PROF. HARALD MISCHAK (Orcid ID : 0000-0003-0323-0306)

Article type : Letter

To the Editor of the European Journal of Clinical Investigation

Re-analysis of “Peptidomic analysis of cartilage and subchondral bone in OA patients”

Dear Editor,

In the May 2019 issue of *European Journal of Clinical Investigation*, Gatenholm *et al.* reported on *"a method for directly analyzing osteochondral samples straight out of the operating room without cell culturing, thereby enabling identification of potential peptide biomarkers to better understand the mechanisms involved in the development of osteoarthritis (OA) and pain"*[1]. Six Samples from patients were investigated, 3 from wounded (WO) and 3 from macroscopically unwounded zones (UOA) of the femur condyle, manifesting OA based on total knee arthroplasty (TKA). Using peptidomics and Tandem Mass Tag (TMT) labeling, the authors in their study reported the identification of 6296 endogenous peptides derived from 915 proteins (889 protein groups) across samples. Out of the total number of obtained peptide sequences, 601 peptides carried TMT labeling and 462 endogenous peptides could be matched and identified in the human database, as provided in Table S1. After performing statistical analysis, 566 peptides differing ($p \leq 0.1$) in unwounded (UOA) and wounded zones from cartilage and subchondral bone in OA patients were identified. However, all significance was lost upon applying multiple testing adjustments.

The content and the results in the manuscript appear to be of high importance, especially towards the identification of peptide biomarkers for diagnosis of osteoarthritis (OA) patients. We aimed at exploiting these published data, specifically integrate them with in house generated datasets acquired in our laboratory[2], as a basis to identify urine peptides significantly deregulated in osteoarthritis.

Towards that end, re-evaluation of the reported findings was initiated. In total, 18 raw files from 6 patients analyzed in triplicates were provided by the authors upon request. According to the manuscript, the samples were labeled with TMT6-plex, with reporter 126, 127, 128 corresponding to healthy cartilage, and reporter 129, 130, 131 to OA tissue. We evaluated the data using Proteome Discoverer 1.4 by applying the same parameters as reported in the manuscript. Using TMT6-plex as fixed modification in the initial data evaluation resulted in a very low number of identifications, likely due to very low labeling efficiency. Therefore, TMT6-plex and oxidation of methionine and proline (to account for hydroxyproline, a frequent modification in collagen) were set as variable modifications. No enzyme specificity was selected, precursor mass was set to 600 - 5000 Da with a minimum peak count of 10. Percolator algorithm was used for the calculation of the FDR level for the peptide spectrum matches (PSMs). Peptides identified with high and medium confidence (FDR <5%.) were considered for further analysis.

In total, 16 out of the 18 raw files were evaluated (2 files could not be processed successfully), resulting in identified of 3073 endogenous peptide sequences. Of these, only 1017 peptide sequences were identified in at least 3 of the 16 datasets, 2056 peptide sequences being identified in less than 3 datasets (Supplementary table 1). All sequences identified were compared to those reported in the manuscript. Overall, 2055 common peptide sequences could be found. Surprisingly, the majority of the peptide sequences reported in the manuscript (4241) could not being confirmed. Since each patient sample was analyzed in triplicate, we next assessed the number of consistent peptides identified in the entire dataset (detectable in >70% of all analyses) or per sample (consistence was defined as detection in at least 2 out of the 3 experiments) . Only 82 peptides were detected in at least 70% (12 or more) of all datasets. A total of 664 endogenous peptides were identified in at least 2 experiments from one sample

(between 40 and 450 peptides per sample, see supplementary table 1). Overall, these data indicate very low consistency.

Even though identical parameters for data evaluation were applied, based on the methodology reported in the original manuscript and adjusted based on further information obtained, the results that were obtained, were highly different, most of the reported findings could not be reproduced. When trying to understand the analytical causes for this discrepancy, multiple different settings were tested which revealed some apparent errors in the methods as reported in the manuscript, as listed below:

1. Although the authors indicate that the samples were first TMT labeled, then combined and the combined sample was subsequently analyzed (see Figure 1), it seems the samples were in fact analyzed separately, likely due to insufficient TMT labeling.
2. The authors described application of the TMT label as fixed modification. However, after evaluating and subsequent discussion with one of the co-authors it became evident that the TMT label has been set as "variable modification" in the data evaluation methods. These changes do have a huge impact on the results returned.

Although this is a small pilot study, which is mentioned as limitation together with issues of small sample amount and difficulties in labeling and quantification of the OA peptides and issues with peptide identification; still it is necessary to be accurate when describing the methods used and the analytical protocols, to allow data replication, re-use and further exploration.

We thus feel it is of outmost importance to share these results for several highly relevant reasons:

1) The results indicate the enormous importance of accurately describing the methods used. The application of methods different from the one described in the manuscript, as apparently was the case here, does have a huge impact on the results obtained and prevents any effort of reproducing the results; basically no similarity exists between the results obtained using the method described in the manuscript and using the method that likely has actually been employed by the authors.

2) The data available also indicate that application of two different software solutions, both using identical data and parameters, and both intended for the same purpose, appears to result in very different outputs (here: lists of peptides). It is unknown which of the two are correct, but it is certain that they cannot both be correct. If only the output, ie. the results after data interpretation by the software is being reported (the list of peptides), possibly even with additional impact by the authors of the study (e.g. selective reporting of only some features), then reproducibility may be completely lost. The same dataset, as a result of applying different software solutions and interpretation by the authors, may give completely different final results. This is very worrying and an easy solution is not evident. Data evaluation using different software solutions is has been shown to be helpful[3], but may not always be practical.

3) This finding further underlines the enormous importance of mandatory sharing of the raw (machine) data, which the authors did. On multiple occasions unfortunately scientists refuse sharing of the raw data, arguing that data protection issues prevent sharing (e.g. [4]). In light of the findings reported here, data sharing should be mandatory for any publication in respected scientific journals. Accepting refusal of data sharing results in potentially accepting major errors in publications, even scientific misconduct as the results published cannot be reproduced by anybody, due to the absence of data. We, as a scientific community, must not support the abuse of data protection to cover up questionable scientific conduct. Even though not mandatory, the authors here did share the raw data and, in this way, enabled uncovering these highly relevant issues. The claims made in the manuscript can probably not be upheld any longer or need at least major revision, however, by sharing all data the authors demonstrated commendable scientific conduct.

Sincerely

Martin Pejchinovski[1] and Harald Mischak[1,2*]

[1]mosaiques-diagnostics GmbH, Rotenburger Str. 20 D-30659 Hannover, GERMANY

[2]University of Glasgow, 126 University Avenue, Glasgow, G12 8TA

*correspondence to:

Prof. Dr.Dr. Harald Mischak

e-mail: mischak@mosaiques-diagnostics.com

Reference List

1. Gatenholm B, Gobom J, Skillback T, Blennow K, Zetterberg H, Brittberg M. Peptidomic analysis of cartilage and subchondral bone in OA patients. *Eur J Clin Invest* 2019; 49: e13082

2. Latosinska A, Siwy J, Mischak H, Frantzi M. Peptidomics and proteomics based on CE-MS as a robust tool in clinical application: The past, the present, and the future. *Electrophoresis* 2019; in press:

3. Latosinska A, Mokou M, Makridakis M *et al*. Proteomics analysis of bladder cancer invasion: Targeting EIF3D for therapeutic intervention. *Oncotarget* 2017;

4. Christensson A, Ash JA, DeLisle RK *et al*. The Impact of the Glomerular Filtration Rate on the Human Plasma Proteome. *Proteomics Clin Appl* 2018; 12: e1700067