Whole-genome sequencing reveals novel genes in ossification of the posterior longitudinal ligament of the thoracic spine in the Chinese population

Chen Liang, Peng Wang, Xiao Liu, Chenlong Yang, Yunlong Ma, Lei Yong, Bin Zhu, Xiaoguang Liu and Zhongjun Liu

Abstract

Background: Ossification of the posterior longitudinal ligament (OPLL) of the spine is a complex, multifactorial disease. Although several genes that are linked to cervical OPLL susceptibility have been reported, specific genetic studies regarding thoracic OPLL are lacking. Whole-genome sequencing has been considered as an efficient strategy to search for disease-causing genes.

Methods: We analysed whole-genome sequences in a cohort of 25 unrelated patients with thoracic OPLL. Bioinformatics analysis and various algorithms were used to predict deleterious variants. Sanger sequencing was used to confirm the variants.

Results: Four deleterious mutations in three genes (c.2716C>T (p.Arg906Cys) in collagen type VI α6 (COL6A6); c.1946G>C (p.Gly649Ala) in collagen type IX α1 (COL9A1); and c.301T>C (p.Ser101Pro) and c.171A>G (p.Ile57Met) in toll-like receptor 1 (TLR1)) were successfully identified. All the variants were confirmed by Sanger sequencing.

Conclusion: The novel deleterious mutations of the three genes may contribute to the development of thoracic OPLL.

Keywords: Ossification of the posterior longitudinal ligament, Thoracic, Whole-genome sequencing, Susceptible gene

Introduction

Ossification of the posterior longitudinal ligament (OPLL) of the spine is characterized by heterotopic bone formation in cervical or thoracic ligaments, causing myelopathy, and OPLL is most prevalent in the East Asian population [1, 2]. As thoracic OPLL (T-OPLL) progresses latently, patients may be underdiagnosed until the advanced stages of the disease when the spinal cord is severely compressed by ectopic ossification. In addition, the surgical treatment of T-OPLL is challenging due to the complicated anatomical structures, and T-OPLL generally has a less favourable prognosis than its cervical counterpart [3, 4]. A comprehensive understanding of the pathogenesis of T-OPLL may provide new insight into the therapeutic approaches for T-OPLL.

Though the aetiology of OPLL remains unclear, it has been widely accepted that this entity is a multifactorial disease influenced by numerous genetic and non-genetic factors [5]. Non-genetic factors include mechanical stress, the degeneration process, diet, and biological rhythm. Compared to the cervical spine, the thoracic spine has less range of motion as it is limited by the thorax. In addition, the thoracic spine is more stable and experiences less mechanical stress than the cervical spine. Thus, the thoracic spine is less susceptible to degeneration. Consequently, we hypothesized that genetic factors play a vital role in the aetiology and development of T-OPLL. Previous genetic studies of OPLL have revealed the following osteogenic gene loci that may be involved in the pathogenesis of cervical OPLL: NPPS, COL6A1, COL11A2, BMP2, BMP4, BMP9, TGF-β1,
TGF-β3, TGFBR2, ESR1, FGFR1, IL-1β, IL-15RA, RUNX2 and RSPO2 [6–18]. A recent study revealed two deleterious variants of COL6A1 and IL17RC in patients with T-OPLL [19]. However, there is inadequate research regarding susceptibility genes for T-OPLL. Therefore, there is an urgent need to identify pathogenic genes to provide novel approaches for future investigation and intervention of T-OPLL. Whole-genome sequencing represents an effective, accurate and reproducible strategy for the identification of specific variants in individual human genomes [20, 21]. The present study analysed data from whole-genome sequencing from 25 sporadic OPLL patients to identify pathogenic genes for T-OPLL. Sanger DNA sequencing was used to validate the variants disclosed by whole-genome sequencing. Single nucleotide polymorphisms (SNPs) were analysed with bioinformatics. The present study reported several pathogenic variants in genes associated with T-OPLL in the Chinese population.

**Materials and methods**

**Patients**

Twenty-five unrelated Chinese patients with T-OPLL were consecutively recruited between 2012 and 2016 from the Department of Orthopaedics at Peking University Third Hospital (PUTH), including 12 male and 13 female patients with an average age of 52.4 years ranging from 29 to 70 years. The present study was conducted with the approval of the Ethics Committee of the PUTH Institutional Review Board. Written informed consent was obtained from all patients whose specimens and clinical information were used for the present study. All patients received radiological examinations, including plain radiographs, computed tomography (CT) and magnetic resonance imaging (MRI). T-OPLL was diagnosed based on clinical and radiological evidence by at least two experienced spinal surgeons. The inclusion criteria were as follows: (a) > 18 years old and (b) both CT and MRI evidence of T-OPLL were available. The exclusion criteria were as follows: (a) metabolic diseases, such as hypophosphataemic rickets, acromegaly, hyperparathyroidism, diffuse idiopathic skeletal hyperostosis and pituitary diseases; (b) administration of medications interfering with bone or calcium metabolism, such as oral contraceptives, calcium, vitamin D and glucocorticoids; (c) concomitant developmental malformations; or (d) family history of hereditary disease [22]. Neurological status and surgical outcome were assessed using the Japanese Orthopaedic Association (JOA) scoring system for thoracic myelopathy [4].

**Whole-genome sequencing**

Genomic DNA was extracted from whole blood with the TIANamp Blood DNA kit (TIANGEN BIOTECH, Beijing, China). Quality and quantity of DNA were assessed with a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Whole-genome sequencing was performed on an Illumina HiSeq X Ten platform using the GenCap custom enrichment kit (MyGenostics, Beijing). The experiment was performed according to the manufacturer’s protocol.

**Bioinformatics analysis**

Bioinformatics analysis started with the raw sequencing data from the Illumina pipeline. Raw data were processed and filtered, and low-quality reads were discarded. Burrows-Wheeler Aligner (BWA) software was used to align the clean reads to the human reference genome (GRCh37/HG19). Genome Analysis Toolkit (GATK) was applied for variant analysis, including local realignment around insertion/deletions (InDels) and base quality score recalibration. Duplicate reads were discarded using Picard. A quality control (QC) system guaranteed high-quality data throughout the entire analysis pipeline.

**Validation and evaluation of mutations**

Sanger DNA sequencing was used to validate the accuracy of all identified variants. Variant frequency was compared with the 1000G SNP database (http://www.1000genomes.org/). Analysis for potential deleterious mutations was performed using various algorithms, including PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/) [23], SIFT (http://sift.jcvi.org/) [24], Mutation Assessor (http://mutationassessor.org/r3/), Mutation Taster (http://www.mutationtaster.org/) [25] and GERP++ prediction (http:// mendel.stanford.edu/SidowLab/downloads/gerp/) [26]. In the present study, the variant frequency was combined with at least four of the five algorithms above to prioritize potential pathogenic variants.

**Sanger sequencing**

DNA was amplified using custom oligonucleotide primers. PCR amplification of all SNPs was conducted using Ex-Taq premix (Takara, Japan). The following reaction conditions were used: 96 °C for 5 min; 35 cycles at 96 °C for 20 s, 52 °C for 30 s and 72 °C for 60 s; and 72 °C for 5 min. PCR assays (total volume of 50 μL) contained 0.25 μL of Ex-Taq premix (5 U/μL), 2 μL of oligonucleotide primer (10 pmol/μL), 4 μL of dNTPs (2.5 mM), 37.75 μL of distilled water and 1 μL of template containing 1 ng of genomic DNA. To perform Sanger sequencing, purification of PCR products using ethanol precipitation was required prior to the cycle sequencing reaction. Purified PCR products were labelled with the BigDye Terminator Kit v3.1 (Applied Biosystems, USA) for cycle sequencing according to the manufacturer’s procedure. After the second purification of the PCR products, sequencing analysis was performed using the
automated sequencer from Applied Biosystems (3730XL DNA Analyzer).

**Statistical analysis**

Statistical analyses were performed using SPSS statistics 23.0. Data for continuous variables are presented as the mean ± standard deviation. Continuous variables were analysed using Student’s t test. Fisher’s exact test was used to identify differences between categorical variables. The statistical significance was defined as a value of $P < 0.05$ based on two-tailed tests.

### Results

#### Variant identification

Whole-genome sequencing was performed on 25 DNA samples resulting in an average of 119,525.02 Mb clean reads from the Illumina HiSeq X Ten sequencer. To filter potential pathogenic variants, rare 1000G_EAS $\leq 0.005$ mutations were identified (based on the BGI database), and damaging variants were predicted by at least four of five algorithms (SIFT, Polyphen2, Mutation Assessor, Mutation Taster and GERP++). As a result, four deleterious variants of three genes were identified in five

| SNP ID   | Gene   | Nucleotide change | Protein change | 1000G_EAS | SIFT | Polyphen2 | Mutation Assessor | Mutation Taster | GERP++ |
|----------|--------|-------------------|----------------|-----------|------|-----------|------------------|-----------------|--------|
| rs200963433 | COL6A6 | c.2716C>T        | p.Arg906Cys   | 0.004     | D    | D         | H                | D               | R      |
| rs201480339 | COL9A1 | c.1946G>C        | p.Gly649Ala   | 0.001     | D    | D         | H                | P               | R      |
| rs200212492 | TLR1   | c.301T>C         | p.Ser101Pro   | 0.001     | D    | D         | H                | D               | R      |
| rs145135062 | TLR1   | c.171A>G         | p.Ile57Met    | 0         | D    | D         | M                | D               | R      |

1000G_EAS, 1000 Genomes_East Asian; SIFT (D, deleterious); Polyphen2 (D, probably damaging); Mutation Taster (D, disease causing, P, polymorphism automatic); Mutation Assessor (predicted functional, H, high; M, medium); GERP++ (R, rejected substitutions)

![Fig. 1 Sanger sequencing chromatograms of four variants in three novel genes. Sanger sequencing confirmed the four variants identified in three novel genes](image-url)
sporadic patients (Table 1), and these variants were confirmed by Sanger sequencing (Fig. 1). These mutations of three genes (COL6A6, COL9A1 and TLR1) accounted for 20.0% of the OPLL patients. The missense mutation of TLR1 was found in three unrelated patients (accounting for ~ 12.0%). The frequency of mutations in COL6A1 and COL9A1 was 4.0% (Table 2). The sequence conservation of the encoded amino acid residues by these SNP mutations was analysed in nine different vertebrate species (Fig. 2). COL9A1 (c.1946G>C), TLR1 (c.301 T>C), TLR1 (c.171A>G) and COL6A6 (c.2716C>T) were evolutionarily conserved in seven, five, two and one vertebrates, respectively.

### Genotype and clinical feature analysis
The genotype/clinical feature correlation was analysed between 5 patients with missense mutations and 20 patients without significant mutations (Table 3). There were no differences between these two groups regarding sex, age and Japanese Orthopaedic Association (JOA) score.

### Discussion
Several lines of evidence have shown that OPLL is a polygenic disease. Genes encoding proteins and transcription factors involved in osteoblast and chondrocyte development and differentiation have been suggested in the pathomechanism of OPLL. Over the past decade, many sibling-pair linkage studies and candidate gene association studies have reported multiple genes or loci linked to OPLL susceptibility [6, 7, 10–20, 27–34] (Table 4). The majority of previous studies have focused on cervical OPLL rather than T-OPLL. Only one genetic study on T-OPLL has been reported, revealing two deleterious variants of COL6A1 and IL17RC [19]. The present study used whole-genome sequencing to identify candidate pathogenic genes for T-OPLL. The data indicated that the novel variants in three genes (COL6A6, COL9A1 and TLR1) may contribute to the pathogenesis of T-OPLL.

The COL6A6 gene on chromosome 3q22.1 has 43 exons with a coding region of 6789 bp, and it encodes COL6A6, a 2262-amino acid protein expressed in various tissues, including the muscle, heart and brain. COL6 family proteins are components of the extracellular matrix (ECM) of many tissues, including the bone, cartilage, muscle, tendon and skin. Similar to COL6A3, the COL6A6 gene encodes multiple von Willebrand factor domains and forms a component of the basal lamina of epithelial cells. COL6A6 may regulate epithelial cell-fibronectin interactions, and variation in this gene may be identified in skin diseases, such as early-onset atopic dermatitis [35]. However, in recent years, COL6A6 has been implicated as a susceptibility gene in musculoskeletal diseases. Gari et al. performed whole-exome sequencing and showed that COL6A6 mutations can influence OA disease [36]. COL6A1 has been accepted as a susceptibility gene of OPLL for many years, but there has been no associative study between COL6A6 and OPLL or T OPLL [14]. COL6A6 encodes a cell-binding protein, which interacts with COL6A1 protein by forming a trimer in collagen type VI. COL6A1 is involved in membranous or endochondral ossification by providing a scaffold for osteoblastic cells, preosteoblastic cells or chondrocytes. Thus, COL6A6 mutation may alter the structure and function of collagen type VI and eventually lead to abnormal ossification. The present study is the first to reveal COL6A6 as a susceptibility gene associated with a predisposition to thoracic OPLL. The function of COL6A6 in the pathogenesis of OPLL will be validated in future studies with respect to cell phenotype and potential signalling pathways.
The **COL9A1** gene is located on chromosome 6q13 and has 41 exons, and it encodes COL9A1, which is involved in synthesizing type IX collagen, a minor collagen component of hyaline cartilage. Previous studies have shown that **COL9A1** is associated with female osteoarthritis (OA), early onset OA and osteoporosis [37–39]. Alfred et al. found that the maturation of cartilage matrix is delayed in COL9A1-deficient mice during bone fracture healing, indicating that endochondral bone formation is impaired during fracture repair in mice [40]. However, the role of COL9A1 in OPLL pathogenesis is unknown. OPLL involves ectopic ossification in the posterior longitudinal ligament, and several histological studies of OPLL have suggested that OPLL develops through a process of endochondral ossification. As COL9A1 is involved in the process of endochondral ossification, it supports the present finding that pathogenic variants in the **COL9A1** gene may be associated with T-OPLL in the Chinese population [41, 42].

The **TLR1** gene is located on chromosome 4p14 and encodes TLR1. TLR1 is a member of the toll-like receptor (TLR) family, which is involved in pathogen recognition and activation of innate immunity. At present, 10 human and 12 murine TLRs have been characterized as TLR1–TLR10 in humans and TLR1–TLR9, TLR11, TLR12 and TLR13 in mice [43]. TLR1 polymorphism is associated with the susceptibility of multiple diseases, including tuberculosis, pancolitis and prostate cancer [44, 45]. Jeong-Eun et al. found that TLRs can be utilized to induce IL-6 expression in adipose-derived stromal cells (ASCs), and they suggested that the TLR/IL-6 signalling pathway may be useful for modulating osteogenic differentiation of ASCs [46]. In addition, chondrocytes can respond to a wide range of TLR ligands, and TLR1 has been reported to play an important role in the pathogenesis of OA [47]. He Hailong et al. explored the differentially expressed genes (DEGs) in OPLL patient ligament cells and stated that **TLR1** may be involved in spinal cord injury in OPLL [48]. The present study is among the first studies to propose **TLR1** as the disease-causing gene of spine OPLL, especially T-OPLL.

There are crosstalk mechanisms between some of the identified genes and signalling pathways to coordinate osteogenesis and bone homeostasis. Wu et al. found that

---

Table 3 Clinical features of patients with or without OPLL gene mutations

|                        | Mutation positive (n = 5) | Mutation negative (n = 20) | p   |
|------------------------|---------------------------|-----------------------------|-----|
| Age                    | 50.8 ± 14.2               | 52.8 ± 10.5                 | 0.517|
| Male/female            | 2/3                       | 8/12                        | 1.000|
| Thoracic JOA score     | 60.0 ± 1.9                | 5.7 ± 2.5                   | 0.316|

Data are presented as mean ± SD or n; JOA, Japanese Orthopedic Association.
knockout or mutation of TGF-β and BMP signalling-related genes in mice results in bone abnormalities of underlying osteoblast differentiation and bone formation [49]. Huang et al. determined that TNF-α/IL-1β decreases BMP-2-induced Runx2 expression through the activation of p38 and ERK1/2 signalling to regulate osteoblastic differentiation [50]. However, there is no study concerning the correlations among COL6A6, COL9A1 and TLR1. In the future, we will focus on the function of these three novel genes and investigate the relations among them. In addition, fibrosis and calcification are involved in the process of OPLL. In the present study, however, no correlation was noted between these novel genes and fibrosis or calcification.

In the present study, the ratio of missense mutations in COL6A6, COL9A1 and TLR1 genes was 20% (5/25), which was consistent with previous studies. Wang et al. identified two deleterious mutations in COL6A1 and IL17RC genes in 7 of 30 patients with T-OPLL, yielding a ratio of 23%. Wei et al. proposed that mutations in PTCH1 and COL17A1 genes may contribute to the development of cervical OPLL with the ratio of mutations equal to 21% (6/28). Given T-OPLL is a polygenic disease, the definitive prevalence of gene mutations associated with T-OPLL still needs further validation in a large-scale cohort.

The prevalence of OPLL is fairly low compared to degenerative disc diseases, such as lumbar disc herniation and cervical spondylosis. In addition, T-OPLL is less common.
in comparison with cervical OPPL. To the best of our knowledge, genetic studies on T-OPLL are rare, and the present study is among the first leading studies to investigate pathogenic genes for T-OPLL. Concerning the rare occurrence of T-OPLL, 25 is an adequate number of patients for genetic study. As a next step, the sample size will be increased to reveal additional T-OPLL susceptibility genes. The limitations of the present study were as follows: (1) no replication study was performed to demonstrate the association of the three recently discovered susceptibility loci with T-OPLL; and (2) no functional study was conducted to determine the mechanism and pathways in which the genes involved affect the development of the disease.

In summary, the four identified polymorphisms in three genes may contribute to susceptibility to T-OPLL in the Chinese population. The present study contributed to a better understanding of the aetiology and pathomechanism of T-OPLL. Further replication and functional studies need to be performed to identify pathogenic genes to reduce the incidence and provide novel therapeutic approaches to T-OPLL.

Acknowledgements
We would like to thank the Central Laboratory of Peking University Third Hospital for technical guidance. We are also grateful for the assistance of Dr. Zhang Xiaxia from the Department of Gastroenterology of Beijing Tiantan Hospital.

Funding
This study was supported by the National Natural Science Foundation of China (81472041).

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors’ contributions
XG, CL, and PW conceived and designed the study; XG, XCL, BZ, and ZL participated in the acquisition of samples and clinical data; LY and YM participated in the statistical analysis and interpretation of data; CL drafted the manuscript; and XG revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
This study was approved by the Ethics Committee of the PUTH Institutional Review Board. Written informed consent was obtained from all patients whose specimens and clinical information were used for this study.

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

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

Author details
1Department of Orthopaedics, Peking University Third Hospital, No. 49, North Garden Road, Haidian District, Beijing 100191, China. 2The Centre for Pain Medicine, Peking University Third Hospital, No. 49, North Garden Road, Haidian District, Beijing 100191, China.
20. Nakajima M, Takahashi A, Tsuji T, et al. A genome-wide association study identifies susceptibility loci for ossification of the posterior longitudinal ligament of the spine. Nat Genet. 2014;46:1012–6.
21. Belkadi A, Bolze A, Itoh Y, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A. 2015;112:5473–8.
22. Chen X, Guo J, Cai T, et al. Targeted next-generation sequencing reveals multiple deleterious variants in OPLL-associated genes. Sci Rep. 2016;6:26962.
23. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9.
24. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40:W452–7.
25. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11:361–2.
26. Davydov EV, Goode DL, Attenello FJ, Hsieh PC. Ossification of the posterior longitudinal ligament: genetics and pathophysiology. Neurosurg Focus. 2011;30:E6.
27. Stapleton CJ, Pham MH, Attenello FJ, Hsieh PC. Ossification of the posterior longitudinal ligament genetics and pathophysiology. Neurosurg Focus. 2011;30:E6.
28. Wei W, He HL, Chen CY, et al. Whole exome sequencing implicates PTCH1 and COL17A1 genes in ossification of the posterior longitudinal ligament of the cervical spine in Chinese patients. Genet Mol Res. 2014;13:794–804.
29. Kobashi G, Ohta K, Washio M, et al. FokI variant of vitamin D receptor gene and factors related to atherosclerosis associated with ossification of the posterior longitudinal ligament of the spine: a multi-hospital case-control study. Spine. 2008;33:E553–8.
30. Furushima K, Shimo-Onoda K, Maeda S, et al. Large-scale screening for candidate genes of ossification of the posterior longitudinal ligament of the spine. J Bone Miner Res. 2002;17:128–37.
31. Chon J, Hong JH, Kim J, et al. Association between BH3 interacting domain death agonist (BID) gene polymorphism and ossification of the posterior longitudinal ligament in Korean population. Mol Biol Rep. 2014;41:895–9.
32. Lim JJ, Shin DA, Jeon YJ, et al. Association of miR-146a, miR-149, miR-196a2, and miR-499 Polymorphisms with Ossification of the Posterior Longitudinal Ligament of the Cervical Spine. PLoS One. 2016;11:e0159756.
33. Xu C, Chen Y, Zhang H, et al. Integrated microRNA-mRNA analyses reveal OPLL specific microRNA regulatory network using high-throughput sequencing. Sci Rep. 2016;6:21580.
34. Heo WI, Park KY, Jin T, et al. Identification of novel candidate variants including COL6A6 polymorphisms in early-onset atopic dermatitis using whole-exome sequencing. BMC Med Genet. 2017;18:8.
35. Gari MA, Alkaff M, Aleshli HS, et al. Identification of novel genetic variations affecting osteoarthritis patients. BMC Med Genet. 2016;17:68.
36. Snegurova TA, Peddie LJ, Stone C, et al. Association of COL1A2, COL2A1 and COL9A1 and primary osteoarthritis in a founder population. Clin Genet. 2005;67:559–60.
37. Mustafa Z, Chapman K, Irven C, et al. Linkage analysis of candidate genes as susceptibility loci for osteoarthritis-suggestive linkage of COL9A1 to female hip osteoarthritis. Rheumatology (Oxford, England). 2000;39:299–306.
38. Wang CJ, Iida K, Egusa H, Hokugo A, Jewett A, Nishimura I. Trabecular bone deterioration in colla1a1+/- mice associated with enlarged osteoclasts adhered to collagen IX-deficient bone. J Bone Miner Res. 2008;23:21580.
39. Opolka A, Ratzinger S, Schubert T, et al. Collagen IX is indispensable for timely maturation of cartilage during fracture repair in mice. Matrix biology. 2007;26:85–95.
40. Dreier R, Opolka A, Grifka J, Bruckner P, Grassel S. Collagen IX-deficiency seriously compromises growth cartilage development in mice. Matrix biology. 2008;27:19–29.
41. Liu H, Zhao Z, Clarke RD, Gao J, Garrett IR, Margerrison EE. Enhanced tissue regeneration potential of juvenile articular cartilage. Am J Pathol. 2013;183:2658–67.
42. Lee CC, Xavlos AM, Ploegh HL. Accessory molecules for Toll-like receptors and their function. Nat Rev Immunol. 2012;12:168–79.
43. Noreen M, Arshad M. Association of TLR1, TLR2, TLR4, TLR6, and TIRAP polymorphisms with disease susceptibility. Immunol Res. 2015;62:234–52.
44. Qi H, Sun L, Wu X, et al. Toll-like receptor 1(TLR1) Gene SNP rs5743618 is associated with increased risk for tuberculosis in Han Chinese children. Tuberculosis (Edinburgh, Scotland). 2015;95:197–203.
45. Huh JE, Lee SY. IL-6 is produced by adipose-derived stromal cells and promotes osteogenesis. Biochim Biophys Acta. 2013;1833:2608–16.