Identification of novel biomarker candidates for immunohistochemical diagnosis to distinguish low-grade chondrosarcoma from enchondroma

Xianyin Lai[1,2]*, Shaoxiong Chen[3]*

[1] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[2] Department of Cellular & Integrative Physiology, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[3] Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA

Correspondence: Xianyin Lai, Ph.D.
Department of Biochemistry and Molecular Biology
Department of Cellular and Integrative Physiology
Indiana University School of Medicine
635 Barnhill Drive, Room 0044
Indianapolis, IN 46202 USA

Email: xlai@iu.edu
Phone: 317-278-0084

Fax: 317-274-4686

∗contributed equally to this work

∗Additional corresponding author: Dr. Shaoxiong Chen, E-mail: chen251@iupui.edu

Abbreviations: FFPE, formalin-fixed paraffin-embedded; IPA, Ingenuity Pathway Analysis; PIBAP, prioritization index of biomarker candidates for assay of plasma/serum specimens; PIBIT, prioritization index of biomarker candidates for immunohistochemistry on tissue specimens; TPP, Trans-Proteomic Pipeline

Keywords:
Biomarker / Enchondroma / Immunohistochemistry / Low-grade chondrosarcoma / Mass spectrometry

Total number of words: 7,182

Abstract:

Chondrosarcoma is the third most common primary bone cancer, requiring surgical resection. However, differentiation of low-grade chondrosarcoma (grade 1) from enchondroma that is benign and only requires regular follow-up is one of the most frequent diagnostic dilemmas facing orthopedic oncologists in clinical management. Although multiple techniques are applied to make the distinction, immunohistochemistry is an important ancillary technique, especially when a histopathological stain of specimen must be obtained in order to guarantee an accurate confirmation. Currently, no adequate immunohistochemical diagnostic protein biomarkers are available to distinguish low-grade chondrosarcoma from enchondroma. To discover novel protein biomarker candidates, a liquid chromatography-tandem mass spectrometry approach was applied to directly compare formalin-fixed, paraffin-embedded low-grade chondrosarcoma with enchondroma tissue samples. The proteomics analysis revealed 17 protein biomarker candidates. A principle was developed to prioritize the candidates using category and ranking. An algorithm, prioritization index of biomarker candidates for immunohistochemistry on tissue specimens (PIBIT), was developed to rank the candidates inside each category. Using the proteomics data and bioinformatics results, the PIBIT revealed periostin as a top candidate. Immunohistochemical staining of periostin in 23 low-grade chondrosarcoma and 31 enchondroma tissue specimens disclosed the specificity 87% and the sensitivity 70%.

1 Introduction

Chondrosarcoma is the third most common primary bone cancer [1, 2]. Based on cellularity, nuclear atypia, and pleomorphism, chondrosarcoma is histologically classified into three grades [3], including low-grade (grade 1), intermediate grade (grade 2), and high-grade (grade 3) [4]. Wide surgical resection is the most common treatment choice for chondrosarcoma [2]. Clinically, it is crucial to distinguish low-grade chondrosarcoma from its benign counterpart enchondroma, because enchondroma only requires regular follow-up [5].

Histologically, distinction of between a low-grade chondrosarcoma from a benign enchondroma relying solely on morphological features is difficult and even impossible in many cases for skilled pathologists because of their similar cytology, cellularity, and cartilaginous matrix [4, 6-8]. Therefore, correlative interpretation of histopathological features, x-ray imaging, computerized tomography, magnetic resonance imaging, clinical examination, and epidemiological information is currently used for making this distinction [5, 9]. Most often, however, a histopathological stain of specimen must be obtained in order to guarantee an accurate confirmation [5]. Therefore, immunohistochemistry is chosen to distinguish a low-grade chondrosarcoma from a benign enchondroma, because it has been proven to be one of the most important ancillary techniques.

Currently, numerous proteins have been immunohistochemically tested to distinguish low-grade chondrosarcoma from enchondroma. However, many of them were carried out on a small number of cases (< 10). Among the proteins tested on ≥ 10 cases, only

12 proteins, ADAM28, CCND1, FERMT2, FOS, ILK, JUN, JUNB, MAPK3, MAPK9, PARVB, PTH1R, and RUNX2 [10-14], possess both sensitivity and specificity ≥ 50%. Only 3 of the 12 proteins, FERMT2, FOS, and MAPK3, have both sensitivity and specificity ≥ 70%. It is unlikely for a biomarker to achieve 100% sensitivity and 100% specificity. An ideal biomarker needs to have at least 90% sensitivity and 90% or more specificity [15]. Obviously, there are not enough promising candidates in the validation at a larger scale to distinguish low-grade chondrosarcoma from enchondroma. Moreover, the majority of the tested proteins were simply chosen because either they were reported to involve signaling pathways in other tumors [13] or they play an important role in cell motility, growth, survival, and ultimately carcinogenesis [12]. They were not chosen from the direct comparison of low-grade chondrosarcoma with enchondroma.

To discover more meaningful protein candidates to meet the 90/90 standard, a global and high-throughput approach should be applied to directly compare low-grade chondrosarcoma with enchondroma. At present, genomics and proteomics have been used only in three and two publications to discover biomarker candidates for the differentiation, respectively [14, 16-19]. The only applied proteomic technique is two-dimensional electrophoresis that has multiple limitations in proteome analysis [20]. Not surprisingly, just a few biomarker candidates have been discovered so far. On the contrary, a shotgun proteomics approach provides better profiling of proteins because of its sensitivity and high-throughput capability [21]. Therefore, a liquid chromatography-tandem mass spectrometry (LC-MS/MS) approach should be pursued for the biomarker discovery.

The goals of the present study are to discover and validate protein biomarker candidates for immunohistochemical diagnosis to distinguish low-grade chondrosarcoma from enchondroma. In this study, formalin-fixed paraffin-embedded (FFPE) tissue samples of low-grade chondrosarcoma and enchondroma were analyzed using a quantitative LC-MS/MS approach. The FFPE samples were chosen because fresh-frozen tissues were not available. One of the discovered protein biomarker candidates was validated with immunohistochemistry. The results indicated that the entire workflow from discovery to validation performed remarkably well, significantly expanding the number of meaningful biomarker candidates reported in previous studies. Also, the results provided a very promising candidate for further validation at an even larger scale.

2 Materials and methods

2.1 Tissue blocks

This study was approved by the Indiana University Institutional Review Board (IRB # 1403932283). Two sets of FFPE tissue blocks of both low-grade chondrosarcoma and enchondroma were obtained from the Indiana University Health Pathology Laboratory. Case Set 1, used for the LC-MS/MS experiment, included five low-grade chondrosarcoma and five enchondroma samples from FFPE tissue blocks prepared between 2007 and 2013. Case Set 2, applied for immunohistochemical stain verification

of specific proteins, contained 23 low-grade chondrosarcoma and 31 enchondroma tissue specimens collected between 2000 and 2014.

2.2 Materials

Urea, DL-Dithiothreitol (DTT), triethylphosphine, iodoethanol, and ammonium bicarbonate (NH_4HCO_3) were purchased from Sigma-Aldrich (St. Louis, MO, USA). LC-MS grade water (H_2O), LC-MS grade 0.1% formic acid in acetonitrile (ACN), and 0.1% formic acid in water (H_2O) were purchased from Burdick & Jackson (Muskegon, MI, USA). Modified sequencing grade porcine trypsin was obtained from Princeton Separations (Freehold, NJ, USA).

2.3 Label-free analysis

2.3.1 Protein extraction from the FFPE samples

Protein extraction from the FFPE samples was performed according to an integrated procedure including deparaffinization, rehydration, cross-link reversion, and protein extraction [22]. Briefly, after an FFPE tissue block was cut into slices (5x2.5x1 mm, lengthxwidthxheight), visible paraffin was removed with a forceps. The slices were placed in a 2.0 mL collection tube and ten volumes (500 µL for 50 mg of tissue slices) of 100 mM NH_4HCO_3 at pH 8.0 were added to the tube. The sample was incubated at 99 °C for 30 min at 900 rpm on an Eppendorf Thermomixer to get the First Extraction. The

tissue slices were homogenized, generating a milk-like solution. The solution was incubated at 99 °C for 60 min at 900 rpm on the Eppendorf Thermomixer to obtain the Second Extraction. The Third Extraction was carried out with incubation at 99 °C for 60 min at 900 rpm. The three extractions were pooled and then stored at -80 °C until analysis. Protein concentration was determined by the Bradford Protein Assay using Bio-Rad protein assay dye reagent concentrate [23].

2.3.2 Protein reduction, alkylation, and digestion for LC-MS/MS

Protein reduction, alkylation, and digestion were carried out using a method previously published by the author [24]. Briefly, a 100 µg aliquot of protein sample was placed in a 2 mL tube and dried by SpeedVac. The sample was reconstituted by 100 µL of lysis buffer (8 M urea, 10 mM DTT solution freshly prepared) and then adjusted to 200 µL by adding 100 µL of water. 200 µL of the reduction/alkylation cocktail consisted of triethylphosphine and iodoethanol was added to the protein solution. The sample was incubated at 35°C for 60 min, dried by SpeedVac, and reconstituted with 100 µL of 100 mM NH_4HCO_3 at pH 8.0. A 150 µL aliquot of a 20 µg/mL trypsin solution was added to the sample and incubated at 35°C for 3 h, after which another 150 µL of trypsin was added, and the solution incubated at 35°C for 3 h.

2.3.3 LC-MS/MS analysis

The digested samples were analyzed using a Thermo Scientific Orbitrap Velos Pro hybrid ion trap-Orbitrap mass spectrometer coupled with a Surveyor autosampler and MS HPLC system (Thermo Scientific). Tryptic peptides were injected onto a C18 reversed phase column (TSKgel ODS-100V, 3 µm, 1.0 mm x 150 mm) at a flow rate of 50 µL/min. The mobile phases A and B were LC-MS grade H_2O with 0.1% formic acid and ACN with 0.1% formic acid, respectively. The gradient elution profile was as follows: 5% B for 6 min, 10-35% B for 156 min, 35-80% B for 10 min, and 80% B for 8 min. The data were collected in the “Data dependent MS/MS” mode of FT-IT (MS-MS/MS) with the ESI interface using normalized collision energy of 35% (CID). Dynamic exclusion settings were set to repeat count 1, repeat duration 30 s, exclusion duration 120 s, and exclusion mass width 10 ppm (low) and 10 ppm (high). Each sample was injected twice. Raw data are available through PeptideAtlas with identifier PASS00645 (http://www.peptideatlas.org/PASS/PASS00645).

2.3.4 Protein identification and quantification

The acquired data were searched against the UniProt protein sequence database of HUMAN (released on 02/19/2014) using X!Tandem algorithms in the Trans-Proteomic Pipeline (TPP, v. 4.6.3) (http://tools.proteomecenter.org/software.php). General parameters were set to: parent monoisotopic mass error set as 10 ppm, cleavage semi set as yes, missed cleavage sites set at 2, and static modification set as + 44.026215

Da on Cysteine. The searched peptides and proteins were validated by PeptideProphet [25] and ProteinProphet [26] in TPP. Only proteins and peptides with protein probability ≥ 0.9000 and peptide probability ≥ 0.8000 were reported. Protein quantification was performed using a label-free quantification software package, IdentiQuantXL™ [27]. Student's t-test was used for the statistical analysis.

2.4 Bioinformatics analysis

In order to assign their subcellular locations, involvement in diseases and disorders, and network connectivity, proteins in each category were individually submitted to QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, ww.qiagen.com/ingenuity) in which the Ingenuity Pathways Knowledge Base is used.

2.5 Immunohistochemistry

The immunohistochemical staining was performed at Indiana University Health Pathology Laboratory according to a published method [28]. The process was carried out with an automated staining instrument Dako Autostainer Plus (Dako, Carpinteria, CA, USA) and its compatible detection kit EnVision™ FLEX+ Mouse (LINKER) (K8022, Dako, Carpinteria, CA, USA). Briefly, FFPE tissues (4 μm thickness) on slides were dried in an oven at 60°C for 20 min and deparafinized using xylene and graded alcohol in water for three times with the Tissue-Tek DRS 2000 automated stainer. Antigen retrieval was carried out by heating sections in EDTA (pH8.0) for 15 min with a pressure

cooker. Slides were cooled and rinsed in running water and placed in Envision Flex Wash Buffer until they were ready to load on the Autostainer Plus. Endogenous peroxidase activity was inactivated by incubation in 3% of H_2O_2 for 5 min. A periostin mouse monoclonal antihuman antibody (TA804575, OriGene Technologies, Inc., Rockville, MD, USA) was used as a primary antibody at 1:1000 dilution and incubated for 30 min at room temperature. EnVision™ FLEX+Mouse (LINKER) was used for signal amplification. EnVision™ FLEX/HRP detection reagent was applied for the coupling reaction. EnVision™ FLEX DAB+ Chromogen was utilized twice to produce a crisp brown endproduct at the site of the target antigen. EnVision™ FLEX Hematoxylin was employed for counterstaining to provide a nuclear staining in blue. The slides were then unloaded from the AutoStainer Plus and placed in the Tissue-Tek DRS stainer to dehydrate through graded alcohols and xylene. Lung adenocarcinoma was chosen as a positive control. Negative control was performed on enchondroma and chondrosarcoma tissue without adding the primary antibody.

The staining of periostin was scored according to percentage of positive staining in the whole section for each case (0 = no positive staining; 1 = 1%-25% positive; 2 = 26%-50% positive; 3 = 51%-100% positive) and its intensity (0 = no staining, 1 = weak, 2 = moderate, 3 = strong staining), as described previously [10, 17]. Because periostin locates in extracellular space, only extracellular matrix staining of atypical chondrocytes and/or stroma closely associated with tumor is considered a positive result. Immunostaining in normal bone marrow spaces should be considered negative. To assess periostin's ability to distinguish low-grade chondrosarcoma from enchondroma, each stain was assigned either positive or negative for the clinical practice when the

sensitivity and specificity analyses were performed using XLSTAT (Addinsoft, New York, NY. http://www.xlstat.com/en/).

3 Results

3.1 Patient clinical information

The clinical data of diagnosis, location, gender, and age at diagnosis of the 64 patients (38 females and 26 males; age range, 14-95 years old; mean, 49 years) are summarized in Supplemental Table 1. The 10 patients in Set 1 ranged in age between 29 and 81 years old (mean, 49 years). Among the 23 low-grade chondrosarcoma and 31 enchondroma tissue specimens in Set 2, the patients in low-grade chondrosarcoma group ranged in age from 27 to 95 years old (mean, 59 years). The male to female ratio was 14/9. Patients in enchondroma group ranged in age from 14 to 71 years old (mean, 41 years). The male to female ratio was 9/22.

3.2 Protein identification and quantification

From the 20 injections, 489 protein groups (unique proteins) with a probability ≥ 0.9000 were identified by 2,661 peptides with a probability ≥ 0.8000. The complete list of identified proteins is available in the Supplemental Table 2, where proteins identified with completely identical peptides are placed into a single protein group. None of decoy proteins was included in the list according to the probability cut-offs. Therefore, the false discovery rate is 0%. Among the 489 proteins, 347 proteins were identified with at least two distinct peptides. To obtain more accurate quantification, multiple filters were

applied to eliminate unqualified peptides for protein quantification [27]. The complete list of quantified proteins is available in the Supplemental Table 3, including 400 protein groups quantified by 1,834 peptides. Among them, 270 proteins were quantified with at least two distinct peptides.

3.3 Bioinformatics analysis

Besides subcellular locations, involvement in diseases and disorders proteins, and network connectivity, IPA provided important information, such as canonical pathways, upstream regulators, regulator effects, etc. To affiliate prioritization of the biomarker candidates, only the highly related results are presented in this publication. The network connectivity and subcellular location are presented in Figures 1-3. Among the 17 proteins, 13 proteins exist in extracellular space. The involvement in diseases and disorders proteins of each protein is presented in Table 1. Almost all were involved with cancer except APOA4 and C1QB; 7 of the 17 proteins were related with connective tissue disorders.

3.4 Immunohistochemistry

Representative immunohistochemistry staining of periostin is presented in Figure 4 at 400× original magnification including negative control, positive control, enchondroma, and low-grade chondrosarcoma. The stain scores for percentage and intensity of periostin in 23 low-grade chondrosarcoma and 31 enchondroma tissue samples are presented in the Supplemental Table 1 in detail. Among the 23 low-grade chondrosarcomas, 14 tissue samples were positively stained, while only 4 of the 31

enchondroma tissue samples were positively stained. The majority of positive cases showed strong and focal staining. The immunohistochemical staining for each sample of low-grade chondrosarcoma is presented in the Supplemental Figure 1. Calculated with XLSTAT, the specificity was 87% with a 95% confidence interval of 70-95% and the sensitivity was 70% with a 95% confidence interval of 49-84%.

4 Discussion

4.1 Principle to prioritize candidate proteins

Once numerous proteins are identified and quantified in biomarker discovery phase, a challenging task is to determine which proteins should be chosen for further validation using alternative approaches. Simply by p value and/or fold change of a protein, the determination normally generates a long list of biomarker candidates. Therefore, it is a dilemma to decide which candidate has a better chance of success in validation and should be validated first since no unlimited source for validation is available. To resolve this problem, we categorize the candidates with different priorities according to couple the most critical factors and then rank the candidates inside each category based on multiple important factors.

The aim of this study was to discover diagnostic biomarker candidates for clinical immunohistochemistry. Thus, the factors considered in candidate selection should be different from other approaches, such as concentration measurement in blood using ELISA or MRM (multiple reaction monitoring). An ideal biomarker should be positively

stained in low-grade chondrosarcoma tissue while negatively stained in enchondroma tissue; also, proteins commonly existing in blood are considered with lower priority than proteins not commonly present in blood.

4.2 Factors in categorizing of protein candidates

The first factor is the difference of protein identification frequency between low-grade chondrosarcoma and enchondroma tissues. Each group had five samples and each sample was injected twice. Therefore, 10 injections were analyzed for each group. If a protein was identified from every injection, its frequency is 1.0; if a protein was not identified from any injection, its frequency is 0.0. The detailed frequency of each protein is included in the Supplemental Table 2. Twenty-six proteins with a frequency difference between low-grade chondrosarcoma and enchondroma ≥ 0.4 were considered potential biomarker candidates. Excluding protein isoforms, 17 unique proteins were considered biomarker candidates. The frequency difference was used to determine candidacy. Also, the frequency difference was used to rank proteins in each biomarker candidate category (described as below).

The second factor is whether a protein commonly existed in blood or skin. Because residual blood could be present in the tissue samples, proteins from blood rather than the tissue cells would be identified as well. Although minimizing contamination has been performed in mass spectrometry experiments, dead skin cells still are present in lab environments. Keratins from skin are very often detected in proteomics [29]. However, these proteins should not be easily excluded from the candidate list, since they are usually expressed in multiple types of tissues and these proteins may be from blood

and/or the analyzed tissues. Therefore, keratins and common plasma proteins in the candidate list were labeled as Category 3, where proteins have lower priority in the candidate list for further validation. Among the 17 proteins, pigment epithelium-derived factor, apolipoprotein A-IV, alpha-1B-glycoprotein, complement factor B, and complement C1q subcomponent subunit B are commonly detected in plasma or serum [30], Type II cytoskeletal 2 epidermal keratin is highly expressed in skin (http://www.genecards.org/cgi-bin/carddisp.pl?gene=KRT2#expression) and also highly detectable in plasma or serum [30].

The third concern is the protein identification frequency in enchondroma tissue. As mentioned above, our goal was to discover proteins positively stained in low-grade chondrosarcoma tissues while negatively stained in enchondroma tissues. If a protein was not identified in any of the enchondroma injections, it was considered a stronger candidate than a protein identified in some of the enchondroma injections. Based on this criterion, proteins with a frequency > 0 are put into Category 2, in which proteins have low priority for further validation. Other proteins are put into Category 1, in which proteins have high priority among the entire list.

4.3 Criteria in ranking of candidate proteins inside categories

Besides the frequency difference described above, the first criterion is the consistency between frequency difference and fold change. To further validate candidates discovered by proteomics, immunohistochemistry is applied. Because these analytical approaches have different sensitivity and reproducibility, selection of greater fold change values is more likely to insure successful biomarker detection across different

approaches. Fold change alone would serve well to rank the candidates when a protein is present in high abundance in one group (disease) but in low abundance or undetectable in the other group (control). When a protein has extremely low abundance in both groups, it is very often identified only in some injections of both groups or only identified in the disease group without detection in the control groups, potentially implying large fold difference between the two groups. However, the fold change is often not as large as expected or is sometimes very small due to detection limitation. Without using alternative and targeting analyses, such as MRM or ELISA, it is not possible to determine the real concentration. Under these circumstances, it is difficult to solely use fold change to determine the priority. On the contrary, if a protein's fold difference is consistent with its frequency difference, this protein can be easily assigned with higher priority than other proteins whose fold difference are not consistent with their frequency difference. The consistency is expressed as fold change × frequency difference. Higher score indicates higher priority.

The second concern is the number of peptides used for protein identification. The more peptides used, the more confident the identification. Thus, a protein identified by multiple peptides has higher priority than a protein identified by a single peptide. Based on this criterion, proteins are further ranked inside the categories.

The third factor is protein subcellular location. From a clinical immunohistochemistry aspect, nuclear proteins are easier to achieve a better immunohistochemistry stain than cytoplasm proteins, cytoplasm proteins are easier than plasma membrane proteins, and plasma membrane proteins are easier than extracellular proteins. Therefore, the assigned priority in each category is nuclear proteins > cytoplasm proteins > plasma

membrane proteins > extracellular proteins. In order to assign their cellular locations, proteins in each category were submitted to IPA. The software revealed that only AEBP1 locates in nucleus, PPIA in cytoplasm, MFI2 in plasma membrane, and all other proteins in extracellular space; KRT2 was the exception, since it was not assigned a specific location. To numerically show the importance and difference of each protein location, 10, 7, 4, 3, 1, and 0 are assigned to nucleus, cytoplasm, plasma membrane, extracellular space, and other (unknown or unassigned), respectively.

The fourth factor is the protein's relationship with targeted diseases and disorders. Because we were looking for biomarkers for chondrosarcoma, if a protein related to connective tissue disorders or cancer, it ranked higher than other proteins. A score was assigned to each related disease or disorder for the final calculation of priority. Proteins involved in connective tissue disorders were assigned 6 points; proteins related to cancer, 3 points; and proteins involved in both disorders, 9 points. IPA analysis was applied to identify the diseases or disorders in which each protein was involved. Among the 17 proteins, 15 were involved with cancer, except APOA4 and C1QB. Seven were related to connective tissue disorders (POSTN, TGFBI, SMOC2, SERPINF1, A1BG, C1QB, and CFB).

The fifth concern is protein connectivity in a network. To find potential links among the candidate proteins, IPA analysis is applied. IPA Networks are generated based on protein connectivity with other proteins. The more connected a protein is, the more important it is and the more influence it has. Therefore, the higher priority a protein is assigned, if a protein has more interconnection with other proteins in a network. When proteins in each category were submitted to IPA, a network was generated for each

category (Figures 1 - 3), indicating direct and indirect connectivity among some of the submitted proteins. The number of a protein's connections in a network were counted. The more a protein has other proteins connected, the higher priority is assigned to the protein.

4.4 Prioritization index of biomarker candidates for immunohistochemistry on tissue specimens (PIBIT)

Multiple factors have been taken into account to rank biomarker candidates selected for immunohistochemical verification. One final score is needed to combine all the six factors. Four of these factors are numeric; i.e., frequency difference, consistency between frequency difference and fold change, number of peptides used for protein identification, and protein connectivity in a network. Protein subcellular location and the protein's involvement in targeted diseases and disorders are not, but subsequently assigned numbers for the final score calculation. Finally, the individual score of each factor is combined in a formula and a signal score is reported to rank proteins in each category.

However, their weight in the calculation should be different. The frequency difference is recorded as a fraction. To normalize the importance of each factor, the weight of frequency difference is assigned as 10. The weight of the consistency between frequency difference and fold change is given 10 as well. Therefore, the final score PIBIT was calculated using the following equation:

$$PIBIT = \sum_{i=0}^{n} (Fi)$$

Where PIBIT = prioritization index of biomarker candidates for immunohistochemistry on tissue specimens, F_0 = frequency difference × 10, F_1 = consistency between frequency difference and fold change (fold change × frequency difference) × 10, F_2 = number of peptides, F_3 = subcellular location, F_4 = involvement in targeted diseases and disorders, and F_5 = connectivity in a network. According to the six factors, the PIBIT of each protein was calculated using the equation. All the factors and detail scores are listed in Table 1. Proteins are prioritized according to their PIBIT within each category.

The formula can be slightly modified to rank biomarker candidates for an assay when plasma/serum samples are used. The only change is the scoring of protein subcellular location. When immunohistochemistry is used to verify biomarker candidates for tissue samples, nuclear proteins have the highest score of 10 and extracellular proteins have a score of 1. On the contrary, when MRM or ELISA is applied to verify biomarker candidates for plasma/serum samples, 1, 3, 4, 7, 10, and 0 are assigned to nuclear proteins, cytoplasm proteins, plasma membrane proteins, extracellular proteins, and other (unknown or unassigned) proteins, respectively. To distinguish the PIBIT, this one is named as PIBAP, representing prioritization index of biomarker candidates for assay of plasma/serum specimens.

Furthermore, this idea can be more extensively applied to other projects whose factors are not identical to the factors discussed in this study, such as sensitivity, specificity, the area under the receiver operating characteristic (ROC) curve (AUC), etc. Basically, each factor and its weight have to be considered and then summed together. The single final score is simple but comprehensive, facilitating the priority determination.

4.5 Selection of periostin for immunohistochemistry verification

Periostin's discovery history, structure, isoforms, expression, functions, role in carcinogenesis and tumor progression-driving, and usage as a prognostic marker and novel therapeutic target have been summarized in a recent publication [31]. Periostin, also called osteoblast-specific factor 2, is an extracellular matrix protein involved in osteology, tissue repair, oncology, cardiovascular and respiratory systems, and in various inflammatory settings [32]. Periostin plays an important role in tumor progression in various types of cancer, including breast cancer, bladder cancer, colon cancer, non-small cell lung carcinoma, head and neck cancer, ovarian cancer, pancreatic ductal adenocarcinoma, melanoma, gastric cancer, oral squamous cell carcinoma, thymoma, and neuroblastoma [31, 33]. Noticeably, the upregulation of periostin in prostate, renal, and penile cancer was usually associated with a more aggressive tumor behavior and advanced stage, while it is weakly expressed in bladder cancer tissues [31]. Generally, it is known that its overexpression in cancer indicates the most malignant phenotypes and the poorest outcomes [34]. Periostin is preferentially expressed in the periosteum that covers a large majority of bones and is responsible for changes in bone diameter and cortical thickness [33]. It is expressed at a high level during embryogenesis and bone growth [33]. Periostin activates the Akt/PKB- and FAK-mediated signaling pathways, leading to increased cell survival, angiogenesis, invasion, metastasis, and epithelial-mesenchymal transition of carcinoma cells [35]. Due to its roles in tumor development, periostin has been speculated as a therapeutic and diagnostic target for cancer [36]. According to the proteomics results and the literature,

periostin is a highly preferred biomarker candidate for further validation. Therefore, periostin was chosen as the first one for verification by immunohistochemistry.

4.6 Other proteins in the category 1

There are six other proteins (TGFBI, AEBP1, PCOLCE, SMOC2, MFI2, CPXM2) listed in the Category 1. TGFBI mediates cell adhesion to extracellular proteins such as collagen, fibronectin and laminins through integrin binding. It is overexpressed in several solid tumors including colon, pancreas, and kidney [37]. AEBP1 has been reported to be upregulated in stroma of mammary tumors and breast cancer cells [38]. PCOLCE stimulates procollagen processing by procollagen C-proteinases and is involved in tumor growth, neurodegenerative diseases, and angiogenesis [39]. SMOC2 enhances the angiogenic effect of basic fibroblast growth factor and vascular endothelial growth factor, mediates mitogenesis in mouse fibroblasts, interacts with vitronectin and cell surface receptors of the integrin family. Moreover, it affects the migration of keratinocytes, the process of metastasis, and pulmonary function [40]. Melanotransferrin is expressed at low levels in normal adult tissues, but at high levels in melanoma tumors, other cancers, and foetal tissues [41]. CPXM2 is a member of the metallocarboxypeptidase gene family and identified from its homology with carboxypeptidase E [42]. It is upregulated in fetal growth restriction [43]. All six were identified with multiple peptides and involved in cancer. TGFBI and SMOC2 are related with connective tissue disorders. This information suggests that the six proteins are good candidates for further biomarker validation.

4.7 Proteins not in the category 1

There are 10 more proteins not in the Category 1 of biomarker candidates. Some have a high PIBIT score, such as decorin and pigment epithelium-derived factor. Decorin is a component of connective tissue and locates at cellular or pericellular matrix. It has become a focus in various areas of cancer research [44]. Decorin has been involved in various biological processes, such as collagen fibrillogenesis, wound healing, myogenesis, stem cell biology, and fibrosis [45]. Pigment epithelium-derived factor is a serine protease inhibitor and has been detected in multiple tissues including brain, spinal cord, liver, bone, eye, heart, lung and plasma [46]. It has been reported to be a potent angiogenic inhibitor to prevent angiogenesis and metastasis, induce tumor cell apoptosis, and prevent cancer cell growth in a range of cancers, such as osteosarcoma, chondrosarcoma, lung, breast, and prostate cancer [47]. Although these proteins are not in the Category 1 of prioritization, they are still valuable biomarker candidates. The category and prioritization are only intended to validate the protein candidates in a quicker and less expensive way by validating candidates who have the best chance of success in validation.

The objective of this research was to discover diagnostic biomarker candidates for clinical immunohistochemistry to distinguish benign (enchondroma) from low-grade chondrosarcoma. The diagnosis is one of the most frequent diagnostic dilemmas in clinical management facing orthopedic oncologists. As a result of the comprehensive mass spectrometry-based analysis used in this study, we have identified 17 biomarker candidates. The best candidate, periostin, was verified with immunohistochemistry on 23 low-grade chondrosarcoma and 31 enchondroma tissue samples, indicating

sensitivity 70% and specificity 87%. Although it does not yet meet the 90/90 standard, better reliable diagnostic immunohistochemistry markers have not yet been identified and verified. Validation of other biomarker candidates in a follow-up study will facilitate establishment of reliable immunohistochemical methods.

The authors have declared no conflict of interest.

5 References

[1] Logie CI, Walker EA, Forsberg JA, Potter BK, Murphey MD. Chondrosarcoma: A Diagnostic Imager's Guide to Decision Making and Patient Management. Semin Musculoskelet Radiol. 2013;17:101-15.

[2] van Driel M, van Leeuwen JP. Cancer and bone: A complex complex. Arch Biochem Biophys. 2014;561C:159-66.

[3] Cintra FF, Etchebehere M, Goncalves JC, Cassone AE, Amstalden EM. Analysis of angiogenic factors and cyclooxygenase-2 expression in cartilaginous tumors- clinical and histological correlation. Clinics (Sao Paulo). 2011;66:1591-6.

[4] Choi BB, Jee WH, Sunwoo HJ, Cho JH, Kim JY, Chun KA, et al. MR differentiation of low-grade chondrosarcoma from enchondroma. Clin Imaging. 2013;37:542-7.

[5] Ferrer-Santacreu EM, Ortiz-Cruz EJ, Gonzalez-Lopez JM, Perez Fernandez E. Enchondroma versus Low-Grade Chondrosarcoma in Appendicular Skeleton: Clinical and Radiological Criteria. J Oncol. 2012;2012:437958.

[6] Zenmyo M, Tanimoto A, Sakakima H, Yokouchi M, Nagano S, Yamamoto T, et al. Gadd45beta expression in chondrosarcoma: a pilot study for diagnostic and biological implications in histological grading. Diagn Pathol. 2010;5:69.

[7] Takeuchi A, Yamamoto Y, Tsuneyama K, Cheng C, Yonekura H, Watanabe T, et al. Endogenous secretory receptor for advanced glycation endproducts as a novel prognostic marker in chondrosarcoma. Cancer. 2007;109:2532-40.

[8] Ryzewicz M, Manaster BJ, Naar E, Lindeque B. Low-grade cartilage tumors: diagnosis and treatment. Orthopedics. 2007;30:35-46; quiz 7-8.

[9] Kaya GC, Demir Y, Ozkal S, Sengoz T, Manisali M, Baran O, et al. Tumor grade-related thallium-201 uptake in chondrosarcomas. Ann Nucl Med. 2010;24:279-86.

[10] Matsuura S, Oda Y, Matono H, Izumi T, Yamamoto H, Tamiya S, et al. Overexpression of A disintegrin and metalloproteinase 28 is correlated with high histologic grade in conventional chondrosarcoma. Hum Pathol. 2010;41:343-51.

[11] Rozeman LB, Hameetman L, Cleton-Jansen AM, Taminiau AH, Hogendoorn PC, Bovee JV. Absence of IHH and retention of PTHrP signalling in enchondromas and central chondrosarcomas. J Pathol. 2005;205:476-82.

[12] Papachristou DJ, Gkretsi V, Rao UN, Papachristou GI, Papaefthymiou OA, Basdra EK, et al. Expression of integrin-linked kinase and its binding partners in chondrosarcoma: association with prognostic significance. Eur J Cancer. 2008;44:2518-25.

[13] Papachristou DJ, Papachristou GI, Papaefthimiou OA, Agnantis NJ, Basdra EK, Papavassiliou AG. The MAPK-AP-1/-Runx2 signalling axes are implicated in chondrosarcoma pathobiology either independently or via up-regulation of VEGF. Histopathology. 2005;47:565-74.
[14] Rozeman LB, Hameetman L, van Wezel T, Taminiau AH, Cleton-Jansen AM, Hogendoorn PC, et al. cDNA expression profiling of chondrosarcomas: Ollier disease resembles solitary tumours and alteration in genes coding for components of energy metabolism occurs with increasing grade. J Pathol. 2005;207:61-71.
[15] Brower V. Biomarkers: Portents of malignancy. Nature. 2011;471:S19-21.
[16] Pansuriya TC, van Eijk R, d'Adamo P, van Ruler MA, Kuijjer ML, Oosting J, et al. Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome. Nat Genet. 2011;43:1256-61.
[17] Pansuriya TC, Oosting J, Krenacs T, Taminiau AH, Verdegaal SH, Sangiorgi L, et al. Genome-wide analysis of Ollier disease: Is it all in the genes? Orphanet J Rare Dis. 2011;6:2.
[18] Vincourt JB, Etienne S, Cottet J, Delaunay C, Malanda CB, Lionneton F, et al. C-propeptides of procollagens I alpha 1 and II that differentially accumulate in enchondromas versus chondrosarcomas regulate tumor cell survival and migration. Cancer Res. 2010;70:4739-48.
[19] Vincourt JB, Vignaud JM, Lionneton F, Sirveaux F, Kawaki H, Marchal S, et al. Increased expression of matrilin-3 not only in osteoarthritic articular cartilage but also in cartilage-forming tumors, and down-regulation of SOX9 via epidermal growth factor domain 1-dependent signaling. Arthritis Rheum. 2008;58:2798-808.
[20] Magdeldin S, Enany S, Yoshida Y, Xu B, Zhang Y, Zureena Z, et al. Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. Clin Proteomics. 2014;11:16.
[21] Abdallah C, Dumas-Gaudot E, Renaut J, Sergeant K. Gel-based and gel-free quantitative proteomics approaches at a glance. Int J Plant Genomics. 2012;2012:494572.
[22] Lai X, Schneider BP. Integrated and convenient procedure for protein extraction from formalin-fixed, paraffin-embedded tissues for LC-MS/MS analysis. Proteomics. 2014.
[23] Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72:248-54.
[24] Lai XY, Bacalla RL, Blazer-Yost BL, Hong D, Mason SB, Witzmann FA. Characterization of the renal cyst fluid proteome in autosomal dominant polycystic kidney disease (ADPKD) patients. Proteomics Clin Appl. 2008;2:1140-52.
[25] Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem. 2002;74:5383-92.
[26] Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem. 2003;75:4646-58.
[27] Lai X, Wang L, Tang H, Witzmann FA. A novel alignment method and multiple filters for exclusion of unqualified peptides to enhance label-free quantification using peptide intensity in LC-MS/MS. J Proteome Res. 2011;10:4799-812.
[28] Idrees MT, Williamson SR, Kieffer TW, Cheng L. The role of OCT4 immunohistochemistry in evaluation of retroperitoneal lymph node dissections: a pilot study. Mod Pathol. 2013;26:1613-9.
[29] Hodge K, Have ST, Hutton L, Lamond AI. Cleaning up the masses: exclusion lists to reduce contamination with HPLC-MS/MS. J Proteomics. 2013;88:92-103.
[30] Lai X, Liangpunsakul S, Crabb DW, Ringham HN, Witzmann FA. A proteomic workflow for discovery of serum carrier protein-bound biomarker candidates of alcohol abuse using LC-MS/MS. Electrophoresis. 2009;30:2207-14.
[31] Nuzzo PV, Buzzatti G, Ricci F, Rubagotti A, Argellati F, Zinoli L, et al. Periostin: A Novel Prognostic and Therapeutic Target For Genitourinary Cancer? Clinical genitourinary cancer. 2014.

[32] Conway SJ, Izuhara K, Kudo Y, Litvin J, Markwald R, Ouyang G, et al. The role of periostin in tissue remodeling across health and disease. Cell Mol Life Sci. 2014;71:1279-88.
[33] Merle B, Garnero P. The multiple facets of periostin in bone metabolism. Osteoporos Int. 2012;23:1199-212.
[34] Ruan K, Bao S, Ouyang G. The multifaceted role of periostin in tumorigenesis. Cell Mol Life Sci. 2009;66:2219-30.
[35] Morra L, Moch H. Periostin expression and epithelial-mesenchymal transition in cancer: a review and an update. Virchows Arch. 2011;459:465-75.
[36] Kudo Y, Siriwardena BS, Hatano H, Ogawa I, Takata T. Periostin: novel diagnostic and therapeutic target for cancer. Histol Histopathol. 2007;22:1167-74.
[37] Pajares MJ, Agorreta J, Salvo E, Behrens C, Wistuba II, Montuenga LM, et al. TGFBI expression is an independent predictor of survival in adjuvant-treated lung squamous cell carcinoma patients. Br J Cancer. 2014;110:1545-51.
[38] Grigoriadis A, Mackay A, Reis JS, Steele D, Iseli C, Stevenson BJ, et al. Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. Breast Cancer Res. 2006;8:R56.
[39] Salza R, Peysselon F, Chautard E, Faye C, Moschcovich L, Weiss T, et al. Extended interaction network of procollagen C-proteinase enhancer-1 in the extracellular matrix. Biochem J. 2014;457:137-49.
[40] Mommaerts H, Esguerra CV, Hartmann U, Luyten FP, Tylzanowski P. Smoc2 modulates embryonic myelopoiesis during zebrafish development. Dev Dyn. 2014.
[41] Suryo Rahmanto Y, Bal S, Loh KH, Yu Y, Richardson DR. Melanotransferrin: search for a function. Biochim Biophys Acta. 2012;1820:237-43.
[42] Xin X, Day R, Dong W, Lei Y, Fricker LD. Identification of mouse CPX-2, a novel member of the metallocarboxypeptidase gene family: cDNA cloning, mRNA distribution, and protein expression and characterization. DNA Cell Biol. 1998;17:897-909.
[43] Sabri A, Lai D, D'Silva A, Seeho S, Kaur J, Ng C, et al. Differential placental gene expression in term pregnancies affected by fetal growth restriction and macrosomia. Fetal Diagn Ther. 2014;36:173-80.
[44] Stander M, Naumann U, Wick W, Weller M. Transforming growth factor-beta and p-21: multiple molecular targets of decorin-mediated suppression of neoplastic growth. Cell Tissue Res. 1999;296:221-7.
[45] Neill T, Torres A, Buraschi S, Owens RT, Hoek JB, Baffa R, et al. Decorin induces mitophagy in breast carcinoma cells via peroxisome proliferator-activated receptor gamma coactivator-1alpha (PGC-1alpha) and mitostatin. J Biol Chem. 2014;289:4952-68.
[46] Chen L, DiPietro LA. Production and function of pigment epithelium-derived factor in isolated skin keratinocytes. Exp Dermatol. 2014;23:436-8.
[47] Alcantara MB, Dass CR. Pigment epithelium-derived factor as a natural matrix metalloproteinase inhibitor: a comparison with classical matrix metalloproteinase inhibitors used for cancer treatment. J Pharm Pharmacol. 2014;66:895-902.

Figure Legends

Figure 1. All seven proteins in Category 1 submitted to IPA are included in the generated network. Top diseases and functions involved in the network include tissue morphology, connective tissue disorders, and hereditary disorder. One protein locates in nucleus, one exists in plasma membrane, and all others are present in extracelluar space. POSTIN has the most network connectivity.

Figure 2. Among the five proteins in Category 2 analyzed with IPA, four proteins are included in the network. Cellular movement, hematological system development and function, and immune cell trafficking are the top three involved diseases and functions. Only PPIA locates in cytoplasm. The other three proteins all exist in extracelluar space. DCN has the most network connectivity.

Figure 3. All six proteins in Category 3 submitted to IPA are included the generated network. The top three 3 diseases and functions involved in the network are lipid metabolism, small molecule biochemistry, and molecular transport. All six proteins all are present in extracellular space. APOA4 has the most network connectivity.

Figure 4. Immunohistochemical staining for periostin of negative control, positive control, enchondroma, and low-grade chondrosarcoma (original magnification ×400). Periostin locates in extralcellular space. Therefore, the matrix staining pattern is observed.

Table

Table 1. The ranking and categorizing of protein biomarker candidates with a frequency difference ≥ 0.4.

									SL		ITDD				
R	C	ID	GN	PFE	PFC	**FD**	**FC**	**#P**	NM	**SC**	CS	CTS	**TS**	**NC**	**PIBIT**
1	1	B1ALD8	POSTN	0	0.8	**0.8**	**8.3**	**5**	E	**1**	3	6	**9**	**14**	**103.4**
2	1	G8JLA8	TGFBI	0	0.6	**0.6**	**2.7**	**8**	E	**1**	3	6	**9**	**5**	**45.2**
3	1	Q8IUX7	AEBP1	0	0.6	**0.6**	**1.1**	**6**	N	**10**	3	0	**3**	**5**	**36.6**
4	1	Q15113	PCOLCE	0	0.6	**0.6**	**1.6**	**5**	E	**1**	3	0	**3**	**8**	**32.6**
5	1	Q9H3U7	SMOC2	0	0.4	**0.4**	**2.8**	**2**	E	**1**	3	6	**9**	**2**	**29.2**
6	1	P08582	MFI2	0	0.4	**0.4**	**2.3**	**5**	P	**4**	3	0	**3**	**1**	**26.2**
7	1	Q8N436	CPXM2	0	0.6	**0.6**	**1.6**	**4**	E	**1**	3	0	**3**	**1**	**24.6**
8	2	P07585	DCN	0.5	1.0	**0.5**	**2.4**	**10**	E	**1**	3	0	**3**	**19**	**50.0**
9	2	P62937	PPIA	0.2	0.6	**0.4**	**-1.0**	**5**	C	**7**	3	0	**3**	**8**	**23.0**
10	2	P07093	SERPINE2	0.3	1.0	**0.7**	**-1.1**	**10**	E	**1**	3	0	**3**	**9**	**22.3**
11	2	Q16674	MIA	0.3	0.7	**0.4**	**1.0**	**1**	E	**1**	3	0	**3**	**2**	**15.0**
12	3	P36955	SERPINF1	0.4	0.8	**0.4**	**1.6**	**12**	E	**1**	3	6	**9**	**7**	**39.4**
13	3	P06727	APOA4	0.2	0.7	**0.5**	**1.1**	**6**	E	**1**	0	0	**0**	**15**	**32.5**
14	3	P04217	A1BG	0	0.4	**0.4**	**1.8**	**5**	E	**1**	3	6	**9**	**2**	**28.2**
15	3	B4E1Z4	CFB	0	0.4	**0.4**	**1.2**	**3**	E	**1**	3	6	**9**	**6**	**27.8**
16	3	D6R934	C1QB	0.4	0.8	**0.4**	**1.0**	**7**	E	**1**	0	6	**6**	**3**	**25.0**
17	3	P35908	KRT2	0.5	0.9	**0.4**	**1.3**	**5**	O	**0**	3	0	**3**	**1**	**18.2**

Note: R, Rank; C, Category; ID, Protein ID; GN, Gene Name; PFE, Protein identification frequency in enchondroma tissues; PFC, Protein identification frequency in low-grade chondrosarcoma tissues; FD, Difference of protein identification frequency between low-grade chondrosarcoma and enchondroma tissues. FC, Fold change; #P, Number of peptides; SL, Subcellular location; NM, Name of the subcellular location, including N for nucleus, C for cytoplasma, P for plasma membrane, E for extracellular space, and O for other; SC, Score of the subcellular location; ITDD, Involvement in targeted diseases and disorders; CS, Cancer score of the involvement in targeted diseases and disorders; CTS, Connective tissue disorder score of the involvement in targeted diseases and disorders; TS, Total score of the involvement in targeted diseases and disorders; NC, Network connectivity; and PIBIT, Prioritization index of biomarker candidates for immunohistochemistry on tissue specimens. The PIBIT is calculated using a formula of FD × 10 + FD × FC × 10 + #P + SC + TS + NC.