

Identification of the missing protein Hyaluronan Synthase 1 in human mesenchymal stem cells derived from adipose tissue or umbilical cord

Luis Felipe Clemente¹, María Luisa Hernández¹, Antonio Ramos-Fernández², Gertrudis Ligeró³, Concha Gil^{1,4}, Fernando Corrales^{5,a} and Miguel Marcilla^{5,a,b}

¹ Proteomics Unit. Faculty of Pharmacy, Complutense University of Madrid. Plaza Ramón y Cajal s/n, 28040. Madrid, Spain.

² Proteobotics SL. Darwin 3, 28049. Madrid, Spain.

³ Andalusian Public Health System Biobank. Avenida Del Conocimiento s/n. 18016, Granada, Spain.

⁴ Department of Microbiology & Parasitology, Faculty of Pharmacy, Complutense University of Madrid (UCM) and Ramón y Cajal Institute of Health Research (IRYCIS). Plaza Ramón y Cajal s/n, 28040 Madrid, Spain.

⁵ Proteomics Unit. Spanish National Biotechnology Centre. Darwin 3, 28049. Madrid, Spain.

^a Equal contribution

^b Corresponding author: mmarcilla@cnb.csic.es

ABSTRACT

Currently, 14% of the human proteome is made up of proteins whose existence is not confirmed by mass spectrometry. We performed a proteomic profiling of human mesenchymal stem cells (hMSCs) derived from adipose tissue or umbilical cord and identified peptides derived from 13 of such missing proteins. Remarkably, we found compelling evidence for the expression of hyaluronan synthase 1 (NX_Q92839-1) and confirmed its identification by the fragmentation of 4 heavy-labeled peptides that coeluted with their endogenous light counterparts. Our data also suggest that hMSCs constitute a promising source for the detection of missing proteins.

Human mesenchymal Stem Cells (hMSCs) are multipotent stromal cells that can be isolated from a number of tissues and organs such as umbilical cord, bone marrow, liver, muscle, adipose tissue or peripheral blood (1). Because of their well-established capacity for tissue repair, they are employed in regenerative medicine, mainly in clinical trials. In addition, they show great promise for the treatment of autoimmune diseases due to their unique immunomodulatory potential (2). For these reasons, there is growing interest in the study of hMSCs and, in particular, in the elucidation of the molecular mechanisms that determine their biological properties.

We performed a large-scale shotgun analysis of these cells in the context of a research project involving the proteomic profiling of hMSCs derived from umbilical cord or adipose tissue. Since hMSCs are undifferentiated cells (1), we reasoned that they could be a good source for the identification of missing proteins as they could express genes that are repressed in normal adult tissues. This letter summarizes our findings in this regard and aims to highlight the potential of hMSCs for the detection of missing proteins.

A detailed description of the methods employed can be found in Supplemental Information S1. In brief, we lysed about 10^7 hMSCs derived from adipose tissue or umbilical cord. After precipitation, we subjected the proteins to two different analytical workflows. First, we fractionated the proteins by SDS-PAGE and performed an in-gel digestion with trypsin of the resulting 14 gel slices. After peptide extraction, we analyzed the samples by LC-MS/MS using a Q Exactive HF mass spectrometer (Thermo Fisher). In the second approach, we performed a trypsin digestion and peptides were resolved by high pH reversed phase chromatography. We collected 6 chromatographic fractions that, together with an aliquot of the unfractionated sample, were analyzed by LC-MS/MS in a 5600 Triple-TOF mass spectrometer (Sciex).

We searched the acquired MS/MS spectra against the last release available of the neXtProt database (v2.15.0; 2018-01-17) using an in-house developed algorithm that combines the outputs of four independent search engines (3, 4): MASCOT (5), X! Tandem (6), OMSSA (7) and

Myrimatch (8). Identifications were filtered at a FDR \leq 1% at the protein level. It must be noted that FDR estimation is based on several imperfect assumptions and, thus, not all proteins surpassing this threshold are confidently identified.

The number of protein hits varied between 4616 and 6202 depending on the origin of the hMSCs and the analytical workflow. Not surprisingly, the analysis performed in the Q Exactive platform after SDS-PAGE separation resulted in a 20% (umbilical cord) and 28% (adipose tissue) increase in protein matches, likely reflecting the higher performance of this mass spectrometer and the more extensive fractionation of the samples (Table 1).

Next, we queried these lists of identifications for the presence of missing proteins. We used the neXtProt peptide uniqueness checker (9) to determine which of the sequences assigned to missing proteins were unique, taking into account all known sequence variation arising from single amino acid substitutions. 13 candidates survived this stringent filtering (Figure 1). We gathered evidence for the presence of 5 and 11 missing proteins in the adipose tissue and the umbilical cord samples respectively. All in all, the overlap between the four datasets was low, and most of the matches relied on a single unique peptide. Nevertheless, we noticed that one of the proteins –Hyaluronan synthase 1 (NX_Q92839-1)– was detected in 3 of the 4 datasets (Figure 1). Furthermore, this identification was supported by 6 different peptide sequences (Table 2 and Supplemental Information S2). The expression of Hyaluronan synthase 1 was particularly evident in the adipose tissue-derived hMSCs where it was identified with 5 different tryptic peptides.

Since 4 of the sequences in this set are of at least 9 amino acids in length, the detection of this missing protein is validated according to the HPP data interpretation guidelines 2.1 (10). However, for further confirmation, we set out to synthesize 5 of the identified peptides isotopically labeled with $^{13}\text{C}_6$ $^{15}\text{N}_4$ Arginine (+10 Da) at the C-terminus. We ruled out the longest one (ACQSYFHCVSCSGPLGLYR, 20 residues long) to avoid problems during solid phase synthesis.

These 5 heavy standards were spiked into tryptic digests of umbilical cord or adipose tissue-derived hMSCs and the resultant mixture was analyzed by parallel reaction monitoring (PRM) in both the Q Exactive and the 5600 Triple TOF platforms. This analysis proved that, with only one exception (LAVEALVR), all the light peptides coeluted with their heavy counterparts (Supplemental Information S3). In agreement with the shotgun analysis, hyaluronan synthase 1 was identified with much more confidence in the extract of hMSCs derived from adipose tissue. Regarding this latter point, only 1 endogenous peptide (LDPMALLELVR) was detected in umbilical cord-derived hMSCs, and exclusively in the analysis performed in the Q Exactive mass spectrometer (Supplemental Information S3). This observation is probably indicative of differential regulation of HAS1 expression between both cells.

It is worth noting that one of the peptides identified in the shotgun analysis (LAVEALVR) failed to coelute with its labeled counterpart despite having very similar MS/MS spectra (Supplemental Information S4). In our opinion, the most likely explanation to this fact is that the endogenous species corresponds to the peptide LAVEAVLR derived from the beta subunit of the CCT chaperone complex (NX_P78371). Both sequences are very similar and differ only in the order of residues 6 and 7, which accounts for the small shift in retention time. Furthermore, the alternative sequence matches better with the MS/MS spectrum, as this shows signals compatible with the daughter ions y_2^+ , y_2^+-17 and b_6^+ (Supplemental Information S4). Whatever the explanation, this example highlights the need of confirming peptide detection claims with SRM/PRM experiments using labeled standards and not only by fragmentation of the corresponding synthetic peptides. Thus, we propose that this sort of verification should be mandatory for reporting the identification of missing proteins in the context of the Human Proteome Project.

Hyaluronan synthase 1 (EC 2.4.1.212) is encoded by the HAS1 gene and belongs to a family of three isozymes that participate in the synthesis of hyaluronic acid, catalyzing the addition of N-acetylglucosamine or glucuronic acid to the nascent polymer (11). The identification of this

enzyme in the proteome of hMSCs is not completely unexpected as mRNA levels of HAS1 are relatively high in bone marrow-derived hMSCs (12). It is likely that this is also the case also for cells of different origins because the hyaluronan receptor CD44 is a positive marker of stemness and the interaction of hMSCs with the extracellular matrix is critical for their physiological function (1).

Considering its well established physiological function and its extensive annotation, the lack of MS-based evidence for hyaluronan synthase 1 in the literature seems surprising. Possibly, the localization of this protein at the cell membrane –dictated by the presence of 7 transmembrane regions– poses a drawback to its efficient extraction and solubilization, hampering its identification by standard proteomic workflows. Anyhow, here we demonstrate the expression of HAS1 at the protein level so that it should no longer be considered a missing protein. Besides, the 4 peptides validated in this work here are ideal targets for the development of PRM/MRM methods. Finally, we provide suggestive evidence of the presence of 12 additional missing proteins that deserve further analysis for their confirmation, indicating that hMSCs may constitute a good model to gain insight into this hidden portion of the human proteome.

ASSOCIATED CONTENTS

S1. Materials and methods.

S2. Annotated MS/MS spectra of the tryptic peptides mapping to hyaluronan synthase 1.

S3. PRM analysis of 5 tryptic peptides derived from hyaluronan synthase 1.

S4. Extracted ion chromatograms and MS/MS spectra of the peptides LAVEALVR[+10] and LAVEAVLR.

AUTHOR INFORMATION

Corresponding Author

^b Miguel Marcilla. Phone: (+34) 91 5854540, Fax: (+34) 91 5854506, Email address: mmarcilla@cnb.csic.es

Present Address

Centro Nacional de Biotecnología. Darwin 3. 28049. Madrid, Spain.

Notes

^a Equal contribution.

ACKNOWLEDGMENTS

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the data set identifier PXD009893. The Proteomics Units of the Spanish National Biotechnology Centre and of the Complutense University of Madrid belong of Proteored-ISCI and are supported by grant PT17/0019 of the PE I+D+i 2013-2016, funded by ISCI and ERDF. The Andalusian Public Health System Biobank is part of the National Bank of Stem Cell Lines (BNLC-ISCI) and is also supported by the above-mentioned grant.

REFERENCES

1. Assis-Ribas, T.; Forni, M. F.; Winnischofer, S. M. B.; Sogayar, M. C.; Trombetta-Lima, M., Extracellular matrix dynamics during mesenchymal stem cells differentiation. *Dev Biol* 2018, 437, (2), 63-74.
2. Sharma, R. R.; Pollock, K.; Hubel, A.; McKenna, D., Mesenchymal stem or stromal cells: a review of clinical applications and manufacturing practices. *Transfusion* 2014, 54, (5), 1418-37.
3. Alpizar, A.; Marino, F.; Ramos-Fernandez, A.; Lombardia, M.; Jeko, A.; Pazos, F.; Paradelo, A.; Santiago, C.; Heck, A. J.; Marcilla, M., A Molecular Basis for the Presentation of Phosphorylated Peptides by HLA-B Antigens. *Mol Cell Proteomics* 2017, 16, (2), 181-193.
4. Ramos-Fernandez, A.; Paradelo, A.; Navajas, R.; Albar, J. P., Generalized method for probability-based peptide and protein identification from tandem mass spectrometry data and sequence database searching. *Mol Cell Proteomics* 2008, 7, (9), 1748-54.
5. Koenig, T.; Menze, B. H.; Kirchner, M.; Monigatti, F.; Parker, K. C.; Patterson, T.; Steen, J. J.; Hamprecht, F. A.; Steen, H., Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. *J Proteome Res* 2008, 7, (9), 3708-17.

6. Craig, R.; Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004, 20, (9), 1466-7.
7. Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H., Open mass spectrometry search algorithm. *J Proteome Res* 2004, 3, (5), 958-64.
8. Tabb, D. L.; Fernando, C. G.; Chambers, M. C., MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J Proteome Res* 2007, 6, (2), 654-61.
9. Schaeffer, M.; Gateau, A.; Teixeira, D.; Michel, P. A.; Zahn-Zabal, M.; Lane, L., The neXtProt peptide uniqueness checker: a tool for the proteomics community. *Bioinformatics* 2017, 33, (21), 3471-3472.
10. Deutsch, E. W.; Overall, C. M.; Van Eyk, J. E.; Baker, M. S.; Paik, Y. K.; Weintraub, S. T.; Lane, L.; Martens, L.; Vandenbrouck, Y.; Kusebauch, U.; Hancock, W. S.; Hermjakob, H.; Aebersold, R.; Moritz, R. L.; Omenn, G. S., Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. *J Proteome Res* 2016, 15, (11), 3961-3970.
11. Itano, N.; Sawai, T.; Yoshida, M.; Lenas, P.; Yamada, Y.; Imagawa, M.; Shinomura, T.; Hamaguchi, M.; Yoshida, Y.; Ohnuki, Y.; Miyauchi, S.; Spicer, A. P.; McDonald, J. A.; Kimata, K., Three isoforms of mammalian hyaluronan synthases have distinct enzymatic properties. *J Biol Chem* 1999, 274, (35), 25085-92.
12. Qu, C.; Rilla, K.; Tammi, R.; Tammi, M.; Kroger, H.; Lammi, M. J., Extensive CD44-dependent hyaluronan coats on human bone marrow-derived mesenchymal stem cells produced by hyaluronan synthases HAS1, HAS2 and HAS3. *Int J Biochem Cell Biol* 2014, 48, 45-54.

FIGURE CAPTIONS

Figure 1. Putative missing proteins identified in extracts of adipose tissue and umbilical cord-derived hMSCs. The proteins in the upper row were detected (green squares) or not (red squares) after LC-MS/MS analysis of the samples described on the left column. The numbers inside the green squares indicate the number of unique peptides identified in each case.

Table 1. Protein identifications after LC-MS/MS analysis of hMSCs.

Dataset	FDR_{PSM}	FDR_{pept}	FDR_{prot}	PSMs	Peptides	Proteins	Protein Groups
Adipose Tissue (RP-QTOF)	0.25%	0.51%	1.00%	105 065	34 887	4616	4166
Umbilical Cord (RP-QTOF)	0.22%	0.49%	1.00%	121 147	41 186	5160	4660
Adipose Tissue (PAGE-Orbitrap)	0.17%	0.48%	1.00%	288 536	54 382	5943	5274
Umbilical Cord (PAGE-Orbitrap)	0.22%	0.47%	1.00%	295 020	60 787	6202	5482

Table 2. Unique peptides mapped to Hyaluronan synthase 1 (NX_Q92839-1). ^a AT: Adipose Tissue. ^b UC: Umbilical Cord.

Peptide	m/z	z	AT ^a Triple TOF	UC ^b Triple TOF	AT ^a Q Exactive	UC ^b Q Exactive	Confirmed
AcQSYFHcVScISGPLGLYR	792.36	3+	+	-	-	-	Not tested
LDPMALLELVR	635.37	2+	+	-	+	-	Yes
GPLDAATAR	436.24	2+	+	-	+	-	Yes
YWVAFNVER	592.30	2+	+	-	-	-	Yes
ADWSGPSR	438.20	2+	+	-	-	-	Yes
LAVEALVR	435.77	2+	-	+	-	-	No

FIGURE 1

	NX_Q92839-1	NX_Q8IVF6-2	NX_I1YAP6-1	NX_A6NFT4-1	NX_A8MZ36-1	NX_P0DN81-1	NX_Q13520-1	NX_Q3ZCT8-1	NX_Q8N9P6-1	NX_Q96HZ4-2	NX_Q9BRU2-1	NX_Q9GZQ6-1	NX_Q9H158-2
Umbilical Cord / Q-exactive (6)		1	1	1		1					1		1
Umbilical Cord / Triple-TOF (6)	1	1					1	1	1			1	
Adipose Tissue / Q-Exactive (3)	2		1							1			
Adipose Tissue / Triple-TOF (3)	5	1			1								

FOR TOC ONLY

