Abstract: The intervertebral discs degeneration (IDD) is one of the leading structural substrates, causing chronic low back pain (LBP). LBP is a common neurological disorder but the LBP genetic predictors have not been sufficiently studied. Fibril collagens are important components of the nucleus pulposus, the anulus fibrosus and the vertebral endplate. Collagen type I is most studied as a structural component of the nucleus pulposus and the anulus fibrosus of the intervertebral disc. Single nucleotide variants (SNVs) of genes encoding alpha-1 and alpha-2 chains of collagen type I are associated with IDD, but the results of genetical studies are not translated into action. (1) The purpose of the study is the analysis of associative genetic and genome-wide studies of the COL1 gene family role in the development of IDD and LBP. The study of the COL1A1 gene’s SNVs association with the IDD is important for the perspective of personalized neurology. A personalized approach can help to identify patients at high risk of the IDD developing and its complications, including intervertebral disc herniation and spinal stenoses in young and working age patients. On the other hand, the role of nutritional support for patients, carriers of the SNV risk alleles in the COL1A1 gene, including collagen hydrolysates and oxyproline preparations has not been sufficiently studied.

Keywords: intervertebral disc degeneration; low back pain; collagen type I; genetics; gene; COL1A1; COL1A2.

Introduction

Chronic low back pain (LBP) is a common presenting disorder [1,2,3]. The intervertebral discs degeneration (IDD), along with the facet joint arthrosis, is one of the key reasons causing LBP among adults [4,5]. The IDD begins in adolescence and it may be asymptomatic for a long time [6,7]. The first symptoms of LBP usually appear about the age of 30 [8]. Chronic LBP mainly affects adults over 40 years old [9,10].

LBP is one of the most common healthcare problems. About 60-90% of the population face LBP throughout life and 25-40% of adults are diagnosed with it annually [11]. In most of the cases, an LBP episode is short-term, about 4% of the working-age population with LBP experience long-term temporary disability and 1% - permanent disability. This is the second most common cause of temporary disability and the fifth most common cause of hospitalization. Most people often deal with LBP episodes which are quite benign in nature. In the vast majority of cases, LBP disappears within 1-2 weeks, but 66-75% of patients suffer from minor aches for about a month after the acute pain episode is relieved.
At the same time, LBP may be the only symptom of a serious illness beginning. Thus, during the first month, among patients with LBP, 4-5% are diagnosed with a clinically significant spinal disc herniation, 4-5% - with spinal stenosis and 1% - with internal organs diseases (kidneys, gynecological problems), rarely oncological and infectious diseases [11].

Among all the backpain varieties LBP prevails. About 80-100% of the population face acute LBP of varying intensity [12]. When analyzing the primary referral to general practitioners about LBP, its causes are identified in the vast majority of cases (70%). "Discogenic" pain and pain associated with facet joints dysfunction affect 20% of patients with LBP. Compressional radiculopathy of the lumbar and sacral vertebrae is observed in 8% of cases. According to epidemiological research conducted in Moscow, 24.9% out of 1,300 primary patients have applied for outpatient medical care mainly because of LPB [13]. Among patients who applied for another reason, back pain has been observed in 3.9% of cases. For the last year, LBP bothered 52.9% of patients, and for the last month — 38.5%. The results of an epidemiological research, which included a survey of more than 46,000 residents of different European countries and Israel, showed that 24% suffer from chronic back pain (of various localization), LBP — 18%, and neck pain - 8% of the population [14].

Thus, the chronic LBP problem is a medical problem of the working-age population.

**Type I collagen**

Pharmacological and non-pharmacological methods of the chronic LBP treatment are actively studied and implemented in neurological clinical practice. However, the methods affecting collagen exchange which form part of the nucleus pulposus', the anulus fibrosus' and the vertebral endplate's structure have not been taken into account until the present (Figures 1-2).

Fibrillar collagens are the most important and most studied with regard to the LBP problem under consideration in humans [15,16].

Fibrillar collagens form the basis of the human body connective tissue and provide its strength and elasticity. Fibrillar collagens are the most common proteins among mammals, making up from 25% to 35% of proteins in the body, i.e., 6% of body weight. The term "collagen" unites a family of closely related fibrillar proteins, which are the main protein element of skin, bones, tendons, cartilage, blood vessels, teeth. Different types of collagens predominate in different tissues. Therefore, this is determined by the role that collagen plays in a particular organ or tissue [15].

![Figure 1](image-url). Schematic representation of the human lumbar intervertebral disc (IVD) in normal conditions and with its degeneration [15].
Figure 2. The relative proportions of the three main components of the normal adult intervertebral disc and the endplate [15].

The collagen (tropocollagen) molecule is made of three peptide chains, each of which contains about 1000 amino acid residues (Figure 3) [17].

Figure 3. Schematic of the triple helical structure of collagen (top), the organization of the twisted structure in a fibril (middle) and fibrils in a collagen fiber (bottom). The presence of a 67-nm gap between adjacent helical structures determines the striation of the fibrils that make up the fiber, visible in an electron microscope [17].

The amino acid composition of collagen is unusual: every third amino acid is Glycine, 20% are Proline and Hydroxyproline residues; 10% - Alanine; the remaining 40% are all
other amino acids. Collagen is the only protein that contains Hydroxyproline. This amino acid is obtained by hydroxylation of the part of Proline residues after the peptide chains formation [18].

Collagen is synthesized and supplied to the extracellular matrix by almost all the cells (fibroblasts, chondroblasts, osteoblasts, odontoblasts, cementoblasts, keratoblasts, etc.). Collagen synthesis and processing is a complex multi-stage process that begins in the cell and ends in the extracellular matrix [15,19]. Collagen synthesis disorders are caused by genes mutations, as well as in the process of translation and posttranslational modification, come with the defective collagens’ appearance. Since about 50% of all the collagen proteins are contained in skeletal tissues, and the remaining 40% in the dermis and 10% in the internal organs stroma, collagen synthesis defects are followed by pathology of both the osteoarticular system and internal organs [19].

Collagen is a pronounced polymorphic protein. For today, 28 types of collagens have been studied, they differ from each other in the primary structure of peptide chains, functions and localization in the body. There are about 30 variants of α-chains forming a triple helix. To denote each type of collagen, a certain formula is used, in which the type of collagen is written in Roman numerals in brackets and Arabic numerals are used to denote α-chains: for example, collagen type II and type III are formed by identical α-chains, their formulas, respectively [α1(II)]3 and [α1(III)]3; collagen types I and IV are heterotrimers and are usually formed by two different types of α-chains, their formulas, respectively [α1(I)]2α2(I) and [α1(IV)]2α2(IV). The index behind the parenthesis denotes the number of identical α-chains [20].

Type I collagen is the most common type of collagen and is expressed in almost all connective tissues. Its primary structure corresponds to the repetition of the amino acid motif (Gly-X-Y), where X and Y are often represented by the amino acids Proline (Pro) or Hydroxyproline (HyPro), respectively (Figure 4). The protein is organized in the form of triple helical chains assembled into fibrils and stabilized due to the formation of intermolecular and interfibrillar cross-links. It is the main protein of bones, skin, tendons, ligaments, intervertebral discs, sclera, cornea and blood vessels. It accounts for approximately 95% of the total collagen content in bone and approximately 80% of all proteins present in bone. Type I collagen is a molecule of heterotrimers. In most cases, it consists of two chains α1 and one chain α2, although the homotrimer α1 exists as a minor form. Each chain consists of more than 1000 amino acids, and the type I collagen molecule length is ~ 300 nm and about 1-5 nm wide (Figure 4) [21].

![Figure 4](https://example.com/figure4.png)

**Figure 4.** Collagen type I chemical structure. (a) Sequence of amino acids - primary structure, (b) left-handed helix - secondary structure; right-handed triple-helix - tertiary structure and (c) staggered - quaternary structure [21].
Type I collagen has three domains: N-terminus non-triple helical domain (N-telopeptide); central triple helical domain; C-terminus non-triple helical domain (C-telopeptide) (Figure 5) [22].

![Figure 5. Molecular structure of fibrillar collagens with the various subdomains as well as the cleavage sites for N- and C-procollagenases (shown is the type I collagen molecule). Whereas they are arranged in tendon in a parallel manner they show a rather network-like supramolecular arrangement in articular cartilage. The primary structure of type I collagen corresponds to a repeat of the amino acid motif (Gly-XY), where X and Y are often represented by the amino acids proline (Pro) or hydroxyproline (HyPro), respectively. The protein is organized in the form of triple helical chains assembled into fibrils and stabilized due to the formation of intermolecular and interfibrillar cross-links [22].](image)

The central domain is the largest, approximately 95% of the entire molecule. The triple helical domain is only possible due to the presence of Glycine repeats (G) -X Y, where X is often a Proline and Y is a Hydroxyproline. Glycine in every third position is necessary for the proper formation of the collagen protein structure [23].

Because of chronological aging and connective tissue diseases, type I collagen undergoes several non-enzymatic posttranslational modifications, such as glycation [24]. This process causes the formation of advanced glycated end-products (AGE), which contribute to an increase in cross-links, and also reduces the diameter and length of individual collagen fibers. Aging can lead to physical changes in matrix proteins of connective tissues.

The changes initiated by biochemical processes have direct consequences for the molecular and structural organization of these proteins, especially the most common component, type I collagen. Molecular changes at the nanoscale level cause microscopic mechanical changes that can affect several tissue functions. Thus, by changing the mechanics or matrix proteins, such as type I collagen (stiffness, length and diameter of type I collagen), aging can, for example, affect the elasticity of many tissues (skin, cardiovascular endothelium, connective tissues) [24]. These changes are the biggest risk factor for the development of many diseases, which can become pathological, including the IDD. Hence, interventions that slow down aging and the IDD will bring great benefits to the health of backbone [25].

**COL1A1 Gene**

The COL1A1 gene is located at chromosomal locus 17q21.33 and is 18 kilobytes (kb) in size and consists of 52 exons (Figure 6) [26].

This gene encodes the pro-alpha 1 chain of type I collagen, the triple helix of which consists of two alpha 1 chains and one alpha 2 chain. Mutations in this gene cause known monogenic diseases (Table 1).

Osteogenesis imperfecta (OI) is a connective tissue disease that in more than 90% of cases is caused by a type I collagen anomaly. Because of the significant phenotypic plasticity, Sillence et al. developed an OI subtypes classification: Type I OI with blue sclera; perinatal lethal OI type II, also known as congenital OO; type III OI, progressively deformed form with normal sclera; and type IV OI with normal sclera [27].
Combined osteogenesis imperfecta and Ehlers-Danlos syndrome type I is an autosomal dominant generalized connective tissue disease characterized by symptoms of osteogenesis imperfecta (brittleness bone, long bone fractures, blue sclera) and Ehlers-Danlos syndrome (joints hyperactivity, softness and elasticity of joints, hyperextensive skin, abnormal wound healing, light bruises, vascular fragility) [28].

Ehlers-Danlos syndrome type I, arthrohalasia EDS differs from other types of Ehlers-Danlos syndrome by the congenital hip dislocation frequency and extreme joint flabbiness with repeated joint subluxations and minimal skin damage [29].

Caffey’s disease is an autosomal dominant disease characterized by an infantile episode of massive formation of a new subcostal bone, which usually includes the diaphysis of the long bones, mandible and collarbones. Painful edema and systemic fever often come with an episode, which usually begins before the age of 5 months and passes by the age of 2 years. Laboratory report include an increased level of alkaline phosphatase, and sometimes an increase in the number of leukocytes and sedimentation rate of erythrocytes. Recurrent episodes are rare [30].

Single nucleotide variants (SNVs) of the COL1A1 gene appear to increase of the osteoporosis risk. Osteoporosis is a condition when bones become increasingly brittle and prone to fractures. Also, SNVs that occur in the control (regulatory) region of the COL1A1 gene probably affect the production of type I collagen, but not its molecule structure. Several studies have shown that women with these SNVs are more likely to have osteoporosis symptoms, especially with low bone density and bone fractures, than women without these SNVs. This change is just one of many factors that can increase the risk of osteoporosis [31].

Reciprocal translocations between chromosomes 17 and 22, where this gene and the platelet growth factor beta gene are located, are associated with a certain type of skin tumor called dermatofibrosarcoma protuberans, as a result of unregulated expression of growth factor [15].

The COL1A1 gene is considered as a candidate gene for the IDD, since the gene is expressed both in the annulus fibrosus (firstly) and in the nucleus pulpous (secondly). Type I collagen is the main annulus fibrosus component of the intervertebral disc.

The collagen type I in the glycosaminoglycan matrix induces proteoglycans synthesis by canine intervertebral disk cells [32]. In mice genetically engineered for reduced type I collagen, IVD tissue was also mechanically inferior compared with control animals [33]. It is therefore plausible that an increased ratio of COL1A1 expression compared with COL1A2 may lead to structural alterations, as well as to healing defects in the annulus fibrosus and other components of the discs in IVDD [15].

Although the mechanism by which genetic changes in type I collagen affect the development of IDD is not fully understood, various population studies have reported that the rs1800012 polymorphism of the COL1A1 gene is associated with an increased risk of IDD. The researchers suggested that this SNV leads to an imbalance between the expression of the COL1A1 and COL1A2 proteins, which causes instability of collagen fibers. In
Sp1 polymorphism (rs1800012), the guanine (G) is substituted by thymidine (T) in the fourth Sp1 binding site in intron 1 of the COL1A1 genes, more specifically, in the promoter +1245 base pair (bp) from the transcription start site [34].

### Table 1. The COL1A1 gene, its characteristics effects.

<table>
<thead>
<tr>
<th>Type I collagen</th>
<th>Description</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene and its localization</td>
<td>COL1A1, 17q21.33</td>
<td>Gene ID:1277 OMIM:120150</td>
</tr>
<tr>
<td></td>
<td>Imperfect osteogenesis. Mutations I with blue sclera</td>
<td>AD (OMIM: 166200)</td>
</tr>
<tr>
<td></td>
<td>Perinatal lethal osteogenesis imperfecta type II, also known as congenital osteogenesis imperfecta</td>
<td>AD (OMIM: 166210)</td>
</tr>
<tr>
<td></td>
<td>Osteogenesis imperfecta type III, progressively deforming form with normal sclera</td>
<td>AD (OMIM: 259420)</td>
</tr>
<tr>
<td></td>
<td>Osteogenesis imperfecta type IV with normal sclera</td>
<td>AD (OMIM: 166220)</td>
</tr>
<tr>
<td>Mutation in the gene and hereditary diseases</td>
<td>Combined osteogenesis imperfecta and Ehlers-Danlos-1 syndrome</td>
<td>AD (OMIM: 619115)</td>
</tr>
<tr>
<td></td>
<td>Ehlers-Danlos syndrome, arthralgia type I</td>
<td>AD (OMIM: 130060)</td>
</tr>
<tr>
<td></td>
<td>Caffi disease</td>
<td>AD (OMIM 1114000)</td>
</tr>
<tr>
<td></td>
<td>Osteoporosis</td>
<td>MU (OMIM 166710)</td>
</tr>
<tr>
<td></td>
<td>Dermatofibrosarcoma protuberans</td>
<td>MU</td>
</tr>
<tr>
<td></td>
<td>Intervertebral disk degeneration</td>
<td>MU</td>
</tr>
<tr>
<td>Expression in tissue</td>
<td>Leather, bones, intervertebral discs, tendons.</td>
<td></td>
</tr>
<tr>
<td>Special domains</td>
<td>Three NC domains and two Col Domains.</td>
<td></td>
</tr>
<tr>
<td>Special neopeptides</td>
<td>N- and C-terminal propeptides and N- and C-terminal degradation peptides.</td>
<td></td>
</tr>
<tr>
<td>Protein structure and function</td>
<td>Type I collagen is a hetero-trimer.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The molecule in most cases consists of two α1 chains and one α2 chain.</td>
<td></td>
</tr>
</tbody>
</table>
The α1 homotrimer can exist in both minor and major forms. Glycines at every third position of the helical domain is crucial for the folding of the helical structure of the protein. An indispensable component for the mechanical competence of the bone extracellular matrix. Key structural component of many other tissues.

<table>
<thead>
<tr>
<th>Binding proteins</th>
<th>Integrins, proteoglycans, etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main function</td>
<td>The main organic component of bones and intervertebral discs.</td>
</tr>
<tr>
<td></td>
<td>Indispensable for the integrity of bones, ligaments, intervertebral discs.</td>
</tr>
</tbody>
</table>

| Biomarkers        | α- and β-CTX-I, NTX, ICTP, PINP, PICP, C1M. |

COL, collagen; CTX-I, C-terminal telopeptide of collagen type 1; NTX, cross-linked collagen type 1 N-telopeptide; ITPC, cross-linked carboxy-terminal telopeptide derived from type 1 collagen; PICP, pro-collagen type 1 amino-terminal proprietyde; C1M, collagen type 1 neoepitope.

In a study on the Netherlands population, it was shown that individuals over 65 years old who are homozygous carriers of the TT rs1800012 genotype had 3.6 times higher susceptibility to IDD than people with the GT or GG genotypes [15,35].

Later, the frequency of rs1800012 polymorphism alleles in young men in Greece was investigated. Comparable results are shown: carriers of the homozygous TT genotype accounted for 33.3% among patients with IDD, but this genotype was not found in the control group. In addition, a significantly smaller number of controls was heterozygotes for this allele: 66.7% in the IDD patients v 41.7% in the controls. Elderly men and women with the TT genotype had a higher risk of IVDD than those with the GG and GT genotypes, with an odds ratio (OR) of 3.6. However, in contrast to the study conducted in the Netherlands, there was a high frequency of GT genotype heterozygous carriers (66.7%) in the main group (IDD patients) compared with the healthy control group (41.7%) [15,35,36].

A study in Finland investigated the predictive role of rs2075555 in the development of IDD in the lumbar spine. A statistically significant association of the studied polymorphism with the development of degenerative changes in the lumbar IVD in Finns was found [15,37].

A study in India showed that the rs1800012 polymorphism does not appear to be associated with IDD in this population. Nevertheless, the frequency of the minor T allele was higher in patients with IDD at the cervical and lumbar spine levels compared with the control group, but the differences did not reach statistical significance. The authors suggested that either another polymorphism of the COL1A1 gene or a polymorphism of the gene encoding a different type of collagen may play a more important role in the development of IDD in the Indian population [15,38].
In 2017, Zhong et al. conducted a meta-analysis of studies on the role of rs1800012 polymorphism as a predictor of IDD among Chinese people. A statistically significant association of the minor T allele homozygous carriage of this polymorphism with the development of IDD, including severe forms, has been demonstrated [15,39].

In 2020, Hanaei et al. studied two groups of patients with IDD in the Iranian population, comparable in age and sex, in order to study the association of SNV rs909102 of the COL1A1 gene. The results showed that the T allele, genotypes CC and TT rs909102 SNV of the COL1A1 gene were more common among patients with IDD; however, the statistical data turned out to be insignificant [15,40].

In summary, we can conclude that COL1A1 is a candidate gene associated with the pathogenesis of IDD in humans. However, the predictive role of the studied SNVs depends on the region of residence, race and ethnicity of the patients.

Conclusion

Thus, the role of type I collagen in the intervertebral discs functioning is important in health and pathology, including their degeneration. The study of a number of COL1A1 gene SNVs association with the IDD is important from the perspective of personalized neurology. A personalized approach can help to identify patients at high risk of the IDD developing and its unfavorable course, including spinal disc herniation and spinal stenoses of young and working age patients. On the other hand, the role of nutritional support for patients, carriers of the SNV risk alleles in the COL1A1 gene, including collagen hydrolysates and oxyproline preparations has not been sufficiently studied.

Author Contributions: Conceptualization, N.A.S., M.A.Z. and V.V.T.; investigation, N.A.Z., V.V.T.; writing—original draft preparation, N.A.Z.; writing—review and editing, V.V.T.; supervision, M.A.Z.; project administration, M.A.Z. and V.V.T. All authors have read and agreed to the published version of the manuscript.

Funding: The study was not sponsored.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References


26) UCSC Genome Browser on Human (GRCh38/hg38) Available online: https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&hgFind=omimGeneAcc&position=120150 (accessed on 8.11.2021)


