], Sofia Tseranidou [3†], Andrea Nüesch [4†], Exarchos Kanelis [5,6†], Ahmad Alminnawi [7,8†], Laura Baumgartner [3†], Estefano Muñoz-Moya [3†], Roger Compte [9†], Francesco Gualdi [10†], Leonidas G. Alexopoulos [5,6†], Liesbet Geris [7,8,11†], Karin Wuertz-Kozak [12,13†], Christine L. Le Maitre [4†], Jérôme Noailly [3†] and Benjamin Gantenbein [1,2†]

[1]Tissue Engineering for Orthopaedics and Mechanobiology, Bone and Joint Program, Department for BioMedical Research (DBMR), Faculty of Medicine, University of Bern, Bern, Switzerland, [2]Department of Orthopaedic Surgery and Traumatology, Inselspital, Bern University Hospital, Medical Faculty, University of Bern, Bern, Switzerland, [3]BCN MedTech, Universitat Pompeu Fabra, Barcelona, Spain, [4]Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield, United Kingdom, [5]ProtATonce Ltd., Athens, Greece, [6]School of Mechanical Engineering, National Technical University of Athens, Zografou, Greece, [7]GIGA In Silico Medicine, University of Liège, Liège, Belgium, [8]Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium, [9]Twin Research and Genetic Epidemiology, St Thomas' Hospital, King's College London, London, United Kingdom, [10]Institut Hospital Del Mar D'Investigacions Mèdiques (IMIM), Barcelona, Spain, [11]Biomechanics Research Unit, KU Leuven, Leuven, Belgium, [12]Department of Biomedical Engineering, Rochester Institute of Technology, Rochester, NY, United States, [13]Spine Center, Schön Klinik München Harlaching Academic Teaching Hospital and Spine Research Institute of the Paracelsus Private Medical University Salzburg (Austria), Munich, Germany

Published in: Frontiers in cell and developmental biology

DOI: *10.3389/fcell.2022.924692*

frontiers | Frontiers in Cell and Developmental Biology

REVIEW
published: 29 June 2022 doi: 10.3389/fcell.2022.924692

OPEN ACCESS

Edited by:
Arjun Singh, Memorial Sloan Kettering Cancer Center, United States

Reviewed by:
Sumedha Bagga, Boston University, United States
Ipsita Mohanty, Children's Hospital of Philadelphia, United States

**Correspondence: Benjamin Gantenbein, Benjamin.Gantenbein@dbmr.unibe.ch*

†ORCID:
Paola Bermudez-Lekerika orcid.org/0000-0002-6858-8213
Katherine Crump orcid.org/0000-0001-5328-667X
Sofia Tseranidou orcid.org/0000-0003-1459-5650
Andrea Nüesch orcid.org/0000-0002-0962-8705
Exarchos Kanelis orcid.org/0000-0002-2059-1480
Ahmad Alminnawi orcid.org/0000-0002-9998-3275
Laura Baumgartner orcid.org/0000-0002-9801-8318 Estefano Muñoz-Moya orcid.org/0000-0001-5222-4071
Roger Compte orcid.org/0000-0002-7034-1092
Francesco Gualdi orcid.org/0000-0003-0449-9884
Leonidas G. Alexopoulos orcid.org/0000-0003-0425-166X
Liesbet Geris orcid.org/0000-0002-8180-1445
Karin Wuertz-Kozak orcid.org/0000-0002-3281-4629
Christine Le Maitre orcid.org/0000-0003-4489-7107
Jérôme Noailly orcid.org/0000-0003-3446-7621 Benjamin Gantenbein orcid.org/0000-0002-9005-0655

‡These authors have contributed equally to this work and share first authorship

Specialty section: This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

Received: 20 April 2022
Accepted: 20 May 2022
Published: 29 June 2022

Immuno-Modulatory Effects of Intervertebral Disc Cells

*Paola Bermudez-Lekerika[1,2†‡], Katherine B. Crump[1,2†‡], Sofia Tseranidou[3†], Andrea Nüesch[4†], Exarchos Kanelis[5,6†], Ahmad Alminnawi[7,8†], Laura Baumgartner[3†], Estefano Muñoz-Moya[3†], Roger Compte[9†], Francesco Gualdi[10†], Leonidas G. Alexopoulos[5,6†], Liesbet Geris[7,8,11†], Karin Wuertz-Kozak[12,13†], Christine L. Le Maitre[4†], Jérôme Noailly[3†] and Benjamin Gantenbein[1,2†]**

[1]*Tissue Engineering for Orthopaedics and Mechanobiology, Bone and Joint Program, Department for BioMedical Research (DBMR), Faculty of Medicine, University of Bern, Bern, Switzerland,* [2]*Department of Orthopaedic Surgery and Traumatology, Inselspital, Bern University Hospital, Medical Faculty, University of Bern, Bern, Switzerland,* [3]*BCN MedTech, Universitat Pompeu Fabra, Barcelona, Spain,* [4]*Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield, United Kingdom,* [5]*ProtATonce Ltd., Athens, Greece,* [6]*School of Mechanical Engineering, National Technical University of Athens, Zografou, Greece,* [7]*GIGA In Silico Medicine, University of Liège, Liège, Belgium,* [8]*Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium,* [9]*Twin Research and Genetic Epidemiology, St Thomas' Hospital, King's College London, London, United Kingdom,* [10]*Institut Hospital Del Mar D'Investigacions Mèdiques (IMIM), Barcelona, Spain,* [11]*Biomechanics Research Unit, KU Leuven, Leuven, Belgium,* [12]*Department of Biomedical Engineering, Rochester Institute of Technology, Rochester, NY, United States,* [13]*Spine Center, Schön Klinik München Harlaching Academic Teaching Hospital and Spine Research Institute of the Paracelsus Private Medical University Salzburg (Austria), Munich, Germany*

Low back pain is a highly prevalent, chronic, and costly medical condition predominantly triggered by intervertebral disc degeneration (IDD). IDD is often caused by structural and biochemical changes in intervertebral discs (IVD) that prompt a pathologic shift from an anabolic to catabolic state, affecting extracellular matrix (ECM) production, enzyme generation, cytokine and chemokine production, neurotrophic and angiogenic factor production. The IVD is an immune-privileged organ. However, during degeneration immune cells and inflammatory factors can infiltrate through defects in the cartilage endplate and annulus fibrosus fissures, further accelerating the catabolic environment. Remarkably, though, catabolic ECM disruption also occurs in the absence of immune cell infiltration, largely due to native disc cell production of catabolic enzymes and cytokines. An unbalanced metabolism could be induced by many different factors, including a harsh microenvironment, biomechanical cues, genetics, and infection. The complex, multifactorial nature of IDD brings the challenge of identifying key factors which initiate the degenerative cascade, eventually leading to back pain. These factors are often investigated through methods including animal models, 3D cell culture, bioreactors, and computational models. However, the crosstalk between the IVD, immune system, and shifted metabolism is frequently misconstrued, often with the assumption that the presence of cytokines and chemokines is synonymous to inflammation or an immune response, which is not true for the intact disc. Therefore, this review will tackle immunomodulatory and IVD cell roles in IDD, clarifying the differences between cellular involvements and implications for therapeutic development and assessing models used to explore inflammatory or catabolic IVD environments.

Keywords: intervertebral disc degeneration, low back pain, inflammation, catabolism, immune-privileged microenvironment, GWAS, artificial intelligence–AI, agent-based model (ABM)

INTRODUCTION

Epidemiology of Intervertebral Disc Degeneration

Low back pain (LBP) is the largest cause of morbidity worldwide, affecting approximately 80% of people from Western countries during their lifetime and resulting in 5 million disability-adjusted life-years in young adults (GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, 2018). Lower intervertebral disc degeneration (IDD) is the cause of around half of all LBP cases in young adults; however not all cases of IDDresult in LBP (Baumgartner et al., 2021c). Although IDD prevalence increases progressively with age, IDD is commonin subjects younger than 30 years old, conveying those various other factors besides age, such as excessive or uneven mechanical load, obesity, genetics, nutrition, trauma, and gender are involved (Hoogendoorn et al., 2000; Paassilta et al., 2001; Pincus et al., 2002; Cheung et al., 2009; Samartzis et al., 2012; Teraguchi et al.,2014; Parenteau et al., 2021). For example, studies have shown that women experience greater pain and disability than men when they are treated for IDD (MacLean et al., 2020). Additionally, LBP prevalence in females after menopause further increases in comparison to men at comparable ages (Wáng et al., 2016). Further, it is unclear whether occupation- related heavy physical loading is an important risk factor for IDD, as studies have contradictory conclusions (Bongers et al., 1990;Videman and Battié, 1999). Some studies have found IDD is significantly more common in athletes compared to the generalpopulation (Swärd et al., 1991). However, various twin studies that have been conducted suggest that occupation or sport relatedrisk factors have only a minor role in IDD, while geneticinfluences were found to play a greater role in predicting degeneration (Battié et al., 2004). On the other hand, obesity is associated with increased IDD severity and extent, likely due to altered biomechanical and/or biological processes such as thosedriven by adipokines (Samartzis et al., 2012; Li W. et al., 2022).Due to the complexity and multifactorial nature of IDD, the initiating and risk factors are poorly understood, which criticallyhampers proper LBP patient stratification and limits the development of personalized therapies.

The Structure of the Intervertebral Disc

The intervertebral disc (IVD) is the largest avascular organ in the human body with blood vessels only present in the outer annulus fibrosus (AF) and boney end plates, with all nutrient and waste exchange taking place via diffusion through the dense extracellular matrix of the disc (Urban 2002). Located between the vertebrae within the spine, the IVD consists of three highly hydrated, major tissues: 1) the nucleus pulposus (NP), 2) the annulus fibrosus (AF), and 3) the cartilage endplate (CEP). The central and proteoglycan-rich NP lies between the cranial and caudal CEPs and is laterally constrained by the peripheral and fiber-reinforced AF (Figure 1). This specialized composition and structure of the IVD ensures both trunk movements and resistance to high mechanical loads. The normal human IVD contains nucleus pulposus cells and annulus fibrosus cells within the NP and AF, respectively, with AF cells becoming more elongated and fibroblast-like towards the periphery. Cells occupy 1% volume of the disc, though are crucial in maintaining the balance between anabolic activity such as the production of proteoglycans and collagens type I and II, and theprocatabolic effects of factors involved in ECM turnover, including metalloproteinases, prostaglandins, and nitric oxide (Kang et al., 1997). Furthermore, mechanical loads are thought to influence ECM homeostasis, where both excessive and insufficient loads lead to catabolism (Vergroesen et al., 2015). Due to the avascularity of the IVD, the environment is hypoxic,where the oxygen tension in the IVD is considered between 1 and5% (Yao et al., 2016; Yao et al., 2017).

In comparison to NP and AF tissues, the CEP receives far less attention in the literature; however, it is a vital tissue when discussing LBP. Lakstins et al. (2021) demonstrated that imperfections and weakness in the CEP can be a better anticipator of pain than IVD degeneration because chemical changes to the CEP are directly related to intervertebral disc degeneration (IDD) (Yao et al., 2016). The CEP is rich in collagen type II (Yao et al., 2016) and performs both mechanical and chemical functions (Roberts et al., 1989; Lakstins et al., 2021). Mechanically, the CEP acts as a physical filter preventing macromolecules from escaping the disc throughthe subchondral bone and is considered important in controllingthe hydration of the disc under mechanical loads (Roberts et al.,1989; Moore 2000; Ruiz Wills et al., 2018). Chemically, the CEPallows metabolites, small molecules and waste to travel between theIVD and neighboring blood vessels in the bony endplates (Robertset al., 1989; Turgut et al., 2003; Yao et al., 2016; Ruiz Wills et al.,2018; Zuo et al., 2019; Sun et al., 2020; Lakstins et al., 2021).

The diffusivity of solutes through the CEP and towards the IVD depends greatly on their size and ionic charge. The healthyIVD is negatively charged due to the high proteoglycan concentration (Moore 2000; Pfannkuche et al., 2020). Therefore, only small, neutrally charged solutes such as glucose and oxygen, as well as cations such as sodium and calcium can penetrate the disc, but small anions such as sulphate and chloride ions can only cross through the CEP. In turn, large, neutrally charged solutes such as antibodies and enzymes usually cannot penetrate the healthy IVD (Moore 2000).

Intervertebral Disc Degeneration

Regarding disc morphology, as the IVD degenerates, it becomes more difficult to distinguish the boundaries between the AF and the NP. This loss of a distinct boundary worsens with age, as the nucleus loses its gel-like quality and becomes more fibrotic (Buckwalter 1995) which was seen as a common degenerative feature across all species (Dahia et al., 2021). Another significant biochemical change during disc degeneration is the loss of proteoglycans, which are necessary to provide the osmotic resistance for the IVD to withstand compressive loads and keep the disc hydrated (Knudson and Knudson, 2001). Such significant changes (loss of water content (Lyons et al., 1981) and disc height (Frobin et al., 2001)) in disc behavior strongly influence other spinal structures and may negatively impact their function and predispose them to injury.

During IDD, the CEP becomes thinner and fissured, with lower collagen and glycosaminoglycan (GAG) content (Turgut et al., 2003; Sun et al., 2020). This change in morphology affects the physiology and the performance of the CEP (Hamilton et al., 2006; Roberts et al., 2006) altering its permeability (Roberts et al., 2006; Yao et al., 2016). Furthermore, the CEP can lose its connection to the vasculature (Moon et al., 2013), which immunohistochemistry has shown leads to blood vessel and nerve fiber infiltration into the IVD through the CEP and subchondral bone and through fissures in the AF (Freemontet al., 1997; Roberts et al., 2006; Binch et al., 2015b; Yao et al., 2016; Lama et al., 2018; Sun et al., 2020). Moreover, the crosstalk between IVD and the bone marrow is facilitated due to the CEPdamage (Dudli et al., 2016), causing possible adjacent "Modic discs". Modic changes (MC) are defined as magnetic resonanceimaging (MRI) signal alterations in the vertebral bone marrow close to a degenerated disc. There are several different types ofMC, with MC1 fibrotic lesion having the highest association with pain, followed by MC2. MC3 are rare and often asymptomatic.MC1 and MC2 are commonly accompanied by persistent inflammatory stimulus. In addition, MC related pain could be related to the neovascularization and neurogenesis due to the increase in growth factor expression by blood vessels and disc cells and inflammatory cytokines (Rätsep et al., 2013; Sun et al., 2013; Dudli et al., 2018) which lead to increased expression of neurotrophic factors (Freemont et al., 2002; Purmessur et al., 2008; Binch et al., 2014) (Figure 1).

During disc degeneration, the balance between anabolism and catabolism is dysregulated, showing decreased synthesis of normal matrix, of collagen type II and aggrecan and increased presence of matrix degrading enzymes, reviewed previously by Baumgartner et al. (2021c). Moreover, several studies have reported decreased NP cell proliferation under catabolic cytokine stimulation (Wang et al., 2013; Li et al., 2019; Linand Lin, 2020). Similarly, during the shift from anabolic to catabolic cell activity in the disc, the presence of these cytokines is also related to NP and AF cell apoptosis, (Huet al., 2017; Yu et al., 2018; Zhang J. et al., 2019; Zhang S.et al., 2019).

These changes have been shown, at least in part, to be modulated by pro-catabolic cytokines in numerous studies,which are often referred to as inflammatory features, in the literature. However, since these factors are produced by native disc cells (NP, AF and CEP) in intact discs, this catabolic response can be easily misconstrued as an inflammatory response. Therefore, the aim of this review is to tackle

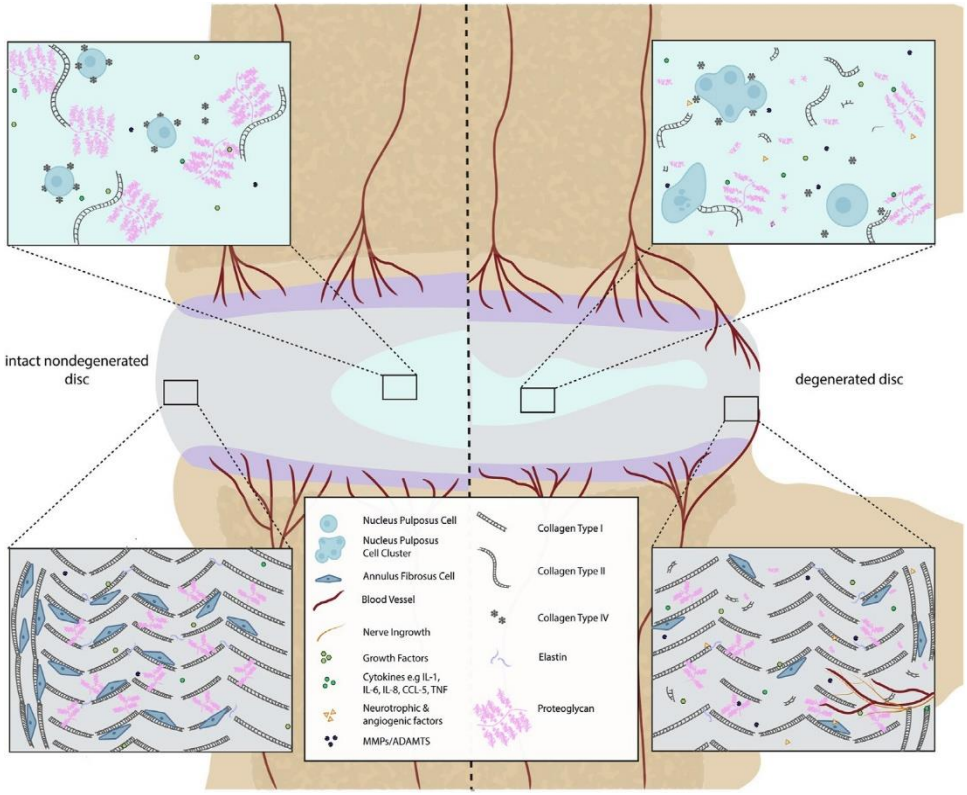


FIGURE 1 | Comparison of a healthy and a degenerated IVD disc (focused on ECM components). In the intact IVD, the NP matrix mostly contains proteoglycans (PG) and non-oriented collagen type II fibers. The proteoglycans contain negatively charged sulfated groups leading to an intradiscal osmotic pressure crucial for the basal hydration of the NP and the biomechanical function of the IVD. Within the degenerated disc, the total content of PG decreases. Small non-aggregating PGs are present. This drop-in PG content negatively affects the swelling capacity of the disc. Additionally, during disc degeneration, the production of catabolic cytokines, matrix-degrading enzymes, and neurotrophic as well as angiogenic factors occur due to cellular changes. This can lead to blood and nerve vessel ingrowth in the AF. TheAF is composed of highly oriented concentric lamella of type I collagen whereas the cell density is higher in intact than in degenerated discs.

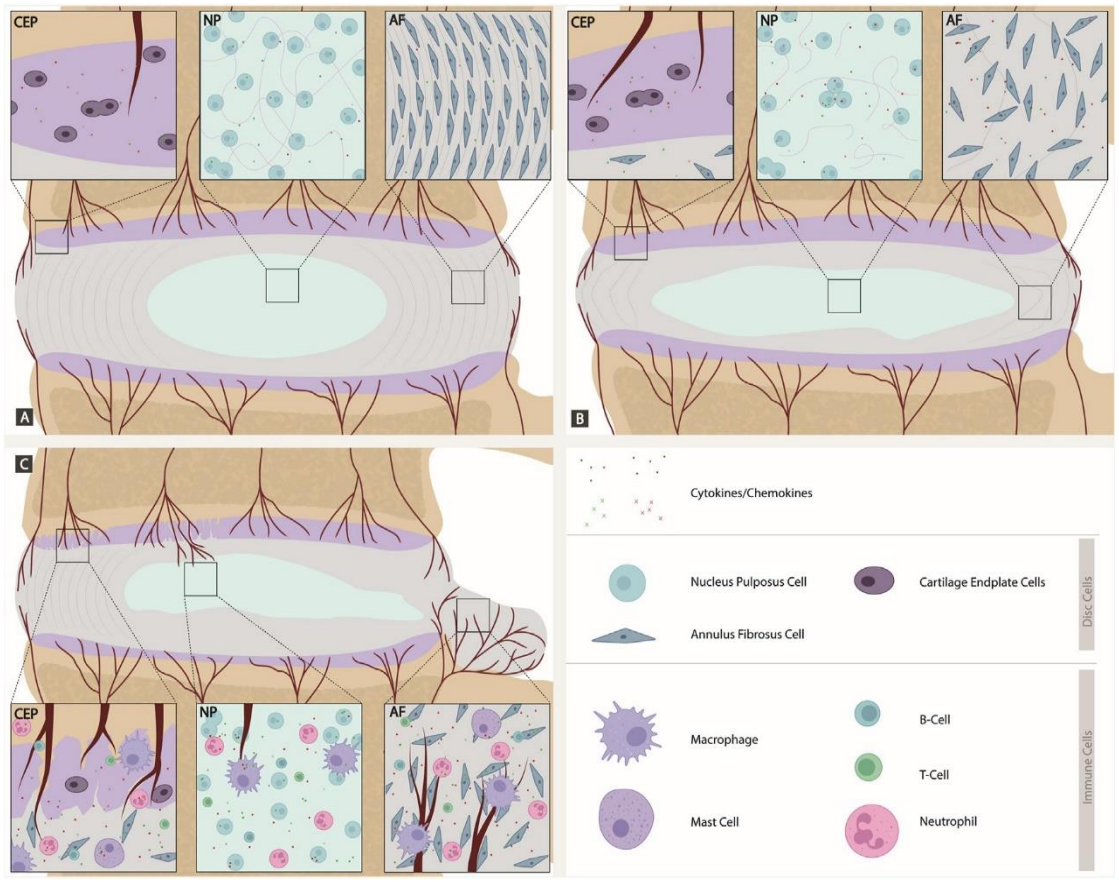


FIGURE 2 | Comparison of a healthy, degenerated, and herniated IVD discs (focused on cellular involvement). (A) Intact IVD: Native disc cells produce a plethora of cytokines and chemokines expressing the corresponding receptors and maintaining homeostasis in a para and autocrine manner. The CEP is intact with blood vessels. The NP has a high number of proteoglycans. AF cells aligned. (B) Degenerated IVD: Shift to catabolic environment. Cytokines are expressed by disc cells themselves. The CEP has a higher amount of blood vessels than in the intact IVD. Proteoglycan number decreases in the NP. In the AF, there is a loss of alignment and support for AF cells. (C) Herniated IVD with crack in CEP: As soon as the AF or CEP is ruptured during injury or disc degeneration, a route for migration of immune cells into the IVDis provided. Immune cells, including T cells, B cells, macrophages, neutrophils and mast cells, contribute to an inflamed environment within the disc, further increasing the cytokine and chemokine expression and leading to a viscous circle of inflammatory driven catabolism. A crack in CEP allows blood vessels to grow into the AF and NP. The AF herniates/bulges, which is where blood vessel in-growth primarily occurs.

immunomodulatory and IVD cell roles in IDD and clarifying the differences between cellular involvements. Furthermore, different *in-silico, in-vivo* and *in-vitro* models used to explore inflammatory or catabolic IVD environments will be discussed.

CROSS-TALK BETWEEN THE IMMUNE SYSTEM AND IVD IN IDD

As mentioned, the IVD is the largest avascular organ with blood vessels only present in the outer AF and bony end plates, so all metabolite exchange takes place via diffusion through the dense extracellular matrix (ECM) of the IVD. The dense ECM of the IVD inhibits blood vessel ingrowth both mechanically by having a high physical pressure, and chemically through high proteoglycan concentration (Johnson et al., 2002; Johnson et al., 2005), which combined with secretory inhibitors prevent nerve and blood vessel ingrowth in non-degenerate discs (Tolofari et al., 2010; Binch et al., 2015a). The AF and the CEP, along with the secretory inhibitors of angiogenesis, are defined as the blood-NP barrier (BNB), which strongly isolates the NP from the circulation and thus the host immune system (Sun et al., 2020).

Where both AF and CEP are intact, the IVD has been described as an immuno-privileged tissue (Sun et al., 2020)with a lack of immune cells (Figure 2). However, this is often confused as the native cells of the disc (i.e., the NP, AF and CEPcells) have been shown to take on roles and markers classically expressed by immune cells (Le Maitre et al., 2005; Jones et al., 2008; Phillips et al., 2013a; Risbud and Shapiro, 2013), and thus have been described by some authors as immune responses or inflammation. However, such activity of native IVD cells is not true inflammation. Therefore, distinguishing which cases of IDD involve an immune response is important as different clinical interventions and treatments would be required.

Native disc cells produce a plethora of cytokines and chemokines which are upregulated during disc degeneration

and have been shown to drive many catabolic events in the IVD (Weiler et al., 2005; Le Maitre et al., 2005; Hoyland et al., 2008; Phillips et al., 2013a; Phillips et al., 2015). A shift to catabolism isat least in part driven by the increased production of cytokines in the disc by the native cells (in an intact disc) and a combination of inflammatory cells and native disc cells following CEP and AF rupture. Phillips et al. (2015) demonstrated that NP cells express a number of cytokine and chemokine receptors and are thus able to respond in a paracrine and autocrine manner (Figure 2). Caused by different, yet not fully understood mechanisms, disc cells upregulate the expression of inflammatory cytokines such as IL-1, TNFα, IL-6, IL-8 and IL-17 amongst others creating a cytokine rich catabolic environment. IL-1 has been shown to drive the catabolic events during disc degeneration (Le Maitre et al., 2005; Weiler et al., 2005; Hoyland et al., 2008; Phillips et al., 2013a; Phillips et al., 2015). Whilst other cytokines and chemokines (e.g., MCP-1, TNFα, IL-8) produced in the disc appear to have limited effects on the native disc cells due to lack of receptor expression *in vivo* (Le Maitre et al., 2007; Phillipset al., 2015), they undoubtedly diffuse to the surrounding tissuesleading to increased inflammation in local tissues, and drive cellular infiltration following AF and CEP rupture and increased sensitisation of nerves (Ye et al., 2022). Such cytokines can stimulate specific intracellular signalingpathways that further enhance the degenerative process(Daniels et al., 2017; Suyama et al., 2018) and upregulate matrix-degrading enzymes known as matrix metalloproteinases (MMPs) and a disintegrant and metalloproteinase with thrombospondin motifs (ADAMTS), specifically MMP- 1, 2, 3,9, 13 and ADAMTS-4, 5 (Baumgartner et al., 2021c). In later phases of IDD, these cytokines can upregulate neurotrophic and angiogenic factors, which could lead to further nerve and bloodvessel ingrowth (Purmessur et al., 2008; Lee et al., 2011; Binch et al., 2014; Krock et al., 2014).

Remarkably, some of these cytokines, such as IL-1, have also been shown to be expressed in cells from non-degenerate discsand display roles in maintaining normal homeostasis (Le Maitreet al., 2005; Phillips et al., 2015). Indeed, if the IL-1 agonists areknocked out during development, IDD can be induced (Gorthet al., 2019). Thus, IL-1 plays a role as a normal regulatory mechanism during IVD homeostasis, which becomes imbalanced during IDD (Le Maitre et al., 2005) (Figure 2). Native NP cells have also been shown to take on other roles normally associated with immune cells. such as phagocytosis: Jones et al. (2008) observed the capacity of bovine NP cells to ingest latex beads at least as efficiently as phagocytic cells and ingested apoptotic cells. This capability could be of great physiological relevance to maintain a healthy disc, as it may prevent inflammation triggered by the release of toxic or immunogenic intracellular content by apoptotic cells (Fadok, 1999). Clearly, the suggestion from some reviews that cytokine production within the disc is solely from immune cells is inaccurate (Ye et al., 2022). However, when the AF or CEP becomes ruptured or fissures occur during injury or disc degeneration this provides a route for blood vessel ingrowth and migration of immune cells into the intervertebral disc. Within these "non-intact" IVDs, immune cells will migrate including T cells ($CD4^+$, $CD8^+$), B cells, macrophages, neutrophils and mast cells (Risbud and Shapiro, 2013)(Figure 2).

These immune cells then contribute to an inflamed environment in the disc, leading to further increases in cytokine and chemokine expression (Phillips et al., 2015). This leads to a viscous circle of inflammatory driven catabolism which acts synergistically with the native IVD cells to cause acceleratedECM breakdown (Figure 3) (Risbud and Shapiro, 2013). Thesecytokines and chemokines play a number of roles within this disc, including direct actions on NP, AF and CEP cells where their receptors are present (Le Maitre et al., 2005; Le Maitre et al., 2007;Phillips et al., 2015). They will also likely diffuse out of the IVD leading to increased cellular migration to the disc (Pattappa et al., 2014), or sensitization of local nerve roots (Leung and Cahill, 2010; Johnson et al., 2015).

THE SHIFT FROM ANABOLISM TO CATABOLISM

A Complex Interplay of Microenvironment, Biomechanics, Genetics and Epigenetics, Bacterial Infection and IVD Cells

As discussed during disc degeneration, there is a shift in metabolism from anabolism (matrix synthesis) to catabolism (matrix degradation), and this shift to catabolism is accompanied by increased production of neurotrophic and angiogenic factors which lead to nerve and blood vessel ingrowth leading to inflammation and increased pain sensation in the disc (Baumgartner et al., 2021c). There remains a poor understanding as to the initiating factors involved in this switch from anabolism to catabolism in disc degeneration, which is likely due to multifactorial processes including the disc microenvironment, biomechanics, genetics and epigenetics, and even bacterial infection of the disc andthe gut microbiome (Figure 3) (Li W. et al., 2022).

Altered Disc Microenvironment

The IVD microenvironment is commonly described as harsh due to its limited nutrition (glucose and oxygen), low pH, and large changes in tissue osmolarity (Urban 2002). These factors not only impair the success of cell therapies (e.g., mesenchymal stromal cell injection) (Loibl et al., 2019; Williams et al., 2021) but can also negatively affect resident IVD cells and thus contribute to the catabolic-inflammatory shift observed during degeneration.

Limited nutrition and tissue acidity are a result of the avascular nature of the IVD (Hukins, 1988). Glucose and oxygen transport into the IVD, as well as the removal of cellular waste products such as lactic acid (which contributes to the drop in tissue pH),are hence dependent on diffusion via the capillaries in the endplate. Degeneration-associated calcification as well as a reduction in the density of the bone marrow contact channels in the endplates might further impair these transport mechanisms (Benneker et al., 2005; Chen et al., 2014). In the centre of the IVD, glucose levels, oxygen and pH can hence drop

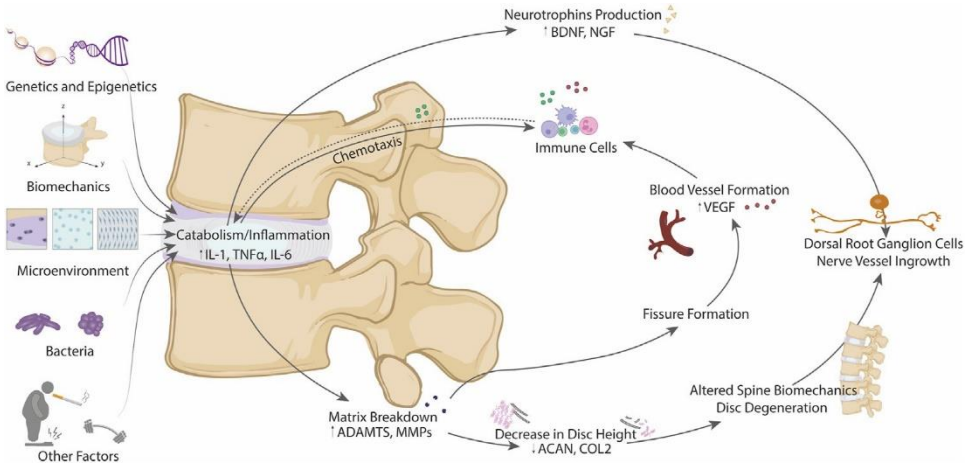


FIGURE 3 | Schematic diagram of the different factors contributing to the metabolic shift from anabolism to catabolism in IDD, including genetics and epigenetics, biomechanics, microenvironment, presence of bacteria and other factors. All these contributors can promote a downstream biochemical effects (matrix breakdown and neurotrophins production) leading to structural and biomechanical alterations, nerve ingrowth and blood vessel formation. Thus, involvement of immune system could be achieved by chemotaxis losing the immuno-privileged state of the IVD.

to 0.5 mM, 1% oxygen, and 6.5, respectively, although even lower values have also been reported (Urban et al., 2004). It is not clear whether glucose deprivation can activate the expression of pro-catabolic factors in NPC (Baumgartner et al., 2021b). However, multiple studies have demonstrated that glucose deprivation impairs NP cell proliferation and survival as well as proteoglycan synthesis/degradation and collagen synthesis(Bibby and Urban, 2004; Johnson et al., 2008; Chang et al., 2017; Saggese et al., 2020), and recent evidence highlights that non-coding RNAs (e.g., circ_0075062) may be involved in these processes (Chang et al., 2021). In contrast, oxygen deprivation alone was shown numerous times to have little effect on IVD cell proliferation or survival and mostly contributes to cell impairment prompted by other microenvironmental factors(Bibby and Urban, 2004; Johnson et al., 2008). Indeed, IVD cells are likely unaffected by hypoxia because of their robust and constitutive expression of hypoxia-inducible factor (HIF) 1, and
more importantly, the inducible subunit HIF-1α (Merceron et al., 2014). Even more, hypoxia and HIF-1α were recently shown to attenuate the endoplasmic reticulum (ER) stress responses in NP cells (Novais et al., 2021). However, research related to rheumatoid arthritis demonstrated that the expression of Toll-like receptor 4 (TLR4) and TNF-α, but also of IL-10, are HIF-1-dependent processes in macrophages, indicating that IVD-infiltrating immune cells may be more affected by the hypoxic microenvironment than IVD cells themselves (Guo and Chen, 2020). Similar to low glucose concentrations, high lactic acid concentrations and the resulting drop in pH promotes cell death and a catabolic shift in mRNA expression (Horner and Urban, 2001; Bibby and Urban, 2004; Neidlinger-Wilke et al., 2012; Gilbert et al., 2016), likely via acid-sensing ion channels (Gilbert et al., 2016). Importantly, this response was morepronounced when cells were simultaneously exposed to low glucose levels (Bibby and Urban, 2004). Therefore, low glucoseand high lactic acid levels, but not hypoxia, contribute to the catabolic shift observed during IVD degeneration. In general, there is a clear need to increase the number of experimental studies where different microenvironmental factors are combined, including both nutritional and pro-inflammatorycues (Baumgartner et al., 2021b).

Aside from low glucose, oxygen, and pH, the osmolality of the IVD is often considered a fourth harsh microenvironmental factor. This mostly refers to the relatively high osmolarity inthe IVD, where 400 mOsm is considered iso-osmotic. However, itis important to note that IVD cells can be exposed to a wide rangeof tissue osmolarity, and these changes are more likely to affectIVD cell behaviour than the iso-osmotic condition. It can drop as low as 300 mOsm with the loss of proteoglycans during degeneration, and increase to approximately 500 mOsm during high mechanical loading (Sadowska et al., 2018). A reduction intissue osmolarity leads to cell swelling (up to 20%) by the solubility-diffusion water transport across the cell membrane (Sadowska et al., 2018; Snuggs et al., 2021). Ample publications have shown that this hypoosmotic shift can activate and/or interplay with pro-inflammatory factors and catabolic responses and, hence, promote IVD inflammation and degeneration (Chen et al., 2002; Wuertz et al., 2007; Walter et al., 2016; Sadowska et al., 2020). Although the underlying mechanisms have not yet been identified, Transient Receptor Potential (TRP) channels and aquaporins may beinvolved (Sadowska et al., 2018; Sadowska et al., 2020; Snuggs et al., 2021). The hyper-osmotic shift in the IVD microenvironment leads to activation of the robustly expressed osmo-sensitive transcription factor TonEBP (tonicity-responsive enhancer binding protein) (Sadowska et al., 2018; Baumgartneret al., 2021c), which protects IVD cell viability underhyperosmotic stress (Tsai et al., 2006; Choi et al., 2018) andcan also be regulated by cytokines (Johnson et al., 2017). These studies on IVD cells highlight that hyperosmolarity is likely not a main contributor to the catabolic shift in the IVD,

whereas hypo-osmolarity seems to have detrimental effects on IVD cells. However, although no studies have specifically investigated the effect of osmolarity on IVD-infiltrating immune cells yet, research on other tissues (e.g., renal medulla, skin, lung epithelium) indicates that increased osmolarity activated macrophage inflammatory responses, which is at least partially TonEBP-dependent (with a thresholdat approximately 360−380 mOsm) (Aramburu and López- Rodríguez, 2019). More research will hence be needed to better understand the role of the IVD microenvironment on macrophage polarization and the behaviour of other infiltratingimmune cells.

Unbalanced Biomechanics and Mechanobiology in Catabolism

Biomechanics is another key contributor in the shift from anabolism to catabolism (Adams et al., 2000). The IVD experiences various forces throughout everyday life, which are necessary to maintain the health of the disc. For instance, the average intradiscal pressure in a healthy IVD ranges from 0.1 MPa (lying prone) to 0.5 MPa (standing flexed forward) (Wilke et al., 1999). However, damage occurs in the disc when it encounters abnormal or excessive forces, leading to catabolism, including increased cytokine production, and matrix degradation (Walter et al., 2011; Vergroesen et al., 2015; Fearing et al., 2018). This damage is believed to cause microinjuries within the disc, which gradually build up over time (Adams and Roughley, 2006; Baumgartner et al., 2021c), and is likely to contribute to the infiltration of immune cells because of the chemo-attraction effect of the pro-inflammatory cytokines released by the native IVD native cells. As the tissue degenerates, the size and composition of the IVD changes, leading to impaired response to any mechanical loading placed on the disc and causing further damage, possibly leading to disc herniation or endplate defects.

As a highly hydrated tissue, the NP provides protection to compressive forces imposed on the IVD (Adams and Roughley, 2006) while the more fibrous, surrounding AF confines the NP swelling pressure and helps the IVD to resist shear and tensile forces (Chu et al., 2018). When the NP loses hydration, the compressive load is transferred to the AF (Adams et al., 1996). Whereas the healthy AF, as a whole, is highly resistant mechanically, aberrant loading can further contribute to fissure formation where the tissue is already weakened by altered turnover, which allows for associated blood vessel growth and immune cell infiltration as discussed earlier (Lama et al., 2018).

At the cell level, specific biomechanical cues have been shown to impair IVD cell response. For example, shear stress has been found to lead to increased nitric oxide, causing downstream reduction in proteoglycan synthesis and increased apoptosis in IVD cells (Liu et al., 2001). Interactions with biochemical signalling was further demonstrated. For instance, it was found that AF and NP cells from a degenerated IVD respond differently to those from a healthy disc, suggesting that mechano-transduction pathways are altered through degeneration (Le Maitre et al., 2008; Le Maitre et al., 2009a; Chu et al., 2018) and can be modulated by cytokines such as IL-1 and IL-4 (Elfervig et al., 2001a; Gilbert et al., 2011). Additionally, a pro-inflammatory environment has been shown to change the mechanobiology of IVD cells. Treatment with inflammatory stimuli, specifically liposaccharide (LPS) or TNF-α, before osmotic loading was shown to increase hydraulic permeability and cell size, disrupt the F-actin cytoskeleton, and increase aquaporin-1, which is a main water channel in NP cells (Maidhof et al., 2014). Recently, Hernandez et al. demonstrated that inhibiting actomyosin contractility in NP cells caused a similar response as TNF-α induced inflammation, while increasing contractility protected the cells against TNF-α. Actomyosin contractility was also shown to regulate nuclear factor kappa-B (NF-κB) and downstream ECM degradation, conveying that mechano-transduction and inflammatory pathways are connected andthe cross-talk could play an important role in IDD (Hernandezet al., 2020b). Thus, altered biomechanics can lead to mechanobiology alterations promoting matrix degradation and impacting the capacity of the disc to sense loads normally, leading to increased catabolism and IDD development.

Genetics and Epigenetics in IDD

Among the different causes for IDD, genetic susceptibility plays a crucial role. High heritability (over 70%) has been systematically reported for IDD (Battié and Videman, 2006; Kepler et al., 2013),as well as specific traits such as herniation (Sambrook et al., 1999) and endplate defects (Munir et al., 2018) and their progression (Williams et al., 2011b). Genetic burden, in such elevated polygenicity presented by IDD, is suggested to carry a larger effect than environmental factors, with the exception of body mass index (BMI). However, BMI itself has high heritability and polygenicity (Robinson et al., 2017), which is partially overlapped with IDD (Zhou et al., 2018).

Genetic associations for IDD have been mostly researched with candidate gene studies [see focalized reviews in Mayer et al.(2013), Feng et al. (2016), Kawaguchi (2018)]. The most representative functional group consists of genes of structural proteins and those regulating its turnover (Table 1).

Among structural components, collagen variants have been extensively assessed, and several collagen types and variants have been associated with IDD. Collagen type IX polymorphisms in alpha 2 and 3 (COL9A2, COL9A3) chains have been found to influence MRI signal intensity in NP (Wrocklage et al., 2000;Solovieva et al., 2006; Zhang et al., 2008; Näkki et al., 2011). Further, Solovieva et al. (2006) stated that the effect of the Trp3 allele in COL9A3 is dependent on an IL-1B polymorphism, reflecting the effect of immune-modulators/catabolic factors on IVD degeneration. Nevertheless, the pathophysiology of this interaction is not yet described. Polymorphisms of collagen type XI (COL11) have been associated with higher risk of herniation (Mio et al., 2007; Videman et al., 2009) and otherdegenerative traits (Noponen-Hietala et al., 2003; Solovieva et al., 2006; Videman et al., 2009; Kalb et al., 2012). Mio et al. (2007) stated that the variant rs1676486, which falls in *cis* elements region lowers COL11A1 expression due to decreased stability of its transcripts/mRNAs.

TABLE 1 | Gene variants of structural/regulatory components of IVD associated with IDD by candidate gene approach.

Structural/Regulatory component	Function	Gene	Gene variant	Molecular level	Contribution to IDD	Referernce
Collagen IX	Cartilage anabolic marker	*COL9A2*	rs7533552	-	Associated with greater disc bulging (L1-L4)	Näkki et al. (2011)
				May contribute to reduced collagen crosslinking	May contribute to disc instability and eventually prolapse in the elderly	Wrocklage et al. (2000)
		COL9A3	Trp3 allele in IL1B 3954 C/T variant	*COL9A3* gene on IDD might be modified by the IL 1β gene polymorphism	Influence MRI signal intensity in NP in the absence of the IL1β 3954 C/T allele	Solovieva et al. (2006)
Collagen XI	Anabolic marker	*COL11A1*	rs1676486	Lower *COL11A1* expression	High risk of herniation	Mio et al. (2007)
		COL11A2	rs2076311	-	Association with (i) disc signal intensity (ii) disc bulging	Videman et al. (2009)
Collagen I	AF anabolic marker	*COL1A1*	rs1800012	-	Not associated with IDD (taken as a single factor)	Anjankar et al. (2015)
				-	Risk factor related to IDD in older people	Pluijm et al. (2004)
				-	Strong association with LDD in young male	Tilkeridis et al. (2005)
Aggrecan	IVD anabolic marker	*ACAN*	*ACAN* VNTR polymorphisms	-	Increased risk of LDD of shorter alleles	Kawaguchi et al. (1999), Gu et al. (2013)
				-	Aggrecan allele with 26 repeats is associated with dark NP MRI intensity	Solovieva et al. (2007)
Cartilage Intermediate Layer Protein	Cartilage-like catabolic marker	*CILP*	rs2073711	TGF-β1 inhibition mediated induction of ECM proteins through direct interaction with TGF-β1	Association between IDD and *CILP* rs2073711 variant in women	Kelempisioti et al. (2011)
			1184T/C	-	The *CILP* SNP 1184T/C is a risk factor for male collegiate athletes	Min et al. (2010)
				-	Upregulation of *CILP* in intervertebral discs increased disc degeneration progressed	Seki et al. (2005)
Metalloproteinase	Catabolic marker	*MMP3*	Combination of the T-C haplotype of IL 1α and the MMP3 minor 5A allele	*IL1* promoted cartilage degradation through the induction of the matrix-degrading enzymes such as MMP1, MMP3, and MMP13	Association between a combination of *IL 1* and *MMP3* gene variations and type II Modic changes among middle-aged Finnish men	Karppinen et al. (2008)
			promoter 5A/6A	Enhanced the degeneration of IVD associated with environmental conditions resulting from the induction of a higher level of MMP3 expression in response to such conditions	accelerate IVD degeneration in the elderly	Takahashi et al. (2001)
			Intron 4 C/T	-	Associated with radiographic progression of LDD	Valdes et al. (2005)
		MMP2	1306C/T	-	Correlation with more severe grades of disc degeneration and thus may be a genetic risk factor related to LDD susceptibility in the young adult population	Dong et al. (2007)
		MMP9	1562 C/T	-	Associated with a high risk of degenerative disc disease in the young adult population in North China	Sun et al. (2009)

(Continued on following page)

TABLE 1 | (*Continued*) Gene variants of structural/regulatory components of IVD associated with IDD by candidate gene approach.

Structural/Regulatory component	Function	Gene	Gene variant	Molecular level	Contribution to IDD	Refererrnce
Interleukin	Catabolic marker	*IL 1β*	3954 C/T	COL9A3 gene polymorphism on IDD might be modified by the IL-1β gene polymorphism	Association between collagen gene polymorphisms and disc degeneration of the lumbar spine is modified or negatively confounded by the IL1β (C3954-T) polymorphism in middle-aged working men	Solovieva et al. (2006)
		IL 1α	Combination of the T-C haplotype of IL1A and the MMP3 minor 5A allele	IL1 promotedcartilage degradation through the induction of matrix-degrading enzymes such as MMP1, MMP3, and MMP13	Association between a combination of IL1 and MMP3 gene variations and type II Modic changes among middle-aged Finnish men	Karppinen et al. (2008)
			889C/T	-	IL1 gene cluster polymorphisms have an effect on the risk of disc degeneration, particularly TT genotype of the IL-1α gene promotes higher risk of disc bulges	Solovieva et al. (2004)
				Significantly increased the transcriptional activity of the IL1A gene and IL-1β protein	IL 1α −889T represented a significant risk factor for the IDD-phenotype	Virtanen et al. (2007)
		IL 6	rs1800797, rs1800796 and rs1800795	-	IL 6 variants are associated with moderate IDD in a sample population of young adults	Kelempisioti et al. (2011)
			597G/A, 174G/C and 15T/A	-	association analysis provided support for a link between the IL 6 sequence variants and IDD	Noponen-Hietala et al. (2005)
		IL 18	rs1420100	-	Association with severe degeneration	Omair et al. (2013)
Thrombospondin	ECM regulation	*THBS2*	rs9406328	lower affinity for MMP binding and thus reduces MMP degradation	Regulation of Intervertebral disc ECM metabolism by the THBS2-MMP system plays an essential role in the etiology and pathogenesis of lumbar disc herniation	Hirose et al. (2008)
A disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS)	Catabolic marker	*ADAMTS5*	rs151058, rs229052, and rs162502	Decreased binding affinity with LRP1 (protein that regulates its degradation by endocytosis)	Genetic polymorphisms of ADAMTS 5 may be associated with susceptibility to LDD	Wu et al. (2014)
Growth differentiation factor 5	Pro-chondrogenic factors	*GDF5*	rs143383	-	5 population cohorts from Northern Europe indicate that a variant in the *GDF5* gene is a risk factor for LDD in women	Williams et al. (2011a)
tSKT		*KIAA1217*	11 KIAA1217 variants in Exon 2, 3, 6, 7, 13, 14, 17 and 19	-	Strong causative candidates for the Vertebral Malformation phenotypes	Al Dhaheri et al. (2020)
			rs16924573	-	Association with lumbar disc herniation	Karasugi et al. (2009)
				-	Association with lumbar disc herniation	Kelempisioti et al. (2011)
FAS receptor and ligand	Cell apoptosis factors	*FAS and FASL*	rs2234767(*FAS*) and rs763110(*FASL*)	-	FAS and FASL may be associated with the presence and severity of LDD	Zhu et al. (2011)

(Continued on following page)

TABLE 1 | (*Continued*) Gene variants of structural/regulatory components of IVD associated with IDD by candidate gene approach.

Structural/Regulatory component	Function	Gene	Gene variant	Molecular level	Contribution to IDD	Referernce
Caspase-9		*CASP-9*	1263A/G	-	Risk factors in the incidence of LBP in Chinese male soldiers	Mu et al. (2013)
			rs1052576	-	Associated with lumbar disc herniation and disc degeneration in the Han population of northern China	Sun et al. (2011)
Tumor necrosis factor related apoptosis-inducing ligand		*TRAIL*	1525 G/A and 1595 C/T	-	Associated with the susceptibility and severity of LDD in the Chinese Han population	Du et al. (2015)
Death receptor 4		*DR4*	rs4871857	-	Associated with the risk and severity of LDD in the Chinese Han population	Tan et al. (2012)

Similarly, a polymorphism found in an intronic region of the collagen 1 gene (COL1A1) that corresponds to a binding site of Specificity protein 1 (Sp1) has been shown to increase the risk of IDD, but the mechanism is not reported yet (Pluijm et al., 2004; Tilkeridis et al., 2005; Anjankar et al., 2015). However, it has been demonstrated that Sp1 downregulates pro-inflammatory cytokine-induced catabolic gene expression in disc cells (Xuet al., 2016). Additionally, Sp1-dependent mechanisms have been reported to modulate mechanically-induced apoptosis and autophagy in IDD (Li et al., 2020). Nonetheless, Sp1 also affects processes in other tissues including differentiation, angiogenesis and chromatin remodeling (Tan and Khachigian, 2009), but its potential effects on disc cells remains yet to be identified. Interestingly, Sp1 expression is inhibited by NF-κB (Tapias et al., 2008), which has been shown to be able to initiate a pro-inflammatory cascade in other tissues, often as a reaction to extracellular stimuli (Hunter and De Plaen, 2014). And althoughit is possible that this dual regulation confounds the potential catabolic effect of Sp1 deprivation with an inflammatory-like response, further studies are needed to distinguish the catabolic and inflammatory responses.

Another interesting gene, Aggrecan gene (ACAN), presents tandem repeat polymorphisms in the CS1 domain. Several studies have reported that lower repeat number can lead to lower chondroitin sulfate (CS), thus linking aggrecan with IVD degeneration (Kawaguchi et al., 1999; Solovieva et al., 2007; Kim et al., 2010; Gu et al., 2013). Low CS reduces the amount of water accumulated to withstand compression loadings, reducing disc mechanical properties. Further, it is possible that lower ACAN and CS reduce the IVD's capability to recover after acute catabolic processes (Kim et al., 2010).

Another protein, cartilage intermediate layer protein (CILP), whose expression is restricted to different cartilage(-like) tissues including IVD, inhibits TGFβ, therefore preventing the ECM anabolism and cell proliferation promoted by TGFβ in IVD (Liu et al., 2021). A polymorphism in the interaction region between CILP and TGFβ has been shown to change their binding affinity, consequently, identifying it as a risk factor for IVD degeneration (Seki et al., 2005; Min et al., 2010; Kelempisiotiet al., 2011). Additionally, CILP is able to inhibit Insulin-like growth factor-1 receptor (IGFR1), acting as an antagonist of Insulin-like growth factor 1 (IGF1), a factor that mediates chondrocyte anabolism and proliferation (Liu et al., 2021). Similarly, tandem repeat polymorphisms in Asporin gene inhibits TGFβ-induced anabolism with likely synergic effects with the CILP variant (Song et al., 2008; Min et al., 2010).

As stated before, pro-inflammatory cytokines are the key factors that start the catabolic shift through increase of matrix-degrading enzymes expression, with IL-1β being one of the most influential cytokines produced by the native IVD cells and immune cells following IVD rupture (Le Maitre et al., 2005; Millward-Sadler et al., 2009; Phillips et al., 2015). As mentioned earlier, combinations of specific IL-1 β and COL9A3 polymorphisms constitute a risk factor for IVD degeneration. Similarly, a combination of MMP3 and IL-1 β polymorphisms also presents a higher risk of IVD degeneration (Karppinen et al., 2008). Additionally, other MMP-3 (Takahashi et al., 2001; Valdes et al., 2005), *MMP-2* (Dong et al., 2007) and *MMP-9* (Sun et al., 2009) polymorphisms have shown greater risk. Likewise, differentsingle nucleotide polymorphisms (SNPs) of IL-1 β and their combinations present increased risk of IDD, possibly due to overactivation under mechanical stress (Solovieva et al., 2004; Virtanen et al., 2007; Karppinen et al., 2008). Other interleukins such as IL-6 (Noponen-Hietala et al., 2005; Kelempisioti et al., 2011) and IL-18 (Omair et al., 2013) have also been reported to increase catabolic processes in the IVD, similar to IL-1 β.

Matrix-degrading enzymes can also be regulated through their degradation. A SNP in the Thrombospondin-2 gene (THBS2), which regulates degradation of MMPs through endocytosis, has shown lower affinity for MMP binding which in turn reduces MMP degradation and increases therisk of IDD (Hirose et al., 2008). In similar fashion, different variants of diverse ADAMTS family members, which degrade aggrecan, were identified to increase risk of degeneration (Rajasekaran et al., 2014; Wu et al., 2014; Liu et al., 2016) and are increased during IVD degeneration (Pockert et al., 2009). A SNP in ADAMTS identified by Wu et al. (2014) decreases binding affinity with the protein that regulates its degradation by endocytosis, LRP1, therefore increasing its catabolic activity.

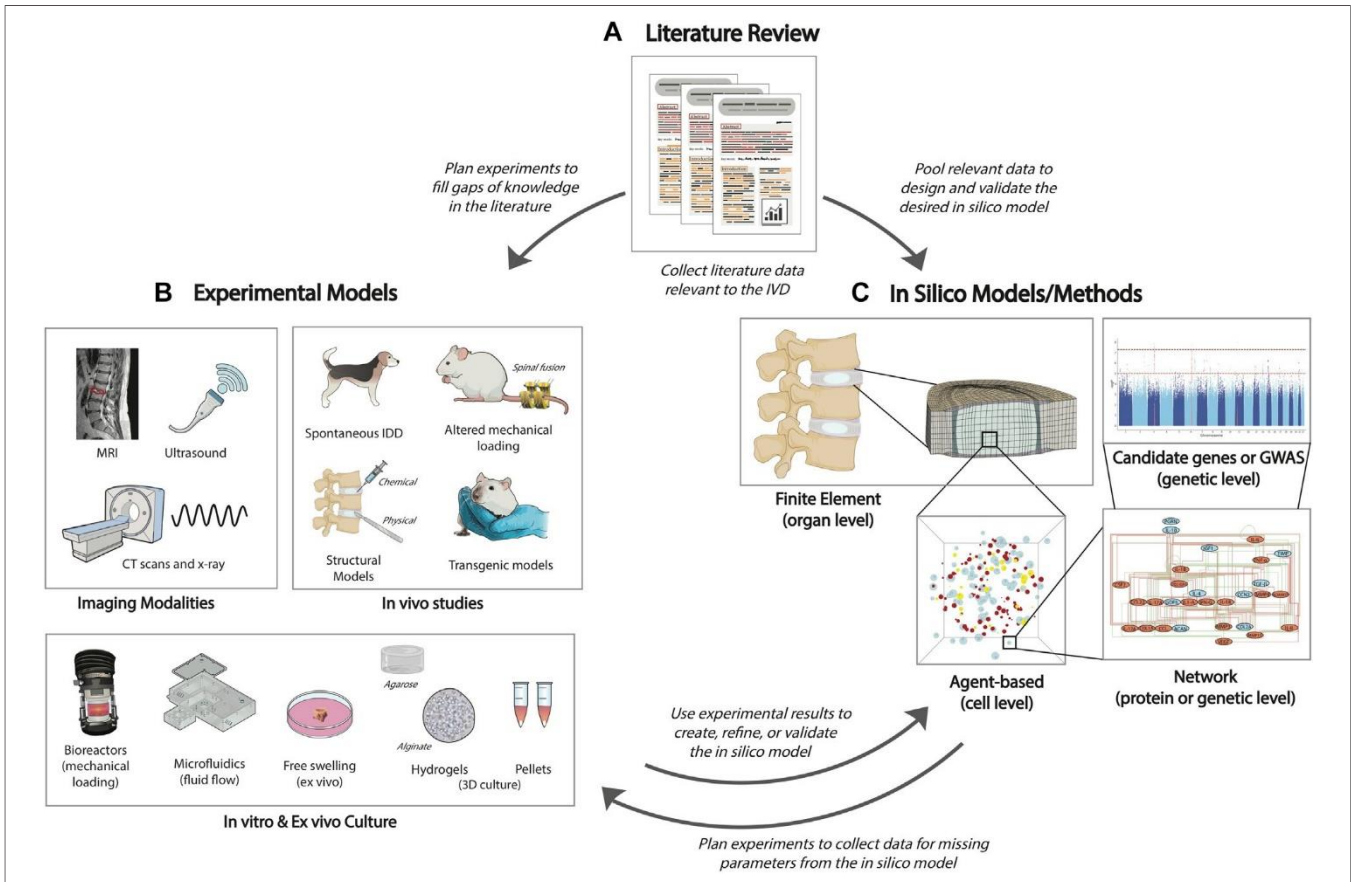


FIGURE 4 | Flow chart of the intersection of experiments and computational modeling. (A) First, a literature review is necessary to determine the current state of the research. Then, the researcher can either perform additional experiments to fill gaps of knowledge in the literature, or use published data to create an *in-silico* model. (B)There are many options for experimental model design, including use of imaging modalities to view the state of the IVD, *in-vivo* animal studies which better examine thecomplexity of IDD, bioreactors and microfluidic devices that allow the investigation of mechanical loading or fluid flow in the IVD, and *in-vitro/ex-vivo* culture of the whole IVD or IVD cells from human or animal tissue. (C) *In-silico* models or methodologies can use published literature or additional experiments to provide deeperinvestigations into complex tissue (FEM), cell (ABM), protein (network modeling), and genetic responses (network modeling, GWAS) as well as explore interactions atmultiple scales which would be difficult and expensive to do experimentally. These models can help identify novel parameters and interactions that should be validated orexplored further through experiments.

A shift toward catabolic processes can also be provokedby perturbing cell differentiation and viability. This has identified different variants of growth factors to be associated with IDD, such as Growth Differentiation Factor5 (GDF5) (Williams et al., 2011a), which is a key regulator of matrix synthesis in the disc (Le Maitre et al., 2009b); SKT gene (KIAA1217) (Karasugi et al., 2009; Kelempisioti et al., 2011; Al Dhaheri et al., 2020), vascular endothelial growth factor (VEGF), and endothelial nitric oxide synthase (eNOS) (Han et al., 2013). Variants in factors that modulate cell apoptosis, such as FAS receptor and its ligand (FASL) (Zhuet al., 2011), Caspase-9 (Sun et al., 2011; Mu et al., 2013),tumor necrosis factor related apoptosis-inducing ligand (TRAIL) (Du et al., 2015), and Death receptor 4 (DR4) (Tan et al., 2012) have been found to be associated with IDD. However, exact mechanisms of how such variants affect cell fate and IDD are still unclear and require further investigation.

In addition to the candidate gene approaches, where the set of genes tested is preselected, a few Genome-wide Association Studies (GWAS) have been performed. GWAS is an agnostic method that tests variants covering the "whole" genome (Duncan et al., 2019). A GWAS performed by FMK Williams et al. found a variant of Parkinson protein 2, E3 ubiquitin protein ligase (PARK2) and one of Proteasome 20S Subunit Beta 9 (PSMB9)to be associated with degenerative discs (Williams et al., 2013). Those genes encode for proteins that aim to tag and degrade unwanted proteins, providing another method to remove matrix- degrading enzymes, altering the metabolic balance. Another GWAS identified a variant of Carbohydrate Sulfo-Transferase
3 (CHST3), a catabolic enzyme that catalyses proteoglycan sulfation, as a susceptibility gene for IDD (Song et al., 2013). The authors suggest that this enzyme interacts with a micro-RNA (miRNA) that targets proteins with important regulatory functions in cell-mediated immune responses, but further analysis is needed to confirm such hypothesis.

Bacterial Infection and Disc Microbiomein IDD

Bacterial contamination has also been proposed as an important regulator of disc cell inflammation and catabolism, particularly in association to Modic changes (Gorth et al., 2015; Granville Smith et al., 2021). Despite the detection of various bacteria within isolated disc tissue, the presence of an IVD microbiome is still controversial as it has been traditionally considered as a sterile, immune privileged structure (Rajasekaran et al., 2020).

In Stirling et al. (2001), first reported the presence of anaerobic bacteria, particularly of *Cutibacterium acnes* (Gilchrist, 1900), previously known as *Propionibacterium acnes,* within the disc tissue of 43 out of 140 patients with *sciatica*. Recently, Granville Smith et al. (2021) performed a PRISMA systematic review identifying 36 articles from 34 research studies investigating bacteria in human IVDs. Bacteria were identified in 27 studies, whereas nine attributed bacterial presence to contamination. *C. acnes*, a Gram-positive anaerobic bacterium that is part of the natural skin microbiome, was the most abundant. Which is also associated with prosthetic joint infection (Jauregui et al., 2021) andwas shown to be able to interact with bone cells (Aubin et al., 2017)and recently disc cells (Capoor et al., 2021). Coagulase-negative (CoNS) bacteria of the genus *Staphylococcus* Rosenbach 1884 were the second most abundant (Granville Smith et al., 2021). Inconsistencies between the identified bacteria and the prevalenceof different bacteria can be partly administered to differences in tissue source (intact or herniated tissues), culture conditions (anaerobic vs. aerobic, culture time, culture media), differences in the methods used to detect bacteria, and differences in the administration of antibiotics. To date, there are few quantitative studies investigating bacterial infection to show whether bacteria are present *in vivo* or represent operative contamination.

Albert et al. (2013a) hypothesized that type 1 Modic changesin the adjacent vertebrae of herniated discs may be due to infection of the disc, highlighting the need to investigate bacteria presence in the disc. Treatment of chronic LBP andModic changes with antibiotics has generated great controversy.It has been shown that in a certain subset of patients, antibiotic treatment was effective to reduce pain as well as disability (Albert et al., 2013a; Gilligan et al., 2021). However, this result has failed to be replicated in subsequent studies (Bråten et al., 2020).

Furthermore, the role of potential bacteria within the discis unknown. However, previous studies have shown that LPS, a main component of Gram-negative bacteria, induces upregulation and production of various proinflammatory cytokines and matrix degrading enzymes in the NP over-activation of the NF-κB pathway (Li et al., 2016). However, most bacteria detected in the disc to date are Gram-positive and the potential influence of those bacteria on the disc remains poorly understood. Recently, Capoor et al. (2021) stimulated human IVD cells with *C. acnes* demonstrating an induction of catabolic cytokine expression by native NP cells, suggesting thatat least in some individuals the increased catabolic cytokines during disc degeneration could be triggered by bacterial infection. Further work is required to understand whether bacteria are present within the disc and whether bacteria could act as a trigger to the catabolic stimuli seen during disc degeneration, and whether the gut microbiome could influence disc degeneration(Li W. et al., 2022).

Translating Knowledge of Initiating Factorsof IDD to Clinics

Understanding the roles and interactions of each of these initiating factors is essential in order to diagnose IDD early and identify suitable treatments. Current treatments alleviate pain but do not regenerate the disc, therefore regenerative strategies are urgently needed in clinics (van Uden et al., 2017). Despite extensive research, tissue engineering strategies have had limited success in translating from preclinical models to beneficial treatments in patients as they fail to address pain (Isa et al., 2022). Precision medicine appears promising to use multiomics profiling to elucidate the pathology of IDD in each patient and prescribe an individualized treatment plan and determine which therapies could be most effective (Isa et al., 2022). However, a more comprehensive understanding of the disc microenvironment, biomechanics, genetics and epigenetics, and even bacterial infection throughout the different stages of IDD and low back pain would be necessary to evaluate which therapies or combination of therapies could be effective in a patient. In particular, there is a lack of studies on cartilage endplate regeneration therapies, despite its importance in the nutrient supply of the IVD (van Uden et al., 2017). Overall, experimental and computational models of IDD, which will be explored in the next chapter, remain critical toward developing novel treatments and regenerative therapies for IDD.

RESEARCH METHODS FOR EXPLORING THE INFLAMMATORY OR CATABOLIC ENVIRONMENT OF THE IVD

Experimental Models in IVD Research

For many years different experimental models have been developed for IVD research to mimic IDD. Several approaches have been used to replicate the physiological state of the IVD as closely as possible, including 3D cell and organ culture models, bioreactors and animal studies (Figure 4B). These approaches are crucial in elucidating the causes and progression of IDD, as wellas in developing and testing novel therapies. Nevertheless, the best strategy to investigate IDD remains unclear, and each culturesystem or animal model offers different advantages and disadvantages that should be considered when planning anexperiment (Table 2).

3D Cell Culture Systems

Over the last decades, three-dimensional (3D) cell culture models have been widely accepted due to the considerable improvements they possess in comparison to two-dimensional (2D) culture, including improved phenotypic retention to that seen *in vivo*, including: cell shape preservation; proliferation rates, and gene and protein phenotypic marker and matrix expression (Jensen and Teng, 2020). Conventional 2D monolayer culture systems

TABLE 2 | Experimental and computational approaches applied for IDD research.

	Model	Culturesystem or methods	Representative studies	Scale	Parametersthat can beprobed	Advantages	Disadvantages	Contributionto IDD
Experimental culture	3D cell	Alginate-based hydrogels	Le Maitre et al. (2005), Hernandez et al. (2020a), Naqvi and Buckley, (2014), Le Maitre et al. (2008), Zhang et al. (2014), Wang et al. (2014), Sun et al. (2015), Öztürk et al. (2016), Thorpe et al. (2018), Lazarus et al. (2021), Rosiak et al. (2021)	Cell	Dynamic loading and catabolism induction, Stimulation with cytokines and catabolic factors	Inexpensive, non-toxic and excellent cell phenotype maintenance, human cells are available	Not recommended for AF cell culture	IVD phenotype maintenance and recovery from catabolism
		Agarose carriers	Kelly et al. (2009), Smith et al. (2011)		Dynamic loading and catabolism induction	long-term 3D culture, human cells are available	Cannot replicate macroscale forces the IVD experiences	Catabolism induction in IVD cells
		Reinforced hydrogels (Silk)	Frauchiger et al. (2017), Wöltje and Böbel, (2017)		Porosity, coating and surface area of scaffold	Biodegradable and resorbable biomaterial with high cytocompatibility	The outer layer can cause immune response	Novel therapeutic approach in IVD repair
		Pellet culture systems	Thorpe et al. (2018), Wangler et al. (2019), Hingert et al. (2019)		Hydrostatic loading and nutrient perturbation	Simple, inexpensive, human cells are available	Chondrogenic phenotype induction and lack of ECM	Effects of hydrostatic loading and nutrient deprivation in IDD
Ex vivo		Bioreactors for mechanical loading	Gantenbein et al. (2006), Vergroesen et al., 2014; Costi et al., 2008, Walter et al. (2011), Paul et al. (2012), Chan et al. (2013), Paul et al. (2013), Salzer et al. (2021), Croft et al. (2021)	Organ/ Tissue	Mechanical loading (compression, torsion, bending, flexion, extension, and asymmetric), environmental control (nutrition, pH, temperature, oxygen level), frequency and duration (static, dynamic, or diurnal)	Mimics physiological conditions, possibility for automation, reproducible, in line with 3R principles ("Replacement, Reduction, and refinement")	Expensive and difficult to build, culture time limited to ∼1 month, no connection to vasculature or immune system, most cannot test large sample sizes at once, human tissue is limited	Determined physiological and catabolic ranges of mechanical loading regimens and test the mechanical properties of suitable biomaterials for IVD replacement
		Perfusion bioreactors, microfluidics, "disc-on-a-chip"	Elfervig et al. (2001b), Chou et al. (2016), de Bournonville et al. (2019), Dai et al. (2019), Hwang et al. (2020), Kim et al. (2021), Mainardi et al. (2021)		Diffusion, shear stress, fluidic pattern, electrical impulses, environmental control (nutrition, pH, temperature, oxygen level)	Possibility for automation, reproducible, extends culture time, in line with 3R principles ("Replacement, Reduction, and refinement")	Cannot replicate macroscale forces the IVD experiences, difficult to design complex systems	Increased cell viability in culture provided platform to investigate cellular response to shear stress and interactions with inflammatory and neurotrophic factors, and test treatments such as electrical stimulation
In vivo		Spontaneous degeneration	Daly et al. (2016)	Whole body/ Organ/ tissue/ Protein/ Genetic	Degeneration progression, therapies/treatments	Occurs naturally (therefore more ethical), immune system and pain response	Long and unpredictable time course, inherent biological and biochemical differences to humans, expensive and complex, ethical considerations	Chondrodystrophic dogs and sand rats have similar pathological changes to human in IDD and have useful in testing cellular therapies and other clinical treatments

(Continued on following page)

TABLE 2 | (*Continued*) Experimental and computational approaches applied for IDD research.

	Model	Culturesystem or methods	Representative studies	Scale	Parametersthat can beprobed	Advantages	Disadvantages	Contributionto IDD
		Altered mechanical loading	Lindblom (1957), Iatridis et al. (1999), Phillips et al. (2002), Ching et al. (2003)		Magnitude, duration, and frequency of loading	Immune system and pain response, repeatable induction of IDD at specific time point	Inherent biological and biochemical differences to humans, expensive and complex, ethical considerations	Models of bending, compression, and spinal fusion have shown that loading changes the mechanical properties of the IVD
		Structural models (physical injury or chemical injection)	Lü et al. (1997), Holm et al. (2004), Ulrich et al. (2007), Yoon et al. (2008), Dudli et al. (2011), Dudli et al. (2014), Alkhatib et al. (2014), Zhang et al. (2020), Zhou et al. (2021)		Degeneration progression, proteoglycan degradation, therapies/treatments	Immune system and pain response, repeatable induction of IDD at specific time point, useful in preclinical trials	Inherent biological and biochemical differences to humans, expensive and complex, ethical considerations, fail to capture pathogenesis of human IDD, viability of native IVD cells preserved	Critical in understanding IDD and developing/ testing novel therapies for clinical application
		Transgenic models	Millecamps et al. (2011), Millecamps et al., 2012;, Li et al. (2009), Wang et al. (2012), phillips et al., 2013, Phillips et al., (2013a,b)Miyagi et al. (2014), Gorth et al. (2018)		Genes and gene pathways	Immune system and pain response, target specific pathways of interest	Inherent biological and biochemical differences to humans	Changes in SPARC, Tg197, CCN2, IL-1rn, cAct, and SMAD3 genes have been identified to contribute to IDD
Computational	Finite element		Jones and Wilcox, (2008), Newell et al. (2017), Ghezelbash et al. (2020), Galbusera et al. (2011b), Malandrino et al. (2011), Volz et al. (2022)	Organ/ Tissue	Loading, environmental perturbations and catabolism induction	Valuable prediction of altered mechanics and transport at the tissue level of the IVD, 3Rs	Challenging comprehensive validation and the cellular and sub-cellular level is not contemplated. Computationally intensive and time consuming	Predictions about metabolic rates, oxygen and lactate transport, osmotic behaviour
	Agent-based	2D or 3D	Baumgartner et al. (2021a), Baumgartner et al. (2021b)	Tissue/ Cellular	Cell behavior and interactions, microenvironment, time, Cell types, environmental perturbations	Dynamic, ability to model heterogenous populations, flexible, stochastic, reveals emergent phenomena	Can be computationally intensive, only as good as the rules inputted	Visual predictions of NP cells expressing TNF-α, IL-1β, or both TNF-α & IL-1β
	Network	Knowledge based or data driven	Shannon et al. (2003), Hu et al. (2004), Batagelj and Andrej, (2004), Kashtan et al. (2004), Yu et al. (2004), Schreiber and Schwöbbermeyer (2005), Wernicke and Rasche (2006), Milenković et al. (2008),	Protein/ Genetic	Protein-protein interaction (PPI) and transcriptomic/ proteomic analysis	Intuitive way to investigate, characterize, and understand interactions between biological components under micro-environmental stimuli	Difficult to construct from available IVD data which has high sample variation, different stages of IVD and type of disc tissue, and the varying methods of analysis	Capture interactions between transcriptome, proteins and their pathways in to understand the critical biochemical factors in IVD regulation. Reveal complex dynamics behind unbalanced metabolism

(Continued on following page)

TABLE 2 | (*Continued*) Experimental and computational approaches applied for IDD research.

Model	Culture system or methods	Representative studies	Scale	Parametersthat can beprobed	Advantages	Disadvantages	Contributionto IDD
		Baumgartner et al. (2021a), Baumgartner et al., 2021b, Xu et al. (2021)					
Genetic analysis	Candidate gene studies, GWAS	Williams et al. (2013), Song et al. (2013), Feng et al. (2016), Kawaguchi (2018), Mayer et al. (2013)	Genetic	Genes and gene pathways	Effective in identifying genes implicated in IDD	Incapable of explaining high heritability in complex diseases due to high polygenicity and unmet "common disease, common variants" hypothesis, and due to other heritable properties as epigenetics	Identified large number of risk genomic loci involved in IDD (Table 1)
Machine learning/ AI/Deep learning	Classification of discs	Rim (2016), Niemeyer et al. (2021)	Whole body/ Organ/ Tissue/ Protein/ Genetic	Imaging, clinical categories, compound structures, gene sequence, protein/ RNA data	Link seemingly unrelated entities of complex/ diverse biological data	Need for algorithm creation and learning. Very subjective score system	Deep learning model for the classification of discs based on MRI with an average sensitivity of 90%
	Simplifying or coupling complex models	Pfaff et al. (2020)			Highly accurate surrogate models, significantly less computational resources and less time-consuming		

lack the spatial architecture of the tissue, inducing a loss of cell phenotype and cell-ECM interactions. In contrast, 3D cultures environments promote extracellular matrix (ECM) deposition, a key factor for the maintenance of NP cell phenotype (Guerreroet al., 2020). Likewise, previous studies using notochordal cells have reported the negative effects of 2D culture (Rastogi et al.,2009) and the necessity of 3D culture system, preferably in hypoxia, and raised osmolality to maintain the phenotype (Gantenbein et al., 2014).

In the last decade many different biomaterials have been used in 3D cell culture of IVD cells, which will be listed according tothe best outcome. Alginate-based hydrogels are commonly usedbecause they are inexpensive, non-toxic and an easy 3D hydrogel model whilst maintaining excellent cell phenotype (Hernandez et al., 2020a). Notably, previous studies have reported IVD phenotype maintenance and recovery from catabolism after3D alginate culture (Le Maitre et al., 2005; Naqvi and Buckley, 2014; Wang et al., 2014; Zhang et al., 2014; Sun et al., 2015). Additionally, alginate has been used with dynamic loading systems (Le Maitre et al., 2008) and for inducing catabolism(Le Maitre et al., 2005; Le Maitre et al., 2008). Moreover, modifiedalginates have shown novel properties and applications in biomedical research (Rosiak et al., 2021). For example, the sulfation of alginate hydrogel has been reported to preserve the phenotype of chondrocytes (Öztürk et al., 2016; Lazaruset al., 2021). Thus, alginate-based hydrogel systems are considered as promising biological constructs for NP cell culture (Thorpe et al., 2018). In contrast, the consistency of alginate-based hydrogels is not recommended for AF cell culture due to the lack of fibrotic structure. In addition, other materials such as agarose have been reported as long-term 3D culture models for inducing catabolism in IVD cells (Smith et al., 2011) and chondrocytes (Kelly et al., 2009). In contrast, other 3D hydrogel models such as fibrin-clots are not a realistic option forIVD 3D culture due to their lack of stiffness, despite the easy modification of this natural hydrogel. However, reinforced hydrogels with resorbable biomaterials, for instance silk, are a novel therapeutic approach in IVD repair due to high cytocompatibility (Frauchiger et al., 2017). Notably, the outer layer of silk filaments can cause immune response due to the presence of sericin (Wöltje and Böbel, 2017). 3D environment features can also be achieved without biomaterials through pellet culture. Notably, pellet culture systems have been utilized to investigate IVD degeneration, specifically examining the effects of

hydrostatic loading and nutrient deprivation (Hingert et al., 2019; Wangler et al., 2019), however they induce a more chondrogenic phenotype rather than NP phenotype (Thorpe et al., 2018). Furthermore, the lack of ECM after the formation of the pellets could influence the response of hydrostatic loading (Zeiter et al., 2009), and they fail to mimic the cell-ECM connections and cell density seen in the IVD, therefore other materials would be better utilised.

In contrast, 3D constructs aim to offer a more physiological interaction between cells with ECM components without the presence of vasculature and innervation as well as the interaction with the immune system. Nevertheless, although biomaterial- based therapies have been developed in the last decades to preventIDD, only a small number of bioengineered therapies are currently undergoing clinical trials (NCT02338271, NCT01290367, NCT01290367 and NCT02412735). Notably,

pain relief does not correlate adequately with functional and structural IVD restoration. Overall, the main clinical challenge isto use clinical signs, patient pain, and disability history alongside advanced imaging techniques to design a sufficient biomaterial approach (Huang et al., 2018; Isa et al., 2022). However, *ex vivo* culture systems seem to be an appealing alternative resulting in a more representative model. Particularly, IVD explants gained attention due to a higher control of the degeneration state, sample geometry and loading (Salzer et al., 2021). Similarly, organ culture bioreactor models allow the presence of the native tissue microenvironment together with a loading system promoting a bridge between *in-vitro* and *in-vivo* models (Gantenbein et al., 2015). Such *ex vivo* culture systems are good model systems investigating intact IVDs. However, they do not enable connection with the vasculature and immune system, although co-culture systems could be developed to model these interactions.

Bioreactors and Microfluidic Devices

Bioreactors are widely accepted as pre-clinically relevant devices that simulate the microenvironment and offer a platform to evaluate the effects of limited nutrition and incorporate more complex parameters, such as mechanical loading and fluid flow, into *in vitro and ex vivo* experiments (Haglund et al., 2011; Illien-Jünger et al., 2014; Walter et al., 2014; Gantenbein et al., 2015; Gantenbein et al., 2019; Pfannkuche et al., 2020).

Various materials or organ culture are used for different bioreactor systems. Freshly isolated IVDs from bovine tails are often used in organ culture studies due to well established operating procedures, as well as their similarities to human discs (Chan and Gantenbein-Ritter, 2012; Saravi et al., 2021). Bioreactors are also used in dynamic 3D cell culture, however the material used must be able to withstand the imposed forces. Consequently, cells are often seeded into hydrogels, such as agarose or alginate, to offer more protection againstmechanical loading (Fernando et al., 2011; Cambria et al., 2020). Initial approaches were only capable of static loading; however, over time, bioreactors have gradually evolved tobecome more complex, incorporating diurnal loading (Gantenbein et al., 2006) and dynamic compression (Paulet al., 2012; Paul et al., 2013; Vergroesen et al., 2014). More recently, bioreactors have advanced past compression to include two- and six-dimensional degrees-of-freedom, allowing for the analysis of torsion, bending, flexion, and extension (Costi et al., 2008; Chan et al., 2013; Croft et al., 2021). Additionally, asymmetrical complex loading has been proposed as a modelto study the effects of scoliosis on disc mechanobiology (Walter et al., 2011). These increasingly complex loading devices are crucial to better understanding the effects of mechanical loading on IDD and how aberrant mechanical loading contributes to the shift to catabolism. Further, these devices are highly clinically relevant as they can test the mechanical viability of novel regenerative therapies that aim to replace or regenerate the NP, AF, and/or CEP.

In addition to mechanical loading, bioreactors are useful in simulating fluid flow to the IVD. Perfusion bioreactors have been developed to allow for *in vitro* perfusion culture of scaffold-based tissue engineering constructs, offering the ability to monitor and control key parameters such as temperature, pH, and fluidic pattern (de Bournonville et al., 2019). Microfluidic, or "organ-on- a-chip", platforms have also been explored to study IDD and havebeen reviewed recently by Mainardi et al. (2021). In 2019, one ofthe first microfluidic "disc-on-a-chip" devices was developed by Dai et al. (2019), permitting continuous media flow to mimic the disc microenvironment, and demonstrating higher cell viability than cells in static culture, allowing for the possibility of long- term organ culture to examine chronic disc degeneration. Microfluidic devices have also been used to investigate mechanical loading in AF cells through fluid-induced shear

stress (Chou et al., 2016). Studies have found that AF cells hada greater response to shear stress when stimulated with IL-1β, suggesting that disc cells are more sensitive to shear during catabolic or inflammatory conditions, possibly affecting IDD development (Elfervig et al., 2001b). More recently, electrical

stimulation was tested as treatment to modulate IL-1β-mediated catabolism in NP cells (Kim et al., 2021). Additionally, a microfluidic platform was used in a co-culture system of AF, NP, and endothelial cells to investigate IDD development from inflammatory and neurotrophic factors, which could be further developed to examine pain mechanisms in IDD (Hwang et al., 2020; Mainardi et al., 2021). The ability to evaluate pain in culture systems of IVD is currently lacking, which is a major issue in translating tissue engineering strategies successfully to clinics (Isa et al., 2022). Therefore, development of a microfluidic platformthat could do this would be an immense step forward toward evaluating new therapies and treatments *in vitro.* Further, microfluidic devices have been proven valuable in testing drug delivery and improving screening strategies (Damiati et al., 2018). Although these systems have been around for less time and are therefore less validated than bioreactors that offer mechanical loading, perfusion and microfluidic systems offer a promising platform to probe inflammatory and catabolic parameters and test new treatments *in vitro*.

In Vivo Animal Models of Disc Degeneration

While *in vitro* and *ex vivo* models of the IVD are highly beneficial and provide insights on components of IDD, they do not offer the same level of complexity as *in vivo* studies,

which may better examine the multifactorial nature of IDD and can include an immune system and pain response (Daly et al., 2016). Many animal models have been used to investigate the IVD, and the advantages and disadvantages have been reviewed extensively before (Lotz 2004; Alini et al., 2008; Showalter et al., 2012; Daly et al., 2016; Jin et al., 2018). However, no perfect model of disc degeneration currently exists, because there are many biological and biochemical differences between discs from animal species and those from humans (Oshimaet al., 1993).

One major difference between human discs and animal discs isthe presence of notochordal cells in the NP. In humans, notochordal cells are present at birth, but rapidly decrease and are gone by adulthood. In most other species, notochordal cellsare present in adulthood. However, cows and sheep and some species of dog, classified as "chondrodystrophic", lose their notochordal cells rapidly, similarly to humans. Although notochordal cells are not well understood, they are often considered as progenitor cells, and therefore their presence in degenerative animal models may lead to results that are minimally relevant to understanding human LDD (Alini et al., 2008). Other biochemical parameters should also be considered, such as water, GAG, and collagen content, as well as how thesefactors change with degeneration and age in animals versus humans, which has been reviewed previously (O'Connell et al.,2007; Beckstein et al., 2008; Miyazaki et al., 2009; Showalter et al., 2012). It should also be noted that rodents have distinctly different aggrecan proteins and do not express the same MMPs as humans, which is considered important in catabolism and tissue remodeling (Barry et al., 1994; Flanneryet al., 1998). However, while rodent models may not be suitablefor translational research and testing new therapies because of themajor differences to humans, they offer a useful platform to elucidate the genetic basis of IDD and catabolic changes due toaging (Masuda and Lotz, 2010; Mainardi et al., 2021). Similarly,larger animal models are not a perfect match toward human IDD (Alini et al., 2008; Gullbrand et al., 2016). Nevertheless, they are furthermore suitable for initial tests of regenerative therapies asthey offer valuable information on the changes in mechanical loading, whether an immune response is initiated, and possibly whether any pain is resolved.

Spontaneous degeneration occurs in mice, sand rats, chondrodystrophic dogs, and baboons; however, these models are unpredictable and often time-consuming (Daly et al., 2016). Therefore, there are various methods, categorized under mechanical or structural, that have been used to induce degeneration in animals.

In rat tails and rabbits, degeneration has been induced through altered mechanical loading, such as bending (Lindblom 1957), compression (Iatridis et al., 1999), or spinal fusion (Phillips et al., 2002; Oswald et al., 2021). In compression, the magnitude, duration, and frequency of loading cause significant changesin IVD mechanical properties, and static loading producesgreater changes than cyclic loading (Ching et al., 2003).

Structural models involve a physical injury or chemical injection to the CEP, AF, or NP (Lotz 2004). Physical injuries are done using either a drill bit, scalpel, or needle. Annular injuries are commonly used and have been shown to cause decreased disc height, higher Pfirrman degeneration scores,and decreased NP GAG content (Yoon et al., 2008). Research has also shown that repetitive injury causes different inflammatory responses in the IVD. Ulrich et al. (2007) found that while a single stab injury in a rat led to localized, short-term pro-inflammatory response, while multiple stab injuries cause a prolonged upregulation of proinflammatory cytokines TNF-α, IL1-b, and IL-8 for up to 28 days after injury. CEP injuries have also been demonstrated to lead to disc degeneration similar to that of humans, characterized by decreased NP proteoglycan content and intradiscal pressure (Holm et al., 2004; Dudli et al., 2014; Zhou et al., 2021), as well as increased catabolic enzyme production and pro-inflammatory gene expression seen following CEP fracture (Dudli et al., 2011; Alkhatib et al., 2014). However injurious degeneration models fail to recapitulate the pathogenesis of human IDD and enable infiltration of inflammatory cells at a much earlier time frame than seen if at all in humans. Chemical injections with papain and chondroitinase ABC or papain are commonly used methods to induce degeneration through degrading proteoglycans in the disc (Daly et al., 2016). Although both cause catabolism, chymopapain was shown to cause greater destruction of the NP and AF proteoglycans, as well as greater spinal instability and disc space narrowing (Lü et al., 1997). However, chondroitinase ABC induced a similar catabolic shift to that seen in human IDD in the IVD of goats (Zhang et al., 2020).

In addition, groups have used transgenic animal models to represent IDD (Jin et al., 2018). The SPARC (secrete protein, acidic, rich in cysteine)-null transgenic mouse has been shown to develop behavioral signs consistent with chronic low back pain due to IDD, such as hypersensitivity to cold, axial discomfort, and motor impairment (Millecamps et al., 2011; Millecamps et al., 2012). Further, the SPARC-null mouse showed age-dependent increased innervation by sensory nerve fibers near the IVD (Miyagi et al., 2014). Gorth et al. (2018) also used Tg197 mice, a TNF-α transgenic line, to investigate the effects of systemic over-expression of TNF-α on IDD, finding that the experienced an increase in annulus tears and herniation with higher vascularity and immune cell infiltration. However, they foundthat intact IVDs remained healthy despite the elevated inflammation. Additionally, knockout technology has been used to create models of notochord-specific CCN2-null mice (Bedore et al., 2013), IL-1 receptor antagonist knockout mice (Phillips et al., 2013a), and β-catenin conditional activation (cAct) mice to examine the signaling pathway roles in disc degeneration (Wang et al., 2012).

Finally, emerging strategies such as tissue-engineered replacement discs have gained substantial attention in the IVD regeneration field. In terms of animal models, significant technical challenges must be addressed including cell source, construct size, culture strategies, and translational models (Gullbrand et al., 2018). Nevertheless, several studies in disc replacement, including prospective randomized comparative trials, have demonstrate advantages such as short-term superiority to spinal fusion (Hellum et al., 2012; Vital andBoissière, 2014) or at least non-inferiority to anterior spinal

interbody fusion (Blumenthal et al., 2005; McAfee et al., 2005; Vital and Boissière, 2014).

Computational Modeling of the IVD

While experimental studies are valuable in determining cell sensitivity to biochemical and mechanical cues, they are not sufficient to capture the full complexity of cell response and interactions with the microenvironment, which is crucial for understanding the transition to catabolism and the initiationof an immune response. In addition, experiments are often expensive and time-consuming. *In-silico* models can use published literature and experimental data to predict multifactorial tissue and ECM regulation at multiple scales(Figure 4). Further, *in silico* models can offer the possibility ofexploring patient-specific IDD and predicting the effects and risksof available therapies prior to being treated (Rijsbergen et al., 2018). Finite element models (FEM) are useful in determining the effect of mechanical loading at the tissue and organ level, whileagent-based models (ABMs) are effective in predicting tissue and cellular level changes due to the microenvironment. At the subcellular level, network modelling provides further insight into the effects of cell signaling pathways and gene variants. Machine learning and deep learning are valuable tools to analyzeand classify clinical images and predict the current and future status of a patient. Each of these *in-silico* tools offers a novel way to explore catabolism or inflammation in IDD, which will be further explained in the following paragraphs.

Finite Element Models in IDD

Finite element models (FEM) have been extensively used to represent the intervertebral disc and to simulate structural changes due to mechanical loading, providing a deeper understanding of each component's role than what can be tested through experiments. The IVD is inhomogeneous, anisotropic, and porous, making it a highly complex structure (Newell et al., 2017). Material properties for each of IVD component, the NP, AF and CEP, are defined and validated through experimental measurements and clinical observations, however comprehensive validation of FE analysis is challenging due to the complex structure and interactions (Ghezelbash et al., 2020). The IVD components are generally modeled as a biphasic material, with an incompressible fluid phase and an elastic solid phase. Other reviews have already been written regarding FEM studies of the IVD (Jones and Wilcox, 2008; Newell et al., 2017; Ghezelbash et al., 2020), so here we will focus on how FEM has been applied to study degeneration and catabolism.

FEMs are a valuable tool to examine the vicious cycle of disc degeneration and aberrant loading. Once a disc is degenerative, the tissue biomechanics are altered, leading to a catabolic environment and causing further damage over time. However, degeneration is not uniform throughout subjects, and there are various ways previous groups have simulated degenerative discs. These include geometrical changes such as decreased height and reduced NP area (Ghezelbash et al., 2020), but also changes in the material properties, including reduced water content, calcified and thinner CEP, and a stiffer NP characterized by a decreased bulk modulus (Galbusera et al., 2011a). These studies predicted that a degenerative IVD experiences higher forces during axial rotation, as well as lower fluid flow and recovery of intradiscal pressure after loading. Models have also shown that as the NP loses fluid, it carries less load under compression, as well as with bending and shear (Ghezelbash et al., 2020). Investigation of the geometry of the IVD has concluded that simplified geometry is less stiff and does not capture the same strain distribution as FEMs based on more complex geometry obtained through segmentation of MRIs, conveying those accurate geometries are essential (Du et al., 2021). However, there is a lack of studies that measure the effects of different patient-specific morphologies, either to observe the mechanical effects of deformation or their implications on nutrient transport. Recently, a coupled and patient-specific mechanoregulated model was developed to predict the effects of spinal fusion on disc degeneration and bone density, demonstrating how FEMs can be used by surgeons to provide insight into which patients could possibly benefit from spinal fusion treatments (Rijsbergen et al., 2018). Similarly, future models could aim to use available clinical data to help develop models that aid doctors in predicting which treatments and surgical interventions would have the best outcome.

Many FEMs of the IVD also simulate osmotic behavior, and ithas been shown that a swelling model with strain-dependent osmotic pressure most accurately represents the IVD, and couldbe applied to investigate crack opening and fissure propagation(Galbusera et al., 2011b). A mechano-transport FEM of the IVD developed by Ruiz Wills et al. (2018) found that CEP permeabilityincreases with aging and degeneration, and that CEP degeneration could be a cause of NP dehydration and play a key role in IDD. Other groups have simulated cell metabolism and nutrient levels in the IVD, predicting that higher cell metabolic rates lead to nutrient depletion and that application of mechanical loading led to decreased glucose levels throughout the IVD (Volz et al., 2022). Additionally, simulations of compression on oxygen and lactate transport within the IVD suggested that degenerative changes including disc height, fluid content, nucleus pressure, and cell density reductions significantly affected transport (Malandrino et al., 2011).

Overall, FEMs have proven valuable in predicting the effects of altered mechanics and transport due to degeneration at the tissue level of the IVD. However, FEMs fail to take into account what is happening at the cellular and sub-cellular level.

Agent-Based Models

Agent-based models (ABMs) are widely used across different spatial scales and research areas. Hence, agents might reflect human beings for socioeconomic studies (Alvarez-Galvez and Suarez-Lledo, 2019) or (sub-)cellular entities in cancer research (Metzcar et al., 2019). They are particularly useful for studying complex biological processes, such as inflammation and tissue degeneration, that are dynamic, spatially heterogeneous, and stochastic. ABMs can represent individual biological cells as computational agents and can simulate how collections of cells within a tissue will respond emergently to literature-derived rules. Previously, ABMs have been shown as valuable in simulating tissue degeneration and inflammation in musculoskeletal and

cardiac tissue, spanning many cell types including immune cells, fibroblasts, and stem cells (Virgilio et al., 2018; Rikard et al., 2019). Thus, ABMs offer much potential to simulate cell dynamics in cartilage tissue (Pearce et al., 2020).

In IVD research, however, ABMs have been used only recently, though initial studies have demonstrated their value in predicting IVD cell responses in a pro-inflammatory environment. Baumgartner et al. (2021c) coupled an ABM with mathematical network models (see Section "*Network Modeling*") to investigate the relative mRNA expression of proteins and proteases in NP cellsof different pro-inflammatory cell states, i.e., immunopositive for TNF-α, IL-1β, both, or none (Baumgartner et al., 2021a; Baumgartner et al., 2021b). The ABM was used to visualize cell states within the NP through predicting how immunopositive cells could be arranged within a 3D environment. Thus, it was assumed that immunopositive cells were organized in clusters based on experimental data of autocrine and paracrine stimulation (Phillipset al., 2013b; Phillips et al., 2015), short half-lives of cytokines according to distantly related studies (Kudo et al., 1990; Oliver et al., 1993; Larson et al., 2006) and diffusion of pro-inflammatory cytokines. While validation is still limited due to the lack of experimental information on the arrangement of immunopositive cells, the ABM presented a novel projection ofhow those cells could be spatially distributed within the NP. In thisregard, ABMs can be useful in identifying novel parameters and interactions implicated in IDD and therefore guiding future experiments. Future work could extend the ABM to simulate AF and CEP cells in addition to NP cells, or to simulateadvanced stages of IDD by including cell migration and adding in immune cells to further investigate the intercellular interactions in the IVD during an immune response.

Network Modeling

Modeling biological networks provides a holistic and intuitiveway to investigate, characterize, and understand the complex interactions between biological components. It is a static diagram represented by nodes (molecules) connected by lines (physical or functional interactions between nodes). The nodes are the stimuli or responses of the network, while the lines indicate either inhibition or activation between nodes directly or indirectly through other signaling pathways. The most frequently studied networks are protein-protein interaction (PPI) networks and the most commonly used software are: Cytoscape (Shannon et al., 2003), VisANT (Hu et al., 2004), TopNet (Yu et al., 2004), MAVisto (Schreiber and Schwöbbermeyer, 2005), FANMOD (Wernicke and Rasche, 2006), Pajek (Batagelj and Andrej, 2004), Mfinder (Kashtan et al., 2004), and GraphCrunch (Milenković et al., 2008).

Biological network modeling usually relies on "bottom-up" approaches, where intracellular interactions are simulated to estimate a final cellular response. Two methods can be used to build a network, either by gathering literature information regarding the pathways and the mechanisms that take part in the IVD degeneration (knowledge-based), or directly from experimental data (data-driven). In IVD degeneration, network modeling tries to capture the interactions between complex sets of proteins and their pathways and reveal the complex dynamics behind the imbalance between anabolic and catabolic processes. Identification of known NP cell high level cell regulatory factors is very important in order to integrate all the single stimuli into an IVD regulatory network model (RNM) for cell regulation, through which hypothesis and testing can be explored.

Regulatory network models can highlight the molecular signatures of the underlying pathological mechanisms that drive a condition. The reductionist view, one gene to onedisease, is not applicable to a highly multifactorial condition such as IDD, thus cell signaling pathway analysis is of high importance in order to understand the system as a whole. A mechanistic understanding of the condition could pave the way for a mechanism-based biomarker selection for the effective andpersonalized treatment of IDD (Baumgartner et al., 2021b).

A common data-driven approach for modelling regulatory networks starts with the acquisition of -omics data from a web-based repository or by generating them. Differentially expressed genes or proteins between healthy and diseased samples are found and a functional enrichment analysis is performed in order to identify the most statistically significant pathways that are presentin the condition. The final step includes experimental verificationof the targets that were identified *in-silico*. However, constructing regulatory networks from IDD samples is a challenging task dueto high sample variation, stage of IVD, type of disc tissue investigated and the chosen method of analysis which couldbe MassSpec, Microarrays or Next Generation RNA sequencing.Xu et al. (2021) created a regulatory network behind IDD by combining transcriptomic and proteomic analysis. Theyhypothesized that post-transcriptional regulation could havean effect on protein content, thus, if a gene presents elevatedmRNA and protein levels, it could be implicated in IDD. Theirresults identified six genes with these characteristics (CHI3L1,KRT19, COL6A2, DPT, TNFAIP6 and COL11A2), two of which were identified as important IDD markers in independent studies.

Another group used transcriptomic data collected from lumbar-degenerated IVDs to build gene regulatory networks, finding differentially expressed genes in chemotactic signaling and matrix-degrading pathways that could later be used to help develop novel pharmacological approaches for IDD treatment (Zamanian et al., 2022). Li H. et al. (2022) constructed a protein-interaction network as well as a disease-gene interaction network that identified two potential therapeutic drugs, entrectinib and larotrectinib, demonstrating how emerging network models canbe leveraged to identify novel treatments.

Recently, a top-down network modeling approach was presented to approximate cell responses of NP cells, where the cell is considered as a "black-box" (Baumgartner et al., 2021a; Baumgartner et al., 2021b). Approximations of cell responses were obtained by directly linking key relevant micro-environmental stimuli with cell responses of interest. Therefore, experimentally obtained data was systematically translated into parameters suitable for systems biology approaches. With this novel approach, interrelated results between NP cells of different pro-inflammatory states, i.e., immunopositive for TNF-α, IL-1β, or both; TNF-α and IL-1β, could be obtained for user-defined stimulus environments. This high-level network modeling methodology

was embedded within an ABM (see Section "*Agent-Based Models*") to visualize a proinflammatory environment and estimate the percentage of cells immunopositive for more than one proinflammatory cytokine, specifically TNF-α and IL-1β.
Considering crucial nutritional and biochemical stimuli, *in-silico* results suggest that pro-inflammatory cytokines are important contributors in catabolic shifts in NP cell responses (Baumgartneret al., 2021b).
Top-down approaches appear promising to tackle highly complex multicellular multifactorial environments, as found in IVD tissues. Amongst others, focus might be set on the integration of more critical stimuli and cell responses in the network model.

METHODS FOR GENETIC ANALYSIS IN IDD

In the past decades, candidate genes and Genome-wide Association Studies (GWAS) have been implemented in the discovery of genetics underpinnings of complex disorders. The former strategy involves testing the association between a particular gene variant and a trait. Therefore, the selection of the studied gene is led by *a priori* knowledge of the biological pathways that are involved in the etiology of the disease. However, the high specificity of candidate gene approaches does not reflect the polygenicity in which multiple genomic loci are involved in the development of the disease (Taboret al., 2002). On the other hand, the aim of GWAS is to investigate relationships between genetic variants and traits spanning the whole genome in order to give an unbiased and comprehensive view on the allelic architecture underlying complex traits (Goddardet al., 2016). Despite GWAS having identified a large number ofrisk genomic loci and provided valuable outcomes on the agnosticgenetic discovery for complex diseases such as IDD (Uffelmannet al., 2021), there are still gaps that have to be filled. GWAS are notcapable of explaining the high heritability of complex diseases. Several reasons could contribute to this limitation, such asepigenetics or epistasis, which is the phenomenon for which the effect of a genetic variant is dependent on the presence of other variants, known as genetic background. Another hypothesis is the "common variants common disease", which considers that the genetic contribution to a disease would come from an elevatednumber of SNPs, each one with a very small contribution difficultto identify (Manolio et al., 2009; Gibson 2010). Moreover, the interpretation of how a specific variant affects the downstream biological pathways is very challenging. First, due to linkage disequilibrium (LD), the phenomenon for which variants closeto each other are inherited together, associated SNPs are often correlated with other neighboring variants. Thus, checking LD is essential to include all potential causal variants. Further, the majority of associations discovered through GWAS (90%) fall innon-coding regions, hindering the interpretation of how such SNPs affect the phenotype (Cano-Gamez and Trynka, 2020).
For all of these reasons, when interpreting GWAS results, one should consider several factors, including the number of different associations that exist at a given locus and their LD correlation. Then it is possible to pinpoint which could be the causal variant and establish the mechanistic effects on the downstream processes (Cannon and Mohlke, 2018). Many approaches have been carried out to tackle the challenge of understanding how many signals are present at a locus, such as applying a thresholdof LD.
Another approach for gathering independent variants, as conceived by Yang et al. (2012), is to perform summary-level statistics conditional analysis. Here, the effect of a lead SNP is tested for association with all the other SNPs at a *locus* to determine the degree of association between them and detect the independent signals (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014).
After different signals at a *locus* have been defined, a subsequent analysis, referred to as "fine mapping", is performed in order to identify the potential causal variant(s). As reviewed by Schaid et al. (2018), several approaches can be adopted to perform this "fine mapping" analysis: The heuristic LD approach considers all of the SNPs that are related to the mainsignal with a value of LD higher than a fixed threshold. Other methods that rely more on statistics consist of jointly analyzing the correlation amongst all of the SNPs at a given *locus* through regression methods. Due to high correlation between variants, penalized regression models have been shown to be the most effective strategy. Recently, Bayesian approaches have been implemented (Jiang et al., 2019) to assess the probability of SNP causality at risk *loci* with success. When a subset of potential causal SNPs is obtained, further analyses procure the annotation of the variant biological effect, often referred as "Variant to Function" (V2F) analysis (Sun et al., 2021). So far,to relate genes to non-coding SNPs, the closest gene is considered.However, recent works (Mountjoy et al., 2021; Forgetta et al., 2022) aimed to improve this method by integrating various data sources and statistical methods to reach a better interpretation of the effect of non-coding variants on complex traits such as IDD. For this, multiple tools such as Variant effect Predictor (VEP) (McLaren et al., 2016) and Functional Mapping and Annotationof Genome-Wide Association Studies (FUMA) (Watanabe et al., 2017) can be implemented to suggest the potential molecularalterations caused by a SNP.
Additionally, an increasing number of data repositories that have collected information regarding gene expression, regulatory elements, and epigenetics, such as the Encyclopedia of DNA Elements (ENCODE) (de Souza 2012),the roadmap Epigenomics Project (Romanoski et al., 2015), and Genotype-Tissue Expression (GTEx) (The GTEX consortium, 2020) can be queried in order to obtain meaningful insights on the functional elements that could be affected by a candidate variant. In this way, one could retrieve information about the effect of the variant on the expression of a given gene. A gene- variant pair in which the variant is correlated to the gene expression is called expression quantitative trait *locus* (eQTL). Nicolae et al. (2010) demonstrated that common SNPs associated with complex diseases are significantly more likely to be eQTL, in comparison to rare SNPs acting on the phenotype by altering the expression of the gene rather than modifying the gene itself.

For instance, a study on ischemic stroke by Amini et al. (2020) showed that in blood, the SNP rs78046578 was significantly correlated to the expression of C-X-C motif chemokine ligand 10 (CXCL10), a small protein whose increased levels in the blood serum have been correlated to patients with IDD and LBP as wellas with Pfirmann grades (Vazirinejad et al., 2014; Yang et al., 2022).
Moreover, Raine et al. (2014) showed that in cartilage, the polymorphism rs8031440 is correlated with the expression of SMAD3. Specifically, carrying the G allele caused a reduction in the expression of the gene. SMAD3 is a key component of the TGF-β signaling pathway, which is highly involved in the anabolism of the extracellular matrix through enhancing the expression of type I collagen (Verrecchia and Mauviel, 2004).It was reported that SMAD3 knockout mice were smaller, had malformed and kyphotic spines, and had reduced levels of collagen and proteoglycans in the disc (Li et al., 2009). Although further evidence is needed, it appears that carrying a variant that leads to a dysregulation of SMAD3 expression couldlead to altered development of the disc and the spine itself.

Machine Learning/AI/Deep Learning

Advances in computer science, computer programming, statistics, mathematics, and modelling allowed the creation of new algorithms that can "learn" and make predictions on new data. Machine learning (ML), AI, and deep learning are relatively new approaches used to tackle the complexity of a disease in orderto identify biomarkers, or therapeutic interventions. The complexity and diversity of biological data (i.e., imaging, clinical categories, compound structures, gene sequence,protein/RNA data) is ideal for ML approaches to link seemingly unrelated entities.
A notable usage of deep learning in IDD is found in the classification of disc degeneration based on MRI images using the Pfirrman score. Although widely used, the Pfirrman score is very subjective and different observers sometimes classify the same image with differing scores (Rim 2016). Consistency in grading is essential for a clinician in order to have a clear idea of the patient's condition, which has led to the development of deep learning models. Niemeyer et al. (2021) have successfully managed to develop a deep learning model for the classification of discs based on MRI data that has an average sensitivity of 90%.
Additionally, ML can be useful in studying disc degeneration through simplifying complex models to decrease computational demand, or through coupling models across several scales to offera more holistic view of IDD. For example, while patient-specific FEMs, which were explained previously in this review, are useful in studying IVD biomechanics, they usually require complex procedures to set up and long computing times to obtain final simulation results. This therefore prevents prompt feedback to clinicians, resulting in studies with minimal sample sizes and severely hindering its suitability for time-sensitive clinical applications. As a response, neural networks have increasingly been employed in complex dynamical systems, resulting in highly accurate surrogate models that can be evaluated with significantly less computational resources and several orders of magnitude faster than conventional finite element solvers (Pfaff et al., 2020).

CONCLUSION

IDD is considered a complex multifactorial and pathological disease which alters biomechanical and biochemical aspects ofthe IVD, resulting in a shifted metabolism associated with increased cytokine and chemokine production by native disc cells driving catabolism, this together with abnormal mechanical loading can result in disc rupture. Following rupture, immune cells are able to invade the disc, allowing crosstalk between the IVD and the immune system. Therefore, a disrupted or herniated disc is needed in order to obtain an entry point for the immune system into the disc. Nevertheless, although catabolic and inflammatory features are different, since IVD cells (NP, AF and CEP) share classical immune cell's roles and markers (Le Maitre et al., 2005; Jones et al., 2008; Risbud and Shapiro, 2013; Phillips et al., 2013) the resulting phenotype in each case remains controversial.
In terms of terminology, immunomodulatory or inflammatory terms should only be used when immune cells are present within the IVD. Consequently, the production of chemokines and cytokines without disc rupture are contributed by the native IVD cells (NP, AF and CEPs) and should be termed catabolic cytokines and chemokines. Hence, catabolic phenotypes which are related to production of inflammatory cytokines and chemokines are incorrectly commonly attributed to inflammatory processesand easily misconstrued in the literature. Further, both processescould also simultaneously appear during IDD, making the recognition of these phenotypes more challenging.
Regarding the shift from anabolism to catabolism, it is expected that many factors are at play, including changes in the microenvironment, biomechanics, genetics, and metabolism. Within the microenvironment, low glucose and high lactic acid levels contribute to a catabolic shift in IDD, while hypo- osmolarity can activate pro-inflammatory and catabolic factors. Altered biomechanics also contributes to this catabolicshift, and aberrant loading can lead to CEP fractures and AF fissures that allow for immune cell infiltration. IDD has shownhigh inheritability, and gene variants in genes of structural
proteins and their turnover as well as cytokines such as IL-1β have been shown to provoke a catabolic shift. While bacterial presence in the IVD is still controversial, some studies indicate that in some cases of IDD, the increased catabolic cytokines could be due to bacterial infection and treatment with antibiotics could be effective to reduce pain although the results are varied. Overall, IDD is highly multifactorial, and each of these factors discussed play a role in the shift tocatabolism within an intact disc and possible immune cell infiltration following AF or CEP rupture.
However, for the best clinical treatment, early diagnosis iscrucial, which means that data analysis must be streamlined and the disc pathology must be classified correctly. Therefore, it is necessary to understand how different methodologies have been used to study the different features of IDD and assign the correct terminology. For that purpose, a wide range of *in-silico, in-vivo* and *in-vitro* models have been discussed in this review to select the best approaches for future IDD studies and thus provide the best clinical output. While 3D cell culture is effective in

investigating individual parameters within IDD, bioreactors and microfluidics studies offer another level of complexity through the addition of mechanics and/or fluid flow. Further, animal models provide even more sophistication as they include interactions between the disc and other tissues, however there are still many biochemical and biological differences in comparison to IDD in humans. This is in part how *in-silico* studies can be useful, as they can predict changes in the IVD based on prior research, without harming humans. Additionally, computational modeling can offer insights into IDD that are difficult or expensive to obtain through experiments. FEMs are useful to determine the biomechanical effects on the disc, which are expected to simulate specific patient models and observe the effects of CEP shape and its implications on nutrient and water transport, as well as on the different NP morphologies. ABMs can offer visual and spatial predictions, and network models provide insight into complex interactions at the protein or genetic level. Moreover, methodologies using candidate genes and GWAS have identified influential gene variants in IDD. Machine learning can then be a useful tool to simplify these models and methodologies, or to streamline and reduce the bias in the classification of IDD.

Hence, this review summarizes the recent advances of cross-disciplinary approaches to identify the mechanisms of the shift of anabolism to catabolism in the progress of IDD and compared them with immunomodulatory features. It demonstrates our current knowledge of the interplay of the immune system, metabolism, genetics, epigenetics, physiology, and mechanics, as well as computational and experimental models used to investigate catabolism and inflammation in the IVD.

AUTHOR CONTRIBUTIONS

PB-L and KC conceptionalisation, major literature research and wrote the main text, AN, KC, PB-L, CM, RC and FG and contributed with writing and reviewing and visualisation, ST, EK, AA, LB and EM-M writing and reviewing, LA, LG, KW-K,CM, JN, LG and BG writing and reviewing conceptionalisation, and funding. All authors approved the final version of the manuscript.

FUNDING

Financial support was received from the Marie Skłodowska Curie International Training Network (ITN) "disc4all" (https://disc4all.upf.edu, accessed on 10 May 2022) grant agreement #955735 (https://cordis.europa.eu/project/id/955735, accessed on 10 May 2022).

ACKNOWLEDGMENTS

We thank the Marie Skłodowska Curie International Training Network (ITN) "disc4all" for financial support. Furthermore, we thank the entire disc4all consortium for enabling this review.

REFERENCES

Adams, M. A., Freeman, B. J. C., Morrison, H. P., Nelson, I. W., and Dolan, P. (2000). Mechanical Initiation of Intervertebral Disc Degeneration. *Spine* 25, 1625–1636. doi:10.1097/00007632-200007010-00005

Adams, M. A., McMillan, D. W., Green, T. P., and Dolan, P. (1996). Sustained Loading Generates Stress Concentrations in Lumbar Intervertebral Discs. *Spine* 21, 434–438. doi:10.1097/00007632-199602150-00006

Adams, M. A., and Roughley, P. J. (2006). What Is Intervertebral Disc Degeneration, and what Causes it? *Spine* 31, 2151–2161. doi:10.1097/01.brs.

Al Dhaheri, N., Wu, N., Zhao, S., Wu, Z., Blank, R. D., Zhang, J., et al. (2020). KIAA1217 : A Novel Candidate Gene Associated with Isolated and Syndromic Vertebral Malformations. *Am. J. Med. Genet.* 182, 1664–1672. doi:10.1002/

Albert, H. B., Lambert, P., Rollason, J., Sorensen, J. S., Worthington, T., Pedersen, M. B., et al. (2013a). Does Nuclear Tissue Infected with Bacteria Following Disc Herniations Lead to Modic Changes in the Adjacent Vertebrae? *Eur. Spine J.* 22, 690–696. doi:10.1007/s00586-013-2674-z

Albert, H. B., Sorensen, J. S., Christensen, B. S., and Manniche, C. (2013b). Antibiotic Treatment in Patients with Chronic Low Back Pain and Vertebral Bone Edema (Modic Type 1 Changes): a Double-Blind Randomized Clinical Controlled Trial of Efficacy. *Eur. Spine J.* 22, 697–707. doi:10.1007/s00586-013-2675-y

Alini, M., Eisenstein, S. M., Ito, K., Little, C., Kettler, A. A., Masuda, K., et al. (2008). Are Animal Models Useful for Studying Human Disc Disorders/degeneration? *Eur. Spine J.* 17, 2–19. doi:10.1007/s00586-007-0414-y

Alkhatib, B., Rosenzweig, D. H., Rosenzweig, D., Krock, E., Roughley, P., Beckman, L., et al. (2014). Acute Mechanical Injury of the Human Intervertebral Disc: Link to Degeneration and Pain. *eCM* 28, 98–111. ; discussion 110-1. doi:10.22203/ecm.v028a08

Alvarez-Galvez, J., and Suarez-Lledo, V. (2019). Using Agent-Based Modeling to Understand the Emergence and Reproduction of Social Inequalities in Health. *Proceedings* 44, 2. doi:10.3390/IECEHS-2-06372

Amini, H., Shroff, N., Stamova, B., Ferino, E., Carmona-Mora, P., Zhan, X., et al. (2020). Genetic Variation Contributes to Gene Expression Response in Ischemic Stroke: an eQTL Study. *Ann. Clin. Transl. Neurol.* 7, 1648–1660.

Anjankar, S. D., Poornima, S., Raju, S., Jaleel, M. A., Bhiladvala, D., and Hasan, Q. (2015). Degenerated Intervertebral Disc Prolapse and its Association of Collagen I Alpha 1 Spl Gene Polymorphism. *Ijoo* 49, 589–594. doi:10.4103/

Aramburu, J., and López-Rodríguez, C. (2019). Regulation of Inflammatory Functions of Macrophages and T Lymphocytes by NFAT5. *Front. Immunol.* 10, 535. doi:10.3389/fimmu.2019.00535

Aubin, G. G., Baud'huin, M., Lavigne, J.-P., Brion, R., Gouin, F., Lepelletier, D., et al. (2017). Interaction of Cutibacterium (Formerly Propionibacterium) Acnes with Bone Cells: a Step toward Understanding Bone and Joint Infection Development. *Sci. Rep.* 7, 42918. doi:10.1038/srep42918

Barry, F. P., Neame, P. J., Sasse, J., and Pearson, D. (1994). Length Variation in the Keratan Sulfate Domain of Mammalian Aggrecan. *Matrix Biol.* 14, 323–328.

Batagelj, V., and Mrvar, A. (2004). "Pajek - Analysis and Visualization of Large Networks," in *Graph Drawing Software*. Editor M. M. P. Jünger (Berlin, Heidelberg: Springer Berlin Heidelberg), 77–103. doi:10.1007/978-3-642-

Battié, M. C., and Videman, T. (2006). Lumbar Disc Degeneration: Epidemiology and Genetics. *JBJS* 88, 3–9. doi:10.2106/JBJS.E.01313

Battié, M. C., Videman, T., and Parent, E. (2004). Lumbar Disc Degeneration. *Spine* 29, 2679–2690. doi:10.1097/01.brs.0000146457.83240.eb

Baumgartner, L., Reagh, J. J., González Ballester, M. A., and Noailly, J. (2021a). Simulating Intervertebral Disc Cell Behaviour within 3D Multifactorial

Environments. *Bioinformatics* 37, 1246–1253. doi:10.1093/bioinformatics/ OBJ

Baumgartner, L., Sadowska, A., Tío, L., González Ballester, M. A., Wuertz-Kozak, K., and Noailly, J. (2021b). Evidence-Based Network Modelling to Simulate Nucleus Pulposus Multicellular Activity in Different Nutritional and Pro- inflammatory Environments. *Front. Bioeng. Biotechnol.* 9, 734258. doi:10. OBJ

Baumgartner, L., Wuertz-Kozak, K., Le Maitre, C. L., Wignall, F., Richardson, S.M., Hoyland, J., et al. (2021c). Multiscale Regulation of the Intervertebral Disc: Achievements in Experimental, In Silico, and Regenerative Research. *Ijms* 22, 703. doi:10.3390/ijms22020703

Beckstein, J. C., Sen, S., Schaer, T. P., Vresilovic, E. J., and Elliott, D. M. (2008). Comparison of Animal Discs Used in Disc Research to Human Lumbar Disc. *Spine* 33, E166–E173. doi:10.1097/BRS.0b013e318166e001

Bedore, J., Sha, W., McCann, M. R., Liu, S., Leask, A., and Séguin, C. A. (2013). Lossof Notochord-Derived CCN2 Results in Impaired Intervertebral Disc Development and Premature Disc Degeneration. *Arthritis & Rheumatism* 65, a–n. doi:10.1002/art.38075

Benneker, L. M., Heini, P. F., Alini, M., Anderson, S. E., and Ito, K. (2005). 2004 Young Investigator Award Winner: Vertebral Endplate Marrow Contact Channel Occlusions and Intervertebral Disc Degeneration. *Spine* 30, 167–173. doi:10.1097/01.brs.0000150833.93248.09

Bibby, S. R. S., and Urban, J. P. G. (2004). Effect of Nutrient Deprivation on the Viability of Intervertebral Disc Cells. *Eur. Spine J.* 13, 695–701. doi:10.1007/ OBJ

Binch, A. L. A., Cole, A. A., Breakwell, L. M., Michael, A. L. R., Chiverton, N., Creemers, L. B., et al. (2015a). Class 3 Semaphorins Expression and Association with Innervation and Angiogenesis within the Degenerate Human Intervertebral Disc. *Oncotarget* 6, 18338–18354. doi:10.18632/oncotarget.4274 Binch, A. L. A., Cole, A. A., Breakwell, L. M., Michael, A. L. R., Chiverton, N., Creemers, L. B., et al. (2015b). Nerves Are More Abundant Than Blood Vessels in the Degenerate Human Intervertebral Disc. *Arthritis Res. Ther.* 17, 370. OBJ

Blumenthal, S., McAfee, P. C., Guyer, R. D., Hochschuler, S. H., Geisler, F. H., Holt, R. T., et al. (2005). A Prospective, Randomized, Multicenter Food and Drug Administration Investigational Device Exemptions Study of Lumbar Total Disc Replacement with the CHARITÉ Artificial Disc versus Lumbar Fusion. *Spine* 30, 1565–1575. discussion E387-91. doi:10.1097/01. OBJ

Binch, A. L., Cole, A. A., Breakwell, L. M., Michael, A. L., Chiverton, N., Cross, A. K., et al. (2014). Expression and Regulation of Neurotrophic and Angiogenic Factors during Human Intervertebral Disc Degeneration. *Arthritis Res. Ther.* 16, 416. doi:10.1186/s13075-014-0416-1

Bongers, P. M., Hulshof, C. T. J., DlJKSTRA, L., Boshuizen, H. C., Groenhout, H. J. M., and Valken, E. (1990). Back Pain and Exposure to Whole Body Vibration in Helicopter Pilots. *Ergonomics* 33, 1007–1026. doi:10.1080/ OBJ

Bournonville, S., Lambrechts, T., Vanhulst, J., Luyten, F. P., Papantoniou, I., and Geris, L. (2019). Towards Self-Regulated Bioprocessing: A Compact Benchtop Bioreactor System for Monitored and Controlled 3D Cell and Tissue Culture. *Biotechnol. J.* 14, 1800545. doi:10.1002/biot.201800545

Bråten, L. C. H., Grøvle, L., Espeland, A., Pripp, A. H., Grotle, M., Helllum, C., et al. (2020). Clinical Effect Modifiers of Antibiotic Treatment in Patients with Chronic Low Back Pain and Modic Changes - Secondary Analyses of a Randomised, Placebo-Controlled Trial (The AIM Study). *Bmc. Musculoskelet. Disord.* 21, 1–11. doi:10.1186/s12891-020-03422-y

Buckwalter, J. A. (1995). Aging and Degeneration of the Human Intervertebral Disc. *Spine* 20, 1307–1314. doi:10.1097/00007632-199506000-00022

Cambria, E., Brunner, S., Heusser, S., Fisch, P., Hitzl, W., Ferguson, S. J., et al. (2020). Cell-Laden Agarose-Collagen Composite Hydrogels for Mechanotransduction Studies. *Front. Bioeng. Biotechnol.* 8, 346. doi:10.3389/ OBJ

Cannon, M. E., and Mohlke, K. L. (2018). Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci. *Am. J. Hum. Genet.* 103, 637–653. OBJ

Cano-Gamez, E., and Trynka, G. (2020). From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying ComplexDiseases. *Front. Genet.* 11, 424. doi:10.3389/fgene.2020.00424

Capoor, M. N., Konieczna, A., McDowell, A., Ruzicka, F., Smrcka, M., Jancalek, R., et al. (2021). Pro-Inflammatory and Neurotrophic Factor Responses of Cells Derived from Degenerative Human Intervertebral Discs to the Opportunistic Pathogen Cutibacterium Acnes. *Ijms* 22, 2347. doi:10.3390/ OBJ

Chan, S. C. W., and Gantenbein-Ritter, B. (2012). Preparation of Intact Bovine Tail Intervertebral Discs for Organ Culture. *JoVE* 60, e3490. doi:10.3791/3490

Chan, S. C. W., Walser, J., Käppeli, P., Shamsollahi, M. J., Ferguson, S. J., and Gantenbein-Ritter, B. (2013). Region Specific Response of Intervertebral Disc Cells to Complex Dynamic Loading: An Organ Culture Study Using a Dynamic Torsion-Compression Bioreactor. *PLoS. ONE.* 8, e72489. doi:10.1371/journal. OBJ

Chang, H., Cai, F., Zhang, Y., Xue, M., Liu, L., Yang, A., et al. (2017). Early-stage Autophagy Protects Nucleus Pulposus Cells from Glucose Deprivation-Induced Degeneration via the P -eIF2α/ATF4 Pathway. *Biomed. Pharmacother.* 89, 529–535. doi:10.1016/j.biopha.2017.02.074

Chang, H., Wang, H., Yang, X., You, K., Jiang, M., Cai, F., et al. (2021). Comprehensive Profile Analysis of Differentially Expressed circRNAs in Glucose Deprivation-Induced Human Nucleus Pulposus Cell Degeneration. *BioMed Res. Int.* 2021, 1–14. doi:10.1155/2021/4770792

Chen, J.-W., Li, B., Yang, Y.-H., Jiang, S.-D., and Jiang, L.-S. (2014). Significance of Hypoxia in the Physiological Function of Intervertebral Disc Cells. *Crit. Rev. Eukaryot. Gene. Expr.* 24, 193–204. doi:10.1615/critreveukaryotgeneexpr. 2014010485

Chen, J., Baer, A. E., Paik, P. Y., Yan, W., and Setton, L. A. (2002). Matrix Protein Gene Expression in Intervertebral Disc Cells Subjected to Altered Osmolarity. *Biochem. Biophysical Res. Commun.* 293, 932–938. doi:10.1016/S0006- OBJ

Cheung, K. M. C., Karppinen, J., Chan, D., Ho, D. W. H., Song, Y.-Q., Sham, P.,et al. (2009). Prevalence and Pattern of Lumbar Magnetic Resonance Imaging Changes in a Population Study of One Thousand Forty-Three Individuals. *Spine* 34, 934–940. doi:10.1097/BRS.0b013e3181a01b3f

Ching, C. T. S., Chow, D. H. K., Yao, F. Y. D., and Holmes, A. D. (2003). The Effectof Cyclic Compression on the Mechanical Properties of the Inter-vertebral Disc:an *In Vivo* Study in a Rat Tail Model. *Clin. Biomech.* 18, 182–189. doi:10.1016/ OBJ

Choi, H., Chaiyamongkol, W., Doolittle, A. C., Johnson, Z. I., Gogate, S. S.,Schoepflin, Z. R., et al. (2018). COX-2 Expression Mediated by Calcium- TonEBP Signaling axis under Hyperosmotic Conditions Serves Osmoprotective Function in Nucleus Pulposus Cells. *J. Biol. Chem.* 293, 8969–8981. doi:10.1074/jbc.RA117.001167

Chou, P.-H., Wang, S.-T., Yen, M.-H., Liu, C.-L., Chang, M.-C., and Lee, O. K.-S. (2016). Fluid-induced, Shear Stress-Regulated Extracellular Matrix and Matrix Metalloproteinase Genes Expression on Human Annulus Fibrosus Cells. *Stem.Cell. Res. Ther.* 7, 34. doi:10.1186/s13287-016-0292-5

Chu, G., Shi, C., Wang, H., Zhang, W., Yang, H., and Li, B. (2018). Strategies for Annulus Fibrosus Regeneration: From Biological Therapies to Tissue Engineering. *Front. Bioeng. Biotechnol.* 6, 90. doi:10.3389/fbioe.2018. OBJ

Costi, J. J., Stokes, I. A., Gardner-Morse, M. G., and Iatridis, J. C. (2008). Frequency-Dependent Behavior of the Intervertebral Disc in Response to Each of Six Degree of Freedom Dynamic Loading. *Spine* 33, 1731–1738. OBJ

Croft, A. S., Roth, Y., Oswald, K. A. C., Ćorluka, S., Bermudez-Lekerika, P., and Gantenbein, B. (2021). *In Situ* Cell Signalling of the Hippo-YAP/TAZ Pathwayin Reaction to Complex Dynamic Loading in an Intervertebral Disc Organ Culture. *Ijms* 22, 13641. doi:10.3390/ijms222413641

Dahia, C. L., Engiles, J. B., Gullbrand, S. E., Iatridis, J. C., Lai, A., Le Maitre, C. L., et al. (2021). A Perspective on the ORS Spine Section Initiative to Develop a Multi-Species JOR Spine Histopathology Series. *JOR. Spine* 4, e1165. doi:10.1002/jsp2.1165

Dai, J., Xing, Y., Xiao, L., Li, J., Cao, R., He, Y., et al. (2019). Microfluidic Disc-On-A-Chip Device for Mouse Intervertebral Disc-Pitching a Next-Generation Research Platform to Study Disc Degeneration. *ACS Biomater. Sci. Eng.* 5, 2041–2051. doi:10.1021/acsbiomaterials.8b01522

Daly, C., Ghosh, P., Jenkin, G., Oehme, D., and Goldschlager, T. (2016). A Review of Animal Models of Intervertebral Disc Degeneration: Pathophysiology, Regeneration, and Translation to the Clinic. *BioMed Res. Int.* 2016, 1–14. doi:10.1155/2016/5952165

Damiati, S., Kompella, U., Damiati, S., and Kodzius, R. (2018). Microfluidic Devices for Drug Delivery Systems and Drug Screening. *Genes* 9, 103. OBJ

Daniels, J., Binch, A. A. L., and Le Maitre, C. L. (2017). Inhibiting IL-1 Signaling Pathways to Inhibit Catabolic Processes in Disc Degeneration. *J. Orthop. Res.* 35, 74–85. doi:10.1002/jor.23363

de Souza, N. (2012). The ENCODE Project. *Nat. Methods.* 9, 1046. doi:10.1038/ OBJ

Dong, D. M., Yao, M., Liu, B., Sun, C. Y., Jiang, Y. Q., and Wang, Y. S. (2007). Association between the -1306C/T Polymorphism of MatrixMetalloproteinase-2 Gene and Lumbar Disc Disease in Chinese Young Adults. *Eur. Spine J.* 16, 1958–1961. doi:10.1007/s00586-007-0454-3

Du, H., Bai, B., Qiu, Y., Yin, S., and Bian, W. (2015). Association between TRAIL Gene Polymorphisms and the Susceptibility and Severity of Lumbar Disc Degeneration. *Int. J. Clin. Exp. Pathol.* 8, 7415–7420.

Du, Y., Tavana, S., Rahman, T., Baxan, N., Hansen, U. N., and Newell, N. (2021). Sensitivity of Intervertebral Disc Finite Element Models to Internal Geometric and Non-geometric Parameters. *Front. Bioeng. Biotechnol.* 9, 660013. doi:10. OBJ

Dudli, S., Fields, A. J., Samartzis, D., Karppinen, J., and Lotz, J. C. (2016). Pathobiology of Modic Changes. *Eur. Spine J.* 25, 3723–3734. doi:10.1007/ OBJ

Dudli, S., Haschtmann, D., and Ferguson, S. J. (2011). Fracture of the Vertebral Endplates, but Not Equienergetic Impact Load, Promotes Disc Degeneration *InVitro*. *J. Orthop. Res.* 30, 809–816. doi:10.1002/jor.21573

Dudli, S., Haschtmann, D., and Ferguson, S. J. (2014). Persistent Degenerative Changes in the Intervertebral Disc after Burst Fracture in an *In Vitro* Model Mimicking Physiological Post-traumatic Conditions. *Eur. Spine J.* 24, 1901–1908. doi:10.1007/s00586-014-3301-3

Dudli, S., Liebenberg, E., Magnitsky, S., Lu, B., Lauricella, M., and Lotz, J. C. (2018). Modic Type 1 Change Is an Autoimmune Response that Requires aProinflammatory Milieu provided by the 'Modic Disc'. *Spine J.* 18, 831–844. OBJ

Duncan, L. E., Ostacher, M., and Ballon, J. (2019). How Genome-wide Association Studies (GWAS) Made Traditional Candidate Gene Studies Obsolete. *Neuropsychopharmacol.* 44, 1518–1523. doi:10.1038/s41386-019-0389-5

Elfervig, M. K., Minchew, J. T., Francke, E., Tsuzaki, M., and Banes, A. J. (2001a). IL-1? Sensitizes Intervertebral Disc Annulus Cells to Fluid-Induced Shear Stress. *J. Cell. Biochem.* 82, 290–298. doi:10.1002/jcb.1153

Elfervig, M. K., Minchew, J. T., Francke, E., Tsuzaki, M., and Banes, A. J. (2001b). IL-1? Sensitizes Intervertebral Disc Annulus Cells to Fluid-Induced Shear Stress. *J. Cell. Biochem.* 82, 290–298. doi:10.1002/jcb.1153

Fadok, V. A. (1999). Clearance: The Last and Often Forgotten Stage of Apoptosis. *J. Mamm. Gl. Biol. Neoplasia* 4, 203–211. doi:10.1023/A:1011384009787 Fearing, B. V., Hernandez, P. A., Setton, L. A., and Chahine, N. O. (2018). Mechanotransduction and Cell Biomechanics of the Intervertebral Disc. *JOR. Spine* 1, e1026. doi:10.1002/jsp2.1026

Feng, Y., Egan, B., and Wang, J. (2016). Genetic Factors in Intervertebral Disc Degeneration. *Genes & Dis.* 3, 178–185. doi:10.1016/j.gendis.2016. OBJ

Fernando, H. N., Czamanski, J., Yuan, T.-Y., Gu, W., Salahadin, A., and Huang, C.-Y. C. (2011). Mechanical Loading Affects the Energy Metabolism of Intervertebral Disc Cells. *J. Orthop. Res.* 29, 1634–1641. doi:10.1002/jor. OBJ

Flannery, C. R., Little, C. B., and Caterson, B. (1998). Molecular Cloning and Sequence Analysis of the Aggrecan Interglobular Domain from Porcine, Equine, Bovine and Ovine Cartilage: Comparison of Proteinase-SusceptibleRegions and Sites of Keratan Sulfate Substitution. *Matrix Biol.* 16, 507–511. OBJ

Forgetta, V., Jiang, L., Vulpescu, N. A., Hogan, M. S., Chen, S., Morris, J. A., et al. (2022). An Effector Index to Predict Target Genes at GWAS Loci. *Hum. Genet.* [Epub ahead of print]. doi:10.1007/s00439-022-02434-z

Frauchiger, D., Tekari, A., Tekari, A., Wöltje, M., Fortunato, G., Benneker, L., et al. (2017). A Review of the Application of Reinforced Hydrogels and Silk as Biomaterials for Intervertebral Disc Repair. *eCM* 34, 271–290. doi:10.22203/ OBJ

Freemont, A. J., Watkins, A., Le Maitre, C., Baird, P., Jeziorska, M., Knight, M. T. N., et al. (2002). Nerve Growth Factor Expression and Innervation of the Painful Intervertebral Disc. *J. Pathol.* 197, 286–292. doi:10.1002/path. OBJ

Freemont, A., Peacock, T., Goupille, P., Hoyland, J., O'Brien, J., and Jayson, M. (1997). Nerve Ingrowth into Diseased Intervertebral Disc in Chronic Back Pain. *Lancet* 350, 178–181. doi:10.1016/s0140-6736(97)02135-1

Frobin, W., Brinckmann, P., Kramer, M., and Hartwig, E. (2001). Height of Lumbar Discs Measured from Radiographs Compared with Degeneration and Height Classified from MR Images. *Eur. Radiol.* 11, 263–269. doi:10. OBJ

Galbusera, F., Schmidt, H., Neidlinger-Wilke, C., and Wilke, H.-J. (2011a). The Effect of Degenerative Morphological Changes of the Intervertebral Disc on theLumbar Spine Biomechanics: a Poroelastic Finite Element Investigation. *Comput. Methods Biomechanics Biomed. Eng.* 14, 729–739. doi:10.1080/ OBJ

Galbusera, F., Schmidt, H., Noailly, J., Malandrino, A., Lacroix, D., Wilke, H.-J.,et al. (2011b). Comparison of Four Methods to Simulate Swelling in Poroelastic Finite Element Models of Intervertebral Discs. *J. Mech. Behav. Biomed. Mater.*4, 1234–1241. doi:10.1016/j.jmbbm.2011.04.008

Gantenbein, B., Frauchiger, D. A., May, R. D., Bakirci, E., Urs, R., and Sibylle, G. (2019). "Developing Bioreactors to Host Joint-Derived Tissues that Require Mechanical Stimulation," in *Encyclopedia of Tissue Engineering and Regenerative Medicine*. Editors R. L. Reis and M. E. Gomes (NY: Elsevier), 261–280. doi:10.1016/b978-0-12-801238-3.65611-8

Gantenbein, B., Calandriello, E., Wuertz-Kozak, K., Benneker, L. M., Keel, M. J., and Chan, S. C. (2014). Activation of Intervertebral Disc Cells by Co-culture with Notochordal Cells, Conditioned Medium and Hypoxia. *BMC Musculoskelet. Disord.* 15, 422. doi:10.1186/1471-2474-15-422

Gantenbein, B., Grünhagen, T., Lee, C. R., van Donkelaar, C. C., Alini, M., and Ito, K. (2006). An *In Vitro* Organ Culturing System for Intervertebral Disc Explantswith Vertebral Endplates. *Spine* 31, 2665–2673. doi:10.1097/01.brs.0000244620. OBJ

Gantenbein, B., Illien-Jünger, S., Chan, S., Walser, J., Haglund, L., Ferguson, S.,et al. (2015). Organ Culture Bioreactors - Platforms to Study Human Intervertebral Disc Degeneration and Regenerative Therapy. *Cscr* 10, 339–352. doi:10.2174/1574888X10666150312102948

GBD 2017 Disease and Injury Incidence and Prevalence Collaborators (2018). Global, Regional, and 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. doi:10.1016/S0140-6736(18)32279-7

Ghezelbash, F., Schmidt, H., Shirazi-Adl, A., and El-Rich, M. (2020). Internal Load-Sharing in the Human Passive Lumbar Spine: Review of *In Vitro* and Finite Element Model Studies. *J. Biomechanics* 102, 109441. doi:10.1016/j.jbiomech. OBJ

Gibson, G. (2010). Hints of Hidden Heritability in GWAS. *Nat. Genet.* 42, 558–560. OBJ

Gilbert, H. T., Hoyland, J. A., Freemont, A. J., and Millward-Sadler, S. J. (2011). The Involvement of Interleukin-1 and Interleukin-4 in the Response of Human Annulus Fibrosus Cells to Cyclic Tensile Strain: an Altered Mechanotransduction Pathway with Degeneration. *Arthritis Res. Ther.* 13, R8. doi:10.1186/ar3229

Gilbert, H. T. J., Hodson, N., Baird, P., Richardson, S. M., and Hoyland, J. A. (2016). Acidic pH Promotes Intervertebral Disc Degeneration: Acid-Sensing Ion Channel -3 as a Potential Therapeutic Target. *Sci. Rep.* 6, 37360. doi:10. OBJ

Gilligan, C. J., Cohen, S. P., Fischetti, V. A., Hirsch, J. A., and Czaplewski, L. G. (2021). Chronic Low Back Pain, Bacterial Infection and Treatment with Antibiotics. *Spine J.* 21, 903–914. doi:10.1016/j.spinee.2021.02.013

Goddard, M. E., Kemper, K. E., MacLeod, I. M., Chamberlain, A. J., and Hayes, B. J. (2016). Genetics of Complex Traits: Prediction of Phenotype, Identification of Causal Polymorphisms and Genetic Architecture. *Proc. R. Soc. B* 283, 20160569. OBJ

Gorth, D. J., Shapiro, I. M., and Risbud, M. V. (2015). Discovery of the Drivers of Inflammation Induced Chronic Low Back Pain: from Bacteria to Diabetes. *Discov. Med.* 20, 177–184.

Gorth, D. J., Shapiro, I. M., and Risbud, M. V. (2019). A New Understanding of the Role of IL-1 in Age-Related Intervertebral Disc Degeneration in a MurineModel. *J. Bone. Min. Res.* 34, 1531–1542. doi:10.1002/jbmr.3714

Gorth, D. J., Shapiro, I. M., and Risbud, M. V. (2018). Transgenic Mice Overexpressing Human TNF-α Experience Early Onset Spontaneous Intervertebral Disc Herniation in the Absence of Overt Degeneration. *Cell. death. Dis.* 10, 1–14. doi:10.1038/s41419-018-1246-x

Granville Smith, I., Danckert, N. P., Freidin, M. B., Wells, P., Marchesi, J. R., and Williams, F. M. K. (2021). Evidence for Infection in Intervertebral Disc Degeneration: a Systematic Review. *Eur. Spine J.* 31, 414–430. doi:10.1007/

Gu, J., Guan, F., Guan, G., Xu, G., Wang, X., Zhao, W., et al. (2013). Aggrecan Variable Number of Tandem Repeat Polymorphism and Lumbar Disc Degeneration. *Spine* 38, E1600–E1607. doi:10.1097/BRS.

Guerrero, J., Häckel, S., Croft, A. S., Albers, C. E., and Gantenbein, B. (2020). The Effects of 3D Culture on the Expansion and Maintenance of Nucleus Pulposus Progenitor Cell Multipotency. *JOR Spine* 4, e1131. doi:10.1002/

Gullbrand, S. E., Malhotra, N. R., Schaer, T. P., Zawacki, Z., Martin, J. T., Bendigo, J. R., et al. (2017). A Large Animal Model that Recapitulates the Spectrum of Human Intervertebral Disc Degeneration. *Osteoarthr. Cartil.* 25, 146–156.

Gullbrand, S. E., Smith, L. J., Smith, H. E., and Mauck, R. L. (2018). Promise, Progress, and Problems in Whole Disc Tissue Engineering. *JOR. Spine* 1, e1015.

Guo, X., and Chen, G. (2020). Hypoxia-Inducible Factor Is Critical forPathogenesis and Regulation of Immune Cell Functions in Rheumatoid Arthritis. *Front. Immunol.* 11, 1668. doi:10.3389/fimmu.2020.01668

Haglund, L., Moir, J., Beckman, L., Mulligan, K. R., Jim, B., Ouellet, J. A., et al. (2011). Development of a Bioreactor for Axially Loaded Intervertebral Disc Organ Culture. *Tissue Eng. Part C. Methods* 17, 1011–1019. doi:10.1089/ten.

Hamilton, D. J., Séguin, C. A., Wang, J., Pilliar, R. M., and Kandel, R. A. (2006). BioEngineering of Skeletal Tissues TeamFormation of a Nucleus Pulposus- Cartilage Endplate Construct *In Vitro*. *Biomaterials* 27, 397–405. doi:10.1016/j.

Han, I. B., Ropper, A. E., Teng, Y. D., Shin, D. A., Jeon, Y. J., Park, H. M., et al. (2013). Association between VEGF and eNOS Gene Polymorphisms and Lumbar Disc Degeneration in a Young Korean Population. *Genet. Mol. Res.* 12, 2294–2305. doi:10.4238/2013.July.8.10

Hellum, C., Johnsen, L. G., JohnsenBerg, L. G., Gjertsen, Ø., Berg, L., Neckelmann, G., et al. (2012). Predictors of Outcome after Surgery with Disc Prosthesis and Rehabilitation in Patients with Chronic Low Back Pain and Degenerative Disc: 2-year Follow-Up. *Eur. Spine J.* 21, 681–690. doi:10.

Hernandez, P. A., Jacobsen, T. D., Barati, Z., and Chahine, N. O. (2020a). Confocal Scanning of Intervertebral Disc Cells in 3D : Inside Alginate Beads and inNative Microenvironment. *JOR Spine* 3, e1106. doi:10.1002/jsp2.1106

Hernandez, P. A., Jacobsen, T. D., and Chahine, N. O. (2020b). Actomyosin Contractility Confers Mechanoprotection against TNFα-Induced Disruption of the Intervertebral Disc. *Sci. Adv.* 6, eaba2368. doi:10.1126/sciadv.aba2368

Hingert, D., Nawilaijaroen, P., Aldridge, J., Baranto, A., and Brisby, H. (2019). Investigation of the Effect of Secreted Factors from Mesenchymal Stem Cells onDisc Cells from Degenerated Discs. *Cells. Tissues. Organs* 208, 76–88. doi:10.

Hirose, Y., Chiba, K., Karasugi, T., Nakajima, M., Kawaguchi, Y., Mikami, Y., et al. (2008). A Functional Polymorphism in THBS2 that Affects Alternative Splicing and MMP Binding Is Associated with Lumbar-Disc Herniation. *Am. J. Hum. Genet.* 82, 1122–1129. doi:10.1016/j.ajhg.2008.03.013

Holm, S., Holm, A. K., Ekström, L., Karladani, A., and Hansson, T. (2004). Experimental Disc Degeneration Due to Endplate Injury. *J. Spinal Disord. Tech.* 17, 64–71. doi:10.1097/00024720-200402000-00012

Hoogendoorn, W. E., van Poppel, M. N. M., Bongers, P. M., Koes, B. W., and Bouter, L. M. (2000). Systematic Review of Psychosocial Factors at Work and Private Life as Risk Factors for Back Pain. *Spine* 25, 2114–2125. doi:10.1097/

Horner, H. A., and Urban, J. P. G. (20012001). 2001 Volvo Award Winner in Basic Science Studies: Effect of Nutrient Supply on the Viability of Cells from the Nucleus Pulposus of the Intervertebral Disc. *Spine* 26, 2543–2549. doi:10.1097/ 00007632-200112010-00006

Hoyland, J. A., Le Maitre, C., and Freemont, A. J. (2008). Investigation of the Role of IL-1 and TNF in Matrix Degradation in the Intervertebral Disc. *Rheumatology* 47, 809–814. doi:10.1093/rheumatology/ken056

Hu, J., Yan, Q., Shi, C., Tian, Y., Cao, P., and Yuan, W. (2017). BMSC Paracrine Activity Attenuates Interleukin-1β-Induced Inflammation and Apoptosis in Rat AF Cells via Inhibiting Relative NF-Kb Signaling and the Mitochondrial Pathway. *Am. J. Transl. Res.* 9, 79–89.

Hu, Z., Mellor, J., Wu, J., and DeLisi, C. (2004). VisANT: an Online Visualization and Analysis Tool for Biological Interaction Data. *Bmc. Bioinforma.* 5, 17.

Huang, Y.-C., Hu, Y., Li, Z., and Luk, K. D. K. (2018). Biomaterials for Intervertebral Disc Regeneration: Current Status and Looming Challenges. *J. Tissue. Eng. Regen. Med.* 12, 2188–2202. doi:10.1002/term.2750

Hunter, C. J., and De Plaen, I. G. (2014). Inflammatory Signaling in NEC: Role of NF-Kb, Cytokines and Other Inflammatory Mediators. *Pathophysiology* 21, 55–65. doi:10.1016/j.pathophys.2013.11.010

Hunter, C. J. (1988). *The Biology of the Intervertebral Disc*. Boca Raton, Florida, US: CRC Press.

Hwang, M. H., Son, H.-G., Kim, J., and Choi, H. (2020). *In Vitro* model of Distinct Catabolic and Inflammatory Response Patterns of Endothelial Cells to Intervertebral Disc Cell Degeneration. *Sci. Rep.* 10, 20596. doi:10.1038/

Iatridis, J. C., Mente, P. L., Stokes, I. A. F., Aronsson, D. D., and Alini, M. (1999). Compression-induced Changes in Intervertebral Disc Properties in a Rat Tail Model. *Spine* 24, 996–1002. doi:10.1097/00007632-199905150-00013

Illien-Jünger, S., Walter, B. A., Mayer, J. E., Hecht, A. C., and Iatridis, J. C. (2014). "Intervertebral Disc Culture Models and Their Applications to Study Pathogenesis and Repair," in *The Intervertebral Disc* (Vienna: Springer), 353–371. doi:10.1007/978-3-7091-1535-0_22

Isa, I. L. M., Mokhtar, S. A., Abbah, S. A., Fauzi, M. B., Devitt, A., and Pandit, A. (2022). Intervertebral Disc Degeneration: Biomaterials and Tissue Engineering Strategies toward Precision Medicine. *Adv. Healthc. Mater.*, e2102530. doi:10.

Jauregui, J. J., Tran, A., Kaveeshwar, S., Nadarajah, V., Chaudhri, M. W., Henn, R. F., et al. (2021). Diagnosing a Periprosthetic Shoulder Infection: A Systematic Review. *J. Orthop.* 26, 58–66. doi:10.1016/j.jor.2021.07.012

Jensen, C., and Teng, Y. (2020). Is it Time to Start Transitioning from 2D to 3D Cell Culture? *Front. Mol. Biosci.* 7, 33. doi:10.3389/fmolb.2020.00033

Jiang, J., Cole, J. B., Freebern, E., Da, Y., VanRaden, P. M., and Ma, L. (2019). Functional Annotation and Bayesian Fine-Mapping Reveals Candidate Genes for Important Agronomic Traits in Holstein Bulls. *Commun. Biol.* 2, 1–12.

Jin, L., Balian, G., and Li, X. J. (2018). Animal Models for Disc Degeneration-An Update. *Histol. Histopathol.* 33, 543–554. doi:10.14670/HH-11-910

Johnson, W. E. B., Caterson, B., Eisenstein, S. M., Hynds, D. L., Snow, D. M., and Roberts, S. (2002). Human Intervertebral Disc Aggrecan Inhibits Nerve Growth *In Vitro*. *Arthritis & Rheumatism* 46, 2658–2664. doi:10.1002/art.

Johnson, W. E. B., Caterson, B., Eisenstein, S. M., and Roberts, S. (2005). Human Intervertebral Disc Aggrecan Inhibits Endothelial Cell Adhesion and Cell Migration *In Vitro*. *Spine* 30, 1139–1147. doi:10.1097/01.brs.0000162624.

Johnson, W. E. B., Stephan, S., and Roberts, S. (2008). The Influence of Serum, Glucose and Oxygen on Intervertebral Disc Cell Growth *In Vitro*: Implications for Degenerative Disc Disease. *Arthritis Res. Ther.* 10, R46.

Johnson, Z. I., Doolittle, A. C., Snuggs, J. W., Shapiro, I. M., Le Maitre, C. L., and Risbud, M. V. (2017). TNF-α Promotes Nuclear Enrichment of the Transcription Factor TonEBP/NFAT5 to Selectively Control Inflammatory but Not Osmoregulatory Responses in Nucleus Pulposus Cells. *J. Biol. Chem.* 292, 17561–17575. doi:10.1074/jbc.M117.790378

Johnson, Z., Schoepflin, Z. R., Schoepflin, Z., Choi, H., Shapiro, I., and Risbud, M. (2015). Disc in Flames: Roles of TNF-α and IL-1β in Intervertebral Disc Degeneration. *eCM* 30, 104–117. doi:10.22203/ecm.v030a08

Jones, A. C., and Wilcox, R. K. (2008). Finite Element Analysis of the Spine: towards a Framework of Verification, Validation and Sensitivity Analysis. *Med. Eng. Phys.* 30, 1287–1304. doi:10.1016/j.medengphy.2008. 09.006

Jones, P., Gardner, L., Menage, J., Williams, G. T., and Roberts, S. (2008). Intervertebral Disc Cells as Competent Phagocytes *In Vitro*: Implications for Cell Death in Disc Degeneration. *Arthritis Res. Ther.* 10, R86. doi:10.1186/ OBJ

Kalb, S., Martirosyan, N. L., Kalani, M. Y. S., Broc, G. G., and Theodore, N. (2012). Genetics of the Degenerated Intervertebral Disc. *World Neurosurg.* 77, 491–501. OBJ

Kang, J. D., Stefanovic-Racic, M., McIntyre, L. A., Georgescu, H. I., and Evans, C. H. (1997). Toward a Biochemical Understanding of Human Intervertebral Disc Degeneration and Herniation. *Spine* 22, 1065–1073. doi:10.1097/00007632- OBJ

Karasugi, T., Semba, K., Hirose, Y., Kelempisioti, A., Nakajima, M., Miyake, A.,et al. (2009). Association of the Tag SNPs in the HumanSKTGene (KIAA1217)with Lumbar Disc Herniation. *J. Bone Mineral Res.* 24, 1537–1543. doi:10.1359/ OBJ

Karppinen, J., Daavittila, I., Solovieva, S., Kuisma, M., Taimela, S., Natri, A., et al. (2008). Genetic Factors Are Associated with Modic Changes in Endplates of Lumbar Vertebral Bodies. *Spine* 33, 1236–1241. doi:10.1097/BRS.0b013e318170fd0e

Kashtan, N., Itzkovitz, S., Milo, R., and Alon, U. (2004). Efficient Sampling Algorithm for Estimating Subgraph Concentrations and Detecting Network Motifs. *Bioinformatics* 20, 1746–1758. doi:10.1093/bioinformatics/bth163

Kawaguchi, Y. (2018). Genetic Background of Degenerative Disc Disease in the Lumbar Spine. *Spine Surg. Relat. Res.* 2, 98–112. doi:10.22603/ssrr.2017- OBJ

Kawaguchi, Y., Osada, R., Kanamori, M., Ishihara, H., Ohmori, K., Matsui, H., et al. (1999). Association between an Aggrecan Gene Polymorphism and Lumbar Disc Degeneration. *Spine* 24, 2456–2460. doi:10.1097/00007632-199912010-00006

Kelempisioti, A., Eskola, P. J., Okuloff, A., Karjalainen, U., Takatalo, J., Daavittila, I., et al. (2011). Genetic Susceptibility of Intervertebral Disc Degeneration Among Young Finnish Adults. *Bmc. Med. Genet.* 12, 12–153. doi:10.1186/1471- OBJ

Kelly, T.-A. N., Ng, K. W., Ateshian, G. A., and Hung, C. T. (2009). Analysis of Radial Variations in Material Properties and Matrix Composition of Chondrocyte-Seeded Agarose Hydrogel Constructs. *Osteoarthr. Cartil.* 17, 73–82. doi:10.1016/j.joca.2008.05.019

Kepler, C. K., Ponnappan, R. K., Tannoury, C. A., Risbud, M. V., and Anderson, D. G. (2013). The Molecular Basis of Intervertebral Disc Degeneration. *Spine J.* 13,318–330. doi:10.1016/j.spinee.2012.12.003

Kim, A.-G., Kim, T.-W., Kwon, W.-K., Lee, K.-H., Jeong, S., Hwang, M.-H., et al. (2021). Microfluidic Chip with Low Constant-Current Stimulation (LCCS) Platform: Human Nucleus Pulposus Degeneration *In Vitro* Model for Symptomatic Intervertebral Disc. *Micromachines* 12, 1291. doi:10.3390/ OBJ

Kim, N. K., Shin, D. A., Han, I. B., Yoo, E. H., Kim, S. H., and Chung, S. S. (2010). The Association of Aggrecan Gene Polymorphism with the Risk of Intervertebral Disc Degeneration. *Acta Neurochir.* 153, 129–133. doi:10. OBJ

Knudson, C. B., and Knudson, W. (2001). Cartilage Proteoglycans. *Seminars Cell & Dev. Biol.* 12, 69–78. doi:10.1006/scdb.2000.0243

Krock, E., Rosenzweig, D. H., Chabot-Doré, A. J., Jarzem, P., Weber, M. H., Ouellet, J. A., et al. (2014). Painful, Degenerating Intervertebral Discs Up-Regulate Neurite Sprouting andCGRPthrough Nociceptive Factors. *J. Cell. Mol. Med.* 18, 1213–1225. doi:10.1111/jcmm.12268

Kudo, S., Mizuno, K., Hirai, Y., and Shimizu, T. (1990). Clearance and Tissue Distribution of Recombinant Human Interleukin 1 Beta in Rats. *Cancer. Res.*50, 5751–5755.

Lakstins, K., Arnold, L., Gunsch, G., Flanigan, D., Khan, S., Gadde, N., et al. (2021). Characterization of the Human Intervertebral Disc Cartilage Endplate at the Molecular, Cell, and Tissue Levels. *J. Orthop. Res.* 39, 1898–1907. doi:10.1002/ OBJ

Lama, P., Le Maitre, C. L., Harding, I. J., Dolan, P., and Adams, M. A. (2018). Nerves and Blood Vessels in Degenerated Intervertebral Discs Are Confined to Physically Disrupted Tissue. *J. Anat.* 233, 86–97. doi:10. OBJ

Larson, J. W., Levicoff, E. A., Gilbertson, L. G., and Kang, J. D. (2006). Biologic Modification of Animal Models of Intervertebral Disc Degeneration. *J. Bone. Jt. Surg. Am.* 88 (Suppl. 2), 83–87. doi:10.2106/JBJS.F.00043

Lazarus, E., Bermudez-Lekerika, P., Farchione, D., Schofield, T., Howard, S., Mambetkadyrov, I., et al. (2021). Sulfated Hydrogels in Intervertebral Discand Cartilage Research. *Cells* 10, 3568. doi:10.3390/cells10123568

Le Maitre, C., Freemont, A. J., and Hoyland, J. (2005). The Role of Interleukin-1 in the Pathogenesis of Human Intervertebral Disc Degeneration. *Arthritis Res. Ther.* 7, R732–R745. doi:10.1186/ar1732

Le Maitre, C., Hoyland, J., and Freemont, A. J. (2007). Catabolic Cytokine Expression in Degenerate and Herniated Human Intervertebral Discs: IL-1β and TNFα Expression Profile. *Arthritis Res. Ther.* 9, R77. doi:10.1186/ar2275

Le Maitre, C. L., Frain, J., Fotheringham, A. P., Freemont, A. J., and Hoyland, J. A. (2008). Human Cells Derived from Degenerate Intervertebral Discs Respond Differently to Those Derived from Non-degenerate Intervertebral Discs Following Application of Dynamic Hydrostatic Pressure. *Biorheology* 45, 563–575. doi:10.3233/BIR-2008-0498

Le Maitre, C. L., Frain, J., Millward-Sadler, J., Fotheringham, A. P., Freemont, A. J., and Hoyland, J. A. (2009a). Altered Integrin Mechanotransduction in Human Nucleus Pulposus Cells Derived from Degenerated Discs. *Arthritis. Rheum.* 60(2), 460–469. doi:10.1002/art.24248

Le Maitre, C. L., Freemont, A. J., and Hoyland, J. A. (2009b). Expression ofCartilage-Derived Morphogenetic Protein in Human Intervertebral Discs andits Effect on Matrix Synthesis in Degenerate Human Nucleus Pulposus Cells. *Arthritis Res. Ther.* 11(5), R137. doi:10.1186/ar2808

Lee, J. M., Song, J. Y., Baek, M., Jung, H.-Y., Kang, H., Han, I. B., et al. (2011). Interleukin-1β Induces Angiogenesis and Innervation in Human Intervertebral Disc Degeneration. *J. Orthop. Res.* 29, 265–269. doi:10.1002/jor.21210

Leung, L., and Cahill, C. M. (2010). TNF-α and Neuropathic Pain - a Review. *J. Neuroinflammation* 7, 27–11. doi:10.1186/1742-2094-7-27

Li, C.-G., Liang, Q.-Q., Zhou, Q., Menga, E., Cui, X.-J., Shu, B., et al. (2009). A Continuous Observation of the Degenerative Process in the Intervertebral Disc of Smad3 Gene Knock-Out Mice. *Spine* 34, 1363–1369. doi:10.1097/BRS. OBJ

Li, H., Li, W., Zhang, L., He, J., Tang, L., Li, Z., et al. (2022). Comprehensive Network Analysis Identified SIRT7, NTRK2, and CHI3L1 as New Potential Markers for Intervertebral Disc Degeneration. *J. Oncol.* 2022, 1–17. doi:10. OBJ

Li, W., Lai, K., Chopra, N., Zheng, Z., Das, A., and Diwan, A. D. (2022). Gut-disc axis: A Cause of Intervertebral Disc Degeneration and Low Back Pain? *Eur. Spine J.* 31(4), 917–992. doi:10.1007/s00586-022-07152-8

Li, X., Lin, F., Wu, Y., Liu, N., Wang, J., Chen, R., et al. (2019). Resveratrol Attenuates Inflammation Environment-Induced Nucleus Pulposus Cell Senescence *In Vitro*. *Biosci. Rep.* 39, 30962260. doi:10.1042/ OBJ

Li, Y., Li, K., Mao, L., Han, X., Zhang, K., Zhao, C., et al. (2016). Cordycepin Inhibits LPS-Induced Inflammatory and Matrix Degradation in the Intervertebral Disc. *PeerJ* 4, e1992. doi:10.7717/peerj.1992

Li, Z., Shao, Z., Chen, S., Huang, D., Peng, Y., Chen, S., et al. (2020). TIGAR Impedes Compression-Induced Intervertebral Disc Degeneration by Suppressing Nucleus Pulposus Cell Apoptosis and Autophagy. *J. Cell. Physiology* 235, 1780–1794. doi:10.1002/jcp.29097

Lin, X., and Lin, Q. (2020). MiRNA-495-3p Attenuates TNF-α Induced Apoptosis and Inflammation in Human Nucleus Pulposus Cells by Targeting IL5RA. *Inflammation* 43, 1797–1805. doi:10.1007/s10753-020-01254-5

Lindblom, K. (1957). Intervertebral-disc Degeneration Considered as a Pressure Atrophy. *J. Bone & Jt. Surg.* 39, 933–945. doi:10.2106/00004623-195739040- OBJ

Liu, G.-Z., Ishihara, H., Osada, R., Kimura, T., and Tsuji, H. (2001). Nitric Oxide Mediates the Change of Proteoglycan Synthesis in the Human Lumbar Intervertebral Disc in Response to Hydrostatic Pressure. *Spine* 26, 134–141. OBJ

Liu, L., He, J., Liu, C., Yang, M., Fu, J., Yi, J., et al. (2021). Cartilage Intermediate Layer Protein Affects the Progression of Intervertebral Disc Degeneration by Regulating the Extracellular Microenvironment (Review). *Int. J. Mol. Med.* 47, 475–484. doi:10.3892/ijmm.2020.4832

Liu, S., Wu, N., Liu, J., Liu, H., Su, X., Liu, Z., et al. (2016). Association betweenADAMTS-4gene Polymorphism and Lumbar Disc Degeneration in Chinese Han Population. *J. Orthop. Res.* 34, 860–864. doi:10.1002/jor. 23081

Loibl, M., Wuertz-Kozak, K., Vadala, G., Lang, S., Fairbank, J., and Urban, J. P. (2019). Controversies in Regenerative Medicine: Should Intervertebral Disc Degeneration Be Treated with Mesenchymal Stem Cells? *JOR Spine* 2, e1043.

Lotz, J. C. (2004). Animal Models of Intervertebral Disc Degeneration. *Spine* 29, 2742–2750. doi:10.1097/01.brs.0000146498.04628.f9

Lü, D.-S., Shono, Y., Oda, I., Abumi, K., and Kaneda, K. (1997). Effects of Chondroitinase ABC and Chymopapain on Spinal Motion SegmentBiomechanics. *Spine* 22, 1828–1834. ; discussion 1834-5. doi:10.1097/00007632-199708150-00006

Lyons, G., Eisenstein, S. M., and Sweet, M. B. E. (1981). Biochemical Changes in Intervertebral Disc Degeneration. *Biochimica Biophysica Acta (BBA) - General Subj.* 673, 443–453. doi:10.1016/0304-4165(81)90476-1

MacLean, M. A., Touchette, C. J., Han, J. H., Christie, S. D., and Pickett, G. E. (2020). Gender Differences in the Surgical Management of Lumbar Degenerative Disease: a Scoping Review. *Spine* 32, 799–816. doi:10.3171/

Maidhof, R., Jacobsen, T., Papatheodorou, A., and Chahine, N. O. (2014). Inflammation Induces Irreversible Biophysical Changes in Isolated Nucleus Pulposus Cells. *PLoS ONE* 9, e99621. doi:10.1371/journal.pone.

Mainardi, A., Cambria, E., Occhetta, P., Martin, I., Barbero, A., Schären, S., et al. (2021). Intervertebral Disc-On-A-Chip as Advanced *In Vitro* Model for Mechanobiology Research and Drug Testing: A Review and Perspective. *Front. Bioeng. Biotechnol.* 9, 826867. doi:10.3389/fbioe.2021.826867

Malandrino, A., Noailly, J., and Lacroix, D. (2011). The Effect of Sustained Compression on Oxygen Metabolic Transport in the Intervertebral Disc Decreases with Degenerative Changes. *PLoS. Comput. Biol.* 7, e1002112.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D.J., et al. (2009). Finding the Missing Heritability of Complex Diseases. *Nature* 461, 747–753. doi:10.1038/nature08494

Masuda, K., and Lotz, J. C. (2010). New Challenges for Intervertebral Disc Treatment Using Regenerative Medicine. *Tissue Eng. Part B Rev.* 16, 147–158. doi:10.1089/ten.TEB.2009.0451

Mayer, J. E., Iatridis, J. C., Chan, D., Qureshi, S. A., Gottesman, O., and Hecht, A. C. (2013). Genetic Polymorphisms Associated with Intervertebral Disc Degeneration. *Spine J.* 13, 299–317. doi:10.1016/j.spinee.2013.01.041

McAfee, P. C., Cunningham, B., Holsapple, G., Adams, K., Blumenthal, S., Guyer, R. D., et al. (2005). A Prospective, Randomized, Multicenter Foodand Drug Administration Investigational Device Exemption Study of Lumbar Total Disc Replacement with the CHARITÉ Artificial Disc versus Lumbar Fusion. *Spine* 30, 1576–1583. doi:10.1097/01.brs.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The Ensembl Variant Effect Predictor. *Genome. Biol.* 17, 1–14. doi:10.

Merceron, C., Mangiavini, L., Robling, A., Wilson, T. L., Giaccia, A. J., Shapiro, I. M., et al. (2014). Loss of HIF-1α in the Notochord Results in Cell Death and Complete Disappearance of the Nucleus Pulposus. *PLoS. ONE* 9, e110768. doi:10.1371/journal.pone.0110768 Metzcar, J., Wang, Y., Heiland, R., and Macklin, P. (2019). A Review of Cell-Based Computational Modeling in Cancer Biology. *JCO Clin. cancer Inf.* 3, 1–13. doi:10.1200/CCI.18.00069

Milenković, T., Lai, J., and Pržulj, N. (2008). GraphCrunch: a Tool for Large Network Analyses. *Bmc. Bioinforma.* 9, 70. doi:10.1186/1471-2105-9-70

Millecamps, I., Tajerian, M., Sage, E. H., and Stone, L. S. (2011). Behavioral Signs of Chronic Back Pain in the SPARC-Null Mouse. *Spine* 36, 95–102. doi:10.1097/

Millecamps, M., Tajerian, M., Naso, L., Sage, H. E., and Stone, L. S. (2012). Lumbar Intervertebral Disc Degeneration Associated with Axial and Radiating LowBack Pain in Ageing SPARC-Null Mice. *Pain* 153, 1167–1179. doi:10.1016/j.

Millward-Sadler, S. J., Costello, P. W., Freemont, A. J., and Hoyland, J. A. (2009). Regulation of Catabolic Gene Expression in Normal and Degenerate Human Intervertebral Disc Cells: Implications for the Pathogenesis of Intervertebral Disc Degeneration. *Arthritis Res. Ther.* 11, R65. doi:10.1186/ar2693

Min, S.-K., Nakazato, K., Yamamoto, Y., Gushiken, K., Fujimoto, H., Fujishiro, H.,et al. (2010). Cartilage Intermediate Layer Protein Gene Is Associated with Lumbar Disc Degeneration in Male, but Not Female, Collegiate Athletes. *Am. J. Sports. Med.* 38, 2552–2557. doi:10.1177/0363546510376714

Mio, F., Chiba, K., Hirose, Y., Kawaguchi, Y., Mikami, Y., Oya, T., et al. (2007). A Functional Polymorphism in COL11A1, Which Encodes the α1 Chain of Type XI Collagen, Is Associated with Susceptibility to Lumbar Disc Herniation. *Am. J. Hum. Genet.* 81, 1271–1277. doi:10.1086/522377

Miyagi, M., Millecamps, M., Danco, A. T., Ohtori, S., Takahashi, K., and Stone, L. S. (2014). ISSLS Prize Winner. *Spine* 39, 1345–1354. doi:10.1097/BRS.0000000000000334

Miyazaki, T., Kobayashi, S., Takeno, K., Meir, A., Urban, J., and Baba, H. (2009). A Phenotypic Comparison of Proteoglycan Production of Intervertebral Disc Cells Isolated from Rats, Rabbits, and Bovine Tails; Which Animal Model Is Most Suitable to Study Tissue Engineering and Biological Repair of Human DiscDisorders? *Tissue Eng. Part A* 15, 3835–3846. doi:10.1089/ten.tea.2009.0250

Moon, S. M., Yoder, J. H., Wright, A. C., Smith, L. J., Vresilovic, E. J., and Elliott, D. M. (2013). Evaluation of Intervertebral Disc Cartilaginous Endplate Structure Using Magnetic Resonance Imaging. *Eur. Spine J.* 22, 1820–1828. doi:10.1007/

Moore, R. J. (2000). The Vertebral End-Plate: what Do We Know? *Eur. Spine J.* 9, 92–96. doi:10.1007/s005860050217

Mountjoy, E., Schmidt, E. M., Carmona, M., Schwartzentruber, J., Peat, G., Miranda, A., et al. (2021). An Open Approach to Systematically Prioritize Causal Variants and Genes at All Published Human GWAS Trait-Associated Loci. *Nat. Genet.* 53, 1527–1533. doi:10.1038/s41588-021-00945-5

Mu, J., Ge, W., Zuo, X., Chen, Y., and Huang, C. (2013). Analysis of Association between IL-1β, CASP-9, and GDF5 Variants and Low-Back Pain in Chinese Male Soldiers. *Spi* 19, 243–247. doi:10.3171/2013.4.SPINE12782

Munir, S., Freidin, M. B., Rade, M., Määttä, J., Livshits, G., and Williams, F. M. K. (2018). Endplate Defect Is Heritable, Associated with Low Back Pain andTriggers Intervertebral Disc Degeneration. *Spine* 43, 1496–1501. doi:10.1097/

Näkki, A., Videman, T., Kujala, U. M., Suhonen, M., Männikkö, M., Peltonen, L.,et al. (2011). Candidate Gene Association Study of Magnetic Resonance Imaging-Based Hip Osteoarthritis (OA): Evidence for COL9A2 Gene as aCommon Predisposing Factor for Hip OA and Lumbar Disc Degeneration.*J. Rheumatol.* 38, 747–752. doi:10.3899/jrheum.100080

Naqvi, S. M., and Buckley, C. T. (2015). Differential Response of Encapsulated Nucleus Pulposus and Bone Marrow Stem Cells in Isolation and Coculture in Alginate and Chitosan Hydrogels. *Tissue Eng. Part A* 21, 288–299. doi:10.1089/

Neidlinger-Wilke, C., Mietsch, A., Rinkler, C., Wilke, H.-J., Ignatius, A., and Urban, J. (2012). Interactions of Environmental Conditions and Mechanical Loads Have Influence on Matrix Turnover by Nucleus Pulposus Cells. *J. Orthop. Res.* 30, 112–121. doi:10.1002/jor.21481

Newell, N., Little, J., Christou, A., Adams, M., Adam, C., and Masouros, S. (2017). Biomechanics of the Human Intervertebral Disc: A Review of Testing Techniques and Results. *J. Mech. Behav. Biomed. Mater.* 69, 420–434.

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E., and Cox, N. J. (2010). Trait-associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS Genet.* 6, e1000888. doi:10.1371/journal.

Niemeyer, F., Galbusera, F., Tao, Y., Kienle, A., Beer, M., and Wilke, H.-J. (2021). A Deep Learning Model for the Accurate and Reliable Classification of DiscDegeneration Based on MRI Data. *Invest. Radiol.* 56, 78–85. doi:10.1097/RLI.

Noponen-Hietala, N., Kyllönen, E., Männikkö, M., Ilkko, E., Karppinen, J., Ott, J., et al. (2003). Sequence Variations in the Collagen IX and XI Genes Are Associated with Degenerative Lumbar Spinal Stenosis. *Ann. Rheumatic Dis.* 62, 1208–1214. doi:10.1136/ard.2003.008334

Noponen-Hietala, N., Virtanen, I., Karttunen, R., Schwenke, S., Jakkula, E., Li, H., et al. (2005). Genetic Variations in IL6 Associate with Intervertebral Disc Disease Characterized by Sciatica. *Pain* 114, 186–194. doi:10.1016/j.pain.2004.

Novais, E. J., Choi, H., Madhu, V., Suyama, K., Anjo, S. I., Manadas, B., et al. (2021). Hypoxia and Hypoxia-Inducible Factor-1α Regulate Endoplasmic Reticulum

Stress in Nucleus Pulposus Cells. *Am. J. Pathology* 191, 487–502. doi:10.1016/j. OBJ

O′Connell, G. D., Vresilovic, E. J., and Elliott, D. M. (2007). Comparison of Animals Used in Disc Research to Human Lumbar Disc Geometry. *Spine* 32, 328–333. doi:10.1097/01.brs.0000253961.40910.c1

Oliver, J. C., Bland, L. A., Oettinger, C. W., Arduino, M. J., McAllister, S. K., Aguero, S. M., et al. (1993). Cytokine Kinetics in an *In Vitro* Whole Blood Model Following an Endotoxin Challenge. *Lymphokine Cytokine Res.* 12, 115–120.

Omair, A., Holden, M., Lie, B. A., Reikeras, O., and Brox, J. I. (2013). Treatment Outcome of Chronic Low Back Pain and Radiographic Lumbar Disc Degeneration Are Associated with Inflammatory and Matrix Degrading Gene Variants: a Prospective Genetic Association Study. *BMC Musculoskelet. Disord.* 14, 1–9. doi:10.1186/1471-2474-14-105

Oshima, H., Ishihara, H., Urban, J. P. G., and Tsuji, H. (1993). The Use ofCoccygeal Discs to Study Intervertebral Disc Metabolism. *J. Orthop. Res.* 11, 332–338. doi:10.1002/jor.1100110304

Oswald, K. A. C., Bigdon, S. F., Croft, A. S., Bermudez-Lekerika, P., Bergadano, A., Gantenbein, B., et al. (2021). Establishment of a Novel Method for Spinal Discectomy Surgery in Elderly Rats in an *In Vivo* Spinal Fusion Model. *MPs* 4, 79. doi:10.3390/mps4040079

Öztürk, E., ArlovAksel, Ø., Aksel, S., Li, L., Ornitz, D. M., Skjåk-Braek, G., et al. (2016). Sulfated Hydrogel Matrices Direct Mitogenicity and Maintenance of Chondrocyte Phenotype through Activation of FGF Signaling. *Adv. Funct. Mat.* 26, 3649–3662. doi:10.1002/adfm.201600092

Paassilta, P., Lohiniva, J., Göring, H. H., Perälä, M., Räinä, S. S., Karppinen, J., et al. (2001). Identification of a Novel Common Genetic Risk Factor for Lumbar Disk Disease. *JAMA* 285, 1843–1849. doi:10.1001/jama.285.14.1843

Parenteau, C. S., Lau, E. C., Campbell, I. C., and Courtney, A. (2021). Prevalence of Spine Degeneration Diagnosis by Type, Age, Gender, and Obesity Using Medicare Data. *Sci. Rep.* 11, 5389. doi:10.1038/s41598-021-84724-6

Pattappa, G., Peroglio, M., Peroglio, M., Sakai, D., Mochida, J., Benneker, L., et al. (2014). CCL5/RANTES Is a Key Chemoattractant Released by Degenerative Intervertebral Discs in Organ Culture. *eCM* 27, 124–136. ; discussion 136. OBJ

Paul, C. P. L., Schoorl, T., Zuiderbaan, H. A., Zandieh Doulabi, B., van der Veen, A.J., van de Ven, P. M., et al. (2013). Dynamic and Static Overloading Induce Early Degenerative Processes in Caprine Lumbar Intervertebral Discs. *PLoS ONE* 8, e62411. doi:10.1371/journal.pone.0062411

Paul, C. P. L., Zuiderbaan, H. A., Zandieh Doulabi, B., van der Veen, A. J., van de Ven, P. M., Smit, T. H., et al. (2012). Simulated-physiological Loading Conditions Preserve Biological and Mechanical Properties of Caprine Lumbar Intervertebral Discs in *Ex Vivo* Culture. *PLoS ONE* 7, e33147. OBJ

Pearce, D., Fischer, S., Huda, F., and Vahdati, A. (2020). Applications of Computer Modeling and Simulation in Cartilage Tissue Engineering. *Tissue. Eng. Regen. Med.* 17, 1–13. doi:10.1007/s13770-019-00216-9

Pfaff, T., Fortunato, M., Sanchez-Gonzalez, A., and Battaglia, P. W. (2020). Learning Mesh-Based Simulation with Graph Networks. doi:10.48550/ OBJ

Pfannkuche, J.-J., Guo, W., Cui, S., Ma, J., Lang, G., Peroglio, M., et al. (2020). Intervertebral Disc Organ Culture for the Investigation of Disc Pathology and Regeneration - Benefits, Limitations, and Future Directions of Bioreactors.*Connect. Tissue Res.* 61, 304–321. doi:10.1080/03008207.2019.1665652

Phillips, F. M., Reuben, J., and Wetzel, F. T. (2002). Intervertebral Disc Degeneration Adjacent to a Lumbar Fusion. An Experimental Rabbit Model. *J. Bone Jt. Surg. Br. volume* 84, 289–294. doi:10.1302/0301-620x. OBJ

Phillips, K. L. E., Chiverton, N., Michael, A. L., Cole, A. A., Breakwell, L. M., Haddock, G., et al. (2013a). The Cytokine and Chemokine Expression Profile of Nucleus Pulposus Cells: Implications for Degeneration and Regeneration of the Intervertebral Disc. *Arthritis Res. Ther.* 15, R213. doi:10.1186/ar4408

Phillips, K. L. E., Cullen, K., Chiverton, N., Michael, A. L. R., Cole, A. A., Breakwell, L. M., et al. (2015). Potential Roles of Cytokines and Chemokines in Human Intervertebral Disc Degeneration: Interleukin-1 Is a Master Regulator of Catabolic Processes. *Osteoarthr. Cartil.* 23, 1165–1177. doi:10.1016/j.joca. 2015.02.017

Phillips, K. L. E., Jordan-Mahy, N., Nicklin, M. J. H., and Le Maitre, C. L. (2013b). Interleukin-1 Receptor Antagonist Deficient Mice Provide Insights into Pathogenesis of Human Intervertebral Disc Degeneration. *Ann. Rheum. Dis.*72(11), 1860–1867. doi:10.1136/annrheumdis-2012-202266

Pincus, T., Burton, A. K., Vogel, S., and Field, A. P. (2002). A Systematic Review of Psychological Factors as Predictors of Chronicity/disability in Prospective Cohorts of Low Back Pain. *Spine* 27, E109–E120. doi:10.1097/00007632- OBJ

Pluijm, S. M. F., Hw, V. E., Bravenboer, N., Uitterlinden, A. G., Smit, J. H., Pols, H.A., et al. (2004). Collagen Type I 1 Sp1 Polymorphism, Osteoporosis, and Intervertebral Disc Degeneration in Older Men and Women. *Ann. RheumaticDis.* 63, 71–77. doi:10.1136/ard.2002.002287

Pockert, A. J., Richardson, S. M., Le Maitre, C. L., Lyon, M., Deakin, J. A., Buttle, D. J., et al. (2009). Modified Expression of the ADAMTS Enzymes and Tissue Inhibitor of Metalloproteinases 3 during Human Intervertebral Disc Degeneration. *Arthritis Rheum.* 60, 482–491. doi:10. OBJ

Purmessur, D., Freemont, A. J., and Hoyland, J. A. (2008). Expression and Regulation of Neurotrophins in the Nondegenerate and Degenerate Human Intervertebral Disc. *Arthritis Res. Ther.* 10, R99. doi:10.1186/ar2487

Raine, E. V. A., Reynard, L. N., van de Laar, I. M. B. H., Bertoli-Avella, A. M., and Loughlin, J. (2014). Identification and Analysis of a SMAD3 Cis-Acting eQTL Operating in Primary Osteoarthritis and in the Aneurysms and Osteoarthritis Syndrome. *Osteoarthr. Cartil.* 22, 698–705. doi:10.1016/j. OBJ

Rajasekaran, S., Kanna, R. M., Senthil, N., Raveendran, M., Ranjani, V., Cheung, K. M. C., et al. (2014). Genetic Susceptibility of Lumbar Degenerative Disc Diseasein Young Indian Adults. *Eur. Spine J.* 24, 1969–1975. doi:10.1007/s00586-014- OBJ

Rajasekaran, S., Soundararajan, D. C. R., Tangavel, C., Muthurajan, R., Sri Vijay Anand, K. S., Matchado, M. S., et al. (2020). Human Intervertebral Discs Harbour a Unique Microbiome and Dysbiosis Determines Health and Disease. *Eur. Spine J.* 29, 1621–1640. doi:10.1007/s00586-020-06446-z

Rastogi, A., Thakore, P., Leung, A., Benavides, M., Machado, M., Morschauser, M. A., et al. (2009). Environmental Regulation of Notochordal Gene Expression in Nucleus Pulposus Cells. *J. Cell. Physiol.* 220, 698–705. doi:10.1002/jcp.21816

Rätsep, T., Minajeva, A., and Asser, T. (2013). Relationship between Neovascularization and Degenerative Changes in Herniated Lumbar Intervertebral Discs. *Eur. Spine J.* 22, 2474–2480. doi:10.1007/s00586-013- OBJ

Rijsbergen, M. V., van Rietbergen, B., Barthelemy, V., Eltes, P., Lazáry, Á., Lacroix, D., et al. (2018). Comparison of Patient-specific Computational Models vs. Clinical Follow-Up, for Adjacent Segment Disc Degeneration and Bone Remodelling after Spinal Fusion. *PLoS ONE* 13, e0200899. doi:10.1371/ OBJ

Rikard, S. M., Athey, T. L., Nelson, A. R., Christiansen, S. L. M., Lee, J.-J., Holmes, J. W., et al. (2019). Multiscale Coupling of an Agent-Based Model of Tissue Fibrosis and a Logic-Based Model of Intracellular Signaling. *Front. Physiol.* 10, 1481. doi:10.3389/fphys.2019.01481

Rim, D. C. (2016). Quantitative Pfirrmann Disc Degeneration Grading System to Overcome the Limitation of Pfirrmann Disc Degeneration Grade. *Korean.J. Spine* 13, 1–8. doi:10.14245/kjs.2016.13.1.1

Risbud, M. V., and Shapiro, I. M. (2013). Role of Cytokines in Intervertebral Disc Degeneration: Pain and Disc Content. *Nat. Rev. Rheumatol.* 10, 44–56. doi:10. OBJ

Roberts, S., Evans, H., Trivedi, J., and Menage, J. (2006). Histology and Pathologyof the Human Intervertebral Disc. *J. Bone. Jt. Surg. Am.* 88 (Suppl. 2), 10–14. OBJ

Roberts, S., Menage, J., and Urban, J. P. G. (1989). Biochemical and Structural Properties of the Cartilage End-Plate and its Relation to the Intervertebral Disc. *Spine* 14, 166–174. doi:10.1097/00007632-198902000-00005

Robinson, M. R., English, G., English, G., Moser, G., Lloyd-Jones, L. R., Triplett, M. A., et al. (2017). Genotype-covariate Interaction Effects and the Heritability of Adult Body Mass Index. *Nat. Genet.* 49, 1174–1181. doi:10. OBJ

Romanoski, C. E., Glass, C. K., Stunnenberg, H. G., Wilson, L., and Almouzni, G. (2015). Roadmap for Regulation. *Nature* 518, 314–316. doi:10.1038/ 518314a

Rosiak, P., Latanska, I., Paul, P., Sujka, W., and Kolesinska, B. (2021). Modification of Alginates to Modulate Their Physic-Chemical Properties and Obtain Biomaterials with Different Functional Properties. *Molecules* 26, 7264. OBJ

Ruiz Wills, C., Foata, B., González Ballester, M. Á., Karppinen, J., and Noailly, J. (2018). Theoretical Explorations Generate New Hypotheses about the Role of the Cartilage Endplate in Early Intervertebral Disk Degeneration. *Front. Physiol.* 9, 1210. doi:10.3389/fphys.2018.01210

Sadowska, A., Altinay, B., Hitzl, W., Ferguson, S. J., and Wuertz-Kozak, K. (2020). Hypo-Osmotic Loading Induces Expression of IL-6 in Nucleus Pulposus Cells of the Intervertebral Disc Independent of TRPV4 and TRPM7. *Front. Pharmacol.* 11, 952. doi:10.3389/fphar.2020.00952

Sadowska, A., Kameda, T., Kameda, T., Krupkova, O., and Wuertz-Kozak, K. (2018). Osmosensing, Osmosignalling and Inflammation: How Intervertebral Disc Cells Respond to Altered Osmolarity. *eCM* 36, 231–250. doi:10.22203/ OBJ

Saggese, T., Thambyah, A., Wade, K., and McGlashan, S. R. (2020). Differential Response of Bovine Mature Nucleus Pulposus and Notochordal Cells to Hydrostatic Pressure and Glucose Restriction. *Cartilage* 11, 221–233. doi:10. OBJ

Salzer, E., Mouser, V. H. M., Tryfonidou, M. A., and Ito, K. (2021). A Bovine Nucleus Pulposus Explant Culture Model. *J. Orthop. Res* [Epub ahead of print]. doi:10.1002/jor.25226

Samartzis, D., Karppinen, J., Chan, D., Luk, K. D. K., and Cheung, K. M. C. (2012). The Association of Lumbar Intervertebral Disc Degeneration on Magnetic Resonance Imaging with Body Mass Index in Overweight and Obese Adults: a Population-Based Study. *Arthritis & Rheumatism* 64, 1488–1496. doi:10.1002/ OBJ

Sambrook, P. N., MacGregor, A. J., and Spector, T. D. (1999). Genetic Influences on Cervical and Lumbar Disc Degeneration: A Magnetic Resonance Imaging Study in Twins. *Arthritis & Rheumatism* 42, 366–372. doi:10.1002/1529-0131(199902)42:2<366::aid-anr20>3.0.co;2-6

Saravi, B., Lang, G., Grad, S., Alini, M., Richards, R. G., Schmal, H., et al. (2021). A Proinflammatory, Degenerative Organ Culture Model to Simulate Early-Stage Intervertebral Disc Disease. *JoVE* 168. doi:10.3791/62100

Schaid, D. J., Chen, W., and Larson, N. B. (2018). From Genome-wide Associationsto Candidate Causal Variants by Statistical Fine-Mapping. *Nat. Rev. Genet.* 19, 491–504. doi:10.1038/s41576-018-0016-z

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological Insights from 108 Schizophrenia-Associated Genetic Loci. *Nature* 511, 421–427. doi:10.1038/nature13595

Schreiber, F., and Schwöbbermeyer, H. (2005). MAVisto: a Tool for the Exploration of Network Motifs. *Bioinformatics* 21, 3572–3574. doi:10.1093/ OBJ

Seki, S., Kawaguchi, Y., Chiba, K., Mikami, Y., Kizawa, H., Oya, T., et al. (2005). A Functional SNP in CILP, Encoding Cartilage Intermediate Layer Protein, Is Associated with Susceptibility to Lumbar Disc Disease. *Nat. Genet.* 37, 607–612. OBJ

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Showalter, B. L., Beckstein, J. C., Martin, J. T., Beattie, E. E., Orías, A. A. E., Schaer, T.P., et al. (2012). Comparison of Animal Discs Used in Disc Research to Human Lumbar Disc. *Spine* 37, E900–E907. doi:10.1097/BRS.0b013e31824d911c

Smith, L., Chiaro, J. A., Chiaro, J., Nerurkar, N., Cortes, D., Horava, S., et al. (2011). Nucleus Pulposus Cells Synthesize a Functional Extracellular Matrix and Respond to Inflammatory Cytokine Challenge Following Long-Term Agarose Culture. *eCM* 22, 291–301. doi:10.22203/ecm.v022a22

Snuggs, J., Bunning, R. A., Bunning, R., and Le Maitre, C. (2021). Osmotic Adaptation of Nucleus Pulposus Cells: the Role of Aquaporin 1, Aquaporin4 and Transient Receptor Potential Vanilloid 4. *eCM* 41, 121–141. doi:10. OBJ

Solovieva, S., Kouhia, S., Leino-Arjas, P., Ala-Kokko, L., Luoma, K., Raininko, R., et al. (2004). Interleukin 1 Polymorphisms and Intervertebral Disc Degeneration. *Epidemiology* 15, 626–633. doi:10.1097/01.ede.0000135179.04563.35

Solovieva, S., Lohiniva, J., Leino-Arjas, P., Raininko, R., Luoma, K., Ala-Kokko, L.,et al. (2006). Intervertebral Disc Degeneration in Relation to the COL9A3 and the IL-1ß Gene Polymorphisms. *Eur. Spine J.* 15, 613–619. doi:10.1007/s00586- OBJ

Solovieva, S., Noponen, N., Männikkö, M., Leino-Arjas, P., Luoma, K., Raininko, R., et al. (2007). Association between the Aggrecan Gene Variable Number of Tandem Repeats Polymorphism and Intervertebral Disc Degeneration. *Spine* 32, 1700–1705. doi:10.1097/BRS.0b013e3180b9ed51

Song, Y.-Q., Cheung, K. M. C., Ho, D. W. H., Poon, S. C. S., Chiba, K., Kawaguchi,Y., et al. (2008). Association of the Asporin D14 Allele with Lumbar-Disc Degeneration in Asians. *Am. J. Hum. Genet.* 82, 744–747. doi:10.1016/j.ajhg. OBJ

Song, Y.-Q., Karasugi, T., Cheung, K. M. C., Chiba, K., Ho, D. W. H., Miyake, A., et al. (2013). Lumbar Disc Degeneration Is Linked to a Carbohydrate Sulfotransferase 3 Variant. *J. Clin. Invest.* 123, 4909–4917. doi:10.1172/ OBJ

Stirling, A., Worthington, T., Rafiq, M., Lambert, P. A., and Elliott, T. S. (2001). Association between Sciatica and *Propionibacterium Acnes*. *Lancet* 357, 2024–2025. doi:10.1016/S0140-6736(00)05109-6

Sun, Q., Crowley, C. A., Huang, L., Wen, J., Chen, J., Bao, E. L., et al. (2022). From GWAS Variant to Function: A Study of ~148,000 Variants for Blood Cell Traits. *Hum. Genet. Genomics Adv.* 3, 100063. doi:10.1016/j.xhgg.2021. OBJ

Sun, Y., Lv, M., Zhou, L., Tam, V., Lv, F., Chan, D., et al. (2015). Enrichment of Committed Human Nucleus Pulposus Cells Expressing Chondroitin Sulfate Proteoglycans under Alginate Encapsulation. *Osteoarthr. Cartil.* 23, 1194–1203. OBJ

Sun, Z., Zhang, M., Zhao, X. H., Liu, Z. H., Gao, Y., Samartzis, D., et al. (2013). Immune Cascades in Human Intervertebral Disc: the Pros and Cons. *Int. J. Clin. Exp. Pathol.* 6, 1009–1014.

Sun, Z.-M., Ling, M., Huo, Y., Chang, Y., Li, Y., Qin, H., et al. (2011). Caspase 9 Gene Polymorphism and Susceptibility to Lumbar Disc Disease in the Han Population in Northern China. *Connect. Tissue Res.* 52, 198–202. doi:10.3109/ OBJ

Sun, Z.-m., Miao, L., Zhang, Y.-g., and Ming, L. (2009). Association between the -1562 C/T Polymorphism of Matrix Metalloproteinase-9 Gene and Lumbar Disc Disease in the Young Adult Population in North China. *Connect. Tissue Res.* 50, 181–185. doi:10.1080/03008200802585630

Sun, Z., Liu, B., and Luo, Z.-J. (2020). The Immune Privilege of the Intervertebral Disc: Implications for Intervertebral Disc Degeneration Treatment. *Int. J. Med. Sci.* 17, 685–692. doi:10.7150/ijms.42238

Suyama, K., Sakai, D., Hirayama, N., Nakamura, Y., Matsushita, E., Terayama, H.,et al. (2018). Effects of interleukin-17A in Nucleus Pulposus Cells and its Small- Molecule Inhibitors for Intervertebral Disc Disease. *J. Cell. Mol. Med.* 22, 5539–5551. doi:10.1111/jcmm.13828

Swärd, L., Hellström, M., Jacobsson, B. O., Nyman, R., and Peterson, L. (1991). Disc Degeneration and Associated Abnormalities of the Spine in Elite Gymnasts. *Spine* 16, 437–443. doi:10.1097/00007632-199104000-00009

Tabor, H. K., Risch, N. J., and Myers, R. M. (2002). Candidate-gene Approaches for Studying Complex Genetic Traits: Practical Considerations. *Nat. Rev. Genet.* 3, 391–397. doi:10.1038/nrg796

Takahashi, M., Haro, H., Wakabayashi, Y., Kawa-uchi, T., Komori, H., and Shinomiya, K. (2001). The Association of Degeneration of the Intervertebral Disc with 5a/6a Polymorphism in the Promoter of the Human Matrix Metalloproteinase-3 Gene. *J. Bone Jt. Surg. Br. volume* 83, 491–495. doi:10. OBJ

Tan, H., Zhao, J., Jiang, J., and Ren, Y. (2012). Association of the Polymorphism of DR4 with the Risk and Severity of Lumbar Disc Degeneration in the Chinese Han Population. *Scand. J. Clin. Laboratory Investigation* 72, 576–579. doi:10. OBJ

Tan, N. Y., and Khachigian, L. M. (2009). Sp1 Phosphorylation and its Regulation of Gene Transcription. *Mol. Cell. Biol.* 29, 2483–2488. doi:10.1128/MCB.01828-08

Tapias, A., Ciudad, C. J., Roninson, I. B., and Noé, V. (2008). Regulation of Sp1 by Cell Cycle Related Proteins. *Cell. Cycle.* 7, 2856–2867. doi:10.4161/cc.7.18.6671

Teraguchi, M., Yoshimura, N., Hashizume, H., Muraki, S., Yamada, H., Minamide, A., et al. (2014). Prevalence and Distribution of Intervertebral Disc Degeneration over the Entire Spine in a Population-Based Cohort: the Wakayama Spine Study. *Osteoarthr. Cartil.* 22, 104–110. doi:10.1016/j.joca. 2013.10.019

The GTEX consortium (2020). The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues. *Science* 369, 1318–1330. doi:10.1126/science. OBJ

Thorpe, A. A., Bach, F. C., Tryfonidou, M. A., Le Maitre, C. L., Mwale, F., Diwan, A. D., et al. (2018). Leaping the Hurdles in Developing Regenerative Treatments for the Intervertebral Disc from Preclinical to Clinical. *JOR. Spine* 1, e1027. OBJ

Tilkeridis, C., Bei, T., Garantziotis, S., and Stratakis, C. A. (2005). Association of a COL1A1 Polymorphism with Lumbar Disc Disease in Young Military Recruits. *J. Med. Genet.* 42, e44. doi:10.1136/jmg.2005.033225

Tolofari, S. K., Richardson, S. M., Freemont, A. J., and Hoyland, J. A. (2010). Expression of Semaphorin 3A and its Receptors in the HumanIntervertebral Disc: Potential Role in Regulating Neural Ingrowth in the Degenerate Intervertebral Disc. *Arthritis Res. Ther.* 12, R1–R11. doi:10. OBJ

Tsai, T.-T., Danielson, K. G., Guttapalli, A., Oguz, E., Albert, T. J., Shapiro, I. M., et al. (2006). TonEBP/OREBP Is a Regulator of Nucleus Pulposus Cell Functionand Survival in the Intervertebral Disc. *J. Biol. Chem.* 281, 25416–25424. doi:10. OBJ

Turgut, M., Uslu, S., Uysal, A. l., Yurtseven, M. E., and Ustün, H. (2003). Changesin Vascularity of Cartilage Endplate of Degenerated Intervertebral Discs in Response to Melatonin Administration in Rats. *Neurosurg. Rev.* 26, 133–138. OBJ

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R.,et al. (2021). Genome-wide Association Studies. *Nat. Rev. Methods Prim.* 1,1–21. doi:10.1038/s43586-021-00056-9

Ulrich, J. A., Liebenberg, E. C., Thuillier, D. U., and Lotz, J. C. (2007). ISSLS Prize Winner: Repeated Disc Injury Causes Persistent Inflammation. *Spine* 32, 2812–2819. doi:10.1097/BRS.0b013e31815b9850

Urban, J. P. G., Smith, S., and Fairbank, J. C. T. (2004). Nutrition of the Intervertebral Disc. *Spine* 29, 2700–2709. doi:10.1097/01.brs.0000146499. 97948.52

Urban, J. P. G. (2002). The Role of the Physicochemical Environment in Determining Disc Cell Behaviour. *Biochem. Soc. Trans.* 30, 858–863. doi:10. OBJ

Valdes, A. M., Hassett, G., Hart, D. J., and Spector, T. D. (2005). Radiographic Progression of Lumbar Spine Disc Degeneration Is Influenced by Variation at Inflammatory Genes. *Spine* 30, 2445–2451. doi:10.1097/01.brs.0000184369. OBJ

van Uden, S., Silva-Correia, J., Oliveira, J. M., and Reis, R. L. (2017). Current Strategies for Treatment of Intervertebral Disc Degeneration: Substitution and Regeneration Possibilities. *Biomater. Res.* 21, 22. doi:10.1186/s40824-017- OBJ

Vazirinejad, R., Ahmadi, Z., Kazemi Arababadi, M., Hassanshahi, G., and Kennedy, D. (2014). The Biological Functions, Structure and Sources of CXCL10 and its Outstanding Part in the Pathophysiology of Multiple Sclerosis. *Neuroimmunomodulation* 21, 322–330. doi:10.1159/000357780

Vergroesen, P.-P. A., Kingma, I., Emanuel, K. S., Hoogendoorn, R. J. W., Welting, T. J., van Royen, B. J., et al. (2015). Mechanics and Biology in Intervertebral Disc Degeneration: a Vicious Circle. *Osteoarthr. Cartil.* 23, 1057–1070. doi:10.1016/j. OBJ

Vergroesen, P.-P. A., van der Veen, A. J., van Royen, B. J., Kingma, I., and Smit, T. H. (2014). Intradiscal Pressure Depends on Recent Loading and Correlates with Disc Height and Compressive Stiffness. *Eur. Spine J.* 23, 2359–2368. doi:10. OBJ

Verrecchia, F., and Mauviel, A. (2004). TGF-β and TNF-α: Antagonistic Cytokines Controlling Type I Collagen Gene Expression. *Cell. Signal.* 16, 873–880. doi:10. 1016/j.cellsig.2004.02.007 Videman, T., and Battié, M. C. (1999). Spine Update. *Spine* 24, 1164–1168. doi:10. 1097/00007632-199906010-00020

Videman, T., Saarela, J., Kaprio, J., Näkki, A., Levälahti, E., Gill, K., et al. (2009). Associations of 25 Structural, Degradative, and Inflammatory Candidate Genes with Lumbar Disc Desiccation, Bulging, and Height Narrowing. *ArthritisRheum.* 60, 470–481. doi:10.1002/art.24268

Virgilio, K. M., Martin, K. S., Peirce, S. M., and Blemker, S. S. (2018). Agent-based Model Illustrates the Role of the Microenvironment in Regeneration in Healthy and Mdx Skeletal Muscle. *J. Appl. Physiology* 125, 1424–1439. doi:10.1152/ japplphysiol.00379.2018

Virtanen, I. M., Karppinen, J., Taimela, S., Ott, J., Barral, S., Kaikkonen, K., et al. (2007). Occupational and Genetic Risk Factors Associated with Intervertebral Disc Disease. *Spine* 32, 1129–1134. doi:10.1097/01.brs. OBJ

Vital, J.-M., and Boissière, L. (2014). Total Disc Replacement. *Orthop. Traumatology Surg. Res.* 100, S1–S14. doi:10.1016/j.otsr.2013.06.018

Volz, M., Elmasry, S., Jackson, A. R., and Travascio, F. (2022). Computational Modeling Intervertebral Disc Pathophysiology: A Review. *Front. Physiol.* 12, 750668. doi:10.3389/fphys.2021.750668

Walter, B. A., Illien-Jünger, S., Nasser, P. R., Hecht, A. C., and Iatridis, J. C. (2014). Development and Validation of a Bioreactor System for Dynamic Loading and Mechanical Characterization of Whole Human Intervertebral Discs in Organ Culture. *J. Biomechanics* 47, 2095–2101. doi:10.1016/j. OBJ

Walter, B. A., Korecki, C. L., Purmessur, D., Roughley, P. J., Michalek, A. J., and Iatridis, J. C. (2011). Complex Loading Affects Intervertebral Disc Mechanics and Biology. *Osteoarthr. Cartil.* 19, 1011–1018. doi:10.1016/j.joca.2011. OBJ

Walter, B., Purmessur, D., Purmessur, D., Moon, A., Occhiogrosso, J., Laudier, D., et al. (2016). Reduced Tissue Osmolarity Increases TRPV4 Expression and Pro-inflammatory Cytokines in Intervertebral Disc Cells. *eCM* 32, 123–136. doi:10. OBJ

Wang, H., Zhou, Y., Huang, B., Liu, L.-T., Liu, M.-H., Wang, J., et al. (2014). Utilization of Stem Cells in Alginate for Nucleus Pulposus Tissue Engineering. *Tissue Eng. Part A* 20, 908–920. doi:10.1089/ten.TEA.2012.0703

Wang, M., Tang, D., Shu, B., Wang, B., Jin, H., Hao, S., et al. (2012). Conditional Activation of β-catenin Signaling in Mice Leads to Severe Defects in Intervertebral Disc Tissue. *Arthritis & Rheumatism* 64, 2611–2623. doi:10. 1002/art.34469

Wang, S.-L., Yu, Y.-L., Tang, C.-L., and Lv, F.-Z. (2013). Effects of TGF-B1 and IL-1β on Expression of ADAMTS Enzymes and TIMP-3 in HumanIntervertebral Disc Degeneration. *Exp. Ther. Med.* 6, 1522–1526. doi:10. 3892/etm.2013.1348

Wáng, Y. X. J., Wáng, J.-Q., and Káplár, Z. (2016). Increased Low Back Pain Prevalence in Females Than in Males after Menopause Age: Evidences Based on Synthetic Literature Review. *Quant. Imaging Med. Surg.* 6, 199–206. doi:10. OBJ

Wangler, S., Menzel, U., Li, Z., Ma, J., Hoppe, S., Benneker, L. M., et al. (2019). CD146/MCAM Distinguishes Stem Cell Subpopulations with Distinct Migration and Regenerative Potential in Degenerative Intervertebral Discs. *Osteoarthr. Cartil.* 27, 1094–1105. doi:10.1016/j.joca.2019.04.002

Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional Mapping and Annotation of Genetic Associations with FUMA. *Nat. Commun.* 8, 1–11. doi:10.1038/s41467-017-01261-5

Weiler, C., Nerlich, A. G., Bachmeier, B. E., and Boos, N. (2005). Expression and Distribution of Tumor Necrosis Factor Alpha in Human Lumbar Intervertebral Discs: A Study in Surgical Specimen and Autopsy Controls. *Spine* 30, 44–53. OBJ

Wernicke, S., and Rasche, F. (2006). FANMOD: a Tool for Fast Network Motif Detection. *Bioinformatics* 22, 1152–1153. doi:10.1093/bioinformatics/ OBJ

Wilke, H. J., Neef, P., Caimi, M., Hoogland, T., and Claes, L. E. (1999). New *In Vivo* Measurements of Pressures in the Intervertebral Disc in Daily Life. *Spine* 24, 755–762. doi:10.1097/00007632-199904150-00005

Williams, F. M. K., Bansal, A. T., van Meurs, J. B., Bell, J. T., Meulenbelt, I., Suri, P., et al. (2013). Novel Genetic Variants Associated with Lumbar Disc Degeneration in Northern Europeans: a Meta-Analysis of 4600 Subjects. *Ann. Rheum. Dis.* 72, 1141–1148. doi:10.1136/annrheumdis- OBJ

Williams, F. M. K., Popham, M., Hart, D. J., de Schepper, E., Bierma-Zeinstra, S., Hofman, A., et al. (2011a). GDF5 Single-Nucleotide Polymorphism Rs143383 Is Associated with Lumbar Disc Degeneration in Northern European Women. *Arthritis & Rheumatism* 63, 708–712. doi:10.1002/ OBJ

Williams, F. M. K., Popham, M., Sambrook, P. N., Jones, A. F., Spector, T. D., and MacGregor, A. J. (2011b). Progression of Lumbar Disc Degeneration over a Decade: a Heritability Study. *Ann. Rheumatic Dis.* 70, 1203–1207. doi:10.1136/ ard.2010.146001

Williams, R. J., Tryfonidou, M. A., Snuggs, J. W., and Le Maitre, C. L. (2021). Cell Sources Proposed for Nucleus Pulposus Regeneration. *JOR. Spine* 4, e1175. OBJ

Wöltje, M., and Böbel, M. (2017). "Natural Biodegradable Medical Polymers," in *Science and Principles of Biodegradable and Bioresorbable Medical Polymers*. Editor X. Zhang (Amsterdam, The Netherlands: Woodhead Publishing), 351–376. doi:10.1016/B978-0-08-100372-5.00012-X

Wrocklage, C., Wassmann, H., and Paulus, W. (2000). COL9A2 Allelotypes in Intervertebral Disc Disease. *Biochem. Biophysical Res. Commun.* 279, 398–400. OBJ

Wu, N., Chen, J., Liu, H., Zhao, L., Liu, S., Liu, J., et al. (2014). The Involvement of ADAMTS-5 Genetic Polymorphisms in Predisposition and Diffusion Tensor Imaging Alterations of Lumbar Disc Degeneration. *J. Orthop. Res.* 32, 686–694. OBJ

Wuertz, K., Urban, J. P. G., Klasen, J., Ignatius, A., Wilke, H.-J., Claes, L., et al. (2007). Influence of Extracellular Osmolarity and Mechanical Stimulation on Gene Expression of Intervertebral Disc Cells. *J. Orthop. Res.* 25, 1513–1522. OBJ

Xu, C., Luo, S., Wei, L., Wu, H., Gu, W., Zhou, W., et al. (2021). Integrated Transcriptome and Proteome Analyses Identify Novel Regulatory Network of Nucleus Pulposus Cells in Intervertebral Disc Degeneration. *Bmc. Med. Genomics.* 14, 1–12. doi:10.1186/s12920-021-00889-z

Xu, K., Wang, X., Zhang, Q., Liang, A., Zhu, H., Huang, D., et al. (2016). Sp1 Downregulates Proinflammatory Cytokine-Induced Catabolic Gene Expression in Nucleus Pulposus Cells. *Mol. Med. Rep.* 14, 3961–3968. doi:10.3892/mmr. OBJ

Yang, J.-E., Zhao, K.-H., Qu, Y., and Zou, Y.-C. (2022). Increased Serum CXCL10 Levels Are Associated with Clinical Severity and Radiographic Progression in Patients with Lumbar Disc Degeneration. *Clin. Chim. Acta* 525, 15–22. doi:10. OBJ

Yang, J., Ferreira, T., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A. F.,et al. (2012). Conditional and Joint Multiple-SNP Analysis of GWAS Summary Statistics Identifies Additional Variants Influencing Complex Traits. *Nat. Genet.* 44, 369–375. doi:10.1038/ng.2213

Yao, Y., Deng, Q., Song, W., Zhang, H., Li, Y., Yang, Y., et al. (2016). MIF Plays a Key Role in Regulating Tissue-specific Chondro-Osteogenic Differentiation Fate of Human Cartilage Endplate Stem Cells under Hypoxia. *Stem. Cell. Rep.* 7, 249–262. doi:10.1016/j.stemcr.2016.07.003

Yao, Y., Song, W., Deng, Q., Zhang, H., Wang, J., Liu, H., et al. (2017). General Regulatory Effects of Hypoxia on Human Cartilage Endplate-Derived Stem Cells: A Genome-wide Analysis of Differential Gene Expression and Alternative Splicing Events. *Mol. Med. Rep.* 16, 3001–3009. doi:10.3892/mmr.2017.6907

Ye, F., Lyu, F. J., Wang, H., and Zheng, Z. (2022). The Involvement of Immune System in Intervertebral Disc Herniation and Degeneration. *JOR. Spine* 5, e1196. doi:10.1002/jsp2.1196

Yoon, S.-H., Miyazaki, M., Hong, S.-W., Tow, B., Morishita, Y., Hu, M., et al. (2008). A Porcine Model of Intervertebral Disc Degeneration Induced by Annular Injury Characterized with Magnetic Resonance Imaging and Histopathological Findings. *Spi* 8, 450–457. doi:10.3171/SPI/2008/8/5/450

Yu, H., Zhu, X., Greenbaum, D., Karro, J., and Gerstein, M. (2004). TopNet: a Tool for Comparing Biological Sub-networks, Correlating Protein Properties with Topological Statistics. *Nucleic Acids Res.* 32, 328–337. doi:10.1093/nar/ OBJ

Yu, W., Fu, J., Liu, Y., Wu, Y., and Jiang, D. (2018). Osteogenic Protein-1 Inhibits Nucleus Pulposus Cell Apoptosis through Regulating the NF-Kb/ROS Pathway in an Inflammation Environment. *Biosci. Rep.* 38, BSR20181530. doi:10.1042/ OBJ

Zamanian, C., Bhandarkar, A. R., Monie, D. D., Moinuddin, F., Vile, R. G., Quiñones-Hinojosa, A., et al. (2022). Systems Neuroimmunology: a Review of Multiomics Methodologies to Characterize Neuroimmunological Interactions in Spinal and Cranial Diseases. *Neurosurg. Focus.* 52, E9. OBJ

Zeiter, S., der Werf, M. V., and Ito, K. (2009). The Fate of Bovine Bone Marrow Stromal Cells in Hydrogels: a Comparison to Nucleus Pulposus Cells and Articular Chondrocytes. *J. Tissue. Eng. Regen. Med.* 3, 310–320. doi:10.1002/ OBJ

Zhang, C., Gullbrand, S. E., Schaer, T. P., Lau, Y. K., Jiang, Z., Dodge, G. R., et al. (2020). Inflammatory Cytokine and Catabolic Enzyme Expression in a Goat Model of Intervertebral Disc Degeneration. *J. Orthop. Res.* 38, 2521–2531. OBJ

Zhang, J., Wang, X., Liu, H., Li, Z., Chen, F., Wang, H., et al. (2019). TNF-α Enhances Apoptosis by Promoting Chop Expression in Nucleus Pulposus Cells:Role of the MAPK and NF-Kb Pathways. *J. Orthop. Res.* 37, 697–705. doi:10. OBJ

Zhang, S., Ding, L., Li, X., and Fan, H. (2019). Identification of Biomarkers Associated with the Recurrence of Osteosarcoma Using ceRNA Regulatory Network Analysis. *Int. J. Mol. Med.* 43, 1723–1733. doi:10.3892/ijmm.2019.4108

Zhang, Y. g., Sun, Z., Liu, J., and Guo, X. (2008). Advances in Susceptibility Genetics of Intervertebral Degenerative Disc Disease. *Int. J. Biol. Sci.* 4, 283–290. OBJ

Zhang, Z., Li, F., Tian, H., Guan, K., Zhao, G., Shan, J., et al. (2014). Differentiationof Adipose-Derived Stem Cells toward Nucleus Pulposus-like Cells Induced by Hypoxia and a Three-Dimensional Chitosan-Alginate Gel Scaffold *In Vitro*. *Chin. Med. J. Engl.* 127, 314–321. doi:10.3760/cma.j.issn.0366-6999.20130584

Zhou, X., Cheung, C.-L., Karasugi, T., Karppinen, J., Samartzis, D., Hsu, Y.-H., et al. (2018). Trans-Ethnic Polygenic Analysis Supports Genetic Overlaps of Lumbar Disc Degeneration with Height, Body Mass Index, and Bone Mineral Density. *Front. Genet.* 9, 267. doi:10.3389/fgene.2018.00267

Zhou, Z., Cui, S., Du, J., Richards, R. G., Alini, M., Grad, S., et al. (2021). One Strike Loading Organ Culture Model to Investigate the Post-traumatic Disc Degenerative Condition. *J. Orthop. Transl.* 26, 141–150. doi:10.1016/j.jot. OBJ

Zhu, G.-b., Jiang, X.-r., Xia, C.-l., Sun, Y.-j., Zeng, Q.-s., Wu, X.-m., et al. (2011). Association of FAS and FAS Ligand Polymorphisms with the Susceptibility and Severity of Lumbar Disc Degeneration in Chinese Han Population. *Biomarkers*16, 485–490. doi:10.3109/1354750X.2011.598563

Zuo, R., Wang, Y., Li, J., Wu, J., Wang, W., Li, B., et al. (2019). Rapamycin Induced Autophagy Inhibits Inflammation-Mediated Endplate Degeneration by Enhancing Nrf2/Keap1 Signaling of Cartilage Endplate Stem Cells. *Stem Cells* 37, 828–840. doi:10.1002/stem.2999

Conflict of Interest: EK and LK were employed by ProtATonce Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those ofthe publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Citation: *Bermudez-Lekerika P, Crump KB, Tseranidou S, Nüesch A, Kanelis E, Alminnawi A, Baumgartner L, Muñoz-Moya E, Compte R, Gualdi F, Alexopoulos LG, Geris L, Wuertz-Kozak K, Le Maitre CL, Noailly J and Gantenbein B (2022) Immuno-Modulatory Effects of Intervertebral Disc Cells. Front. Cell Dev. Biol. 10:924692. doi: 10.3389/fcell.2022.924692*

Copyright © 2022 Bermudez-Lekerika, Crump, Tseranidou, Nüesch, Kanelis, Alminnawi, Baumgartner, Muñoz-Moya, Compte, Gualdi, Alexopoulos, Geris, Wuertz-Kozak, Le Maitre, Noailly and Gantenbein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academicpractice. No use, distribution or reproduction is permitted which does not complywith these terms.

GLOSSARY

ABM Agent-based model

ADAMTS a disintegrant and metalloproteinase with thrombospondin motifs

AF Annulus fibrosus AI

Artificial intelligence

BNB Blood NP barrier

BMI Body mass index

CEP Cartilage endplate

CLIP Cartilage Intermediate layer protein

CoNS Coagulase-negative Staphylococci

CS Chondroitin sulfate

CXCL10 C-X-C motif chemokine ligand 10

DR Death receptor

ECM Extracellular Matrix

ENCODE Encyclopedia of DNA Elements

ENOS Endothelial nitric oxide synthase

eQTL expression quantitative trait *locus* ER

Endoplasmic Reticulum

FEM Finite element model

FUMA Functional Mapping and Annotation of Genome-Wide Association Studies

GAG Glycosaminoglycan

GCN Graph convolutional networks GDF

Growth differentiation factor GTEx

Genotype-Tissue Expression GWAS

Genome-wide Association StudiesHIF

Hypoxia-inducible factor

IDD Intervertebral disc degeneration

IGF Insulin growth factor

IGFR Insulin growth factor receptor

IL Interleukin

IVD Intervertebral disc

LBP Low Back Pain

LD Linkage disequilibrium LDD

Lumbar Disc DegenerationLPS

Lipopolysaccharide miRNA micro-RNA

MC Modic change

ML Machine learning

MMP matrix metalloproteinase

MRI Magnetic resonance imaging

NFκ-B nuclear factor kappa-B

NP Nucleus pulposus

SNP Single nucleotide polymorphism

SP1 Specificity Protein 1 (syn. Sp1 transcription factor)

SPARC Secrete protein, acidic, rich in cysteine THBS2

Thrombospondin-2 gene

TonEBP Tonicity-responsive enhancer binding protein

TGF Tumor growth factor

TNF-α Tumor necrosis factor alpha

TRAIL TNF-α-related apoptosis-inducing ligand

TRP Transient receptor potential

V2F Variant to Function

VEP Variant effect Predictor

VEGF Vascular endothelial growth factor

B. Cartilaginous endplates: A comprehensive review on a neglected structure in intervertebral disc research

This chapter is based on:

Cartilaginous endplates: A comprehensive review on a neglected structure in the intervertebral disc research

Paola Bermudez-Lekerika 1,2†‡, Katherine B. Crump 1,2†‡, Sofia Tseranidou 3†, Andrea Nüesch 4†, Exarchos Kanelis 5,6†, Ahmad Alminnawi 7,8†, Laura Baumgartner 3†, Estefano Muñoz-Moya 3†, Roger Compte 9†, Francesco Gualdi 10†, Leonidas G. Alexopoulos 5,6†, Liesbet Geris 7,8,11†, Karin Wuertz-Kozak 12,13†, Christine L. Le Maitre 4†, Jérôme Noailly 3† and Benjamin Gantenbein 1,2

1Tissue Engineering for Orthopaedics and Mechanobiology, Bone and Joint Program, Department for BioMedical Research (DBMR), Faculty of Medicine, University of Bern, Bern, Switzerland, 2Department of Orthopaedic Surgery and Traumatology, Inselspital, Bern University Hospital, Medical Faculty, University of Bern, Bern, Switzerland, 3BCN MedTech, Universitat Pompeu Fabra, Barcelona, Spain, 4Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield, United Kingdom, 5ProtATonce Ltd., Athens, Greece, 6School of Mechanical Engineering, National Technical University of Athens, Zografou, Greece, 7GIGA In Silico Medicine, University of Liège, Liège, Belgium, 8Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium, 9Twin Research and Genetic Epidemiology, St Thomas' Hospital, King's College London, London, United Kingdom, 10Institut Hospital Del Mar D'Investigacions Mèdiques (IMIM), Barcelona, Spain, 11Biomechanics Research Unit, KU Leuven, Leuven, Belgium, 12Department of Biomedical Engineering, Rochester Institute of Technology, Rochester, NY, United States, 13Spine Center, Schön Klinik München Harlaching Academic Teaching Hospital and Spine Research Institute of the Paracelsus Private Medical University Salzburg (Austria), Munich.

Published in: JOR Spine

DOI: *10.1002/jsp2.1294*

Received: 29 April 2023 | Revised: 15 September 2023 | Accepted: 26 September 2023

DOI: 10.1002/jsp2.1294

REVIEW

JOR Spine OPEN ACCESS

Cartilaginous endplates: A comprehensive review on a neglected structure in intervertebral disc research

Katherine B. Crump [1,2,3] | Ahmad Alminnawi [4,5,6] | Paola Bermudez-Lekerika [1,2,3] | Roger Compte [7] | Francesco Gualdi [8] | Terence McSweeney [9] | Estefano Muñoz-Moya [10] | Andrea Nüesch [11] | Liesbet Geris [4,5,6] | Stefan Dudli [12,13] | Jaro Karppinen [9,14,15] | Jérôme Noailly [10] | Christine Le Maitre [11] | Benjamin Gantenbein [1,2]

[1]Tissue Engineering for Orthopaedics & Mechanobiology, Bone & Joint Program, Department for BioMedical Research (DBMR), Medical Faculty, University of Bern,Bern, Switzerland

[2]Department of Orthopaedic Surgery and Traumatology, Inselspital, Bern University Hospital, Medical Faculty, University of Bern, Bern, Switzerland

[3]Graduate School for Cellular and Biomedical Sciences (GCB), University of Bern, Bern, Switzerland

[4]GIGA In Silico Medicine, University of Liège, Liège, Belgium

[5]Skeletal Biology and Engineering Research Center, KU Leuven, Leuven, Belgium

[6]Biomechanics Research Unit, KU Leuven, Leuven, Belgium

[7]Twin Research & Genetic Epidemiology, St. Thomas' Hospital, King's College London, London, UK

[8]Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain

[9]Research Unit of Health Sciences and Technology, University of Oulu, Oulu, Finland

[10]BCN MedTech, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

[11]Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield, UK

[12]Center of Experimental Rheumatology, Department of Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland [13]Department of Physical Medicine and Rheumatology, Balgrist University Hospital, Balgrist Campus, University of Zurich, Zurich, Switzerland[14]Finnish Institute of Occupational Health, Oulu, Finland

[15]Rehabilitation Services of South Karelia Social and Health Care District, Lappeenranta, Finland

Correspondence
Benjamin Gantenbein, Tissue Engineering for Orthopaedics & Mechanobiology, Bone & Joint Program, Department for BioMedical Research (DBMR), Medical Faculty, Universityof Bern, CH-3008 Bern, Switzerland.
Email: benjamin.gantenbein@unibe.ch

Abstract

The cartilaginous endplates (CEP) are key components of the intervertebral disc (IVD)necessary for sustaining the nutrition of the disc while distributing mechanical loadsand preventing the disc from bulging into the adjacent vertebral body. The size, shape, and composition of the CEP are essential in maintaining its function, and

Abbreviations: AC, articular cartilage; ACAN, Aggrecan; AF, annulus fibrosus; ALP, alkaline phosphatase; BEP, bony endplate; BMP, bone morphogenetic protein; BV/TV, bone volume fraction; CEP, cartilage endplate; CFD, computational fluid dynamics; COL2, collagen type II gene; ECM, extracellular matrix; eDAPS, endplate-modified disc-like angle ply structure; ERK, extracellular signal-regulated kinase; EZH2, enhancer of zeste homologue 2; FCD, fixed charge density; FE, finite element; FGF, fibroblast growth factor; GAG, glycosaminoglycan; ICD, International Classification of Diseases; IL, interleukin; IVD, intervertebral disc; LBP, low back pain; MAPK, mitogen-activated protein kinase; MC1/2/3, Modic type 1/2/3 changes; MIF, macrophage migrationinhibitory factor; miRNA, microRNA; μ(CT), microcomputed tomography; MMP, matrix metalloproteinase; MRI, magnetic resonance imaging; NP, nucleus pulposus; PEG, polyethylene glycol; PTHrP, parathyroid hormone-related protein; ROS, reactive oxidative species; Shh, Sonic hedgehog; TEPS, total endplate score; TGF-β, transforming growth factor beta; TIMP, tissue inhibitors ofmetalloproteinases; TLR2, toll-like receptor 2; TLR4, toll-like receptor 4; UTE, ultrashort time to echo; VB, vertebral body; WT, weight.

Katherine B. Crump and Ahmad Alminnawi contributed equally to this work and share first authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium,provided the original work is properly cited.
© 2023 The Authors. *JOR Spine* published by Wiley Periodicals LLC on behalf of Orthopaedic Research Society.

Funding information
HORIZON EUROPE Marie Sklodowska-CurieActions, Grant/Award Number: 955735

degeneration of the CEP is considered a contributor to early IVD degeneration. Inaddition, the CEP is implicated in Modic changes, which are often associated with lowback pain. This review aims to tackle the current knowledge of the CEP regarding itsstructure, composition, permeability, and mechanical role in a healthy disc, how theychange with degeneration, and how they connect to IVD degeneration and low back pain. Additionally, the authors suggest a standardized naming convention regarding the CEP and bony endplate and suggest avoiding the term vertebral endplate. Currently, there is limited data on the CEP itself as reported data is often a combination of CEP and bony endplate, or the CEP is considered as articular cartilage. However, it is clearthe CEP is a unique tissue type that differs from articular cartilage, bony endplate, andother IVD tissues. Thus, future research should investigate the CEP separately to fullyunderstand its role in healthy and degenerated IVDs. Further, most IVD regeneration therapies in development failed to address, or even considered the CEP, despite its keyrole in nutrition and mechanical stability within the IVD. Thus, the CEP should be con-sidered and potentially targeted for future sustainable treatments.

KEYWORDS
biologic therapies, biomechanics, degeneration, pre-clinical models

INTRODUCTION

What is the CEP?

The intervertebral disc (IVD) provides the spine with flexibility and operational mechanical support. Depending on the speed or dyna- mism of the mechanical loads, it can store or dissipate energy and allows movement in the vertebral column. The IVD comprises three anatomic regions: a gelatinous core, the nucleus pulposus (NP); the annulus fibrosus (AF), a fibrocartilage that confines the NP laterally; two cartilaginous endplates (CEP) that are thin hyaline-like cartilage layers, covering the cranial and caudal ends of NP and the inner part of the AF. The human CEP is 0.1–1.6 mm (~0.06 in) thick. It separatesthe IVD from the adjacent endplates of the vertebral bone, that is, thebony endplates (BEP).[1-4] The CEP thickness varies greatly even within healthy IVDs, according to age, location in the spine (disc level), posi-tion in the IVD (cranial, caudal), and region in the tissue (central, peripheral). Its extracellular matrix consists of mainly type II collagen,proteoglycans, and water.[5]

The CEP plays a key mechanical role in preventing the disc from bulging into the adjacent vertebral body (VB)[6] and providing cranial and caudal anchorage for the fibers of the inner AF and NP of theinnermost part of the AF (Table 1).[7–9] In addition, the CEP provides a key path for the diffusion of nutrients from the peripheral vasculatureto the IVD and waste out of the IVD, which is crucial as it is the larg- est avascular tissue in the human body.[10,11] While AC relies on diffu-sion from the subchondral bone and via synovial fluid for nutrition,[12,13] the CEP relies on diffusion from neighboring blood ves- sels. Solutes, including oxygen and glucose, have been hypothesized to be predominantly transported into the disc through the CEP and their availability is regulated by the bone marrow contact channels that cross the BEP.[14] Often, the combined CEP and BEP are referredto as the vertebral endplate; however, the term vertebral endplate is also used interchangeably to refer purely to the BEP. Although theCEP and BEP have been recognized as distinct tissues since the 1930s, many studies do not distinguish the CEP from the BEP,[15] forexample, when reporting fluid transport in the IVD[16] or radiologicalsigns of IVD degeneration[17] even if the authors acknowledge that the vertebral endplate is a bilayer of cartilage and bone.[18,19] Arguably, it isdifficult to isolate the CEP from the BEP experimentally, and this mustbe done very carefully.[20,21] Clinically, the distinction of the two tis-sues on medical images is also very challenging. Thus, when reporting methods, it should be clearly stated what tissue or construct (CEP, BEP, or a combination thereof) is being used and a clear consistent nomenclature should be used, to avoid confusion.[22] Further, it is the authors' recommendation to define the CEP and the BEP indepen-dently where possible, or otherwise explicitly define the vertebral endplate as a construct of two tissues.

Developmental biology of the CEP

The CEP, AF, and vertebral bodies develop from the mesoderm, spe-cifically from the sclerotome.[23] AC is also derived from the mesoderm (Table 1).[24,25] In contrast, the NP develops from the notochord.[26] The mesoderm (paraxial) undergoes somitogenesis promoted by precise and cyclic temporal and spatial regulation of Notch and Wnt, and fibroblast growth factor (FGF) signaling pathways, respectively.[27] The Sonic Hedgehog (Shh) spatiotemporal regulation led by the notochord further differentiates somite to sclerotome and simultaneously

TABLE 1 Differences between the cartilaginous tissues CEP, AC, NP, and AF regarding: (A) general differences, (B) biochemical composition,and (C) mechanics and permeability.

	Cartilaginous endplate (CEP)	Articular cartilage (AC)	Nucleus pulposus (NP)	Annulus fibrosus (AF)
A. General				
Tissue origin	Mesoderm (sclerotome)[26]	Mesoderm[24,25]	Notochord[26]	Mesoderm (sclerotome)[26]
Vascularity	Vascular at birth, avascular in adulthood[26,29,30]	Avascular[13,37] regardlessof the stage of development[39]	Avascular[38]	Vascular at outer one-thirdof the AF and avascular at inner AF[36]
Nutrition source	Diffusion from neighboring blood vessels[53] subchondral bone[12]	Diffusion from synovial fluid[12,13] and the	Anaerobic glycolysis,[51] diffusion from CEP[50,51,53]	Diffusion from neighboringblood vessels and CEP[50]
Imaging*	Hypointense on T2-weighted MR indistinguishable from bony endplate. Specific sequences suchas UTE assist in differentiating CEP from bony endplate on MR[114,116,222]	Hypointense on T2-weighted MR and distinguishable from adjacent soft tissue and bone[113]	Hyperintense on T2-weighted MR, clear delineation from CEP in healthy IVD and gradient in boundary between AF and NP[114]	Hypointense on T2-weighted MR, indistinguishable fromCEP, gradient in boundary between AFand NP in healthy IVD[114]
B. Biochemical composition				
Cell density	15 10^6 cells/mL[67]	14–15 10^6 cells/mL[66]	4 10^6 cells/mL[67]	9 10^6 cells/mL[67]
Water content	1.585–1.666 mg water/mg dry wt,[4] and 22.1%–62.4%[9,60]	70.7%[64]	70%–90%[62]	50%–70%[62]
Proteoglycan	7.2%–13.4% sulfated GAG μg/mg dry weight.[4] and 4.37%–18.48% μg/mg dry weight[60]	5%–15% GAG by dry weight[61]	30%–50% GAG of dry weight[62]	10% GAG of dry weight[62]
Collagen	681 ± 171 μg/mg dry weight,[4] and 329.0–886.9 μg/mg dry wt[60]	60%–70% collagen by dry weight[61]	20% of dry weight[62]	70% of dry weight,[62] aligned with alternate orientations of an average of ±30°[63]
Cell morphology	Rounded and slightly elongated in the direction of the collagen fibers[65]	Superficial and Mid zone: rounded and slightly elongated in the direction of the collagen fibers Deep Zone: round[24]	Fibrochondrocyte-like cells[76]	Rounded chondrocyte-like cells (inner AF) and elongated, fusiform, fibroblast-like cells (outer AF)[77]
Gene markers	ERK, BMP, ACAN, COL1A1, COL2A1[51]	GDF10, CYTL1, IBSP, FBLN1,[78] ACAN, PTN[79]	PAX1, FOXF1, HBB, CA12, OVOS2,[78] KRT19,[80] ACAN, VCAN, TNMD, BASP1, TNFAIP6, FOXF1, FOXF2 and AQP1[79]	COL1, VCAN, PTN, TNMD, BASP1, TNFAIP6, FOXF1, FOXF2, and AQP1[79]
Pericellular matrix	Randomly arranged[9]	Columnar organization[9]	Single cells in lacunae[74]	Single cell, paired, or multiple cells in contiguity[75]
C. Mechanics and permeability				
Primary mechanical function	Resist fluid flow in and out of the disc and maintain a uniform stress distribution across the IVD,[86,223,224] and prevent the disc from herniating or bulging[9,223,224]	Distribute load during joint movement and provide lubricated (low friction) movement[37]	Withstand compressive loads to the IVD and maintain the BEP–CEP interface through fluid pressure[1]	Confine the NP laterally,[7] and anchor for the IVD to the VB.[48]
Primary permeability function	Main gateway of nutrients and waste into and out of the disc and waste,[6,10,15,50,52,54] prevent loss of large proteoglycan molecules from the disc[223]	N/A	N/A	Secondary gateway of nutrients and waste into and out of the disc[10,15,50]

(Continues)

TABLE 1 (Continued)

Cartilaginous endplate (CEP)		Articular cartilage (AC)	Nucleus pulposus (NP)	Annulus fibrosus (AF)
Permeability	1.27 10^{16} and 1.66 10^{14} m^4/Ns[4,10,53]	(0.76 ± 0.42) 10^{14} m^4/Ns[61]	0.67 ± 0.09 10^{15} m^4/Ns[87]	0.23 ± 0.19 10^{15} m^4/Ns[87]
Tensile modulus	0.5–21.8 MPa[60]	1–30 MPa[89]	1–1.66 MPa[90]	2.56–12.29 MPa[90]
Bone interaction	Parallel collagen fibers that make it weak and susceptible to detachment[48]	Anchored by perpendicular collagenfibers making them strongly attached[48]	N/A	Outer AF extends to the VB anchoring the disc tothe vertebral rim[48]

*Multiple imaging modalities are relevant in distinguishing these structures, expecially when degenerative features are present, e.g. CT in CEP sclerosis.

promotes spine segmentation.[28] After sclerotome segmentation, cells proliferate, condense, and undergo chondrogenesis to form the vertebral bone (through the endochondral bone process), the AF, and CEP. This complex regulation is governed by the coordinated action of Shh, SOX5/6/9, Pax1/9, and bone morphogenetic protein (BMP) pathways.[23]

At birth, the human CEP is thicker and takes up approximatelyhalf of the intervertebral space, which is reduced to about 5% by adulthood.[26] Additionally, blood vessels are present at infancy, but are replaced over time by cartilaginous ECM and almost disappear by skeletal maturity.[26,29,30] In humans, the CEP acts as a growth plate for the vertebrae, but this is lost after teenage years, so that only a thin layer of hyaline cartilage remains.[26,29] This is different in many animalssuch as sheep or bovine, in which the growth plate has been shown topersist into adulthood, and is separated from the CEP by the BEP.[31]The shape of the lumbar (L4–L5) CEP also changes with age, starting with a biconvex shape at infancy, but evolves to a concave shape beginning around the age of 2 or 3 years when children start to walk.[32–35] Weight bearing and movement have been shown to influ- ence the shape of the vertebrae and IVDs.[35] The CEP is vascularized during fetal development; however, by the age of 10 there is a sub- stantial decrease in blood vessels, which are lost by adulthood.[36] Simi-larly, blood vessels have been shown to be present in the outer third of the AF up until the age of two, but decrease by age 30 unless thereis damage that allows for revascularization.[36] In contrast, AC[13,37] and the NP[38] are avascular regardless of the stage of development.[39] Throughout the 20s and after adulthood, calcification is observed, often in focal points in the CEP. These calcified sections can driverevascularization and bone formation which occurs following activa- tion of matrix metalloproteinases (MMPs) which degrade the ECM.[29,30] Furthermore, oxidative stress has been shown to induce CEP calcification through the p38/extracellular signal-regulated kinase(ERK)/p6 pathway,[40] which acts in conjunction with mitogen- activated protein kinase (MAPK) stimulation. This pathway is also involved in cartilage calcification in osteoarthritis,[41] and is implicated in embryonical endochondral ossification in coordination with the transforming growth factor beta (TGF-β) and BMP families.[40,42] Simi-larly, AF and CEP have both been demonstrated to be the source ofpathological fibrocartilage in the NP.[43] Thus, the ability of CEP to move toward bone phenotype (calcification) and to equivalent AF cellfate (fibrocartilage) suggests that CEP cells maintain developmental-like plasticity, and consistent tissue homeostasis to maintain their healthy phenotype. Similarly, this de-differentiation capacity has been shown in articular cartilage (AC), where significant differential expression in the ERK and BMP pathway genes is observed.[44]

HEALTHY CEP

Structure and composition of the CEP

Throughout life, the composition and anatomy of the CEP and BEP continuously change. During early life, ossification of the VB occurs. While the vertebra-sided part of the endplates becomes ossified forming the BEP in young adults, the disc-sided part remains cartilaginous forming the CEP.[30] The BEP is a layer of porous, coalesced trabecular bone containing pockets of vascularized bone marrow enabling the two-way transport of nutrients and cellular metabolic products.[45] In adults, the structure is avascular but has a base that contains a dense network of capillaries formed by terminal branches of metaphyseal and nutrient arteries.[46] The thicker peripheral section of the BEP forms a junction between the CEP and AF with the vertebral body.[47]

The structural integration of the CEP into the BEP, AF, and NP var-ies.[1] At the bone interface there is minor integration, and the bone–cartilage junction is seen as a straight line with no gaps with collagenfibers of the CEP aligned parallel to the bone.[1,48,49] This is different than AC–bone interfaces, where the collagen fibers of the cartilage are per-pendicular to the bone, anchoring the tissue types together.[48] However,in a healthy disc, the fluid pressure of the NP maintains the CEP andthe BEP pressed together and thus, under normal loading (i.e., compression), the limited integration of the CEP and BEP is enough to maintain the BEP–CEP interface.[1,48] Also at the BEP, there are capil-laries that penetrate the pores of the subchondral bone and terminate by looping before the CEP junction.[50,51] The capillaries are denser atthe center of the vertebral endplates above the NP, which is where theIVD is thickest.[52] As the mature IVD is considered as avascular, thenutrient and waste exchange occur by diffusion from these capillaries.[53]Although diffusion can occur through the outer AF, the CEP is consid-ered as the main gateway for nutrients into the disc and waste out ofit.[6,10,15,50,52,54] The diffusion distance between the CEP and the cells inthe center of the NP can reach 8 mm,[14] which provides a shorter routefor diffusion than through the AF.

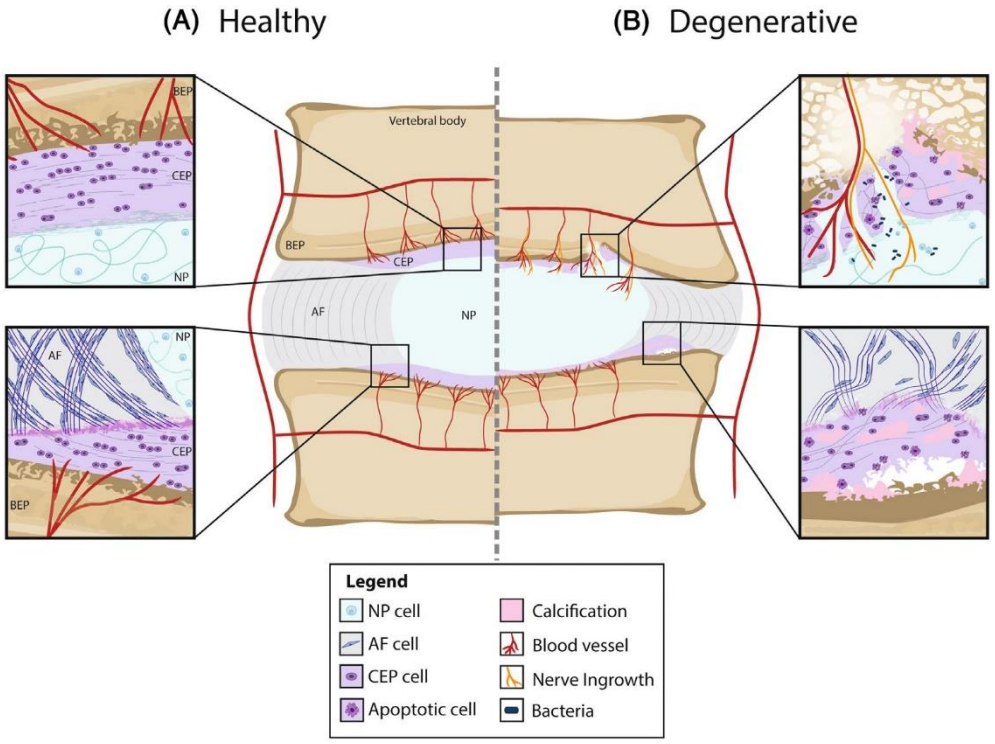


FIGURE 1 (A) *Healthy CEP*. In the healthy CEP, the collagen fibers of the AF continue into the CEP parallel to the bone (bottom). The collagen fibers of the NP penetrate at least partially into the CEP (top). The BEP–CEP junction is seen as a straight line with no gaps. CEP chondrocytes are rounded, and slightly elongated in the direction of the collagen fibers. The healthy CEP is avascular but has a base that contains a dense network of capillaries formed by terminal branches of metaphyseal and nutrient arteries. (B) *Degenerative CEP*. A degenerated CEP shows loss ofthickness, fibrosis, calcification, and apoptotic cells. Fissures allow for blood vessels and nerve ingrowth as well as bacteria entering the NP (top). The adjacent BEP can show sclerosis and signs of MC (top). Avulsions of the CEP from the BEP can also occur (bottom). Integration between the CEP and the NP or AF can also become weaker. These degenerative changes can be identified histologically.[74] Note that details regarding the other IVD tissues are not included in the image.

In addition, collagen fibers from the outer AF extend into adjacent vertebrae and serve as anchor for the IVD to the rim of the vertebral bone.[48] The outer AF connects directly into the bone while the inner AF and NP connect to the CEP.[1,48,55] The integration at the inner AF,however, is more complex as the collagen fibers of the AF lamellae are continuous with those of the CEP (Figure 1).[49,56] Additionally, SEM analysis of ovine discs has shown that the collagen fibers of the AF branch, which strengthens the annulus–EP anchorage by increas- ing the interface area over which shear forces are distributed.[57] Thesefibers also intertwine and sometimes merge with the fibrils of theCEP. The strength of this connection is essential for the CEP ability toresist tensile loading.[1,49] Collagen fibers in the CEP differ from both the NP and AF.[58] The collagen network of the CEP is denser than that of the NP; and tends to be arranged mainly parallel to the vertebrae, although not as highly oriented as in the AF.[9,58] Additionally, the colla-gen fibers at the inner CEP are stronger and more interconnectedthan those at the outer CEP, which could play a role in the anisotropicflow resistance of the CEP.[59] The highly convoluted collagen fiberswithin the NP penetrate at least partially into the CEP, providing the resistance to tensile force at the NP-CEP interface.[7–9]

Interestingly, the caudal CEP is usually thinner than cranial one[9] and both are rich in type II collagen, proteoglycans, and water.[9]

Measurements of cadaveric human lumbar CEPs properties found that collagen content was 681 ± 171 μg/mg dry weight,[4] or within a rangeof 32.9%–88.69% dry weight.[60] In comparison, the collagen contentof AC has a smaller range of 60%–70%[61] while that of the NP was around 20%.[62] Collagen content of the AF was around 70%[62] and hasa specific alternating aligned orientation of an average ±30°.[63] The average sulfated glycosaminoglycan (GAG) content was 103 ± 31 μg/mg dry weight,[4] or within a range of 4.37% and 18.48% dry weight.[60] The GAG content is similar to that of AC, which ranges between 5%and 15%,[61] and that of the AF which is 10%.[62] The NP GAG content,on the other hand, is much higher, ranging between 30% and 50%.[62] CEP hydration was 1.585–1.666 mg water/mg dry weight, and CEP porosity was 0.648 ± 0.069.[4] The water content of the CEP, which ranges between 22.1% and 62.4%,[9,60] is much lower than that of AC, at 70.7%,[64] the NP, at 70%–90%, and the AF, at 50%–70%.[62] Chon-drocytes are distributed throughout the CEP and are responsible for maintaining the ECM and, thus, providing stability to the tissue.[2] Macroscopically, CEP chondrocytes are typically rounded, although slightly elongated in the direction of the collagen fibers, more similarto the chondrocytes of the mid or superficial zone of AC than to the chondrocytes of the deep zone of AC.[24,65] The cell density of the CEP is ~15 10^6 cells/mL with the highest density closest to the

vertebral bone. Although similar to the cell density of AC (1.4–1.5 10^7 cells/mL)[66] this is nearly four times the cell density of theNP (4 10^6 cells/mL) and two times that of the AF (9 10^6 cells/mL).[67] Recent single-cell analysis of human IVD cells have demon- strated that there are different IVD cell clusters non-randomly distrib-uted in the AF, NP, and CEP.[68] The interactions between these IVDtissues at a mechanistic level are essential to understand the pivotal role of CEP chondrocytes in a healthy CEP structure. The main com- ponents of the pericellular matrix around the CEP chondrocytes are hyaluronan, proteoglycans, and type VI collagen,[69] while interstitial collagen is mainly composed of type II collagen[70] and proteoglycans, where the type chain, length, and quantity of glycosaminoglycans (GAGs) determine the water content.[71] Type X collagen, a calcium-binding collagen which is a key marker of hypertrophy, increases with age and is associated with increased calcification of the CEP.[72] Addi- tionally, depletion of one of the c*ollagen type II* gene (COL2) alleles has been shown to also promote calcification of the CEP in mice.[73]Unlike AC, the CEP has a randomly arranged pericellular matrix that does not follow a columnar organization.[9] In comparison, the NP has single cells in lacunae,[74] and the AF has single cell, paired, or multiple cells in contiguity.[75] In addition, the CEP is less hydrated and has lower GAG content than AC,[9] which is why it should be considered a different and unique tissue. Negatively charged proteoglycans repre- sent approximately 15% of the dry weight of CEP tissue.[4] CEP cellsexhibit an elongated morphology aligned to the collagen-rich ECM and are arranged parallel to the VB, similar to superficial zone chon- drocytes of AC, while deep cartilage AC cells present a round mor-phology and are arranged perpendicular to the adjacent bone.[24,65] In contrast, the NP has fibrochondrocyte-like cells.[76] The cells of the AF differ between the inner AF, where cells are rounded andchondrocyte-like, and the outer AF, where cells are elongated, fusi- form, and fibroblast-like.[77] Furthermore, AC cells present a decreased expression of ECM genes (*ACAN*, *COL1A1*, *COL2A1* genes) relative toCEP cell expression.[65] Other possible gene markers suggested for theCEP include *ERK*, *BMP*.[51] Genetic markers of AC have been proposed as *GDF10*, *CYTL1*, *IBSP*, *FBLN1*,[78] *ACAN*, and *PTN*[79] while NP markers are considered as *PAX1*, *FOXF1*, *HBB*, *CA12*, *OVOS2*,[78] *KRT19*,[80] *ACAN*, *VCAN*, *TNMD*, *BASP1*, *TNFAIP6*, *FOXF1*, *FOXF2*, and *AQP1*.[79]Genetic markers of the AF have been indicated as *COL1*, *VCAN*, *PTN*, *TNMD*, *BASP1*, *TNFAIP6*, *FOXF1*, *FOXF2*, and *AQP1*.[79]

| Mechanics and permeability within the CEP

| Effects of pressurization on CEP mechanics and permeability

Mechanical forces are crucial in maintaining cartilage homeostasis.[81] Chondrocytes respond to the mechanical environment, which contrib-utes to the regulation of cell metabolism. As with other cartilage tis-sues, mechanical loading, such as compressive, tensile, and shear forces as well as pressure from fluid flow, is essential for the functionof the CEP.[82] While overloading of the disc will result in vertebral endplate fractures or another injury, lack of mechanical stimuli will also impair disc homeostasis.[83] As it sits above and below the NP, theCEP acts as a mechanical barrier that adds resistance to the flow offluid from the IVD to the VB, allowing for the pressurization of inter- stitial fluid in response to compression while preventing the disc frombulging into the adjacent VB.[58,84,85] This fluid pressurization helps tomaintain a uniform stress distribution across the IVD.[86] Thus, the per-meability of the CEP is essential in maintaining the intradiscal pres- sure. In contrast, AC mechanically distributes load during joint movement and provides lubricated (low friction) movement.[37] However, it does not play an important role in permeability.

The reported permeability of cartilaginous tissue varies sig-nificantly. AC has been reported to have a permeability of 0.76 ± 0.42 10^{-14} m^4/Ns[61] while the NP has a permeabilityof 0.67 ± 0.09 10^{-15} m^4/Ns,[87] and the AF has a permeability of 0.23 ± 0.19 10^{-15} m^4/Ns.[87] The permeability of the CEP ranges between 1.27 10^{-16} and 1.66 10^{-14} m^4/Ns,[4,10,53] depending on the CEP location in the IVD, the animal model, and the region of the sample within a CEP. Rodriguez et al. also reported a permeability of 1.19 10^{-10} m^4/Ns in human CEPs, and explained the significantly lower measured permeability as due to inhomogeneities and focal carti- lage lesions common in degenerated human samples.[52] Despite the considerable variation in the reported CEP permeability, it is stillat least an order of magnitude less than that of the BEP(~2.21 10^{-9} m^4/Ns).[4,51,52,88] However, a more accurate representa- tion would consider the CEP permeability as a gradient which exponen- tially increases from the NP toward the BEP according to proteoglycan content.[10] Early studies using disulfine blue dye also demonstrated higher permeability across the CEP at the center than across the lateral regions.[15] However, despite this, the central region of the NP experi-ences low nutrition and high lactate concentrations due to the large dif-fusion distances across discs, particularly in the lumbar region.[54] Further, there are differences in permeabilities in the CEPs of the same IVD, in which the cranial CEP is significantly more permeable than the caudal one[9] possibly due to the differences in loading experience.

| Effects of biochemical composition on CEP mechanics and permeability

The biochemical composition is fundamental in determining the mate-rial properties of the CEP, and therefore the permeability and response to mechanical loading.[60,85] It has been proposed that thereis an optimal range of biochemical composition that balances both the biomechanical and nutritional demands of the CEP.[60] To this extent, the CEP must be stiff enough to hold the disc together but porousenough to allow for solute transport. Therefore, the CEP tensile prop-erties have been found to be inversely related to the transport proper- ties.[60] Additionally, within the bovine CEP, the biochemical composition, and therefore the biomechanical properties, were foundto vary in the central region located next to the NP, compared withthe lateral CEP, which is stiffer and thus could withstand a more sig- nificant portion of loading.[85]

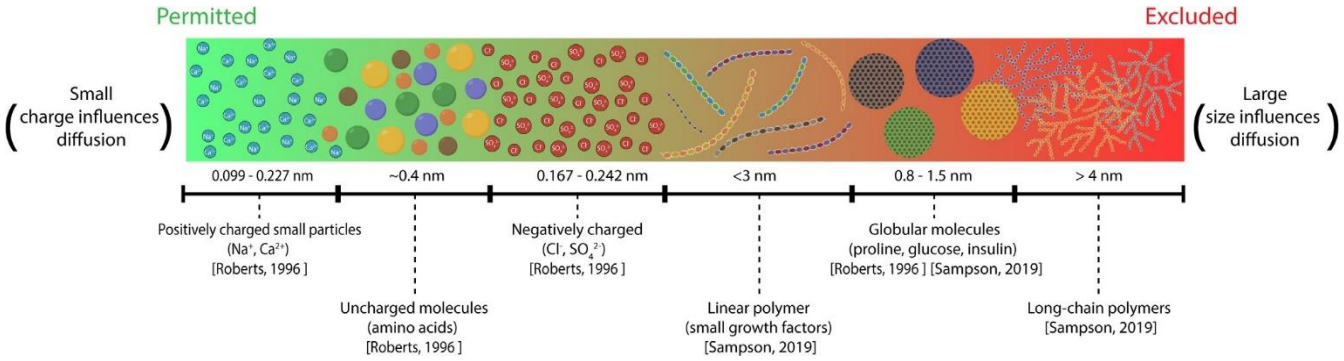


FIGURE 2 Healthy CEP diffusivity to different types of solutes based on their charge for small molecules and/or on their size and shape forlarge molecules. The molecules and diffusivities are based on Roberts et al.[6] and Sampson et al.[4]

The tensile stresses within the CEP occur through Poisson's effectswhen the CEP is pressed against the BEP by the NP pressure, and/orthrough direct peripheral pulling by the inner AF fibers that blend withthe CEP. At the inner AF and NP, the collagen is not as highly orientedas the outer AF, and therefore does not exert high tensile forces on theCEP,[1] although the central CEP will experience transverse shear and some tensile stress when the NP bulges laterally during compression.[49]The tensile modulus of the CEP was found to be 5.9 ± 5.7 MPa, rangingfrom 0.5 to 21.8 MPa.[60] In comparison, the tensile modulus of AC ranges from 1 to 30 MPa,[89] NP from 1 to 1.66 MPa, and AF from 2.56 to 12.29.[90] The tensile modulus of the CEP positively correlates to thecollagen content; however, water and GAG content have been shown to have minimal effects.[60] In contrast, water content is expected to playa more prominent role than GAG content in the extrinsic viscoelastic, or poro-elastic, properties, that is, aggregate modulus and hydraulic per-meability, and diffusion of the CEP, alongside other cartilage tissues including the NP, AF, and AC.[9,58,85] Yet, GAG content is considered more important in osmotic properties, rather than elastic properties inthe CEP.[9] In particular, multiphysics models showed that this character-istic might provide the GAG with an important role in the effective con-trol of the fluxes of fluid between the BEP and the NP, by the CEP.[10]Interestingly, GAG quantity has been shown to not correlate with watercontent,[9] suggesting that the type and quality of the GAGs are moreimportant than the quantity.

CEP transport properties depend on the porosity and collagen, GAG, and water content of the CEP matrix,[6,51] while solute transportinto and out of the IVD depends on solute size,[50] shape, weight,[6] and charge (Figure 2). For small molecules, net charge is the determiningfactor of diffusivity through the CEP.[5] The charge of the molecules are important due to the Donnan osmosis effect, in which small posi-tive ions from the interstitial fluid migrate into proteoglycan-rich tis- sues.[91] Water enters the tissue to equilibrate the chemical potential, and the tissue swells as much as allowed by the collagen network and surrounding tissue constraints. Thus, electrical charge of small parti-cles impacts the diffusivity.[92,93] Further, multiphysics models suggest that proteoglycan content has a greater effect than collagen content on the macroscopic hydraulic permeability of the CEP.[10] Neverthe-less, when biopsies or cores of CEP are used it is challenging to control potential GAG loss and tissue swelling in altered osmotic envi-ronments in vitro compared with those seen in vivo. Thus, future studies where GAG release into media is prevented and osmotic pres-sure is controlled are essential to further the understanding of CEP permeability.[94]

Conversely, for large electrically uncharged molecules, size is the determining factor of diffusivity as the space between GAG chains of aggrecan is only ~3–4 nm.[6] Due to their ability to bend, linear poly- mers have a relatively high diffusivity compared with spherical mole- cules of the same molecular weight.[4] However, long-chain polymers have a relatively low diffusivity compared with globular molecules of the same, or even higher, molecular weight. For example, starch, a globular 10 kDa molecule, diffuses more than polyethylene glycol (PEG), a long-chain 4 kDa polymer.[6]

Additionally, dynamic loading influences convective solutetransport of large solutes, particularly in less porous CEPs.[4,58,95,96] However, dynamic compression has been shown to have a minimal effect on small molecule, such as glucose or lactate, transport as thisacts primarily through diffusion.[4,94] In contrast, static compression can decrease diffusivity and inhibit nutrient transport because the tissue gets compacted, and thus porosity decreases.[58,82]

In silico investigation of CEP multiphysics

In silico simulations, such as finite element (FE) and computational fluid dynamics (CFD) analyses, have been widely used to determine the multiphysics mechanisms involved in biological tissues such as the ones of the IVD, in response to external mechanical loads. Models areuseful in understanding the particularity of CEP mechanics. However,despite its importance, the CEP is often disregarded in in silico simula- tions, simplified to a boundary condition of the IVD,[97] or given homogenized properties.[98] Since experimental data on the CEP is lim-ited, equations and material properties determined for the IVD or AC are often used to represent the CEP.[10] Osmo-poroelastic models have been demonstrated to be generic enough to use for cartilaginous tissues made up of proteoglycans and collagen fibers, however, CEP-specific data should be used when possible.

Nevertheless, models have given insight into how the IVD gains water during 8 h of rest at night much faster than it loses water during 16 h of activity during the day. The compression of the CEP against the BEP is considered to close the porosity of the CEP and limit water *out-flow* when the IVD is compressed, while opening the porosity andfavoring fluid *in-flow* into the IVD when unloaded.[99] Coined as the "intervertebral disc valve theory," this mechanism might explain functional anisotropic, or direction-dependent, flow resistance mechanismin the CEP[100] and has been backed up by experimental measurements[100] and independent permeability measurements.[53] Further, in silico simulations used composition-dependent CEP modeling and incorporated cranio-caudal gradients of compositions measured in a healthy human disc to support the theory.[2] These simulations found that CEP porosity changes additionally induced by the composition gradients reinforced the resistance to the mass flow of water that reaches the BEP when the NP is pressurized.[10] However, there are contradictory findings on whether the favored direction of flow at the CEP is *in-flow* or *out-flow*.[94,101]

IMPORTANCE OF CEP IN IVD DEGENERATION

The integrity of the CEP and BEPs are key to the homeostasis of the motion segment as they form an interface to exchange nutrients and metabolites from the IVD to the external circulation and play an essential role in the mechanical stability of the motion segment. Thus, pathological processes that occur to the CEP and BEP can alter the mechanical and nutritional environment of the IVD, triggering degeneration. Although most studies have focused on changes to the BEP with degeneration, which are visible on MRI images, these are likely associated with CEP changes as well. Indeed, cells derived from CEP adjacent to degenerated discs have very similar properties (morphology, immune phenotyping, proliferation, and gene expression) to bone marrow mesenchymal cells from the same patients.[21] BEP defects have been associated strongly and independently with IVD degeneration,[102] where they have been hypothesized to be an initiat- ing factor to degeneration of the disc.[33,103–105]

Imaging endplate defects

BEP defects include several key recognizable features, which are normally identified on magnetic resonance imaging (MRI) images (Figure 3). Computed tomography (CT) has been used to identify the presence of endplate sclerosis in MC but is not usually suitable for clinical diagnosis or epidemiological studies.[106] Generally, the term "endplate changes" in literature describing clinical T1- and T2-weighted MRI features refers to changes seen in the bone marrow adjacent to the CEP, typically referred to as Modic changes (MC). Atypical changes that affect the endplate which are detectable from MRIs can be classified into three categories: focal, corner, and erosive.[107] Focal changes, in which Schmorl's nodes are included, are defined as local hollow regions on the endplate with NP protrusion into the subchondral bone; while corner defects are changes in anterior or posterior end of the BEP with the compromise of the vertebral trabeculae. Finally erosive defects are characterized by an irregular extensive alteration of the endplate on T2-weighted images.[107] However, some features, such as endplate changes in the upper lumbar spine may have a developmental rather than degenerative origin.[108] Moreover, in a large population-based study BEP damage was strictly associated with MC, rather than other endplate defects.[109]

Modic changes, although by definition connected to the CEP, are at best an indirect reflection of CEP status and do not correlate well with histology.[110,111] The CEP itself is not clearly visible with conven-tional MRI sequences. Cartilage has short T2 values, so the CEP signalis not detected or is very hypointense. Additionally, its size ranges from 0.1 to 1.6 mm,[1] which is close to the typical pixel size of a sagit-tal T2-weighted MRI of 0.5 mm^2. This poses a significant challenge for identifying the role, if any, of the CEP in LBP through imaging.[112] In contrast, AC is distinguishable from adjacent soft tissue and bone.[113] In healthy discs, the position of the CEP can be inferred as a line ofhypointensity between the hyper-intense NP and the VB on T2-weighted MRI but it is impossible to distinguish this from the BEP (see fig. in Law et al. (2013)[114]). However, this feature is obscured when the AF is hypointense, which is characteristic of the early stages of degeneration.

The terms "Schmorl's nodes" and "endplate damage/disruption" are often used to describe specific MRI features at the boundary between the IVD and the VB seen on MRI, making them more specificto the CEP. However, the multiple proposed classification schemes for these imaging features reflect a poorly defined phenotype.[32] Considering the limitations of T1- and T2-weighted sequences and CTfor analyzing CEP features, alternative quantitative MRI approaches have been implemented.[115] Ultrashort time to echo (UTE) MRI is themost widely reported as an effective means of visualizing the CEP andcan show in vivo the delineation of the BEP and CEP.[114,116] Outsideof the clinical setting, both CT and MRI are widely used in ex vivoanimal or human studies. For example, MRI has been used in ex vivo human cadaveric spine segments to characterize the structure of the CEP in detail.[117] microcomputed tomography (μCT) has also been used in such settings to confirm compositional characteristics of the CEP seen in MRI such as sclerosis,[118] or for detailed morphologicaland biochemical characterization using contrast enhanced μCT.[119] These ex vivo imaging approaches are needed to better understand the role of early CEP changes in the pathogenesis of IVD degenera-tion, while clinically applicable sequences and accompanying analysescapable of detecting early CEP changes are needed for identifyingat-risk patients and early intervention targets.

Significance in low back pain

The evidence for innervation of the vertebral bone marrow extendingto the endosteal surface gives a biological basis for the endplate as a source of nociception in vertebrogenic pain.[120] Areas of vertebral

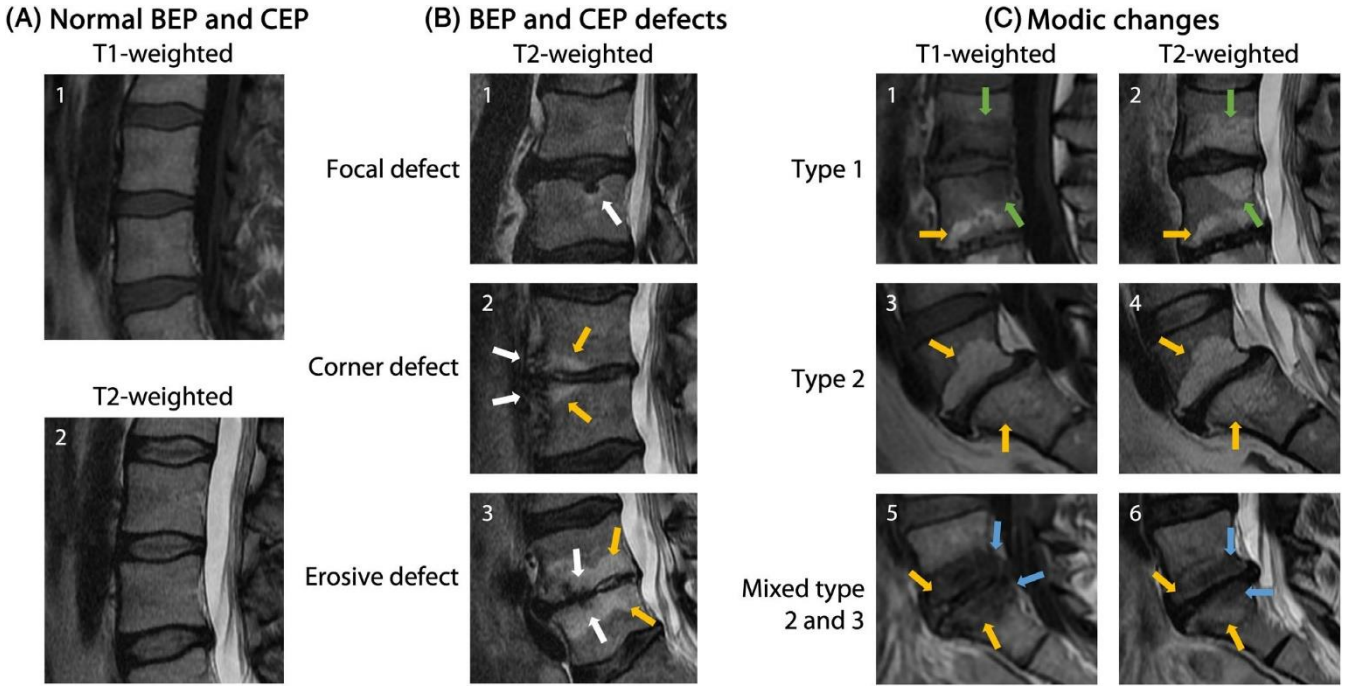


FIGURE 3 Examples of CEP and BEP appearance in standard of care MRI from 47-year-old participants in the Northern Finland Birth Cohort 1966. (A.1–A.2) Characteristic appearance of healthy CEP and BEP on T1- and T2-weighted MRI. (B.1) Schmorl's node as an example of a focal defect at the L3–L4 IVD caudal BEP (white arrow); (B.2) corner defects at the anterior edges of the L4–L5 IVD cranial and caudal BEP (white arrows) with accompanying type 2 MC (orange arrows); (B.3) erosive defects at the L4–L5 IVD cranial and caudal BEP (white arrows) with accompanying type 2 MC (orange arrows); (C.1–C.2) type 1 MC extending from the L3–L4 IVD cranial and caudal BEP (green arrows) with type 2 MC also visible in the L4 vertebral body (orange arrows); (C.3–C.4) type 2 MC extending from the L5–S1 IVD cranial and caudal BEP (orange arrows); (C.5–C.6) type 2 and type 3 MC (orange and blue arrows, respectively) extending from the L5–S1 IVD cranial and caudal BEP.

endplate damage may trigger neoinnervation,[120–122] with concomitant bone marrow pathologies (MC) also shown to be innervated.[120,123–125] While these sensory fibers generally terminate in or near the endosteal surface of the BEP, the CEP has also been shownto have small vascular spaces containing nerve fibers in some cases[121]and the CEP specifically can be accessed by nerves and blood vesselsin the case of damage.[126]

Endplate sensory fibers may be activated mechanically and chem-ically in the case of endplate damage. Disc/vertebra crosstalk as a consequence of the breakdown of the barrier provided by the CEP can contribute to nerve irritation with exposure to proinflammatory and neurogenic factors and the by-products of NP anaerobic metabolism such as lactic acid.[18] Additionally, damage to the endplate alters the distribution of IVD stress and the response to spinal loading,[18,45] contributing to local mechanical nerve activation. Associated changesto the paraspinal muscle quality may further interfere with segmental biomechanics and play an aggravating role in endplate nociception.[127] Although established as a plausible source of nociception, as with other spine image phenotypes, the observed relationship between vis-ible defects of the vertebral endplate and the experience of pain is unclear. Multiple grading schemes for qualitative grading have been put forward,[128,129] but no consensus exists for endplate MRI image phenotype nomenclature,[32] which makes it more challenging to interpret and aggregate pain association study results. There is also a limitation in the ability to specifically detect CEP changes using clinical diagnostic tools (e.g., standard T1- and T2-weighted MRI sequences), and these cannot distinguish changes specifically associated with neoinnervation. Alternative sequences such as UTE[115] are not widely used in the clinical setting, and few methods for quantitative image analysis of the endplate have been tested.[112,130]

| Modic changes: Definition, prevalence,natural course, and pain association

MC are MRI signal intensity changes of the vertebral bone marrow around a degenerated IVD[104,131] and independently associate with chronic low back pain (LBP).[132–136] A meta-analysis showed that MC prevalence in LBP patients is about seven times higher than in the non-clinical population (43% vs. 6%). MC occur predominantly in the lower lumbar spine.[137] There are three interconvertible types ofMC depending on their appearance on T1- and T2-weighted MRI. Modic type 1 changes (MC1) (Figure 3) are hypointense on T1-weighted images and hyperintense on T2-weighted images and represent edema, fibrovascular granulation tissue, infiltration of immune cells, and expansion of profibrotic stromal cells.[131,138–140] Modic type 2 changes (MC2) are hyperintense on T1- and T2-weighted images and represent fatty marrow conversion with

presence of fibrotic tissue.[131,140] Modic type 3 changes (MC3) are hypointense on T1-weighted and T2-weighted images and are sclerotic changes.[131,141] The reported prevalence for MC1, MC2, and mixed type (MC1/2, MC2/3) are highly variable. Median prevalence is around 15%–20% for MC1, 25%–65% for MC2, and <5% for MC3.[132,142] Mixed type MC1/2 are also frequent (15%–20%), whereas mixed type MC2/3 are rare. The prevalence generally increases with age and peaks in the 60s.[143,144] MC can inter-convert over time.[131,145–148] MC1 are the least stable, where within 4 years, most MC1 either convert to MC2, increase in size, or resolve. MC2 are more stable, while MC3 are a terminal stage. Smaller MC lesions are more likely to resolve than larger lesions.[145]

The association of chronic LBP with MC has been extensively studied and reviewed..[149] About half of the studies report a significantassociation with odds ratio ranging from 1.53 (95% CI: 1.02–2.29) to 83.10 (95% CI: 4.85–1424) Only one study reported a significant negative association with MC2 with an odds ratio of 3.2 (CI: 5.39 to 0.1). The association of MC with discography concordant pain has a specificity of >94% in five of six studies with a OR of 4.01 (1.52–10.61) in a meta-analysis.[18,149] Larger lesions had a stronger associa- tion with discography-concordant pain.[150] Overall, these data showan association of MC with LBP, in particular of MC1.

Increased innervation of the CEP in MC1 and MC2 is believed tocause increased pain sensitization at MC levels and is often referred to as vertebrogenic LBP.[120,124] Low back vertebral endplate pain (DM54.51) has recently been added to the International Classification of Diseases (ICD-11) as a subclassification of patients with LBP and MCs.

Association of BEP changes with CEPchanges and MC

Despite the importance of CEP and BEP in the onset of spinal pathologies, the relationship between CEP and BEP is poorly understood due to the difficulty of evaluating the CEP with imaging techniques. A study on cadaveric lumbar spines using ultrashort time-to-echo MRI was used to enable investigation of CEP morphology within IVDs withBEP lesions, demonstrating abnormalities of the CEP were statistically associated with BEP lesions.[151] After needle induction of IVD degeneration in a rabbit animal model, the CEP progressively thickened and showed increased collagen accumulation.[152] Along with changes in cartilaginous tissue, the bone interface was modified, specifically an increase in bone volume fraction. These findings suggest that CEP could have a role in the development of BEP lesions or vice versa. Ithas been shown that poor CEP composition can affect disc healthwith and without defects in the VB.[153] However, a precise under- standing of the sequelae of changes in the CEP and BEP is still miss- ing. There is comparably little data available about the integrity and damage of the CEP in MC. Fields et al.[120] showed that in cadaveric human spines, CEP damage is associated with histological changes consistent with MC. Still, not all specimens with histopathological changes had MC on MRI. In another human cadaveric study, Heggliet al.[125] showed that CEP and BEP damage are strongly associated with MC2. Supporting evidence for CEP damage in MC stems from studies assessing CEP fragments in surgically removed herniated disc tissue at the MC level. CEP fragments can co-herniate with disc tissuein cases of avulsion-type herniations, where the CEP is torn from the BEP. These CEP avulsions were found to associate with MC.[154,155]

Mechanisms of CEP damage in Modicchanges

While damage to the BEP is believed to be caused mainly through mechanical cues, mechanisms of CEP damage are poorly understood.[156] In MC, local biological reactions seem to contribute to CEP damage. In an animal model of MC1, immune reactions in the MC1 bone marrow caused damage to the adjacent endplate. This is noteworthy because it demonstrates that MC1 are not just reactive changes to disc degeneration, but that MC1 themselves can cause CEP damage and maintain the cross-talk of the bone marrow with the adjacent disc.[157] Recent studies confirm the possibility that activated immune cells in MC1 bone marrow can lead to CEP damage.[158] Increased lactate dehydrogenase activity and increased concentrationof C-reactive protein and of complement factors in MC bone marrow indicate also a humoral immune response related to local tissue damage.[125,159] On a cellular level, CEP cells at MC levels express more tumor necrosis factor (TNF), a disintegrin and metalloproteinase with thrombospondin motifs-5 (ADAMTS-5), macrophage migration inhibitory factor (MIF), and its receptor CD74.[124,160,161] TNF upregulates MIF in CEP cells, and MIF upregulates the secretion of proinflamma- tory cytokines by CEP cells, through an autocrine mechanism involv- ing CD74. The existence of this positive inflammatory feedback loop suggests that the CEP has the capability to escalate inflammation in MC independently from the disc and the bone marrow. Together, these data evidence that CEP damage in MC is not a pure mechanical mechanism but that the local inflammatory processes in the bone mar- row and of the CEP itself can lead to progressive CEP damage.

Occult disc infection, mainly with the *Cutibacterium acnes* (Gilchrist 1900) and other coagulase-negative staphylococci are dis- cussed as a potential etiology of MC1, at least in a subset of patients.[144,162] This is based on reports in which disc tissue from microdiscectomy of herniated discs was analyzed for the presence of *C. acnes*.[144,163–165] *C. acnes* has been isolated from discs adjacent to MC1 and the presence of this bacteria was predictive for the development of new MC1 after 1 year. *C. acnes* is assumed to migrate to structurally damaged discs through hematogenous spread from a distant infection or from the skin and other epithelial surfaces through the blood after innocuous lesions, for example, tooth brushing.[166,167] Disc herniation and endplate damage, both structural damages pre- sent in MC, can represent disc damage which allow bacteria to enter discs. Once in the disc, the low oxygen tension and low pH in the discfavor the proliferation of *C. acnes*. Furthermore, *C. acnes* is unlikely tocolonize the MC1 bone marrow because of too high oxygen tension in the bone marrow. Rat and rabbit models have demonstrated the biological plausibility that *C. acnes* injected into disc can trigger

MC1-like changes.[168–170] For example, injecting a *C. acnes* strain, which had been isolated from a human MC1 disc triggered hallmarks of MC1 within 2 weeks after injection (i.e., MC1-like MRI changes in the bone marrow, disc degeneration, fibrotic-inflammatory bone mar-row changes, and almost complete resorption of the CEP).[170]

The exact mechanism of how intradiscal *C. acnes* causes CEP resorption is still unclear. It has been shown that disc cells respond to *C. acnes* with the release of proinflammatory and neurotrophic factors through a toll-like receptor 2 (TLR2)-dependent pathway.[170–173] Stimulation of TLR2 on disc cells also upregulates matrix proteases that can degrade the disc and CEP matrix. Additionally, *C. acnes* secrete different virulence factors such as proteinases, hyaluronidases, lipases, and neuraminidases.[174,175] Proteinases and hyaluronidases can directly degrade ECM components of the CEP. Lipases hydrolyze triacylglycerides into glycerol and free fatty acids. Free fatty acids, in particular saturated free fatty acids, are highly proinflammatory by signaling through toll-like receptor 4 (TLR4).[176,177] Neuraminidases disrupt the ternary complex of TLRs with the membrane components CD24 and SiglecG/10 and abolish the inhibition of TLR signaling by SiglecG/10.[178] Together, *C. acnes* has the capacity to cause CEP damage, yet the precise mechanism remains unclear.

Role of CEP damage in Modic changes

Damage to the CEP compromises its function as a sieve for cells and macromolecules. In MC, where endplate damages are present, the degenerating disc can cross-talk with the adjacent bone marrow.[157] Proinflammatory and pro-osteoclastic cytokines that are produced at higher rates from MC discs can more easily escape into the adjacent bone marrow.[156,157,179] Marrow-sided leukocytes can in turn aggravate degenerative changes in the disc, even without infiltrating the disc.[156] Consequently, CEP damage in MC facilitates an inflammatory cross-talk between the disc and the bone marrow that contributes to the rapid degeneration of MC segments.[104] Furthermore, increased concentrations of cytokines and chemokines produced during disc degeneration[180,181] could more easily diffuse out of the IVD into the adjacent bone marrow leading to chemotaxis gradients and activation of immune cells.

Microscopic changes to the CEP

During IVD degeneration, key structural and cellular changes occur within all areas of the IVD, with key features identified within the NP, AF, CEP, and BEP (Figure 1). While macroscopic structural changes can be identified by imaging techniques such as MRI, these often fail to identify changes to the CEP. Furthermore, cellular and fine ECM changes require microscopic examination to be identified. (Figure 4) During the development of the recently published standardized histopathology scoring system for human IVD degeneration,[74] each of the regions (NP, AF, CEP, and BEP) were scored independently with equal weighting, demonstrating the recognition of the critical role of the CEP and BEP in IVD degeneration.[74] Within the CEP the degenerative features identified histologically included scoring for cellularity, lesions and ECM structure.[74] Cellularity changes seen during degeneration included: abnormal cellular clusters, empty lacunae, extensive neovascularization, and presence of apoptotic, necrotic, and senescent cells.[74] Changes within ECM can be identified histologically, and include loss of endplate thickness; avulsions from BEP; cracks and fissures; loss of normal matrix staining; fibrosis and calcification.[74] These visible microscopic changes within the CEP are due to alterations in the biomechanical and cellular regulation of the CEP.

Additionally, through histological analysis Huang et al.[182] identi- fied physical microdamage in 40% of degenerated CEPs. Further, pain and disability scores were significantly higher in these microdamaged CEPs than in degenerated IVDs without damaged CEPs. Six main pat- terns of microdamage were used to classify the CEPs, including fis- sures, vascular mimicry, NP herniation into the CEP, NP herniation and incorporation of bone tissue into the CEP, incorporation of bone tissue into the CEP, and traumatic nodes.[182]

Another methodology used to categorize the endplate is the total endplate score, or "TEPS."[183] MRI scans are used to quantify the damage, which was given a score based on diffusion patterns and the presence of endplate breaks or defects, focal thinning, MC, and irregularities or sclerosis. Five visually distinct diffusion patterns were identified through the study as well, that reflected disc marrow contact, focal leakage into the subchondral bone or NP, and pooling of liquid. These diffusion patterns correlated with degeneration level as well as the TEPS. As the TEPS increased, the diffusion pattern increased toward pooling. The TEPS was shown to correlate with Pfirmann's grading regardless of the age and levels of disc. Unfortunately, there was no distinction between CEP and BEP in that study.

Biomechanical regulation of CEP degeneration

There are two main theories about the contribution of mechanical loading to IVD degeneration; the overload theory, which states that excessive mechanical loading damages the IVD over time, and the immobilization theory, which states that low mobility causes the IVD to adapt and leads to tissue weakness and degeneration.[83] There is evidence that both overloading and immobilization contribute to IVD degeneration, and it is considered that there is a range of mechanical loading in which the disc remains healthy. Outside of this range, the metabolism shifts to catabolism.

The CEP is essential in maintaining the mechanics of the disc. Damage to the CEP barrier alters the hydration and could allow water to escape from the NP under loading, leading to NP decompression and degeneration.[1] FE simulations of the disc, performed in the 90s, suggested that initial failure always occurred in the endplates, demonstrating their vulnerability within the disc under loading.[184] Nevertheless, experimental testing of disc failure show that although most failures happen at the CEP–bone interface or disc–CEP interface, some primary failures also occur within the subchondral bone,[48]

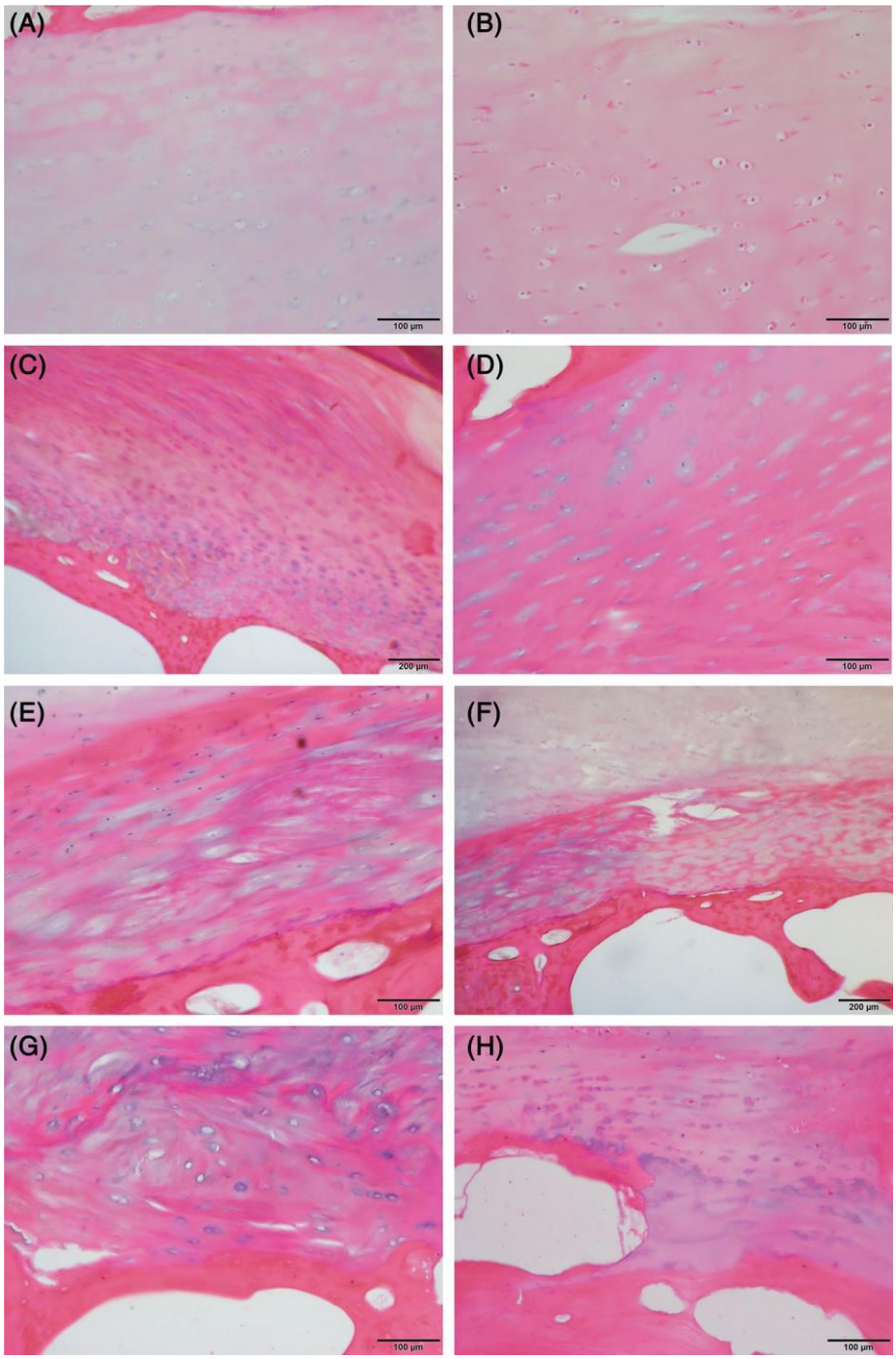


FIGURE 4 Hematoxylin and Eosin staining of human cartilaginous end plates(CEP) demonstrating key histological features of non-degenerate and degenerate CEP. (A) Non-degenerate CEPwith BEP top left of image, CEP within region of the NP. (B) CEP within the region of NP demonstrating excellent maintenance of eosin staining. (C) CEP/ AF enthesis within a non-degenerate region, image shows BEP at bottom left and CEP within the middle connecting into AF. (D) CEP within region of AF tissue demonstrating change in matrix organization at bottom of image.
(E) Abnormal CEP demonstrating clear disorganization and fibrosis. (F) Abnormal CEP showing fissures and disorganization of the CEP. (G) Abnormal CEP showing disorganization of the extracellular matrix.
(H) Abnormal BEP/CEP enthesis with boney evulsion shown. Scale bars as indicated: 100 μm (A, B, D, E, G, and H), 200 μm (C and F).

which was not predicted by such early FE models.[184] Additionally, endplate herniation through the CEP is a key feature seen in patients and is the most common type of herniation.[185,186] However, after endplate damage, additional failures could propagate to the annulus or lead to delamination.[184] This highlights the importance of CEP integrity; preventing CEP damage early on could also avert damage to the rest of the disc and prevent disc degeneration.

Cyclic compression has been shown to lead to the devel- opment of microstructural voids at the CEP-BEP border.[187] Microdamage accumulation occurred more often in flexed-joint postures in comparison to neutral postures. Following the damage, there was also a decrease in type I and II collagen content. Thus, overuse injuries could occur in the CEP and lead to altered biochemical content.

The IVD is most vulnerable to failure under bending, in which stretching puts the disc under tension.[1,48] Studies have found that in cadaveric thoracic IVD samples under tension, 71% of primary failures occur at the CEP-BEP interface.[48] The next most common failure was at the IVD-CEP interface (21%), while the rest failed within the sub-chondral bone. When pulled in tension from the vertebral bone on each side, the tensile failure strength of the CEP-BEP interface is 0.4 MPa,[48,49] while failure strain, calculated from force–displacement

data normalized to the initial specimen height, was found to be 38.5% ± 20.3% with the highest principal strains occurring in the mid-AF.[48] The rate of loading also affects the failure; slower rates cause disc pressurization with localized stretching and failure in the AF, while under rapid loading the AF does not have time to stretch andthus pulls the CEP from the bone.[188] Further, the position of the disc affects the failure location; neutral discs, which would be seen duringnormal standing, tear at the CEP-inner AF interface, while flexed discs,which would be seen during bending movements, fail at the outer AF-endplate interface or in the BEP.[188,189]

The thickness, porosity, and curvature of the CEP also influence the biomechanics.[48,49] Further, Thompson Grade and bone densityhave been correlated with failure strength.[48] Additionally, bone vol-ume fraction (bone volume (BV)/total volume (TV)) has been shown to positively correlate to the failure strength, with stress increased at the higher BV/TV end of the VB.[45,48,190] In humans, the cranial endplates have a lower BV/TV than caudal endplates, and therefore they tendto fail before the caudal endplate fail.[45] The microarchitectural fea-tures of endplate concavity are also significant predictors of failure strength.[190] Specifically, when the concavity of the CEP is wider, more voluminous, and less steep, it is capable of tolerating higher loads before failure.[190] Overall, CEPs are stronger when they are thicker and denser with a higher concave curvature that allows more space for the NP.[48,49,190,191] Thickness has even been suggested tobe used as a clinical risk measure for avulsion-type herniation.[48] It should be noted that quadrupeds have a different curvature than humans, as well as additional material properties such as BV/TV, andtherefore it is essential to consider the animal model used during mechanical tests.[190,191]

Animal models of the CEP

As human CEP samples are not always available, it is important to consider animal models which can be used to elucidate the properties and functions of the CEP. Interestingly, rats and mice have only a CEP with no BEP.[31] Rabbits and goats have a very thin CEP (1–3 cell layers) with a larger BEP. Larger animals such as dogs, cows, and sheep have both; however, the CEPs are thinner and the BEPs thicker than those of humans.[20] Bovine and canine CEPs have been shown to have similar biochemical content to human CEPs, although canine CEPs have been shown to have significantly more sulfated GAG than those of humans.[20] Further, bovine CEPs have been shown to have more proteoglycans in the outer AF–EP region in comparison to theNP–EP region similar to the human CEP, while canine CEPs show the opposite pattern. Bovine CEP cells are rounded and organized in stacked columns, in contrast to canine cells which have no organiza- tion and human CEP cells which are along the collagen fibers parallelto the disc. Thus, the molecular similarities of the bovine and humanCEPs make the bovine a more suitable model for investigating mechanics and transport in the CEP.[20] Additionally, baboon CEPs have been demonstrated to have similar biochemical content, includ-ing GAG, water and collagen, as those of humans, and thus could alsobe a good model to study mechanics with the CEP.[86] Rabbits have also been validated as a model to investigate initial endplate failure, although rabbit endplates have a higher BV/TV and a steeper, nar-rower concavity which should be considered when translating resultsto humans.[190] Particular care should be taken with diffusivity studies, taking into account the difference in thickness and CEP:BEP ratio compared with the human endplate.[20] Overall, no single animal model provides a complete representation of the human CEP and caution should be taken when extrapolating data.[31]

Cellular regulation of CEP degeneration

With increasing degeneration of the CEP, its composition undergoes several changes that could reduce its permeability and limit nutrient transport. In tissue adjacent to degenerated discs, the calcium concen-tration was shown to be higher. There is a delicate balance between the nutrient demand and the nutrient supply, which is imposed by the cell population and controlled by the permeability of the CEP.[51,54] The clas-sical paradigm is that any reduction in CEP permeability leads to nutri-ent retention and accumulation of lactate that in turn decreases the pHof the disc. A decrease in pH is directly related to hydrogen ion concen-tration. CEP permeability reduction or loss would impair disc oxygenation,[15] and subsequently, reduce cell survivability and activity, being possibly a major regulator of IVD cell populations.[51]

Increased levels of calcium showed to enhance the cleavage of aggrecan by ADAMTS5.[192] Not only was a decrease in aggrecan observed, but also a change in its composition changing from a 1:1 ratio of keratan sulfate to chondroitin sulfate to a 3:1 ratio.[71] Those compositional changes result in a decrease in the net hydrophilic property of the tissue. Furthermore, a positive correlation between the degenerative state of the tissue and increased denaturation oftype II collagen has been shown.[193] Moreover, a mouse spondylosismodel demonstrated that with increased age, apoptosis of chondro- cytes in the CEP lead to a markable decrease in cell density. Subse-quently, the disappearance of the CEP structure.[194] This was corroborated in human CEP samples. In addition to decreased celldensity and a higher rate of MC in degenerated CEP, expression ofMMP3, MMP9, interleukin-1 alpha (IL-1α) and IL-1β was increased.[195] In addition, tissue inhibitors of metalloproteinases-3 (TIMP3) also showed increased expression in degenerated CEP, suggesting a com- pensatory mechanism to regulate the increased ADAMTS expression.Interestingly, TIMP1 and TIMP2, but not TIMP3, were overexpressedin degenerated AF and NP compared with non-degenerate tissues.[196] CEP chondrocytes have also been demonstrated to have a differentresponse than AC chondrocytes to the same stimuli.[197] In response tohypertrophic stimuli such as Wnt agonist, CEP chondrocytes did not undergo the morphological changes seen in AC chondrocytes. However, they did show hypertrophic gene and protein expression and a decrease in proteoglycans. Oxidative damage-induced stress was also shown to induce apoptosis and promote calcification in the human CEP.[198] Neidlinger-Wilke et al.[199] showed in an in vitro modelstimulating NP cells with conditioned media of CEPs a significant increase in IL-6, IL-8, and MMP3, as well as MMP13. Aggrecan andtype II collagen were significantly decreased in NP cells exposed

to the CEP-conditioned media.[199] Those findings indicate the interactions between the CEP and the NP tissue via molecular factors influencing the pathophysiology of disc degeneration. Limited studies have been performed on the expression of cytokines within the isolated CEP. In addition, there is strong evidence of genetic regulation of CEP cell fate through non-coding RNA, which includes microRNA, short interference RNA and circular RNA. Proliferation, apoptosis, migration, and autophagy of CEP cells are the processes shown to be targeted by those RNAs and promoters of its degeneration.[200–203] There is also evidence for epigenetic roles in CEP degeneration regulation and in extension, IVD degeneration. Overexpression of histone methyltransferase enhancer of zeste homologue 2 (EZH2) in CEP cells produces reduced expression of *COL2*, *ACAN*, and *SOX-9* genes and increased *ADAMTS5* and *MMP13* genes in rat CEP cells.[204]

Furthermore, reduced CEP permeability, and consequently reduced nutrition and lowered pH, can downregulate both catabolic and anabolic gene expression of the NP cells negatively affecting ECM homeostasis.[51,54] Specifically, the mRNA expression for *ACAN*, *COL2A1*, and *MMP2* in the NP reduces.[51] Additionally, increased Ca^{2+} deposition leads to the activation of calcium-sensing receptor (CaSR), causing increased catabolism through the suppression of collagen and GAG synthesis and can induce calcification of CEP tissue through upregulationof alkaline phosphatase (ALP).[192] Further, deposition of Ca^{2+} in the IVDhas been associated with increased parathyroid hormone-related pro-tein (PTHrP) signaling, which drives calcification.[205]

Calcification and influence on permeability

Increased Ca^{2+} deposition in the CEP is seen with increased IVD degeneration in humans.[192,205] Calcification of the CEP is associated with decreased nutrient transport into the disc and waste transport out of the disc, which leads to nutrient starvation disc cells and a decreased pH within the disc, respectively.[2] Further, calcification can lead to lower porosity, hydration, and permeability.[4] With lower poros-ity and hydration, diffusion is impaired, and thus, dynamic loading hasa greater effect on nutrient transport.[4,94] Static compression, however, reduces the CEP porosity and leads to reduced oxygen and greater lactate accumulation in the disc, limiting nutrient transport and gas exchange.[82,94] Altered nutrient transportation through the CEP has thus been suggested to be a significant factor in the pathogenesis of IVD degeneration.[206] Preserving sufficient metabolite transport through the CEP is essential for the IVD to maintain its ECM and biochemical environment.[52] Degenerative changes in the CEP could lead to up to 70% decrease in CEP permeability and, ultimately cell death.[9] Overall, within calcified discs, dynamic compression improves disc nutrition while static compression impairs nutrition and leads to further degeneration. However, calcification is not always present in degenerated discs. When the CEP degenerates, GAG and collagen concentra- tion decrease, which causes higher porosity and thus increased solute transport.[60,207] Dynamic compression has been shown to have less effect on higher porosity CEPs.[4,54] Nevertheless, the decreased matrix content also decreases the tensile modulus of the CEP, losing its ability to withstand mechanical forces.[60]

Degeneration has not been found to be correlated to GAG or water content separately, but rather to fixed charge density (FCD), which is a property related to both. Higher FCD hinders transport through creating steric and ionic barriers. It has been found to bedirectly proportional to IVD degeneration and inversely proportional to CEP permeability.[9] This agrees with findings that CEPs with low permeability have high levels of collagen, aggrecan, and minerals which can physically block the solutes.[51] Shirazi-Adl et al.[50] demon-strated that IVD cells start dying when CEP permeability decreasesbelow 30% and the death rate increases exponentially as CEP perme- ability decreases further. The NP is the tissue most severely affected by CEP permeability changes.[10] While the NP periphery is adjacent to the CEP, its center can be as far as 8–10 mm in the IVD.[51] While there is also some diffusion through the outer AF, it is not enough tocompensate for an impermeable CEP due to calcification.[4]

While calcification and dehydration have been shown to reduce the permeability of the CEP,[6,51] the effect of calcification in the CEP has also become a topic of debate recently. On one side, there is the hypothesis that calcification prevents fluid from flowing into the IVD to transport nutrients. Opposingly, there is the hypothesis that fluid movement, and thus the effect of calcification, has a negligible effect on nutrient transport. Supporters of the first hypothesis claim that calcification leads to a reduction of the pore volumes creating a physical impermeable barrier that obscures the fluid path.[6] That, in turn, leads to reduced nutrient sup-ply to the cells and, ultimately IVD degeneration.[6,50,54] Supporters of thelatter hypothesis claim that advection, or the movement of fluid, throughthe CEP has minimal effect on nutrient supply since the nutrient concen-tration in the IVD is controlled by diffusion. Diffusion occurs regardless of fluid velocity,[88] especially when mechanical loading is present.[208] Further,some studies claim that severe nutritional deprivation does not appearuntil calcification causes a 50% blockage.[10] This level of blockage onlyoccurs at late stages of degeneration, which contrasts other hypothesesthat the depletion of CEP ECM might promote early degenerative mecha- nisms in the IVD through local cell starvation in the NP.[10] The lack ofblockage at early stages of IVD degeneration suggests that calcification asa physical barrier has minimal effect on IVD degeneration, however, cell catabolism induced by Ca^{2+} could induce IVD degeneration.[52]

Nonetheless, there is evidence that there is a strong positive corre-lation between CEP porosity and hydration.[4,6] Within extreme IVD degeneration, the CEP permeability barrier would be negligible due to the infiltration of blood vessels into the disc after damage. Therefore, both hypotheses could be correct depending on the age of the subject, degeneration level of the CEP, degree of calcification, and vascularization. This topic requires more intensive research to fully understand the effects of calcification on the IVD. Additionally, it is important to consider the poromechanical interactions among the BEP, the CEP and theNP to understand the mechanotransport of nutrients.[10]

EMERGING THERAPIES

Nutritional supply controls the population and activity of the IVD cellsto synthesize and maintain the ECM. Thus, for any disc cell therapy to succeed, maintenance of the nutrition supply is crucial. Furthermore,

even a successful attempt to biologically repair the IVD can fail in the long run if the nutrition-population balance is not maintained.

Most therapies aim to surgically remove the source of LBP or re-establish IVD mechanical functions overlooking the sustainability of such operations, which is part of why they have a low success rate.[51] Intradiscal biological therapy is a non-invasive alternative that consists of injecting genes, growth factors, and other molecules to the IVDthat aim to boost the cell population and produce ECM to restore a physiological environment.[51] Yet, increasing the cell number without increasing the nutrient supply is unsustainable because it will lead to the supply–demand disturbance, as previously mentioned.[51] A FE analysis study demonstrated that cell injection could lead to increased and accelerated degeneration in the IVD due to a higher supply–demand disturbance that is directly related to the state of the CEP, that is, whether it is healthy, calcified, or thinned.[54] Thus, the harsh nutrient environment of the IVD, particularly concerning the state of the CEP, must be accounted for to make any cell therapy beneficial.They suggested reducing the CEP thickness to enhance nutrition in the IVD, but this could compromise the mechanical stability ofthe IVD.

Treatments targeting the CEP are often focused on enhancing permeability. One possible method would be decalcification of the CEP through injecting compounds that can bind calcium.[209] Other methods suggested include enzymatic treatment, such as trypsin or hyaluronidase, to remove large proteoglycans from the CEP.[209] However, chemical injection with enzymes is also used to induce degeneration in animal models.[180] Therefore, the use of enzymes for CEP therapies should be strictly restricted to the CEP. Excessive/ untargeted protease activity as in antiquated chemonucleolysis can even trigger MC within 6 weeks.[210] Dolor et al.[207] treated the CEP with MMP8 to reduce the matrix and enhance solute uptake and nutrient diffusion. MMP8 is selective for type II collagen and aggre-can, which are the two main matrix components of the CEP. However,these are also the main components of the NP, and therefore it is cru-cial that using MMP8 or another enzyme does not induce degenera- tion within the NP or AF due to off-target digestion or matrix fragments triggering a catabolic response. One method Dolor et al.[207]considered to avoid this was using targeted delivery through injectionand linking the enzyme to bulky nanoparticles that cannot migrate toother tissues. Nevertheless, this treatment was only performed in human cadaveric CEPs, so it is unknown whether a catabolic responsewill be produced testing in vivo.

Although therapies addressing the CEP are still preliminary, stud-ies have shown that incorporating the CEP into a tissue engineereddisc improves the performance.[211] While development of functionalNP and AF replacements is important, these tissues must be inte-grated into the CEP to allow for successful and functional IVD replacement.[209,212] Studies have shown that using direct contact co-culture of AF[212] or NP cell-seeded scaffolds[213] with chondrocyte- seeded scaffolds produced native interface characteristics. However, although type I collagen, type II collagen, and aggrecan distribution were like native tissue, the apparent mechanical strength was 57-times weaker than in native tissue segments, which means it would not function well under daily mechanical loads.[212] Obtaining comparable mechanical properties of a native disc has been a large problem in tissue engineered scaffolds.[214,215] Gullbrand et al.[211] have tested endplate-modified disc-like angle ply structures (eDAPS) as replacement discs in rat and goat animal models. In the eDAPS, the endplate was made up of acellular, porous polyE-caprolactone (PCL) foams which was combined with the NP and AF components.[211] They showed that after 20 weeks with external fixation, native cells from neighboring tissues could migrate into the CEP structure and start producing matrix components and sparse vascularization,[211] which is a focus area of CEP treatments.[209] However, it should be noted that while vascularization is important for the BEP and CEP, it can lead to increased degeneration and pain if angiogenesis occurs in the NP.[216] Although Gullbrand et al.[211] observed improvements in tensile properties, the failure strain of the eDAPS was only 50% that of native values. However, it was shown that the constructs which included a CEP structure outperformed those without.[217]

Bioprinting is also a popular technique for developing tissue engi-neered constructs but has the same issue of sub-optimal mechanical properties.[214] Printing using bioinks with reinforcement structures such as carbon fibers or alumina platelets has been considered for rec-reating load-bearing tissues such as the CEP, although low printing resolutions can limit the functionality.[214] Using decellularized ECM is another option for tissue engineered constructs, which addresses the problem of low printing resolutions, and could help with the design of 3D printed scaffolds.[215] Nevertheless, mechanics of chemically modi-fied decellularized ECM also have weak mechanical properties that do not approach the Young's modulus of native tissues.[215] Further, scal-ability and reproducibility alongside high manufacturing costs are lim-iting in the bio fabrication of IVD constructs.[214]

Recently, Liu et al.[21] have identified the presence of progenitorcells in the CEP. After culturing in agarose, cells isolated from degen-erated human CEPs were found to be positive for stem cell markers *OCT-4*, *NANOG*, *and SOX-2* as well as common BM-MSCs markers *CD105*, *CD73*, *CD90*, *CD44*, *CD166*, and *Stro-1*.[21] Further, the group found that NP cells that were stimulated with CEP progenitor cellsisolated from healthy subjects showed a decrease in apoptotic rate by releasing exosomes that activate the PI3K/AKT signaling pathway.[218] Thus, CEP progenitor cell-derived exosomes could be a possible ther-apeutic tool in the treatment of IVD degeneration.

Several studies have also found potential targets for the treat-ment of the CEP with regard to IVD degeneration, although addi- tional research needs to be done before testing. For example, the HIF1A/MIF pathways has been shown to play a role in promoting chondrogenesis, while also inhibiting osteogenesis.[219] EZH2 inhibi- tion has also shown promise as a therapeutic target to combat CEP degeneration through upregulation of SOX9.[204] Managing oxidative stress and damage could be another novel therapy target. For this, research has found inhibiting ROS reduced apoptosis in CEP cells under oxidative stress[198] and, similarly, enhancing the NrF/Keap1 pathway in CEP cells increases antioxidants that can combat damage from ROS.[220] Certain microRNAs (miRNAs), specifically miR-495-3p[202] and miR-34a,[201] have been found to play a role in ECM

degradation and chondrocyte apoptosis, respectively. Therefore, silencing these miRNAs could be a novel treatment for CEP and IVD degeneration.

CONCLUSION

The CEP is a unique tissue distinct from other cartilaginous tissues in morphology, gene expression, and mechanical and transport properties. It is an essential component of the IVD and is considered to play a key role in the early stages of IVD degeneration due to its fundamental role in nutrient transport.[10,221] However, most data regarding IVD degeneration have focused on the NP and AF tissues. Further, research that does include the CEP often does not distinguish it from the BEP. New MRI techniques[114,116] as well as the standardization of histopathology scoring in the CEP[74] allow for the characterization of the CEP separate from the IVD or BEP, and thus future research should aim to investigate the CEP as an independent tissue type. In particular, the role of the CEP in MC should be investigated further as most research is limited to the BEP or the combined BEP and CEP.

Additionally, the CEP is often considered to be the same as AC particularly in modeling and simulations.[10] However, as detailed in this review, the CEP differs from AC in function, cellular response, biochemical content, and material properties.[9,24,60,65–67,89,197] Thus, future research should aim to characterize the CEP itself, and avoid assumptions that the CEP will behave and/or respond the same as AC.

Much is still unknown about the mechanics and transport properties of the CEP, and reported values show a large variation. The wide range of reported values is due to various testing methods, environmental conditions, species, and degree of degeneration,[4,10,52,53,60] and thus highlights the need for standardized, reproducible methods and guidelines for investigating the CEP. Likewise, the authors recommend to standardize specifying the terms cartilaginous endplate (CEP) and bony endplate (BEP) and advise avoiding the term vertebral endplate. Further, it is essential that researchers clearly state which tissue is being investigated, whether it is CEP, BEP, or a combination of both.

Calcification should also be a focus of future research, as there is controversy regarding the role it plays in CEP permeability. It is accepted that calcification occurs in degeneration. However, it is unclear whether disc damage occurs at an early stage due to impaired nutrient transport induced by calcification, cellular level changes caused by excess Ca^{2+} in the environment,[6,10,51,52,88,208] and/or due to early depletion of the CEP ECM (possibly related with MC) disrupting the functional fluid exchange between the vertebrae and NP under mechanical loads.[10]

Overall, much is still unknown about the CEP and the mechanisms of CEP degeneration. Additional research is needed to elucidate the mechanical and transport properties, gene expression, cellular response, and how these traits change with degeneration and age. Understanding the CEP is essential to develop therapies that target or include the CEP. Notably, the CEP should be considered in any treatment of the IVD, as the nutrient and waste transport must be functional for any therapy targeting the NP or AF to be successful. Thus, any long-lasting and sustainable therapy aiming to reverse IVD degeneration should target the CEP first or simultaneously with the NP and AF to rescue the IVD from a pathological environment.

AUTHOR CONTRIBUTIONS

Katherine B. Crump, Ahmad Alminnawi, Liesbet Geris, Christine Le Maitre, and Benjamin Gantenbein contributed to the main conception and design of this review. Katherine B. Crump, Ahmad Alminnawi, Paola Bermudez-Lekerika, Roger Compte, Francesco Gualdi, Terence McSweeney, Estefano Muñoz-Moya, Andrea Nüesch, Stefan Dudli, and Christine Le Maitre contributed to the literature review and drafted the text. Liesbet Geris, Stefan Dudli, Jaro Karppinen, Jérôme Noailly, Christine Le Maitre, and Benjamin Gantenbein provided scientific guidance. Liesbet Geris, Jaro Karppinen, Jérôme Noailly, ChristineLe Maitre, and Benjamin Gantenbein sourced funding. All authors edited the text and approved the final version of the manuscript. Katherine B. Crump, Ahmad Alminnawi, Paola Bermudez-Lekerika, Terence McSweeney, Andrea Nüesch, and Christine Le Maitre contributed to the figures.

ACKNOWLEDGMENTS

This research review was financed by the Marie Skłodowska Curie International Training Network (ITN) "disc4all" (https://disc4all.upf.edu, accessed on 3 September 2023) grant agreement #955735 (https://cordis.europa.eu/project/id/955735, accessed on 3 September 2023).

CONFLICT OF INTEREST STATEMENT

Benjamin Gantenbein and Christine Le Maitre are editorial board members of JOR Spine and co-author of this article. They were excluded from editorial decision-making related to the acceptance of this article for publication in the journal. All other authors have no conflicts of interest to declare in relation to this article.

ORCID

Katherine B. Crump https://orcid.org/0000-0001-5328-667X
Ahmad Alminnawi https://orcid.org/0000-0002-9998-3275
Paola Bermudez-Lekerika https://orcid.org/0000-0002-6858-8213
Roger Compte https://orcid.org/0000-0002-7034-1092
Francesco Gualdi https://orcid.org/0000-0003-0449-9884
Terence McSweeney https://orcid.org/0000-0003-0966-210X
Estefano Muñoz-Moya https://orcid.org/0000-0001-5222-4071
Andrea Nüesch https://orcid.org/0000-0002-0962-8705
Liesbet Geris https://orcid.org/0000-0002-8180-1445
Stefan Dudli https://orcid.org/0000-0001-7351-713X
Jaro Karppinen https://orcid.org/0000-0002-2158-6042 *Jérôme Noailly* https://orcid.org/0000-0003-3446-7621 *Christine Le Maitre* https://orcid.org/0000-0003-4489-7107 *Benjamin Gantenbein* https://orcid.org/0000-0002-9005-0655

REFERENCES

1. Lama P, Zehra U, Balkovec C, et al. Significance of cartilage endplate within herniated disc tissue. *Eur Spine J*. 2014;*23*:1869-1877.
2. Roberts S, Menage J, Urban JP. Biochemical and structural proper- ties of the cartilage end-plate and its relation to the intervertebral disc. *Spine*. 1989;*14*:166-174.
3. Berg-Johansen B, Han M, Fields AJ, et al. Cartilage endplate thick- ness variation measured by ultrashort Echo-time MRI is associated with adjacent disc degeneration. *Spine*. 2018;*43*:E592-E600.
4. Sampson SL, Sylvia M, Fields AJ. Effects of dynamic loading on sol- ute transport through the human cartilage endplate. *J Biomech*. 2019;*83*:273-279.
5. Urban JP, Smith S, Fairbank JC. Nutrition of the intervertebral disc. *Spine*. 2004;*29*:2700-2709.
6. Roberts S, Urban JP, Evans H, Eisenstein SM. Transport properties of the human cartilage endplate in relation to its composition and calcification. *Spine*. 1996;*21*:415-420.
7. Wade KR, Robertson PA, Broom ND. A fresh look at the nucleus- endplate region: new evidence for significant structural integration. *Eur Spine J*. 2011;*20*:1225-1232.
8. Wade KR, Robertson PA, Broom ND. On how nucleus-endplate inte- gration is achieved at the fibrillar level in the ovine lumbar disc. *J Anat*. 2012;*221*:39-46.
9. DeLucca JF, Cortes DH, Jacobs NT, Vresilovic EJ, Duncan RL, Elliott DM. Human cartilage endplate permeability varies with degeneration and intervertebral disc site. *J Biomech*. 2016;*49*: 550-557.
10. Ruiz Wills C, Foata B, González Ballester MA, Karppinen J, Noailly J. Theoretical explorations generate new hypotheses about the role of the cartilage endplate in early intervertebral disk degeneration. *Front Physiol*. 2018;*9*:1210.
11. Oegema TR. Biochemistry of the intervertebral disc. *Clin Sports Med*. 1993;*12*:419-439.
12. Wang Y, Wei L, Zeng L, He D, Wei X. Nutrition and degeneration of articular cartilage. *Knee Surg Sports Traumatol Arthrosc*. 2013;*21*: 1751-1762.
13. O'Hara BP, Urban JP, Maroudas A. Influence of cyclic loading on the nutrition of articular cartilage. *Ann Rheum Dis*. 1990;*49*:536-539.
14. Benneker LM, Heini PF, Alini M, Anderson SE, Ito K. 2004 young investigator award winner: vertebral endplate marrow contact chan- nel occlusions and intervertebral disc degeneration. *Spine*. 2005;*30*: 167-173.
15. Nachemson A, Lewin T, Maroudas A, Freeman MA. In vitro diffusion of dye through the end-plates and the annulus fibrosus of human lumbar inter- vertebral discs. *Acta Orthop Scand*. 1970;*41*:589-607.
16. Sivan S, Neidlinger-Wilke C, Würtz K, Maroudas A, Urban JP. Diurnal fluid expression and activity of intervertebral disc cells. *Bior- heology*. 2006;*43*:283-291.
17. Kushchayev SV, Glushko T, Jarraya M, et al. ABCs of the degenera- tive spine. *Insights Imaging*. 2018;*9*:253-274.
18. Lotz JC, Fields AJ, Liebenberg EC. The role of the vertebral end plate in low back pain. *Global Spine J*. 2013;*3*:153-164.
19. Herrero CF, Garcia SB, Garcia LV, Aparecido Defino HL. Endplates changes related to age and vertebral segment. *Biomed Res Int*. 2014; *2014*:545017.
20. Lakstins K, Arnold L, Gunsch G, Khan S, Moore S, Purmessur D. Characterization of bovine and canine animal model cartilage end- plates and comparison to human cartilage endplate structure, matrix composition, and cell phenotype. *JOR Spine*. 2020;*3*:e1116.
21. Liu LT, Huang B, Li CQ, Zhuang Y, Wang J, Zhou Y. Characteristics of stem cells derived from the degenerated human intervertebral disc cartilage endplate. *PLoS One*. 2011;*6*:e26285.
22. Gradiš̌nik L, Maver U, Gole B, et al. The endplate role in degenerative disc disease research: the isolation of human chondrocytes from verte- bral endplate-An optimised protocol. *Bioengineering (Basel)*. 2022;*9*:137.
23. Alkhatib B, Ban GI, Williams S, Serra R. IVD development: nucleus pulposus development and sclerotome specification. *Curr Mol Biol Rep*. 2018;*4*:132-141.
24. Sophia Fox A, Bedi A, Rodeo SA. The basic science of articular carti- lage: structure, composition, and function. *Sports Health*. 2009;*1*: 461-468.
25. Alsalameh S, Amin R, Gemba T, Lotz M. Identification of mesenchy- mal progenitor cells in normal and osteoarthritic human articular car- tilage. *Arthritis Rheum*. 2004;*50*:1522-1532.
26. Roberts S, Evans H, Trivedi J, Menage J. Histology and pathology of the human intervertebral disc. *J Bone Joint Surg Am*. 2006;*88*(Suppl 2):10-14.
27. Tam PP, Trainor PA. Specification and segmentation of the paraxial mesoderm. *Anat Embryol (Berl)*. 1994;*189*:275-305.
28. Borycki AG, Mendham L, Emerson CP. Control of somite patterning by sonic hedgehog and its downstream signal response genes. *Devel- opment*. 1998;*125*:777-790.
29. Oda J, Tanaka H, Tsuzuki N. Intervertebral disc changes with aging of human cervical vertebra. From the neonate to the eighties. *Spine*. 1988;*13*:1205-1211.
30. Moore RJ. The vertebral endplate: disc degeneration, disc regenera- tion. *Eur Spine J*. 2006;*15*(Suppl 3):S333-S337.
31. Zhang Y, Lenart BA, Lee JK, et al. Histological features of endplates of the mammalian spine: from mice to men. *Spine*. 2014;*39*:E312-E317.
32. Zehra U, Siyaji ZK, Mallow GM, et al. 10 – Vertebral endplate abnor- malities, defects, and changes. In: Samartzis D, Karppinen JI, Williams FM, eds. *Spine Phenotypes*. Academic Press; 2022:203-222.
33. Zehra U, Cheung JPY, Bow C, Lu W, Samartzis D. Multidimensional vertebral endplate defects are associated with disc degeneration, Modic changes, facet joint abnormalities, and pain. *J Orthop Res*. 2019;*37*:1080-1089.
34. Taylor JR. Growth of human intervertebral discs and vertebral bod- ies. *J Anat*. 1975;*120*:49-68.
35. Donisch EW, Trapp W. The cartilage endplates of the human verte- bral column (some considerations of postnatal development). *Anat Rec*. 1971;*169*:705-716.
36. Fournier DE, Kiser PK, Shoemaker JK, Battié MC, Séguin CA. Vascu- larization of the human intervertebral disc: a scoping review. *JOR Spine*. 2020;*3*:e1123.
37. Kuettner KE. Biochemistry of articular cartilage in health and dis- ease. *Clin Biochem*. 1992;*25*:155-163.
38. Sun Z, Liu B, Luo ZJ. The immune privilege of the intervertebral disc: implications for intervertebral disc degeneration treatment. *Int J Med Sci*. 2020;*17*:685-692.
39. Shapiro F. Epiphyseal and physeal cartilage vascularization: a light microscopic and tritiated thymidine autoradiographic study of carti- lage canals in newborn and young postnatal rabbit bone. *Anat Rec*. 1998;*252*:140-148.
40. Hirao M, Tamai N, Tsumaki N, Yoshikawa H, Myoui A. Oxygen ten- sion regulates chondrocyte differentiation and function during endo- chondral ossification. *J Biol Chem*. 2006;*281*:31079-31092.
41. Bernabei I, So A, Busso N, Nasi S. Cartilage calcification in osteoar- thritis: mechanisms and clinical relevance. *Nat Rev Rheumatol*. 2023; *19*:10-27.
42. Wu M, Chen G, Li Y-P. TGF-β and BMP signaling in osteoblast, skel- etal development, and bone formation, homeostasis and disease. *Bone Res*. 2016;*4*:6009.
43. Rufai A, Benjamin M, Ralphs JR. The development of fibrocartilage in the rat intervertebral disc. *Anat Embryol (Berl)*. 1995;*192*:53-62.
44. Ma B, Leijten JC, Wu L, et al. Gene expression profiling of dediffer- entiated human articular chondrocytes in monolayer culture. *Osteoarth Cartil*. 2013;*21*:599-603.
45. Zhao FD, Pollintine P, Hole BD, Adams MA, Dolan P. Vertebral frac- tures usually affect the cranial endplate because it is thinner and supported by less-dense trabecular bone. *Bone*. 2009;*44*:372-379.
46. Wallace AL, Wyatt BC, McCarthy ID, Hughes SP. Humoral regula- tion of blood flow in the vertebral endplate. *Spine*. 1994;*19*:1324- 1328.

47. Paietta RC, Burger EL, Ferguson VL. Mineralization and collagen orientation throughout aging at the vertebral endplate in the human lumbar spine. *J Struct Biol*. 2013;*184*:310-320.
48. Berg-Johansen B, Fields AJ, Liebenberg EC, Li A, Lotz JC. Structure-function relationships at the human spinal disc-vertebra Interface. *J Orthop Res*. 2017;*36*:192-201.
49. Fields AJ, Ballatori A, Liebenberg EC, Lotz JC. Contribution of the endplates to disc degeneration. *Curr Mol Biol Rep*. 2018;*4*:151-160.
50. Shirazi-Adl A, Taheri M, Urban JP. Analysis of cell viability in intervertebral disc: effect of endplate permeability on cell population. *J Biomech*. 2010;*43*:1330-1336.
51. Wong J, Sampson SL, Bell-Briones H, et al. Nutrient supply and nucleus pulposus cell function: effects of the transport properties of the cartilage endplate and potential implications for intradiscal bio- logic therapy. *Osteoarthr Cartil*. 2019;*27*:956-964.
52. Rodriguez AG, Slichter CK, Acosta FL, et al. Human disc nucleus properties and vertebral endplate permeability. *Spine*. 2011;*36*: 512-520.
53. Accadbled F, Laffosse JM, Ambard D, Gomez-Brouchet A, de Gauzy JS, Swider P. Influence of location, fluid flow direction, and tissue maturity on the macroscopic permeability of vertebral end plates. *Spine*. 2008;*33*:612-619.
54. Wu Y, Cisewski S, Sachs BL, Yao H. Effect of cartilage endplate on cell based disc regeneration: a finite element analysis. *Mol Cell Bio- mech*. 2013;*10*:159-182.
55. Green TP, Adams MA, Dolan P. Tensile properties of the annulus fibrosus II. Ultimate tensile strength and fatigue life. *Eur Spine J*. 1993;*2*:209-214.
56. Rodrigues SA, Wade KR, Thambyah A, Broom ND. Micromechanics of annulus-end plate integration in the intervertebral disc. *Spine J*. 2012;*12*:143-150.
57. Rodrigues SA, Thambyah A, Broom ND. How maturity influences annulus-endplate integration in the ovine intervertebral disc: a micro- and ultra-structural study. *J Anat*. 2017;*230*:152-164.
58. Wu Y, Cisewski SE, Wegner N, et al. Region and strain-dependent diffusivities of glucose and lactate in healthy human cartilage end- plate. *J Biomech*. 2016;*49*:2756-2762.
59. Balkovec C, Adams MA, Dolan P, McGill SM. Annulus fibrosus can strip hyaline cartilage end plate from subchondral bone: a study of the intervertebral disk in tension. *Global Spine J*. 2015;*5*:360-365.
60. Fields AJ, Rodriguez D, Gary KN, Liebenberg EC, Lotz JC. Influence of biochemical composition on endplate cartilage tensile properties in the human lumbar spine. *J Orthop Res*. 2014;*32*:245-252.
61. Mow VC, Kuei SC, Lai WM, Armstrong CG. Biphasic creep and stress relaxation of articular cartilage in compression? Theory and experiments. *J Biomech Eng*. 1980;*102*:73-84.
62. Pattappa G, Li Z, Peroglio M, Wismer N, Alini M, Grad S. Diversity of intervertebral disc cells: phenotype and function. *J Anat*. 2012;*221*: 480-496.
63. Holzapfel GA, Schulze-Bauer CA, Feigl G, Regitnig P. Single lamellar mechanics of the human lumbar anulus fibrosus. *Biomech Model Mechanobiol*. 2005;*3*:125-140.
64. Lüsse S, Claassen H, Gehrke T, et al. Evaluation of water content by spatially resolved transverse relaxation times of human articular car- tilage. *Magn Reson Imaging*. 2000;*18*:423-430.
65. Lakstins K, Arnold L, Gunsch G, et al. Characterization of the human intervertebral disc cartilage endplate at the molecular, cell, and tis- sue levels. *J Orthop Res*. 2021;*39*:1898-1907.
66. Stockwell RA. The interrelationship of cell density and cartilage thick- ness in mammalian articular cartilage. *J Anat*. 1971;*109*:411-421.
67. Maroudas A, Stockwell RA, Nachemson A, Urban J. Factors involved in the nutrition of the human lumbar intervertebral disc: cellularity and diffusion of glucose in vitro. *J Anat*. 1975;*120*:113-130.
68. Gan Y, He J, Zhu J, et al. Spatially defined single-cell transcriptional profiling characterizes diverse chondrocyte subtypes and nucleus pulposus progenitor progenitors in human intervertebral discs. Bone Res. 2021; 9:37.
69. Lee GM, Poole CA, Kelley SS, Chang J, Caterson B. Isolated chon- drons: a viable alternative for studies of chondrocyte metabolism in vitro. *Osteoarthr Cartil*. 1997;*5*:261-274.
70. Roberts S, Menage J, Duance V, Wotton S, Ayad S. 1991 Volvo award in basic sciences. Collagen types around the cells of the inter- vertebral disc and cartilage end plate: an immunolocalization study. *Spine*. 1991;*16*:1030-1038.
71. Bishop PB, Pearce RH. The proteoglycans of the cartilaginous end- plate of the human intervertebral disc change after maturity. *J Orthop Res*. 1993;*11*:324-331.
72. Aigner T, Gresk-otter KR, Fairbank JC, von der Mark K, Urban JP. Variation with age in the pattern of type X collagen expression in normal and scoliotic human intervertebral discs. *Calcif Tissue Int*. 1998;*63*:263-268.
73. Sahlman J, Inkinen R, Hirvonen T, et al. Premature vertebral endplate ossification and mild disc degeneration in mice after inactivation of one allele belonging to the Col2a1 gene for type II collagen. *Spine*. 2001;*26*:2558-2565.
74. Le Maitre CL, Dahia CL, Giers M, et al. Development of a standard- ized histopathology scoring system for human intervertebral disc degeneration: an Orthopaedic Research Society Spine Section Initiative. *JOR Spine*. 2021;*4*:e1167.
75. Cao L, Guilak F, Setton LA. Pericellular matrix mechanics in the anu- lus fibrosus predicted by a three-dimensional finite element model and in situ morphology. *Cell Mol Bioeng*. 2009;*2*:306-319.
76. Tang X, Jing L, Chen J. Changes in the molecular phenotype of nucleus pulposus cells with intervertebral disc aging. *PLoS One*. 2012;*7*:e52020.
77. van den Akker GG, Surtel DA, Cremers A, et al. Novel immortal cell lines support cellular heterogeneity in the human annulus Fibrosus. *PLoS One*. 2016;*11*:e0144497.
78. Minogue BM, Richardson SM, Zeef LA, Freemont AJ, Hoyland JA. Characterization of the human nucleus pulposus cell phenotype and evaluation of novel marker gene expression to define adult stem cell differentiation. *Arthritis Rheum*. 2010;*62*:3695-3705.
79. Minogue BM, Richardson SM, Zeef LA, Freemont AJ, Hoyland JA. Transcriptional profiling of bovine intervertebral disc cells: implica- tions for identification of normal and degenerate human interverteb- ral disc cell phenotypes. *Arthritis Res Ther*. 2010;*12*:R22.
80. Rutges J, Creemers LB, Dhert W, et al. Variations in gene and pro- tein expression in human nucleus pulposus in comparison with annu- lus fibrosus and cartilage cells: potential associations with aging and degeneration. *Osteoarthr Cartil*. 2010;*18*:416-423.
81. Hodgkinson T, Kelly DC, Curtin CM, O'Brien FJ. Mechanosignalling in cartilage: an emerging target for the treatment of osteoarthritis. *Nat Rev Rheumatol*. 2022;*18*:34934171.
82. Fearing BV, Hernandez PA, Setton LA, Chahine NO. Mechanotrans- duction and cell biomechanics of the intervertebral disc. *JOR Spine*. 2018;*1*:e1026.
83. Stokes IA, Iatridis JC. Mechanical conditions that accelerate interver- tebral disc degeneration: overload versus immobilization. *Spine*. 2004;*29*:2724-2732.
84. Kuo J, Zhang L, Bacro T, Yao H. The region-dependent biphasic vis- coelastic properties of human temporomandibular joint discs under confined compression. *J Biomech*. 2010;*43*:1316-1321.
85. Wu Y, Cisewski SE, Sachs BL, et al. The region-dependent biome- chanical and biochemical properties of bovine cartilaginous endplate. *J Biomech*. 2015;*48*:3185-3191.
86. Setton LA, Zhu W, Weidenbaum M, Ratcliffe A, Mow VC. Compres- sive properties of the cartilaginous end-plate of the baboon lumbar spine. *J Orthop Res*. 1993;*11*:228-239.
87. Périé D, Korda D, Iatridis JC. Confined compression experiments on bovine nucleus pulposus and annulus fibrosus: sensitivity of the experiment in the determination of compressive modulus and hydraulic permeability. J Biomech. 2005;38:2164-2171.

88. Malandrino A, Lacroix D, Hellmich C, Ito K, Ferguson SJ, Noailly J. The role of endplate poromechanical properties on the nutrient availability in the intervertebral disc. *Osteoarthr Cartil*. 2014;*22*: 1053-1060.
89. Akizuki S, Mow VC, Müller F, Pita JC, Howell DS, Manicourt DH. Tensile properties of human knee joint cartilage: I. Influence of ionic conditions, weight bearing, and fibrillation on the tensile modulus. *J Orthop Res*. 1986;*4*:379-392.
90. Malandrino A, Planell JA, Lacroix D. Statistical factorial analysis on the poroelastic material properties sensitivity of the lumbar interver- tebral disc under compression, flexion and axial rotation. *J Biomech*. 2009;*42*:2780-2788.
91. Huyghe JM. Intra-extrafibrillar mixture formulation of soft charged hydrated tissues. *J Theor Appl Mech*. 1999;*3*:519-536.
92. Jackson AR, Eismont A, Yu L, et al. Diffusion of antibiotics in inter- vertebral disc. *J Biomech*. 2018;*76*:259-262.
93. Zhu Q, Gao X, Brown MD, Eismont F, Gu W. Effects of diurnal load- ing on the transport of charged antibiotics into intervertebral discs. *J Biomech*. 2019;*87*:177-182.
94. Huang CY, Gu WY. Effects of mechanical compression on metabo- lism and distribution of oxygen and lactate in intervertebral disc. *J Biomech*. 2008;*41*:1184-1196.
95. Schmidt H, Galbusera F, Rohlmann A, Shirazi-Adl A. What have we learned from finite element model studies of lumbar intervertebral discs in the past four decades? *J Biomech*. 2013;*46*:2342-2355.
96. Zhu Q, Jackson AR, Gu WY. Cell viability in intervertebral disc under various nutritional and dynamic loading conditions: 3d finite element analysis. *J Biomech*. 2012;*45*:2769-2777.
97. Hassan CR, Lee W, Komatsu DE, Qin Y-X. Evaluation of nucleus pul- posus fluid velocity and pressure alteration induced by cartilage end- plate sclerosis using a poro-elastic finite element analysis. *Biomech Model Mechanobiol*. 2020;*20*:281-291.
98. Riches PE, McNally DS. A one-dimensional theoretical prediction of the effect of reduced end-plate permeability on the mechanics of the intervertebral disc. *Proc Inst Mech Eng H*. 2005;*219*:329-335.
99. Ayotte DC, Ito K, Perren SM, Tepic S. Direction-dependent constric- tion flow in a poroelastic solid: the intervertebral disc valve. *J Biomech Eng*. 2000;*122*:587-593.
100. Ayotte DC, Ito K, Tepic S. Direction-dependent resistance to flow in the endplate of the intervertebral disc: an ex vivo study. *J Orthop Res*. 2001;*19*:1073-1077.
101. Swider P, Accadbled F, Laffosse JM, Sales de Gauzy J. Influence of fluid-flow direction on effective permeability of the vertebral end plate: an analytical model. *Comput Methods Biomech Biomed Engin*. 2012;*15*:151-156.
102. Rade M, Määttä JH, Freidin MB, Airaksinen O, Karppinen J, Williams FMK. Vertebral endplate defect as initiating factor in inter- vertebral disc degeneration: strong association between endplate defect and disc degeneration in the general population. *Spine*. 2018; *43*:412-419.
103. Holm S, Holm AK, Ekström L, Karladani A, Hansson T. Experimental disc degeneration due to endplate injury. *J Spinal Disord Tech*. 2004; *17*:64-71.
104. Farshad-Amacker NA, Hughes A, Herzog RJ, Seifert B, Farshad M. The intervertebral disc, the endplates and the vertebral bone mar- row as a unit in the process of degeneration. *Eur Radiol*. 2017;*27*: 2507-2520.
105. Dudli S, John Ferguson S, Haschtmann D. Severity and pattern of posttraumatic intervertebral disc degeneration depends on the type of injury. *Spine J*. 2014;*14*:1256-1264.
106. Kuisma M, Karppinen J, Haapea M, Lammentausta E, Niinimäki J, Tervonen O. Modic changes in vertebral endplates: a comparison of MR imaging and multislice CT. *Skeletal Radiol*. 2009;*38*:141-147.
107. Feng Z, Liu Y, Yang G, Battié MC, Wang Y. Lumbar vertebral end- plate defects on magnetic resonance images: classification, distribu- tion patterns, and associations with Modic changes and disc degeneration. *Spine*. 2018;*43*:919-927.
108. Li Y, Samartzis D, Campbell DD, et al. Two subtypes of interverteb- ral disc degeneration distinguished by large-scale population-based study. *Spine J*. 2016;*16*:79-89.
109. Määttä JH, Rade M, Freidin MB, Airaksinen O, Karppinen J, Williams FMK. Strong association between vertebral endplate defect and Modic change in the general population. *Sci Rep*. 2018;*8*: 16630.
110. Davies BM, Atkinson RA, Ludwinski F, Freemont AJ, Hoyland JA, Gnanalingham KK. Qualitative grading of disc degeneration by mag- netic resonance in the lumbar and cervical spine: lack of correlation with histology in surgical cases. *Br J Neurosurg*. 2016;*30*:414-421.
111. Benneker LM, Heini PF, Anderson SE, Alini M, Ito K. Correlation of radiographic and MRI parameters to morphological and biochemical assessment of intervertebral disc degeneration. *Eur Spine J*. 2005; *14*:27-35.
112. Fields AJ, Battié MC, Herzog RJ, et al. Measuring and reporting of vertebral endplate bone marrow lesions as seen on MRI (Modic changes): recommendations from the ISSLS degenerative spinal phe- notypes group. *Eur Spine J*. 2019;*28*:2266-2274.
113. Loeuille D, Olivier P, Mainard D, Gillet P, Netter P, Blum A. Review: magnetic resonance imaging of normal and osteoarthritic cartilage. *Arthritis Rheum*. 1998;*41*:963-975.
114. Law T, Anthony MP, Chan Q, et al. Ultrashort time-to-echo MRI of the cartilaginous endplate: technique and association with interver- tebral disc degeneration. *J Med Imaging Radiat Oncol*. 2013;*57*: 427-434.
115. Fields AJ, Han M, Krug R, Lotz JC. Cartilaginous end plates: quanti- tative MR imaging with very short echo times-orientation depen- dence and correlation with biochemical composition. *Radiology*. 2015;*274*:482-489.
116. Bonnheim NB, Wang L, Lazar AA, et al. Deep-learning-based bio- marker of spinal cartilage endplate health using ultra-short echo time magnetic resonance imaging. *Quant Imaging Med Surg*. 2023; *13*:2807-2821.
117. Moon SM, Yoder JH, Wright AC, Smith LJ, Vresilovic EJ, Elliott DM. Evaluation of intervertebral disc cartilaginous endplate structure using magnetic resonance imaging. *Eur Spine J*. 2013;*22*: 1820-1828.
118. Ling Z, Li L, Chen Y, et al. Changes of the end plate cartilage are associated with intervertebral disc degeneration: a quantitative mag- netic resonance imaging study in rhesus monkeys and humans. *J Orthop Translat*. 2020;*24*:23-31.
119. Newton MD, Hartner SE, Timmons S, et al. Contrast-enhanced μCT of the intervertebral disc: a comparison of anionic and cationic con- trast agents for biochemical and morphological characterization. *J Orthop Res*. 2017;*35*:1067-1075.
120. Fields AJ, Liebenberg EC, Lotz JC. Innervation of pathologies in the lumbar vertebral end plate and intervertebral disc. *Spine J*. 2014;*14*: 513-521.
121. Brown MF, Hukkanen MV, McCarthy ID, et al. Sensory and sympa- thetic innervation of the vertebral endplate in patients with degen- erative disc disease. *J Bone Joint Surg Br*. 1997;*79*:147-153.
122. Ni S, Ling Z, Wang X, et al. Sensory innervation in porous endplates by Netrin-1 from osteoclasts mediates PGE2-induced spinal hyper- sensitivity in mice. *Nat Commun*. 2019;*10*:5643.
123. Wang Y, Videman T, Battié MC. ISSLS prize winner: lumbar vertebral endplate lesions: associations with disc degeneration and back pain history. *Spine*. 2012;*37*:1490-1496.
124. Ohtori S, Inoue G, Ito T, et al. Tumor necrosis factor- immunoreactive cells and PGP 9.5-immunoreactive nerve fibers in vertebral endplates of patients with discogenic low back pain and Modic type 1 or type 2 changes on MRI.Spine. 2006;31:1026-1031.

125. Heggli I, Laux CJ, Mengis T, et al. Modic type 2 changes are fibroinflammatory changes with complement system involvement adjacent to degenerated vertebral endplates. *JOR Spine*. 2022;*6*:e1237.
126. Nerlich AG, Schaaf R, Wälchli B, Boos N. Temporo-spatial distribution of blood vessels in human lumbar intervertebral discs. *Eur Spine J*. 2007;*16*:547-555.
127. Bailey JF, Fields AJ, Ballatori A, et al. The relationship between endplate pathology and patient-reported symptoms for chronic low back pain depends on lumbar paraspinal muscle quality. *Spine*. 2019; *44*:1010-1017.
128. Zehra U, Bow C, Lotz JC, et al. Structural vertebral endplate nomenclature and etiology: a study by the ISSLS spinal phenotype focus group. *Eur Spine J*. 2018;*27*:2-12.
129. Wang Y, Videman T, Battié MC. Lumbar vertebral endplate lesions: prevalence, classification, and association with age. *Spine*. 2012;*37*: 1432-1439.
130. Wang Y, Videman T, Niemeläinen R, Battié MC. Quantitative measures of Modic changes in lumbar spine magnetic resonance imag- ing: intra- and inter-rater reliability. *Spine*. 2011;*36*:1236-1243.
131. Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR. Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. *Radiology*. 1988;*166*:193-199.
132. Jensen TS, Karppinen J, Sorensen JS, Niinimäki J, Leboeuf-Yde C. Vertebral endplate signal changes (Modic change): a systematic liter- ature review of prevalence and association with non-specific low back pain. *Eur Spine J*. 2008;*17*:1407-1422.
133. Määttä JH, Wadge S, MacGregor A, Karppinen J, Williams FM. ISSLS prize winner: vertebral endplate (Modic) change is an independent risk factor for episodes of severe and disabling low back pain. *Spine*. 2015;*40*:1187-1193.
134. Kuisma M, Karppinen J, Niinimäki J, et al. Modic changes in endplates of lumbar vertebral bodies: prevalence and association with low back and sciatic pain among middle-aged male workers. *Spine*. 2007;*32*:1116-1122.
135. Järvinen J, Karppinen J, Niinimäki J, et al. Association between changes in lumbar Modic changes and low back symptoms over a two-year period. *BMC Musculoskelet Disord*. 2015;*16*:98.
136. Saukkonen J, Määttä J, Oura P, et al. Association between Modic changes and low back pain in middle age: a northern Finland birth cohort study. *Spine*. 2020;*45*:1360-1367.
137. Karchevsky M, Schweitzer ME, Carrino JA, Zoga A, Montgomery D, Parker L. Reactive endplate marrow changes: a systematic morpho- logic and epidemiologic evaluation. *Skeletal Radiol*. 2005;*34*: 125-129.
138. Lenz GP, Assheuer J, Lenz W, Gottschlich KW. New aspects of lumbar disc disease. MR imaging and histological findings. *Arch Orthop Trauma Surg*. 1990;*109*:75-82.
139. Assheuer J, Lenz G, Lenz W, Gottschlich KW, Schulitz KP. Fat/water separation in the NMR tomogram. The imaging of bone marrow reactions in degenerative intervertebral disk changes. *Rofo*. 1987; *147*:58-63.
140. Dudli S, Karol A, Giudici L, et al. CD90-positive stromal cells associate with inflammatory and fibrotic changes in Modic changes. *Osteoarthr Cartil Open*. 2022;*4*:100287.
141. Perilli E, Parkinson IH, Truong LH, Chong KC, Fazzalari NL, Osti OL. Modic (endplate) changes in the lumbar spine: bone micro- architecture and remodelling. *Eur Spine J*. 2015;*24*:1926-1934.
142. Zhang YH, Zhao CQ, Jiang LS, Chen XD, Dai LY. Modic changes: a systematic review of the literature. *Eur Spine J*. 2008;*17*:1289-1299.
143. Tarukado K, Ono T, Tono O, et al. Does Modic change progress with age? *Spine*. 2017;*42*:1805-1809.
144. Albert HB, Briggs AM, Kent P, Byrhagen A, Hansen C, Kjaergaard K. The prevalence of MRI-defined spinal pathoanatomies and their association with Modic changes in individuals seeking care for low back pain. *Eur Spine J*. 2011;*20*:1355-1362.
145. Jensen TS, Bendix T, Sorensen JS, Manniche C, Korsholm L, Kjaer P. Characteristics and natural course of vertebral endplate signal (Modic) changes in the Danish general population. *BMC Musculoske- let Disord*. 2009;*10*:81.
146. Kuisma M, Karppinen J, Niinimäki J, et al. A three-year follow-up of lumbar spine endplate (Modic) changes. *Spine*. 2006;*31*:1714-1718.
147. Mitra D, Cassar-Pullicino VN, McCall IW. Longitudinal study of vertebral type-1 end-plate changes on MR of the lumbar spine. *Eur Radiol*. 2004;*14*:1574-1581.
148. Vital JM, Gille O, Pointillart V, et al. Course of Modic 1 six months after lumbar posterior osteosynthesis. *Spine*. 2003;*28*:715-720.
149. Herlin C, Kjaer P, Espeland A, et al. Modic changes-their associations with low back pain and activity limitation: a systematic literature review and meta-analysis. *PLoS One*. 2018;*13*:e0200677.
150. Weishaupt D, Zanetti M, Hodler J, et al. Painful lumbar disk derangement: relevance of endplate abnormalities at MR imaging. *Radiology*. 2001;*218*:420-427.
151. Chen KC, Tran B, Biswas R, et al. Evaluation of the disco-vertebral junction using ultrashort time-to-echo magnetic resonance imaging: inter-reader agreement and association with vertebral endplate lesions. *Skeletal Radiol*. 2016;*45*:1249-1256.
152. Ashinsky BG, Bonnevie ED, Mandalapu SA, et al. Intervertebral disc degeneration is associated with aberrant endplate remodeling and reduced small molecule transport. *J Bone Miner Res*. 2020;*35*:1572- 1581.
153. Bonnheim NB, Wang L, Lazar AA, et al. The contributions of carti- lage endplate composition and vertebral bone marrow fat to inter- vertebral disc degeneration in patients with chronic low back pain. *Eur Spine J*. 2022;*31*:1866-1872.
154. Feng ZY, Hu XJ, Zheng QQ, Battié MC, Chen Z, Wang Y. Cartilaginous endplate avulsion is associated with Modic changes and end- plate defects, and residual back and leg pain following lumbar discectomy. *Osteoarthr Cartil*. 2021;*29*:707-717.
155. Schmid G, Witteler A, Willburger R, Kuhnen C, Jergas M, Koester O. Lumbar disk herniation: correlation of histologic findings with mar- row signal intensity changes in vertebral endplates at MR imaging. *Radiology*. 2004;*231*:352-358.
156. Dudli S, Boffa DB, Ferguson SJ, Haschtmann D. Leukocytes enhance inflammatory and catabolic degenerative changes in the interverteb- ral disc after endplate fracture in vitro without infiltrating the disc. *Spine*. 2015;*40*:1799-1806.
157. Dudli S, Sing DC, Hu SS, et al. ISSLS prize in basic science 2017: intervertebral disc/bone marrow cross-talk with Modic changes. *Eur Spine J*. 2017;*26*:1362-1373.
158. Heggli I, Habib M, Scheer J, et al. Poster No. 1086. Activated neutrophils degrade human cartilage endplates. In: Annual Meeting of the Orthopaedic Research Society 2023, Dallas.
159. Dudli S, Heggli I, Laux CJ, et al. Role of C-reactive protein in the bone marrow of Modic type 1 changes. *J Orthop Res*. 2022;*41*: 1115-1122.
160. Chen S, Huang Y, Zhou ZJ, et al. Upregulation of tumor necrosis fac- tor α and ADAMTS-5, but not ADAMTS-4, in human intervertebral cartilage endplate with Modic changes. *Spine*. 2014;*39*:E817-E825.
161. Xiong C, Huang B, Cun Y, Aghdasi BG, Zhou Y. Migration inhibitory factor enhances inflammation via CD74 in cartilage end plates with Modic type 1 changes on MRI. *Clin Orthop Relat Res*. 2014;*472*: 1943-1954.
162. Dudli S, Fields AJ, Samartzis D, Karppinen J, Lotz JC. Pathobiology of Modic changes. *Eur Spine J*. 2016;*25*:3723-3734.
163. Rollason J, McDowell A, Albert HB, et al. Genotypic and antimicro- bial characterisation of *Propionibacterium acnes* isolates from surgi- cally excised lumbar disc herniations. *Biomed Res Int*. 2013;*2013*: 530382.

164. Albert HB, Lambert P, Rollason J, et al. Does nuclear tissue infected with bacteria following disc herniations lead to Modic changes in the adjacent vertebrae? *Eur Spine J*. 2013;*22*:690-696.
165. Stirling A, Worthington T, Rafiq M, Lambert PA, Elliott TS. Association between sciatica and *Propionibacterium acnes*. *Lancet*. 2001; *357*:2024-2025.
166. Fritzell P, Bergström T, Welinder-Olsson C. Detection of bacterial DNA in painful degenerated spinal discs in patients without signs of clinical infection. *Eur Spine J*. 2004;*13*:702-706.
167. Urquhart DM, Zheng Y, Cheng AC, et al. Could low grade bacterial infection contribute to low back pain? A systematic review. *BMC Med*. 2015;*13*:13.
168. Shan Z, Zhang X, Li S, Yu T, Liu J, Zhao F. *Propionibacterium acnes* incubation in the discs can result in time-dependent Modic changes: a long-term rabbit model. *Spine*. 2017;*42*:1595-1603.
169. Chen Z, Zheng Y, Yuan Y, et al. Modic changes and disc degenera- tion caused by inoculation of *Propionibacterium acnes* inside inter- vertebral discs of rabbits: a pilot study. *Biomed Res Int*. 2016;*2016*: 9612437.
170. Dudli S, Liebenberg E, Magnitsky S, Miller S, Demir-Deviren S, Lotz JC. *Propionibacterium acnes* infected intervertebral discs cause vertebral bone marrow lesions consistent with Modic changes. *J Orthop Res*. 2016;*34*:1447-1455.
171. Jiao Y, Yuan Y, Lin Y, et al. *Propionibacterium acnes* induces discogenic low back pain via stimulating nucleus pulposus cells to secrete pro-algesic factor of IL-8/CINC-1 through TLR2-NF-κB p65 path- way. *J Mol Med (Berl)*. 2018;*97*:25-35.
172. Capoor MN, Konieczna A, McDowell A, et al. Pro-inflammatory and neurotrophic factor responses of cells derived from degenerative human intervertebral discs to the opportunistic pathogen *Cutibacter- ium acnes*. *Int J Mol Sci*. 2021;*22*:2347.
173. Dudli S, Miller S, Demir-Deviren S, Lotz JC. Inflammatory response of disc cells against *Propionibacterium acnes* depends on the pres- ence of lumbar Modic changes. *Eur Spine J*. 2018;*27*:1013-1020.
174. McDowell A, Barnard E, Nagy I, et al. An expanded multilocus sequence typing scheme for *Propionibacterium acnes*: investigation of 'pathogenic', 'commensal' and antibiotic resistant strains. *PLoS One*. 2012;*7*:e41480.
175. McDowell A, Perry AL, Lambert PA, Patrick S. A new phylogenetic group of *Propionibacterium acnes*. *J Med Microbiol*. 2008;*57*: 218-224.
176. Schäffler A, Schölmerich J. Innate immunity and adipose tissue biology. *Trends Immunol*. 2010;*31*:228-235.
177. Falcocchio S, Ruiz C, Pastor FIJ, Saso L, Diaz P. *Propionibacterium acnes* GehA lipase, an enzyme involved in acne development, can be successfully inhibited by defined natural substances. *J Mol Catal B: Enzym*. 2006;*40*:132-137.
178. Paulson JC, Kawasaki N. Sialidase inhibitors DAMPen sepsis. *Nat Biotechnol*. 2011;*29*:406-407.
179. Schroeder GD, Markova DZ, Koerner JD, et al. Are Modic changes associated with intervertebral disc cytokine profiles? *Spine J*. 2017; *17*:129-134.
180. Bermudez-Lekerika P, Crump KB, Tseranidou S, et al. Immuno-modulatory effects of intervertebral disc cells. *Front Cell Dev Biol*. 2022;*10*:924692.
181. Phillips KL, Chiverton N, Michael AL, et al. The cytokine and chemokine expression profile of nucleus pulposus cells: implications for degeneration and regeneration of the intervertebral disc. *Arthritis Res Ther*. 2013;*15*:R213.
182. Huang B, Liu J, Wei X, et al. Damage to the human lumbar cartilage endplate and its clinical implications. *J Anat*. 2021;*238*:338-348.
183. Rajasekaran S, Venkatadass K, Naresh Babu J, Ganesh K, Shetty AP. Pharmacological enhancement of disc diffusion and differentiation of healthy, ageing and degenerated discs: results from in-vivo serial post-contrast MRI studies in 365 human lumbar discs. *Eur Spine J*. 2008;*17*:626-643.
184. Natarajan RN, Ke JH, Andersson GB. A model to study the disc degeneration process. *Spine*. 1994;*19*:259-265.
185. Rajasekaran S, Bajaj N, Tubaki V, Kanna RM, Shetty AP. ISSLS prize winner: the anatomy of failure in lumbar disc herniation: an in vivo, multimodal, prospective study of 181 subjects. *Spine*. 2013;*38*: 1491-1500.
186. Wilke HJ, Kienle A, Maile S, Rasche V, Berger-Roscher N. A new dynamic six degrees of freedom disc-loading simulator allows to provoke disc damage and herniation (Georg Schmorl Preis 2015). *Eur Spine J*. 2016;*25*:1363-1372.
187. Zehr JD, Quadrilatero J, Callaghan JP. Initiation and accumulation of loading induced changes to native collagen content and microstruc- tural damage in the cartilaginous endplate. *Spine J*. 2023;S1529- 9430:03282-5. Online ahead of print.
188. Veres SP, Robertson PA, Broom ND. ISSLS prize winner: how loading rate influences disc failure mechanics: a microstructural assessment of internal disruption. *Spine*. 2010;*35*:1897-1908.
189. Mörl F, Günther M, Riede JM, Hammer M, Schmitt S. Loads distrib- uted in vivo among vertebrae, muscles, spinal ligaments, and inter- vertebral discs in a passively flexed lumbar spine. *Biomech Model Mechanobiol*. 2020;*19*:2015-2047.
190. Dudli S, Enns-Bray W, Pauchard Y, et al. Larger vertebral endplate concavities cause higher failure load and work at failure under high- rate impact loading of rabbit spinal explants. *J Mech Behav Biomed Mater*. 2018;*80*:104-110.
191. Langrana NA, Kale SP, Edwards WT, Lee CK, Kopacz KJ. Measurement and analyses of the effects of adjacent end plate curvatures on vertebral stresses. *Spine J*. 2006;*6*:267-278.
192. Grant MP, Epure LM, Bokhari R, Roughley P, Antoniou J, Mwale F. Human cartilaginous endplate degeneration is induced by calcium and the extracellular calcium-sensing receptor in the intervertebral disc. *Eur Cell Mater*. 2016;*32*:137-151.
193. Antoniou J, Goudsouzian NM, Heathfield TF, et al. The human lum- bar endplate. Evidence of changes in biosynthesis and denaturation of the extracellular matrix with growth, maturation, aging, and degeneration. *Spine*. 1996;*21*:1153-1161.
194. Ariga K, Miyamoto S, Nakase T, et al. The relationship between apoptosis of endplate chondrocytes and aging and degeneration of the intervertebral disc. *Spine*. 2001;*26*:2414-2420.
195. Zhang JF, Wang GL, Zhou ZJ, Fang XQ, Chen S, Fan SW. Expression of matrix metalloproteinases, tissue inhibitors of metalloproteinases, and interleukins in vertebral cartilage endplate. *Orthop Surg*. 2018; *10*:306-311.
196. Le Maitre CL, Freemont AJ, Hoyland JA. Localization of degradative enzymes and their inhibitors in the degenerate human intervertebral disc. *J Pathol*. 2004;*204*:47-54.
197. Lakstins K, Yeater T, Arnold L, Khan S, Hoyland JA, Purmessur D. Investigating the role of culture conditions on hypertrophic differen- tiation in human cartilage endplate cells. *J Orthop Res*. 2021;*39*: 1204-1216.
198. Han Y, Li X, Yan M, et al. Oxidative damage induces apoptosis and promotes calcification in disc cartilage endplate cell through ROS/-MAPK/NF-κB pathway: implications for disc degeneration. *Biochem Biophys Res Commun*. 2017;*516*:1026-1032.
199. Neidlinger-Wilke C, Boldt A, Brochhausen C, et al. Molecular interactions between human cartilaginous endplates and nucleus pulpo- sus cells: a preliminary investigation. *Spine*. 2014;*39*:1355-1364.
200. Jacobs WR. Science for combatting tuberculosis. *Bull N Y Acad Med*. 1996;*73*:46-52.
201. Chen H, Wang J, Hu B, et al. MiR-34a promotes Fas-mediated carti- lage endplate chondrocyte apoptosis by targeting Bcl-2. *Mol Cell Bio- chem*. 2015;*406*:21-30.
202. Zhang J, Hu S, Ding R, et al. CircSNHG5 sponges Mir-495-3p and modulates CITED2 to protect cartilage endplate from degradation. *Front Cell Dev Biol*. 2021;*9*:668715.

203. Zhao QH, Tian JW, Wang L, Dong SH, Xia T, Liu CY. Effect of siRNA-Cox-2 on the growth inhibition and apoptosis of cartilage endplate chondrocytes. *Zhonghua Yi Xue Za Zhi*. 2011;*91*:1031- 1035.
204. Jiang C, Guo Q, Jin Y, et al. Inhibition of EZH2 ameliorates cartilage endplate degeneration and attenuates the progression of interver- tebral disc degeneration via demethylation of Sox-9. *EBioMedicine*. 2019;*48*:619-629.
205. Bach FC, de Rooij KM, Riemers FM, et al. Hedgehog proteins and parathyroid hormone-related protein are involved in intervertebral disc maturation, degeneration, and calcification. *JOR Spine*. 2019;*2*: e1071.
206. Grunhagen T, Shirazi-Adl A, Fairbank JC, Urban JP. Intervertebral disk nutrition: a review of factors influencing concentrations of nutrients and metabolites. *Orthop Clin North Am*. 2011;*42*:465-477.
207. Dolor A, Sampson SL, Lazar AA, Lotz JC, Szoka FC, Fields AJ. Matrix modification for enhancing the transport properties of the human cartilage endplate to improve disc nutrition. *PLoS One*. 2019;*14*: e0215218.
208. Malandrino A, Noailly J, Lacroix D. The effect of sustained compres- sion on oxygen metabolic transport in the intervertebral disc decreases with degenerative changes. *PLoS Comput Biol*. 2011;*7*: e1002112.
209. Boyd LM, Carter AJ. Injectable biomaterials and vertebral endplate treatment for repair and regeneration of the intervertebral disc. *Eur Spine J*. 2006;*15*(Suppl 3):S414-S421.
210. Masaryk TJ, Boumphrey F, Modic MT, Tamborrello C, Ross JS, Brown MD. Effects of chemonucleolysis demonstrated by MR imag- ing. *J Comput Assist Tomogr*. 1986;*10*:917-923.
211. Gullbrand SE, Ashinsky BG, Bonnevie ED, et al. Long-term mechanical function and integration of an implanted tissue- engineered intervertebral disc. *Sci Transl Med*. 2018;*468*:10.
212. Chong JE, Santerre JP, Kandel RA. Generation of an in vitro model of the outer annulus fibrosus-cartilage interface. *JOR Spine*. 2020;*3*: e1089.
213. Arana CJ, Diamandis EP, Kandel RA. Cartilage tissue enhances pro- teoglycan retention by nucleus pulposus cells in vitro. *Arthritis Rheum*. 2010;*62*:3395-3403.
214. De Pieri A, Byerley AM, Musumeci CR, Salemizadehparizi F, Vanderhorst MA, Wuertz-Kozak K. Electrospinning and 3D bioprinting for intervertebral disc tissue engineering. *JOR Spine*. 2020;*3*:e1117.
215. Vernengo AJ, Grad S, Eglin D, Alini M, Li Z. Bioprinting tissue ana- logues with decellularized extracellular matrix bioink for regenera- tion and tissue models of cartilage and intervertebral discs. *Adv Funct Mater*. 2020;*30*:1909044.
216. Williams RJ, Tryfonidou MA, Snuggs JW, Le Maitre CL. Cell sources proposed for nucleus pulposus regeneration. *JOR Spine*. 2021;*4*: e1175.
217. Martin JT, Gullbrand SE, Kim DH, et al. In vitro maturation and in vivo integration and function of an engineered cell-seeded disc- like angle ply structure (DAPS) for total disc arthroplasty. *Sci Rep*. 2017;*7*:15765.
218. Luo L, Jian X, Sun H, et al. Cartilage endplate stem cells inhibit inter- vertebral disc degeneration by releasing exosomes to nucleus pulpo- sus cells to activate Akt/autophagy. *Stem Cells*. 2021;*39*:467-481.
219. Yao Y, Deng Q, Song W, et al. MIF plays a key role in regulating tissue- specific chondro-osteogenic differentiation fate of human cartilage endplate stem cells under hypoxia. *Stem Cell Rep*. 2016;*7*: 249-262.
220. Li Y, Chen L, Gao Y, Zou X, Wei F. Oxidative stress and interverteb- ral disc degeneration: pathophysiology, signaling pathway, and ther- apy. *Oxid Med Cell Longev*. 2022;*2022*:1984742.
221. Urban JP, Roberts S. Degeneration of the intervertebral disc. *Arthri- tis Res Ther*. 2003;*5*:120-130.
222. Fields AJ, Johansen BB, Metz LN, et al. Alterations in intervertebral disc composition, matrix homeostasis and biomechanical behavior in the UCD-T2DM rat model of type 2 diabetes. *J Orthop Res*. 2015; *33*:738-746.
223. Ferguson SJ, Steffen T. Biomechanics of the aging spine. *Eur Spine J*. 2003;*12*(Suppl 2):S97-S103.
224. Li YH, Wu HL, Li Z, et al. Species variation in the cartilaginous end- plate of the lumbar intervertebral disc. *JOR Spine*. 2022;*5*:e1218.

References

[1] Z. Sun, B. Liu, and Z.-J. Luo, "The Immune Privilege of the Intervertebral Disc: Implications for Intervertebral Disc Degeneration Treatment," *Int J Med Sci*, vol. 17, no. 5, pp. 685–692, 2020, doi: 10.7150/ijms.42238.

[2] T. Kadow, G. Sowa, N. Vo, and J. D. Kang, "Molecular Basis of Intervertebral Disc Degeneration and Herniations: What Are the Important Translational Questions?," *Clin Orthop Relat Res*, vol. 473, no. 6, pp. 1903–1912, Jun. 2015, doi: 10.1007/s11999-014-3774-8.

[3] J. P. G. Urban and S. Roberts, "Degeneration of the intervertebral disc.," *Arthritis Res Ther*, vol. 5, no. 3, pp. 120–30, 2003, doi: 10.1186/ar629.

[4] N. Kos, L. Gradisnik, and T. Velnar, "A Brief Review of the Degenerative Intervertebral Disc Disease," Dec. 01, 2019, *NLM (Medline)*. doi: 10.5455/medarh.2019.73.421-424.

[5] V. K. Viswanathan, A. P. Shetty, and S. Rajasekaran, "Modic changes - An evidence-based, narrative review on its patho-physiology, clinical significance and role in chronic low back pain," *J Clin Orthop Trauma*, vol. 11, no. 5, pp. 761–769, Sep. 2020, doi: 10.1016/j.jcot.2020.06.025.

[6] S. Çevik and H. Yılmaz, "Evaluation of the Relationship Between Clinical Symptoms and Modic Changes," *Cureus*, Feb. 2020, doi: 10.7759/cureus.6970.

[7] R. Rahme and R. Moussa, "The Modic Vertebral Endplate and Marrow Changes: Pathologic Significance and Relation to Low Back Pain and Segmental Instability of the Lumbar Spine," *American Journal of Neuroradiology*, vol. 29, no. 5, pp. 838–842, May 2008, doi: 10.3174/ajnr.A0925.

[8] J. Xin, Y. Wang, Z. Zheng, S. Wang, S. Na, and S. Zhang, "Treatment of Intervertebral Disc Degeneration," Jul. 01, 2022, *Sociedade Brasileira de Matematica Aplicada e Computacional*. doi: 10.1111/os.13254.

[9] S. Kirnaz *et al.*, "Innovative Biological Treatment Methods for Degenerative Disc Disease," *World Neurosurg*, vol. 157, pp. 282–299, Jan. 2022, doi: 10.1016/j.wneu.2021.09.068.

[10] B. Jordan, "Genes and Non-Mendelian Diseases: Dealing with Complexity," *Perspect Biol Med*, vol. 57, no. 1, pp. 118–131, 2014, doi: 10.1353/pbm.2014.0002.

[11] P. Kreitmaier, G. Katsoula, and E. Zeggini, "Insights from multi-omics integration in complex disease primary tissues," *Trends in Genetics*, vol. 39, no. 1, pp. 46–58, Jan. 2023, doi: 10.1016/j.tig.2022.08.005.

[12] J. Piñero *et al.*, "The DisGeNET knowledge platform for disease genomics: 2019 update," *Nucleic Acids Res*, vol. 48, no. D1, pp. D845–D855, May 2019, doi: 10.1093/nar/gkz1021.

[13] A. Moayyeri, C. J. Hammond, D. J. Hart, and T. D. Spector, "The UK Adult Twin Registry (TwinsUK Resource)," *Twin Research and Human Genetics*, vol. 16, no. 1, pp. 144–149, Feb. 2013, doi: 10.1017/thg.2012.89.

[14] J. M. Walker, "M e t h o d s i n M o l e c u l a r B i o l o g y ™ Series Editor." [Online]. Available: www.springer.com/series/7651

[15] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," May 05, 2017, *BioMed Central Ltd.* doi: 10.1186/s13059-017-1215-1.

[16] "Finishing the euchromatic sequence of the human genome," *Nature*, vol. 431, no. 7011, pp. 931–945, Oct. 2004, doi: 10.1038/nature03001.

[17] R. J. Klein *et al.*, "Complement factor H polymorphism in age-related macular degeneration," *Science (1979)*, vol. 308, no. 5720, pp. 385–389, Apr. 2005, doi: 10.1126/science.1109557.

[18] T. A. Manolio *et al.*, "Finding the missing heritability of complex diseases," Oct. 08, 2009. doi: 10.1038/nature08494.

[19] V. Tam, N. Patel, M. Turcotte, Y. Bossé, G. Paré, and D. Meyre, "Benefits and limitations of genome-wide association studies," Aug. 01, 2019, *Nature Publishing Group*. doi: 10.1038/s41576-019-0127-1.

[20] K. J. Mitchell, "What is complex about complex disorders?," *Genome Biol*, vol. 13, no. 1, p. 237, 2012, doi: 10.1186/gb-2012-13-1-237.

[21] F. W. Sabb *et al.*, “Challenges in phenotype definition in the whole-genome era: Multivariate models of memory and intelligence,” Nov. 24, 2009. doi: 10.1016/j.neuroscience.2009.05.013.
[22] S. Liu *et al.*, “Illuminating links between cis-regulators and trans-acting variants in the human prefrontal cortex,” *Genome Med*, vol. 14, no. 1, Dec. 2022, doi: 10.1186/s13073-022-01133-8.
[23] Z. Zhu *et al.*, “Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets,” *Nat Genet*, vol. 48, no. 5, pp. 481–487, May 2016, doi: 10.1038/ng.3538.
[24] M. D. Gallagher and A. S. Chen-Plotkin, “The Post-GWAS Era: From Association to Function,” May 03, 2018, *Cell Press*. doi: 10.1016/j.ajhg.2018.04.002.
[25] E. Uffelmann *et al.*, “Genome-wide association studies,” Dec. 01, 2021, *Springer Nature*. doi: 10.1038/s43586-021-00056-9.
[26] C. N. Chow, C. W. Yang, and W. C. Chang, “Databases and prospects of dynamic gene regulation in eukaryotes: A mini review,” Jan. 01, 2023, *Elsevier B.V.* doi: 10.1016/j.csbj.2023.03.032.
[27] F. J. Martin *et al.*, “Ensembl 2023,” *Nucleic Acids Res*, vol. 51, no. 1 D, pp. D933–D941, Jan. 2023, doi: 10.1093/nar/gkac958.
[28] N. de Souza, “Genomics: The ENCODE project,” 2012. doi: 10.1038/nmeth.2238.
[29] Roadmap Epigenomics Consortium *et al.*, “Integrative analysis of 111 reference human epigenomes,” *Nature*, vol. 518, no. 7539, pp. 317–329, Feb. 2015, doi: 10.1038/nature14248.
[30] “The GTEx Consortium atlas of genetic regulatory effects across human tissues,” *Science (1979)*, vol. 369, no. 6509, pp. 1318–1330, Sep. 2020, doi: 10.1126/science.aaz1776.
[31] R. Andersson *et al.*, “An atlas of active enhancers across human cell types and tissues,” *Nature*, vol. 507, no. 7493, pp. 455–461, 2014, doi: 10.1038/nature12787.
[32] A. P. Boyle *et al.*, “Annotation of functional variation in personal genomes using RegulomeDB,” *Genome Res*, vol. 22, no. 9, pp. 1790–1797, Sep. 2012, doi: 10.1101/gr.137323.112.
[33] E. Wingender *et al.*, “The TRANSFAC system on gene expression regulation,” 2001. [Online]. Available: www.gene-regulation.de/
[34] A. Sandelin, W. Alkema, P. Engström, W. W. Wasserman, and B. Lenhard, “JASPAR: An open-access database for eukaryotic transcription factor binding profiles,” *Nucleic Acids Res*, vol. 32, no. DATABASE ISS., Jan. 2004, doi: 10.1093/nar/gkh012.
[35] X. Yun, L. Xia, B. Tang, H. Zhang, F. Li, and Z. Zhang, “3CDB: A manually curated database of chromosome conformation capture data,” *Database*, vol. 2016, 2016, doi: 10.1093/database/baw044.
[36] E. Cano-Gamez and G. Trynka, “From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases,” May 13, 2020, *Frontiers Media S.A.* doi: 10.3389/fgene.2020.00424.
[37] S. L. Spain and J. C. Barrett, “Strategies for fine-mapping complex traits,” Oct. 15, 2015, *Oxford University Press*. doi: 10.1093/hmg/ddv260.
[38] C. Kanduri, C. Bock, S. Gundersen, E. Hovig, and G. K. Sandve, “Colocalization analyses of genomic elements: Approaches, recommendations and challenges,” *Bioinformatics*, vol. 35, no. 9, pp. 1615–1624, May 2019, doi: 10.1093/bioinformatics/bty835.
[39] F. Hormozdiari *et al.*, “Colocalization of GWAS and eQTL Signals Detects Target Genes,” *Am J Hum Genet*, vol. 99, no. 6, pp. 1245–1260, Dec. 2016, doi: 10.1016/j.ajhg.2016.10.003.
[40] D. J. McCarthy *et al.*, “Choice of transcripts and software has a large effect on variant annotation,” *Genome Med*, vol. 6, no. 3, p. 26, 2014, doi: 10.1186/gm543.
[41] T. Nalpathamkalam, A. Derkach, A. D. Paterson, and D. Merico, “Genetic Analysis Workshop 18 single-nucleotide variant prioritization based on protein impact, sequence conservation, and gene annotation,” *BMC Proc*, vol. 8, no. S1, p. S11, Jun. 2014, doi: 10.1186/1753-6561-8-S1-S11.
[42] W. McLaren *et al.*, “The Ensembl Variant Effect Predictor,” *Genome Biol*, vol. 17, no. 1, Jun. 2016, doi: 10.1186/s13059-016-0974-4.
[43] M. Kircher, D. M. Witten, P. Jain, B. J. O’roak, G. M. Cooper, and J. Shendure, “A general framework for estimating the relative pathogenicity of human genetic variants,” *Nat Genet*, vol. 46, no. 3, pp. 310–315, 2014, doi: 10.1038/ng.2892.
[44] P. C. Ng, “SIFT: predicting amino acid changes that affect protein function,” *Nucleic Acids Res*, vol. 31, no. 13, pp. 3812–3814, Jul. 2003, doi: 10.1093/nar/gkg509.

[45] I. Adzhubei, D. M. Jordan, and S. R. Sunyaev, “Predicting functional effect of human missense mutations using PolyPhen-2,” *Curr Protoc Hum Genet*, no. SUPPL.76, 2013, doi: 10.1002/0471142905.hg0720s76.
[46] A. S. Dunham, P. Beltrao, and M. AlQuraishi, “High-throughput deep learning variant effect prediction with Sequence UNET,” *Genome Biol*, vol. 24, no. 1, p. 110, May 2023, doi: 10.1186/s13059-023-02948-3.
[47] K. Watanabe, E. Taskesen, A. Van Bochoven, and D. Posthuma, “Functional mapping and annotation of genetic associations with FUMA,” *Nat Commun*, vol. 8, no. 1, Dec. 2017, doi: 10.1038/s41467-017-01261-5.
[48] C. A. de Leeuw, J. M. Mooij, T. Heskes, and D. Posthuma, “MAGMA: Generalized Gene-Set Analysis of GWAS Data,” *PLoS Comput Biol*, vol. 11, no. 4, Apr. 2015, doi: 10.1371/journal.pcbi.1004219.
[49] J. Yang, S. H. Lee, M. E. Goddard, and P. M. Visscher, “GCTA: A tool for genome-wide complex trait analysis,” *Am J Hum Genet*, vol. 88, no. 1, pp. 76–82, Jan. 2011, doi: 10.1016/j.ajhg.2010.11.011.
[50] C. Benner, C. C. A. Spencer, A. S. Havulinna, V. Salomaa, S. Ripatti, and M. Pirinen, “FINEMAP: Efficient variable selection using summary data from genome-wide association studies,” *Bioinformatics*, vol. 32, no. 10, pp. 1493–1501, May 2016, doi: 10.1093/bioinformatics/btw018.
[51] C. Wallace, “Statistical testing of shared genetic control for potentially related traits,” *Genet Epidemiol*, vol. 37, no. 8, pp. 802–813, Dec. 2013, doi: 10.1002/gepi.21765.
[52] M. C. Casotti *et al.*, “Translational Bioinformatics Applied to the Study of Complex Diseases,” Feb. 01, 2023, *MDPI*. doi: 10.3390/genes14020419.
[53] S. Leonelli, “The challenges of big data biology,” *Elife*, vol. 8, Apr. 2019, doi: 10.7554/eLife.47381.
[54] M. C. Battié and T. Videman, “Lumbar Disc Degeneration: Epidemiology and Genetics,” *Journal of Bone and Joint Surgery*, vol. 88, no. suppl_2, pp. 3–9, Apr. 2006, doi: 10.2106/JBJS.E.01313.
[55] N. M. Khan, M. E. Diaz-Hernandez, S. M. Presciutti, and H. Drissi, “Network Analysis Identifies Gene Regulatory Network Indicating the Role of RUNX1 in Human Intervertebral Disc Degeneration,” *Genes (Basel)*, vol. 11, no. 7, p. 771, Jul. 2020, doi: 10.3390/genes11070771.
[56] Q. Yan *et al.*, “Bioinformatics-Based Research on Key Genes and Pathways of Intervertebral Disc Degeneration,” *Cartilage*, vol. 13, no. 2_suppl, pp. 582S-591S, Dec. 2021, doi: 10.1177/1947603520973247.
[57] R. Albert, “Scale-free networks in cell biology,” Nov. 01, 2005. doi: 10.1242/jcs.02714.
[58] S. Peri *et al.*, “Human protein reference database as a discovery resource for proteomics,” *Nucleic Acids Res*, vol. 32, no. DATABASE ISS., Jan. 2004, doi: 10.1093/nar/gkh070.
[59] J. Garcia-Garcia, E. Guney, R. Aragues, J. Planas-Iglesias, and B. Oliva, “Biana: a software framework for compiling biological interactions and analyzing networks,” 2010. [Online]. Available: http://www.biomedcentral.com/1471-2105/11/56
[60] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Szló Barabá, “The human disease network,” 2007. [Online]. Available: www.pnas.org/cgi/content/full/
[61] A. Hamosh, A. F. Scott, J. Amberger, D. Valle, and V. A. McKusick, “Online Mendelian Inheritance in Man (OMIM),” *Hum Mutat*, vol. 15, no. 1, pp. 57–61, 2000, doi: 10.1002/(SICI)1098-1004(200001)15:1<57::AID-HUMU12>3.0.CO;2-G.
[62] S. Choobdar *et al.*, “Assessment of network module identification across complex diseases,” *Nat Methods*, vol. 16, no. 9, pp. 843–852, Sep. 2019, doi: 10.1038/s41592-019-0509-5.
[63] F. Conte and P. Paci, “Alzheimer’s disease: insights from a network medicine perspective,” *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-20404-3.
[64] E. Guney, J. Garcia-Garcia, and B. Oliva, “GUILDify: A web server for phenotypic characterization of genes through biological data integration and network-based prioritization algorithms,” *Bioinformatics*, vol. 30, no. 12, pp. 1789–1790, Jun. 2014, doi: 10.1093/bioinformatics/btu092.
[65] W. Chang *et al.*, “Identification of novel hub genes associated with liver metastasis of gastric cancer,” *Int J Cancer*, vol. 125, no. 12, pp. 2844–2853, Dec. 2009, doi: 10.1002/ijc.24699.
[66] J. Menche *et al.*, “Uncovering disease-disease relationships through the incomplete interactome,” *Science (1979)*, vol. 347, no. 6224, p. 841, Feb. 2015, doi: 10.1126/science.1257601.

[67] R. Mirnezami, J. Nicholson, and A. Darzi, “Preparing for Precision Medicine,” *New England Journal of Medicine*, vol. 366, no. 6, pp. 489–491, Feb. 2012, doi: 10.1056/NEJMp1114866.
[68] Q. Yao *et al.*, “Global Prioritization of Disease Candidate Metabolites Based on a Multi-omics Composite Network,” *Sci Rep*, vol. 5, Nov. 2015, doi: 10.1038/srep17201.
[69] K. Athanasopoulou, G. N. Daneva, P. G. Adamopoulos, and A. Scorilas, “Artificial Intelligence: The Milestone in Modern Biomedical Research,” Dec. 01, 2022, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/biomedinformatics2040049.
[70] R. S. Vilhekar and A. Rawekar, “Artificial Intelligence in Genetics,” *Cureus*, Jan. 2024, doi: 10.7759/cureus.52035.
[71] F. F. Peleke, S. M. Zumkeller, M. Gültas, A. Schmitt, and J. Szymański, “Deep learning the cis-regulatory code for gene expression in selected model plants,” *Nat Commun*, vol. 15, no. 1, p. 3488, Apr. 2024, doi: 10.1038/s41467-024-47744-0.
[72] M. Kang, S. Kim, D.-B. Lee, C. Hong, and K.-B. Hwang, “Gene-specific machine learning for pathogenicity prediction of rare BRCA1 and BRCA2 missense variants,” *Sci Rep*, vol. 13, no. 1, p. 10478, Jun. 2023, doi: 10.1038/s41598-023-37698-6.
[73] H. Y. Xiong *et al.*, “The human splicing code reveals new insights into the genetic determinants of disease,” *Science (1979)*, vol. 347, no. 6218, May 2015, doi: 10.1126/science.1254806.
[74] A. S. Edwards, B. Kaplan, and T. Jie, “A Primer on Machine Learning,” Apr. 01, 2021, *Lippincott Williams and Wilkins*. doi: 10.1097/TP.0000000000003316.
[75] T. Jiang, J. L. Gradus, and A. J. Rosellini, “Supervised Machine Learning: A Brief Primer,” *Behav Ther*, vol. 51, no. 5, pp. 675–687, Sep. 2020, doi: 10.1016/j.beth.2020.05.002.
[76] J. M. Bernardo *et al.*, “Generative or Discriminative? Getting the Best of Both Worlds.”
[77] G. D. Smith, W. H. Ching, P. Cornejo-Páramo, and E. S. Wong, “Decoding enhancer complexity with machine learning and high-throughput discovery,” Dec. 01, 2023, *BioMed Central Ltd*. doi: 10.1186/s13059-023-02955-4.
[78] M. R. Amin, A. Yurovsky, Y. Tian, and S. Skiena, “DeepAnnotator: Genome Annotation with Deep Learning,” in *ACM-BCB 2018 - Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Association for Computing Machinery, Inc, Aug. 2018, pp. 254–259. doi: 10.1145/3233547.3233577.
[79] A. Saxe, S. Nelli, and C. Summerfield, “If deep learning is the answer, what is the question?,” Jan. 01, 2021, *Nature Research*. doi: 10.1038/s41583-020-00395-8.
[80] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” May 27, 2015, *Nature Publishing Group*. doi: 10.1038/nature14539.
[81] H. Y. Xiong *et al.*, “The human splicing code reveals new insights into the genetic determinants of disease,” *Science (1979)*, vol. 347, no. 6218, Jan. 2015, doi: 10.1126/science.1254806.
[82] M. W. Libbrecht and W. S. Noble, “Machine learning applications in genetics and genomics,” May 18, 2015, *Nature Publishing Group*. doi: 10.1038/nrg3920.
[83] L. Peña-Castillo *et al.*, “A critical assessment of Mus musculus gene function prediction using integrated genomic evidence,” *Genome Biol*, vol. 9, no. Suppl 1, p. S2, 2008, doi: 10.1186/gb-2008-9-s1-s2.
[84] O. G. Troyanskaya, K. Dolinski, A. B. Owen, R. B. Altman, and D. Botstein, “A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*),” *Proceedings of the National Academy of Sciences*, vol. 100, no. 14, pp. 8348–8353, Jul. 2003, doi: 10.1073/pnas.0832373100.
[85] D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins, “Next-Generation Machine Learning for Biological Networks,” Jun. 14, 2018, *Cell Press*. doi: 10.1016/j.cell.2018.05.015.
[86] A. Grover and J. Leskovec, “node2vec,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 855–864. doi: 10.1145/2939672.2939754.
[87] B. Perozzi, R. Al-Rfou, and S. Skiena, “DeepWalk: Online learning of social representations,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 2014, pp. 701–710. doi: 10.1145/2623330.2623732.

[88] Y. Dong, N. V. Chawla, and A. Swami, "Metapath2vec: Scalable representation learning for heterogeneous networks," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2017, pp. 135–144. doi: 10.1145/3097983.3098036.
[89] T. Trouillon, J. Welbl, S. Riedel, Uké. Gaussier, and G. Bouchard, "Complex Embeddings for Simple Link Prediction," 2016. [Online]. Available: https://github.com/ttrouill/complex
[90] B. Yang, W. Yih, X. He, J. Gao, and L. Deng, "Embedding Entities and Relations for Learning and Inference in Knowledge Bases," Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6575
[91] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," Sep. 2016, [Online]. Available: http://arxiv.org/abs/1609.02907
[92] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, "Modeling Relational Data with Graph Convolutional Networks," Mar. 2017, [Online]. Available: http://arxiv.org/abs/1703.06103
[93] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, and O. Yakhnenko, "Translating Embeddings for Modeling Multi-relational Data."
[94] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, "RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space," Feb. 2019, doi: https://doi.org/10.48550/arXiv.1902.10197.
[95] X. M. Zhang, L. Liang, L. Liu, and M. J. Tang, "Graph Neural Networks and Their Current Applications in Bioinformatics," *Front Genet*, vol. 12, Jul. 2021, doi: 10.3389/fgene.2021.690049.
[96] K. Jha, S. Saha, and H. Singh, "Prediction of protein–protein interaction using graph neural networks," *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-12201-9.
[97] Z. Gao, Y. Pan, P. Ding, and R. Xu, "A knowledge graph-based disease-gene prediction system using multi-relational graph convolution networks.," *AMIA Annu Symp Proc*, vol. 2022, pp. 468–476, 2022, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/37128437
[98] K. Opap and N. Mulder, "Recent advances in predicting gene–disease associations," *F1000Res*, vol. 6, p. 578, Apr. 2017, doi: 10.12688/f1000research.10788.1.
[99] A. R. Sonawane, S. T. Weiss, K. Glass, and A. Sharma, "Network Medicine in the Age of Biomedical Big Data," *Front Genet*, vol. 10, no. APR, Apr. 2019, doi: 10.3389/fgene.2019.00294.
[100] A. Hur, N. Janjua, and M. Ahmed, "A Survey on State-of-the-art Techniques for Knowledge Graphs Construction and Challenges ahead," in *2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, IEEE, Dec. 2021, pp. 99–103. doi: 10.1109/AIKE52691.2021.00021.
[101] C. Peng, F. Xia, M. Naseriparsa, and F. Osborne, "Knowledge Graphs: Opportunities and Challenges," *Artif Intell Rev*, vol. 56, no. 11, pp. 13071–13102, Nov. 2023, doi: 10.1007/s10462-023-10465-9.
[102] S. K. Mohamed, A. Nounu, and V. Nováček, "Biological applications of knowledge graph embedding models," *Brief Bioinform*, vol. 22, no. 2, pp. 1679–1693, Mar. 2021, doi: 10.1093/bib/bbaa012.
[103] Y. Dai, S. Wang, N. N. Xiong, and W. Guo, "A Survey on Knowledge Graph Embedding: Approaches, Applications and Benchmarks," *Electronics (Basel)*, vol. 9, no. 5, p. 750, May 2020, doi: 10.3390/electronics9050750.
[104] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge Graph Embedding: A Survey of Approaches and Applications," *IEEE Trans Knowl Data Eng*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017, doi: 10.1109/TKDE.2017.2754499.
[105] G. Wang, Y. Ding, Z. Xie, Y. Ma, Z. Zhou, and W. Qian, "RotatGAT: Learning Knowledge Graph Embedding with Translation Assumptions and Graph Attention Networks," in *2022 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2022, pp. 1–7. doi: 10.1109/IJCNN55064.2022.9892206.
[106] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, pp. 107–117, Apr. 1998, doi: 10.1016/S0169-7552(98)00110-X.
[107] L. Wei, "Combining Graph Attention Mechanism and PageRank to Learn Graph-level Representations," 2022.

[108] S. Nunes, R. T. Sousa, and C. Pesquita, “Predicting Gene-Disease Associations with Knowledge Graph Embeddings over Multiple Ontologies,” May 2021, doi: https://doi.org/10.48550/arXiv.2105.04944.
[109] W. Choi and H. Lee, “Identifying disease-gene associations using a convolutional neural network-based model by embedding a biological knowledge graph with entity descriptions,” *PLoS One*, vol. 16, no. 10, p. e0258626, Oct. 2021, doi: 10.1371/journal.pone.0258626.
[110] S. K. Mohamed, V. Nováček, and A. Nounu, “Discovering protein drug targets using knowledge graph embeddings,” *Bioinformatics*, vol. 36, no. 2, pp. 603–610, Jan. 2020, doi: 10.1093/bioinformatics/btz600.
[111] X. Zhong and J. C. Rajapakse, “Graph embeddings on gene ontology annotations for protein–protein interaction prediction,” *BMC Bioinformatics*, vol. 21, no. S16, p. 560, Dec. 2020, doi: 10.1186/s12859-020-03816-8.
[112] C. Ruiz, M. Zitnik, and J. Leskovec, “Identification of disease treatment mechanisms through the multiscale interactome,” *Nat Commun*, vol. 12, no. 1, p. 1796, Mar. 2021, doi: 10.1038/s41467-021-21770-8.
[113] A. Fernández-Torras, M. Duran-Frigola, M. Bertoni, M. Locatelli, and P. Aloy, “Integrating and formatting biomedical data as pre-calculated knowledge graph embeddings in the Bioteque,” *Nat Commun*, vol. 13, no. 1, p. 5304, Sep. 2022, doi: 10.1038/s41467-022-33026-0.
[114] R. Oughtred *et al.*, “The BioGRID interaction database: 2019 update,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D529–D541, Jan. 2019, doi: 10.1093/nar/gky1079.
[115] L. Salwinski, “The Database of Interacting Proteins: 2004 update,” *Nucleic Acids Res*, vol. 32, no. 90001, pp. 449D – 451, Jan. 2004, doi: 10.1093/nar/gkh086.
[116] K. Luck *et al.*, “A reference map of the human binary protein interactome,” *Nature*, vol. 580, no. 7803, pp. 402–408, Apr. 2020, doi: 10.1038/s41586-020-2188-x.
[117] E. Wingender, “The TRANSFAC system on gene expression regulation,” *Nucleic Acids Res*, vol. 29, no. 1, pp. 281–283, Jan. 2001, doi: 10.1093/nar/29.1.281.
[118] H. Hermjakob *et al.*, “IntAct: An open source molecular interaction database,” *Nucleic Acids Res*, vol. 32, no. DATABASE ISS., Jan. 2004, doi: 10.1093/nar/gkh052.
[119] R. Goel, H. C. Harsha, A. Pandey, and T. S. K. Prasad, “Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis,” *Mol. BioSyst.*, vol. 8, no. 2, pp. 453–463, 2012, doi: 10.1039/C1MB05340J.
[120] M. Kanehisa, “KEGG: Kyoto Encyclopedia of Genes and Genomes,” *Nucleic Acids Res*, vol. 28, no. 1, pp. 27–30, Jan. 2000, doi: 10.1093/nar/28.1.27.
[121] J. Schellenberger, J. O. Park, T. M. Conrad, and B. Ø. Palsson, “BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions,” *BMC Bioinformatics*, vol. 11, no. 1, p. 213, Dec. 2010, doi: 10.1186/1471-2105-11-213.
[122] A. Ruepp *et al.*, “CORUM: the comprehensive resource of mammalian protein complexes,” *Nucleic Acids Res*, vol. 36, no. Database, pp. D646–D650, Dec. 2007, doi: 10.1093/nar/gkm936.
[123] P. V. Hornbeck *et al.*, “PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse,” *Nucleic Acids Res*, vol. 40, no. D1, pp. D261–D270, Jan. 2012, doi: 10.1093/nar/gkr1122.
[124] A. Vinayagam *et al.*, “A Directed Protein Interaction Network for Investigating Intracellular Signal Transduction,” *Sci Signal*, vol. 4, no. 189, Sep. 2011, doi: 10.1126/scisignal.2001699.
[125] J. Hastings, “Primer on Ontologies,” in *Methods in Molecular Biology*, vol. 1446, Humana Press Inc., 2017, pp. 3–13. doi: 10.1007/978-1-4939-3743-1_1.
[126] M. Ashburner *et al.*, “Gene Ontology: tool for the unification of biology,” *Nat Genet*, vol. 25, no. 1, pp. 25–29, May 2000, doi: 10.1038/75556.
[127] L. M. Schriml *et al.*, “Human Disease Ontology 2018 update: classification, content and workflow expansion,” *Nucleic Acids Res*, vol. 47, no. D1, pp. D955–D962, Jan. 2019, doi: 10.1093/nar/gky1032.
[128] T. Groza *et al.*, “The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease,” *The American Journal of Human Genetics*, vol. 97, no. 1, pp. 111–124, Jul. 2015, doi: 10.1016/j.ajhg.2015.05.020.

[129] B. T. McInnes, T. Pedersen, and J. Carlis, “Using UMLS Concept Unique Identifiers (CUIs) for word sense disambiguation in the biomedical domain.,” *AMIA Annu Symp Proc*, vol. 2007, pp. 533–7, Oct. 2007, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/18693893
[130] J. Piñero *et al.*, “The DisGeNET knowledge platform for disease genomics: 2019 update,” *Nucleic Acids Res*, vol. 48, no. D1, pp. D845–D855, Nov. 2019, doi: 10.1093/nar/gkz1021.
[131] A. S. Brown and C. J. Patel, “A standard database for drug repositioning,” *Sci Data*, vol. 4, no. 1, p. 170029, Mar. 2017, doi: 10.1038/sdata.2017.29.
[132] O. Ursu *et al.*, “DrugCentral: Online drug compendium,” *Nucleic Acids Res*, vol. 45, no. D1, pp. D932–D939, Jan. 2017, doi: 10.1093/nar/gkw993.
[133] S. M. Corsello *et al.*, “The Drug Repurposing Hub: a next-generation drug library and information resource,” *Nat Med*, vol. 23, no. 4, pp. 405–408, Apr. 2017, doi: 10.1038/nm.4306.
[134] M. E. Sharp, “Toward a comprehensive drug ontology: extraction of drug-indication relations from diverse information sources,” *J Biomed Semantics*, vol. 8, no. 1, p. 2, Dec. 2017, doi: 10.1186/s13326-016-0110-0.
[135] D. S. Wishart, “DrugBank: a comprehensive resource for in silico drug discovery and exploration,” *Nucleic Acids Res*, vol. 34, no. 90001, pp. D668–D672, Jan. 2006, doi: 10.1093/nar/gkj067.
[136] A. Bordes, N. Usunier, A. Garcia-Durán, J. Weston, and O. Yakhnenko, “Translating Embeddings for Modeling Multi-relational Data.”
[137] P. J. Thul and C. Lindskog, “The human protein atlas: A spatial map of the human proteome,” *Protein Science*, vol. 27, no. 1, pp. 233–244, Jan. 2018, doi: 10.1002/pro.3307.
[138] W. H. Organization, “International classification of diseases : [9th] ninth revision, basic tabulation list with alphabetic index,” 1978, *World Health Organization*.
[139] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” Jan. 2012, [Online]. Available: http://arxiv.org/abs/1201.0490
[140] A. Paszke *et al.*, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” Dec. 2019, [Online]. Available: http://arxiv.org/abs/1912.01703
[141] M. Larralde *et al.*, “althonos/pronto: v2.5.4,” Apr. 2023, *Zenodo*. doi: 10.5281/zenodo.7814219.
[142] U. Raudvere *et al.*, “g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update),” *Nucleic Acids Res*, vol. 47, no. W1, pp. W191–W198, Jul. 2019, doi: 10.1093/nar/gkz369.
[143] R. Albert, “Scale-free networks in cell biology,” Nov. 01, 2005. doi: 10.1242/jcs.02714.
[144] P. Bermudez-Lekerika *et al.*, “Immuno-Modulatory Effects of Intervertebral Disc Cells,” *Front Cell Dev Biol*, vol. 10, Jun. 2022, doi: 10.3389/fcell.2022.924692.
[145] J. A. Buckwalter, “Aging and Degeneration of the Human Intervertebral Disc,” *Spine (Phila Pa 1976)*, vol. 20, no. 11, pp. 1307–1314, Jun. 1995, doi: 10.1097/00007632-199506000-00022.
[146] S. Chen, S. Liu, K. Ma, L. Zhao, H. Lin, and Z. Shao, “TGF-β signaling in intervertebral disc health and disease,” Aug. 01, 2019, *W.B. Saunders Ltd*. doi: 10.1016/j.joca.2019.05.005.
[147] C.-G. Li *et al.*, “A Continuous Observation of the Degenerative Process in the Intervertebral Disc of Smad3 Gene Knock-Out Mice,” *Spine (Phila Pa 1976)*, vol. 34, no. 13, pp. 1363–1369, Jun. 2009, doi: 10.1097/BRS.0b013e3181a3c7c7.
[148] N. V. Vo, R. A. Hartman, T. Yurube, L. J. Jacobs, G. A. Sowa, and J. D. Kang, “Expression and regulation of metalloproteinases and their inhibitors in intervertebral disc aging and degeneration,” *The Spine Journal*, vol. 13, no. 3, pp. 331–341, Mar. 2013, doi: 10.1016/j.spinee.2012.02.027.
[149] W. Lin, L. Xu, and G. Li, “Molecular Insights Into Lysyl Oxidases in Cartilage Regeneration and Rejuvenation,” *Front Bioeng Biotechnol*, vol. 8, Apr. 2020, doi: 10.3389/fbioe.2020.00359.
[150] R. Zhao *et al.*, “Lysyl oxidase inhibits TNF-α induced rat nucleus pulposus cell apoptosis via regulating Fas/FasL pathway and the p53 pathways,” *Life Sci*, vol. 260, p. 118483, Nov. 2020, doi: 10.1016/j.lfs.2020.118483.
[151] Fatima Zohra Smaili, “OPA2Vec: combining formal and informal contentof biomedical ontologies to improvesimilarity-based prediction,” *Bioinformatics*, vol. 33, no. 16, pp. 1–7, Aug. 2017, doi: 10.1093/bioinformatics/xxxxxx.

[152] M. Kulmanov, F. Z. Smaili, X. Gao, and R. Hoehndorf, “Semantic similarity and machine learning with ontologies,” *Brief Bioinform*, vol. 22, no. 4, Jul. 2021, doi: 10.1093/bib/bbaa199.
[153] I. Osman, S. Ben Yahia, and G. Diallo, “Ontology Integration: Approaches and Challenging Issues,” *Information Fusion*, vol. 71, pp. 38–63, Jul. 2021, doi: 10.1016/j.inffus.2021.01.007.
[154] P. M. Visscher, L. Yengo, N. J. Cox, and N. R. Wray, “Discovery and implications of polygenicity of common diseases,” *Science (1979)*, vol. 373, no. 6562, pp. 1468–1473, Sep. 2021, doi: 10.1126/science.abi8206.
[155] W. S. Bush and J. H. Moore, “Chapter 11: Genome-Wide Association Studies,” *PLoS Comput Biol*, vol. 8, no. 12, Dec. 2012, doi: 10.1371/journal.pcbi.1002822.
[156] S. L. Edwards, J. Beesley, J. D. French, and M. Dunning, “Beyond GWASs: Illuminating the dark road from association to function,” Nov. 07, 2013, *Cell Press*. doi: 10.1016/j.ajhg.2013.10.012.
[157] R. Edgar, M. Domrachev, and A. E. Lash, “Gene Expression Omnibus: NCBI gene expression and hybridization array data repository,” 2002. [Online]. Available: http://www.ninds.nih.gov/
[158] J. Lonsdale *et al.*, “The Genotype-Tissue Expression (GTEx) project,” *Nat Genet*, vol. 45, no. 6, pp. 580–585, Jun. 2013, doi: 10.1038/ng.2653.
[159] E. Sollis *et al.*, “The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource,” *Nucleic Acids Res*, vol. 51, no. D1, pp. D977–D985, Jan. 2023, doi: 10.1093/nar/gkac1010.
[160] B. Li and M. D. Ritchie, “From GWAS to Gene: Transcriptome-Wide Association Studies and Other Methods to Functionally Understand GWAS Discoveries,” Sep. 30, 2021, *Frontiers Media S.A.* doi: 10.3389/fgene.2021.713230.
[161] J. Pérez-Granado, J. Piñero, and L. I. Furlong, “Benchmarking post-GWAS analysis tools in major depression: Challenges and implications,” *Front Genet*, vol. 13, Oct. 2022, doi: 10.3389/fgene.2022.1006903.
[162] E. Mountjoy *et al.*, “An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci,” *Nat Genet*, vol. 53, no. 11, pp. 1527–1533, Nov. 2021, doi: 10.1038/s41588-021-00945-5.
[163] J. Malone *et al.*, “Modeling sample variables with an Experimental Factor Ontology,” *Bioinformatics*, vol. 26, no. 8, pp. 1112–1118, Mar. 2010, doi: 10.1093/bioinformatics/btq099.
[164] L. Prokunina and M. E. Alarcón-Riquelme, “Regulatory SNPs in complex diseases: Their identification and functional validation,” May 2004. doi: 10.1017/S1462399404007690.
[165] N. Kerimov *et al.*, “A compendium of uniformly processed human gene expression and splicing quantitative trait loci,” *Nat Genet*, vol. 53, no. 9, pp. 1290–1299, Sep. 2021, doi: 10.1038/s41588-021-00924-w.
[166] G. Alanis-Lobato, M. A. Andrade-Navarro, and M. H. Schaefer, “HIPPIE v2.0: Enhancing meaningfulness and reliability of protein-protein interaction networks,” *Nucleic Acids Res*, vol. 45, no. D1, pp. D408–D414, Jan. 2017, doi: 10.1093/nar/gkw985.
[167] M. J. Machiela and S. J. Chanock, “LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants,” *Bioinformatics*, vol. 31, no. 21, pp. 3555–3557, Dec. 2015, doi: 10.1093/bioinformatics/btv402.
[168] A. Auton *et al.*, “A global reference for human genetic variation,” Sep. 30, 2015, *Nature Publishing Group*. doi: 10.1038/nature15393.
[169] K. Watanabe, E. Taskesen, A. Van Bochoven, and D. Posthuma, “Functional mapping and annotation of genetic associations with FUMA,” *Nat Commun*, vol. 8, no. 1, May 2017, doi: 10.1038/s41467-017-01261-5.
[170] A. Franke *et al.*, “Genome-wide meta-analysis increases to 71 the number of confirmed Crohn’s disease susceptibility loci,” *Nat Genet*, vol. 42, no. 12, pp. 1118–1125, Dec. 2010, doi: 10.1038/ng.717.
[171] N.-P. Andreou, “Inflammatory bowel disease pathobiology: the role of the interferon signature,” *Ann Gastroenterol*, vol. 33, no. 2, pp. 125–133, 2020, doi: 10.20524/aog.2020.0457.
[172] S. Munir, M. Rade, J. H. Määttä, M. B. Freidin, and F. M. K. Williams, “Intervertebral Disc Biology: Genetic Basis of Disc Degeneration,” *Curr Mol Biol Rep*, vol. 4, no. 4, pp. 143–150, May 2018, doi: 10.1007/s40610-018-0101-2.

[173] K. G. Becker, “The common variants/multiple disease hypothesis of common complex genetic disorders,” *Med Hypotheses*, vol. 62, no. 2, pp. 309–317, 2004, doi: 10.1016/S0306-9877(03)00332-3.
[174] M. L. Freedman *et al.*, “Principles for the post-GWAS functional characterization of cancer risk loci,” May 2011. doi: 10.1038/ng.840.
[175] E. Mountjoy *et al.*, “An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci,” *Nat Genet*, vol. 53, no. 11, pp. 1527–1533, May 2021, doi: 10.1038/s41588-021-00945-5.
[176] K. Xu *et al.*, “Sp1 downregulates proinflammatory cytokine-induced catabolic gene expression in nucleus pulposus cells,” *Mol Med Rep*, vol. 14, no. 4, pp. 3961–3968, Oct. 2016, doi: 10.3892/mmr.2016.5730.
[177] H. Li, W. Li, B. Liang, J. Wei, D. Yin, and Q. Fan, “Role of AP-2α/TGF-β1/Smad3 axis in rats with intervertebral disc degeneration,” *Life Sci*, vol. 263, Dec. 2020, doi: 10.1016/j.lfs.2020.118567.
[178] J.-W. Chen, B. Li, Y.-H. Yang, S.-D. Jiang, and L.-S. Jiang, “Significance of Hypoxia in the Physiological Function of Intervertebral Disc Cells,” *Crit Rev Eukaryot Gene Expr*, vol. 24, no. 3, pp. 193–204, 2014, doi: 10.1615/CritRevEukaryotGeneExpr.2014010485.
[179] X. Meng *et al.*, “Hypoxia-inducible factor (HIF)-1alpha knockout accelerates intervertebral disc degeneration in mice,” 2018. [Online]. Available: www.ijcep.com/
[180] S. Gianola *et al.*, “Effectiveness of treatments for acute and subacute mechanical non-specific low back pain: a systematic review with network meta-analysis,” *Br J Sports Med*, vol. 56, no. 1, pp. 41–50, Jan. 2022, doi: 10.1136/bjsports-2020-103596.
[181] S. Chen *et al.*, “Global, regional and national burden of low back pain 1990–2019: A systematic analysis of the Global Burden of Disease study 2019,” *J Orthop Translat*, vol. 32, pp. 49–58, Jan. 2022, doi: 10.1016/j.jot.2021.07.005.
[182] I. Urits *et al.*, “Low Back Pain, a Comprehensive Review: Pathophysiology, Diagnosis, and Treatment,” *Curr Pain Headache Rep*, vol. 23, no. 3, p. 23, Mar. 2019, doi: 10.1007/s11916-019-0757-1.
[183] J. P. G. Urban and S. Roberts, “Degeneration of the intervertebral disc.,” *Arthritis Res Ther*, vol. 5, no. 3, pp. 120–30, 2003, doi: 10.1186/ar629.
[184] J. H. Määttä, A. MacGregor, J. Karppinen, and F. M. K. Williams, “The relationship between Modic changes and intervertebral disc degeneration,” *BMC Musculoskelet Disord*, vol. 17, no. 1, p. 371, Dec. 2016, doi: 10.1186/s12891-016-1198-1.
[185] K. Luoma, T. Vehmas, M. Grönblad, L. Kerttula, and E. Kääpä, “Relationship of Modic type 1 change with disc degeneration: a prospective MRI study,” *Skeletal Radiol*, vol. 38, no. 3, pp. 237–244, Mar. 2009, doi: 10.1007/s00256-008-0611-8.
[186] A. de Roos, H. Kressel, C. Spritzer, and M. Dalinka, “MR imaging of marrow changes adjacent to end plates in degenerative lumbar disk disease,” *American Journal of Roentgenology*, vol. 149, no. 3, pp. 531–534, Sep. 1987, doi: 10.2214/ajr.149.3.531.
[187] J. H. Määttä *et al.*, “Vertebral endplate change as a feature of intervertebral disc degeneration: a heritability study,” *European Spine Journal*, vol. 23, no. 9, pp. 1856–1862, Sep. 2014, doi: 10.1007/s00586-014-3333-8.
[188] S. Dudli, A. J. Fields, D. Samartzis, J. Karppinen, and J. C. Lotz, “Pathobiology of Modic changes,” *European Spine Journal*, vol. 25, no. 11, pp. 3723–3734, Nov. 2016, doi: 10.1007/s00586-016-4459-7.
[189] X.-L. Ma, J.-X. Ma, T. Wang, P. Tian, and C. Han, “Possible role of autoimmune reaction in Modic Type I changes,” *Med Hypotheses*, vol. 76, no. 5, pp. 692–694, May 2011, doi: 10.1016/j.mehy.2011.01.035.
[190] H. B. Albert, P. Kjaer, T. S. Jensen, J. S. Sorensen, T. Bendix, and C. Manniche, “Modic changes, possible causes and relation to low back pain,” *Med Hypotheses*, vol. 70, no. 2, pp. 361–368, Jan. 2008, doi: 10.1016/j.mehy.2007.05.014.
[191] Philippa Marrack and John Kappler, “Autoimmune disease: why and where it occurs”, doi: 10.1038/90935.
[192] S. Verdi *et al.*, “TwinsUK: The UK Adult Twin Registry Update,” *Twin Research and Human Genetics*, vol. 22, no. 6, pp. 523–529, Dec. 2019, doi: 10.1017/thg.2019.65.

[193] J. H. Määttä, M. Rade, M. B. Freidin, O. Airaksinen, J. Karppinen, and F. M. K. Williams, “Strong association between vertebral endplate defect and Modic change in the general population,” *Sci Rep*, vol. 8, no. 1, p. 16630, Dec. 2018, doi: 10.1038/s41598-018-34933-3.
[194] Y. Wang, T. Videman, and M. C. Battié, “Modic changes: prevalence, distribution patterns, and association with age in white men,” *The Spine Journal*, vol. 12, no. 5, pp. 411–416, May 2012, doi: 10.1016/j.spinee.2012.03.026.
[195] S. Munir, M. B. Freidin, M. Rade, J. Määttä, G. Livshits, and F. M. K. Williams, “Endplate Defect Is Heritable, Associated With Low Back Pain and Triggers Intervertebral Disc Degeneration,” *Spine (Phila Pa 1976)*, vol. 43, no. 21, pp. 1496–1501, Nov. 2018, doi: 10.1097/BRS.0000000000002721.
[196] P. N. Sambrook, A. J. MacGregor, and T. D. Spector, “Genetic influences on cervical and lumbar disc degeneration: A magnetic resonance imaging study in twins,” *Arthritis Rheum*, vol. 42, no. 2, pp. 366–372, Feb. 1999, doi: 10.1002/1529-0131(199902)42:2<366::AID-ANR20>3.0.CO;2-6.
[197] A. J. MacGregor, T. Andrew, P. N. Sambrook, and T. D. Spector, “Structural, psychological, and genetic influences on low back and neck pain: A study of adult female twins,” *Arthritis Care Res (Hoboken)*, vol. 51, no. 2, pp. 160–167, Apr. 2004, doi: 10.1002/art.20236.
[198] G. Molenberghs, D. Renard, and G. Verbeke, “A review of generalized linear mixed models.” [Online]. Available: http://www.numdam.org/
[199] C. B. Dean and J. D. Nielsen, “Generalized linear mixed models: a review and some extensions,” *Lifetime Data Anal*, vol. 13, no. 4, pp. 497–512, Dec. 2007, doi: 10.1007/s10985-007-9065-x.
[200] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, “G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences,” *Behav Res Methods*, vol. 39, no. 2, pp. 175–191, May 2007, doi: 10.3758/BF03193146.
[201] R. Song, B. Wang, Q. Yao, Q. Li, X. Jia, and J. Zhang, “The Impact of Obesity on Thyroid Autoimmunity and Dysfunction: A Systematic Review and Meta-Analysis,” *Front Immunol*, vol. 10, Oct. 2019, doi: 10.3389/fimmu.2019.02349.
[202] J. H. Määttä, S. Wadge, A. MacGregor, J. Karppinen, and F. M. K. Williams, “ISSLS prize winner: Vertebral endplate (modic) change is an independent risk factor for episodes of severe and disabling low back pain,” *Spine (Phila Pa 1976)*, vol. 40, no. 15, pp. 1187–1193, Aug. 2015, doi: 10.1097/BRS.0000000000000937.
[203] M. Karchevsky, M. E. Schweitzer, J. A. Carrino, A. Zoga, D. Montgomery, and L. Parker, “Reactive endplate marrow changes: a systematic morphologic and epidemiologic evaluation,” *Skeletal Radiol*, vol. 34, no. 3, pp. 125–129, Mar. 2005, doi: 10.1007/s00256-004-0886-3.
[204] A. J. Teichtahl *et al.*, “Lumbar disc degeneration is associated with modic change and high paraspinal fat content – a 3.0T magnetic resonance imaging study,” *BMC Musculoskelet Disord*, vol. 17, no. 1, p. 439, Dec. 2016, doi: 10.1186/s12891-016-1297-z.
[205] M. J. Lacagnina, C. J. Heijnen, L. R. Watkins, and P. M. Grace, “Autoimmune regulation of chronic pain,” *Pain Rep*, vol. 6, no. 1, p. e905, Jan. 2021, doi: 10.1097/PR9.0000000000000905.
[206] Y.-H. Zhang, C.-Q. Zhao, L.-S. Jiang, X.-D. Chen, and L.-Y. Dai, “Modic changes: a systematic review of the literature,” *European Spine Journal*, vol. 17, no. 10, pp. 1289–1299, Oct. 2008, doi: 10.1007/s00586-008-0758-y.
[207] B. N. Green, C. D. Johnson, J. Snodgrass, M. Smith, and A. S. Dunn, “Association Between Smoking and Back Pain in a Cross-Section of Adult Americans,” *Cureus*, Sep. 2016, doi: 10.7759/cureus.806.
[208] R. A. DEYO and J. E. BASS, “Lifestyle and Low-Back Pain,” *Spine (Phila Pa 1976)*, vol. 14, no. 5, pp. 501–506, May 1989, doi: 10.1097/00007632-198905000-00005.
[209] I. M. da Cruz Fernandes, R. Z. Pinto, P. Ferreira, and F. S. Lira, “Low back pain, obesity, and inflammatory markers: exercise as potential treatment,” *J Exerc Rehabil*, vol. 14, no. 2, pp. 168–174, Apr. 2018, doi: 10.12965/jer.1836070.035.
[210] L. Wang, F.-S. Wang, and M. E. Gershwin, “Human autoimmune diseases: a comprehensive update,” *J Intern Med*, vol. 278, no. 4, pp. 369–395, Oct. 2015, doi: 10.1111/joim.12395.
[211] P. K. Gregersen and T. W. Behrens, “Genetics of autoimmune diseases — disorders of immune homeostasis,” *Nat Rev Genet*, vol. 7, no. 12, pp. 917–928, Dec. 2006, doi: 10.1038/nrg1944.

[212] J. Karppinen *et al.*, "Genetic Factors Are Associated With Modic Changes in Endplates of Lumbar Vertebral Bodies," *Spine (Phila Pa 1976)*, vol. 33, no. 11, pp. 1236–1241, May 2008, doi: 10.1097/BRS.0b013e318170fd0e.
[213] R. M. Kanna *et al.*, "Prevalence, Patterns, and Genetic Association Analysis of Modic Vertebral Endplate Changes," *Asian Spine J*, vol. 11, no. 4, pp. 594–600, Aug. 2017, doi: 10.4184/asj.2017.11.4.594.
[214] M. Freidin *et al.*, "Genome-wide meta-analysis identifies genetic locus on chromosome 9 associated with Modic changes," *J Med Genet*, vol. 56, no. 7, pp. 420–426, Jul. 2019, doi: 10.1136/jmedgenet-2018-105726.
[215] G. S. Cooper and B. C. Stroehla, "The epidemiology of autoimmune diseases," *Autoimmun Rev*, vol. 2, no. 3, pp. 119–125, May 2003, doi: 10.1016/S1568-9972(03)00006-5.
[216] Y. Jiao, Y. Lin, Y. Zheng, Y. Yuan, Z. Chen, and P. Cao, "The bacteria-positive proportion in the disc tissue samples from surgery: a systematic review and meta-analysis," *European Spine Journal*, vol. 28, no. 12, pp. 2941–2950, Dec. 2019, doi: 10.1007/s00586-019-06062-6.
[217] D. M. Urquhart *et al.*, "Could low grade bacterial infection contribute to low back pain? A systematic review," *BMC Med*, vol. 13, no. 1, p. 13, Dec. 2015, doi: 10.1186/s12916-015-0267-x.
[218] I. Granville Smith, N. P. Danckert, M. B. Freidin, P. Wells, J. R. Marchesi, and F. M. K. Williams, "Evidence for infection in intervertebral disc degeneration: a systematic review," *European Spine Journal*, vol. 31, no. 2, pp. 414–430, Feb. 2022, doi: 10.1007/s00586-021-07062-1.
[219] H. B. Albert, P. Kjaer, T. S. Jensen, J. S. Sorensen, T. Bendix, and C. Manniche, "Modic changes, possible causes and relation to low back pain," *Med Hypotheses*, vol. 70, no. 2, pp. 361–368, Jan. 2008, doi: 10.1016/j.mehy.2007.05.014.
[220] M. M. Georgy, F. Vaida, M. Stern, and K. Murphy, "Association between Type 1 Modic Changes and *Propionibacterium Acnes* Infection in the Cervical Spine: An Observational Study," *American Journal of Neuroradiology*, vol. 39, no. 9, pp. 1764–1767, Sep. 2018, doi: 10.3174/ajnr.A5741.
[221] H. B. Albert *et al.*, "Does nuclear tissue infected with bacteria following disc herniations lead to Modic changes in the adjacent vertebrae?," *European Spine Journal*, vol. 22, no. 4, pp. 690–696, 2013, doi: 10.1007/s00586-013-2674-z.
[222] J. Aghazadeh *et al.*, "Modic changes in the adjacent vertebrae due to disc material infection with Propionibacterium acnes in patients with lumbar disc herniation," *European Spine Journal*, vol. 26, no. 12, pp. 3129–3134, Dec. 2017, doi: 10.1007/s00586-016-4887-4.
[223] M. N. Capoor *et al.*, "Propionibacterium acnes biofilm is present in intervertebral discs of patients undergoing microdiscectomy," *PLoS One*, vol. 12, no. 4, p. e0174518, Apr. 2017, doi: 10.1371/journal.pone.0174518.
[224] Z. Chen *et al.*, "Modic Changes and Disc Degeneration Caused by Inoculation of Propionibacterium acnes inside Intervertebral Discs of Rabbits: A Pilot Study," *Biomed Res Int*, vol. 2016, 2016, doi: 10.1155/2016/9612437.
[225] S. Dudli, E. Liebenberg, S. Magnitsky, S. Miller, S. Demir-Deviren, and J. C. Lotz, "Propionibacterium acnes infected intervertebral discs cause vertebral bone marrow lesions consistent with Modic changes," *Journal of Orthopaedic Research*, vol. 34, no. 8, pp. 1447–1455, Aug. 2016, doi: 10.1002/jor.23265.
[226] H. B. Albert, J. S. Sorensen, B. S. Christensen, and C. Manniche, "Antibiotic treatment in patients with chronic low back pain and vertebral bone edema (Modic type 1 changes): A double-blind randomized clinical controlled trial of efficacy," *European Spine Journal*, vol. 22, no. 4, pp. 697–707, Apr. 2013, doi: 10.1007/s00586-013-2675-y.
[227] C. J. Gilligan, S. P. Cohen, V. A. Fischetti, J. A. Hirsch, and L. G. Czaplewski, "Chronic low back pain, bacterial infection and treatment with antibiotics," Jun. 01, 2021, *Elsevier Inc.* doi: 10.1016/j.spinee.2021.02.013.
[228] S. Iyer, P. K. Louie, M. T. Nolte, and F. M. Phillips, "The Relationship Between Low-Grade Infection and Degenerative Disk Disease," *Journal of the American Academy of Orthopaedic Surgeons*, vol. 27, no. 14, pp. 509–518, Jul. 2019, doi: 10.5435/JAAOS-D-18-00257.

[229] F. Angum, T. Khan, J. Kaler, L. Siddiqui, and A. Hussain, “The Prevalence of Autoimmune Disorders in Women: A Narrative Review,” *Cureus*, May 2020, doi: 10.7759/cureus.8094.
[230] X. Liu, C. Wu, C. Li, and E. Boerwinkle, “dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs,” *Hum Mutat*, vol. 37, no. 3, pp. 235–241, Mar. 2016, doi: 10.1002/humu.22932.
[231] M. Ghoussaini *et al.*, “Open Targets Genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics,” *Nucleic Acids Res*, vol. 49, no. D1, pp. D1311–D1320, Jan. 2021, doi: 10.1093/nar/gkaa840.
[232] Z. Qu, F. Zhang, W. Chen, T. Lin, and Y. Sun, “High-dose TGF-β1 degrades human nucleus pulposus cells via ALK1-Smad1/5/8 activation,” *Exp Ther Med*, Aug. 2020, doi: 10.3892/etm.2020.9088.
[233] D. R. Eyre and H. Muir, “Types I and It Collagens in Intervertebral Disc INTERCHANGING RADIAL DISTRIBUTIONS IN ANNULUS FIBROSUS,” 1976.
[234] J. P. H. J. Rutges *et al.*, “Increased MMP-2 activity during intervertebral disc degeneration is correlated to MMP-14 levels,” *Journal of Pathology*, vol. 214, no. 4, pp. 523–530, Mar. 2008, doi: 10.1002/path.2317.
[235] F. M. K. Williams *et al.*, “Novel genetic variants associated with lumbar disc degeneration in northern Europeans: a meta-analysis of 4600 subjects,” *Ann Rheum Dis*, vol. 72, no. 7, pp. 1141–1148, Jul. 2013, doi: 10.1136/annrheumdis-2012-201551.
[236] S. Chen, S. Liu, K. Ma, L. Zhao, H. Lin, and Z. Shao, “TGF-β signaling in intervertebral disc health and disease,” *Osteoarthritis Cartilage*, vol. 27, no. 8, pp. 1109–1117, Aug. 2019, doi: 10.1016/j.joca.2019.05.005.
[237] J.-W. Kim *et al.*, “Activation of Hypoxia-Inducible Factor-1α Signaling Pathway Has the Protective Effect of Intervertebral Disc Degeneration,” *Int J Mol Sci*, vol. 22, no. 21, p. 11355, Oct. 2021, doi: 10.3390/ijms222111355.
[238] Y. Wang, M. Che, J. Xin, Z. Zheng, J. Li, and S. Zhang, “The role of IL-1β and TNF-α in intervertebral disc degeneration,” Nov. 01, 2020, *Elsevier Masson SAS*. doi: 10.1016/j.biopha.2020.110660.
[239] E. P. M. Urovitz and V. L. Fornasier, “Autoimmunity in Degenerative Disk Disease A Histopathologic Study.” [Online]. Available: http://journals.lww.com/clinorthop
[240] M. S. Ellulu, I. Patimah, H. Khaza’ai, A. Rahmat, and Y. Abed, “Obesity & inflammation: The linking mechanism & the complications,” *Archives of Medical Science*, vol. 13, no. 4, pp. 851–863, 2017, doi: 10.5114/aoms.2016.58928.
[241] S. Biswas, K. D. Chaudhuri, P. Mitra, and K. S. Rao, “Relation Predictions in Comorbid Disease Centric Knowledge Graph Using Heterogeneous GNN Models,” 2023, pp. 343–356. doi: 10.1007/978-3-031-34960-7_24.
[242] P. Lecca, “Machine Learning for Causal Inference in Biological Networks: Perspectives of This Challenge,” *Frontiers in Bioinformatics*, vol. 1, 2021, doi: 10.3389/fbinf.2021.746712.
[243] R. Moraffah, M. Karami, R. Guo, A. Raglin, and H. Liu, “Causal Interpretability for Machine Learning-Problems, Methods and Evaluation.”