



## OPEN ACCESS

## EDITED BY

Bikul Das,  
KaviKrishna Laboratory, India

## REVIEWED BY

Rajkumar P. Thummer,  
Indian Institute of Technology Guwahati,  
India  
Jayshree Advani,  
National Institutes of Health (NIH),  
United States

## \*CORRESPONDENCE

Ahmed Moustafa,  
✉ amoustafa@aucegypt.edu  
Asma Amleh,  
✉ aamleh@aucegypt.edu

## SPECIALTY SECTION

This article was submitted to Stem  
Cell Research,  
a section of the journal  
Frontiers in Cell and Developmental  
Biology

RECEIVED 09 October 2022

ACCEPTED 16 February 2023

PUBLISHED 01 March 2023

## CITATION

Ghabriel M, El Hosseiny A, Moustafa A and  
Amleh A (2023), Computational  
comparative analysis identifies potential  
stemness-related markers for  
mesenchymal stromal/stem cells.  
*Front. Cell Dev. Biol.* 11:1065050.  
doi: 10.3389/fcell.2023.1065050

## COPYRIGHT

© 2023 Ghabriel, El Hosseiny, Moustafa  
and Amleh. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original author(s)  
and the copyright owner(s) are credited  
and that the original publication in this  
journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Computational comparative analysis identifies potential stemness-related markers for mesenchymal stromal/stem cells

Myret Ghabriel<sup>1</sup>, Ahmed El Hosseiny<sup>1,2,3</sup>, Ahmed Moustafa<sup>1,2,3\*</sup> and Asma Amleh<sup>1,2\*</sup>

<sup>1</sup>Biotechnology Graduate Program, American University in Cairo, New Cairo, Egypt, <sup>2</sup>Department of Biology, American University in Cairo, New Cairo, Egypt, <sup>3</sup>Systems Genomics Laboratory, American University in Cairo, New Cairo, Egypt

Mesenchymal stromal/stem cells (MSCs) are multipotent cells that reside in multiple tissues are capable of self-renewal and differentiation into various cell types. These properties make them promising candidates for regenerative therapies. MSC identification is critical in yielding pure populations for successful therapeutic applications; however, the criteria for MSC identification proposed by the International Society for Cellular Therapy (ISCT) are inconsistent across different tissue sources. This study aimed to identify potential markers to be used together with the ISCT criteria to provide a more accurate means of MSC identification. Thus, we carried out a computational comparative analysis of the gene expression in human and mouse MSCs derived from multiple tissues to identify the differentially expressed genes that are shared between the two species. We show that six members of the proteasome degradation system are similarly expressed across MSCs derived from bone marrow, adipose tissue, amnion, and umbilical cord. Additionally, with the help of predictive models, we found that the expression profile of these genes correctly validated the identity of the MSCs across all the tissue sources tested. Moreover, using genetic interaction networks, we showed a possible link between these genes and antioxidant enzymes in the MSC antioxidant defense system, thereby pointing to their potential role in prolonging the life span of MSCs. According to our findings, members of the proteasome degradation system may serve as stemness-related markers.

## KEYWORDS

mesenchymal stem cells, proteasome, stemness-related markers, transcriptomics, gene interaction networks

## Introduction

Mesenchymal stem cells (MSCs) are multipotent adult stem cells that can be isolated from a variety of tissues such as bone marrow (BM) (Friedenstein et al., 1968), adipose tissue (AT) (Zuk et al., 2002), amnion (AM) (Alviano et al., 2007), umbilical cord (UC) (Erices et al., 2000). Due to the myriad sources of MSCs, the International Society for Cellular Therapy (ISCT) proposed minimum criteria by which these MSCs can be identified. These criteria include 1) plastic-adherence of cells *in vitro*, 2) expression of specific cell surface markers (CD105, CD90, and CD73), and lack of expression of others (CD45, CD14, CD19, CD34, CD11b, CD79alpha, and HLA-DR), and 3) ability to differentiate into osteoblasts,

chondroblasts, and adipocytes *in vitro* (Dominici et al., 2006). Unfortunately, growing evidence shows that these criteria are not consistent across different tissues and species since they define only general functional and morphological characteristics (Peister et al., 2004). As a result, scientists have resorted to using additional “stemness” or “stemness-related” genes as markers to aid in the correct identification of MSCs (Zhao et al., 2017). Proper identification of MSCs is crucial to producing pure populations, thereby increasing their use in regenerative therapies. Thus, MSCs are an attractive tool for regenerative therapies, as their ease of isolation and ability to differentiate into multiple lineages make them ideal candidates for this purpose.

MSCs play a critical role in tissue maintenance, regeneration, and homeostasis *in vivo* (Minguell et al., 2001). Generally, MSCs remain quiescent, relying on glycolysis to produce energy for their metabolic needs (Sart et al., 2015); however, upon tissue injury or loss, MSCs are activated to regenerate the damaged tissue and exit quiescence in favor of a more proliferative state. This highly proliferative state must maintain the balance between replenishing downstream lineages and replenishing the stem cell pool. As they begin to proliferate, energy demands increase, and glycolysis shifts to oxidative phosphorylation; this shift is accompanied by an increase in the production of reactive oxygen species (ROS) (Liang and Ghaffari, 2014). Oxidative phosphorylation is, indeed, a much more efficient means of generating ATP than glycolysis and can produce up to fifteen times more ATP. However, this is a double-edged sword since excess ROS can impair self-renewal and proliferation of MSCs (Ko et al., 2012; Choo et al., 2014).

For this reason, MSCs have an active antioxidant defense system. It has been demonstrated that MSCs constitutively express high levels of antioxidant enzymes, such as superoxide dismutases, catalases, and glutathione peroxidases (Valle-Prieto and Conget, 2010). These enzymes repair oxidatively damaged proteins, but some become oxidatively modified or damaged irreversibly. The cell has systems that recognize and remove these irreversibly damaged proteins and consequently prevent their buildup (Jung and Grune, 2008). One of these systems is the proteasome degradation system, which plays a vital role in the degradation of oxidized and damaged proteins, preventing their accumulation and subsequent cellular dysfunction (Chondrogianni et al., 2014).

The 26S proteasome is a multicatalytic degradation complex composed of a core particle (the 20S) and one or two regulatory particles (19S). The 20S core comprises four rings, two of which are composed of seven alpha subunits, while the other two rings are composed of seven beta subunits. The 19S regulator is comprised of a base (containing six ATPase and two non-ATPase subunits) and a lid (containing up to 10 non-ATPase subunits) (Tanaka, 2009). The proteasome’s primary function in the cell is to degrade unneeded or damaged proteins by proteolysis; this can be carried out in either a ubiquitin-dependent manner through the 26S pathway or a ubiquitin-independent way through the 20S pathway. Recently, the proteasome has gained a lot of attention, and it has been shown to play an essential role in preserving the self-renewal and stemness of human MSCs. Kapetanou and others showed that senescence and loss of stemness in human MSCs are accompanied

by a sharp decline in proteasome content and activity (Kapetanou et al., 2017). Furthermore, they showed that the expression of some proteasome subunits is possibly affected by pluripotency factors such as Oct4. Taken together, these observations support their hypothesis of a relationship between proteostasis and stem cell function, where proteostasis is critical in maintaining proper protein levels, leading to efficient multipotency and self-renewal maintenance.

In this study, we carried out a comparative analysis of publicly available RNA-Seq data of MSCs derived from different tissues of origin (BM, AT, UC, AM) and various species (human, mouse) to yield a list of common differentially expressed genes (DEGs). We provide evidence that members of the proteasome show similar patterns of expression across all MSC samples. Furthermore, we offer a possible relationship between the proteasome and antioxidant enzymes in protecting MSCs from oxidative stress, highlighting their importance in MSC survival. Finally, we demonstrate that six proteasomal degradation systems can be used as supplementary stemness-related markers for MSC identification through predictive models.

## Materials and methods

### RNA-seq datasets and processing

RNA-seq datasets were obtained from the Gene Expression Omnibus (GEO) (Barrett et al., 2012) and Array Express databases (Athar et al., 2018). We collected transcriptomic data for human MSCs derived from umbilical cord (h\_UC\_MSCs), amnion (h\_AM\_MSCs), bone marrow (h\_BM\_MSCs), and adipose tissue (h\_AT\_MSCs) and their tissue-specific counterparts (h\_UC\_TSCs, h\_AM\_TSCs, h\_BM\_TSCs, and h\_AT\_TSCs). Mouse transcriptomic data included bone marrow-derived MSCs (m\_BM\_MSCs) and adipose tissue-derived MSCs (m\_AT\_MSCs), along with their tissue-specific counterparts (m\_BM\_TSCs and m\_AT\_TSCs) (Supplementary Table S1). The tissue-specific counterpart cells are all cells other than the MSCs found in the MSCs tissue of origin. Those cells served as a reference for identifying differentially expressed genes that distinguished the MSCs from the other pool of cells belonging to the tissue to allow proper comparison between the different MSCs without any interfering background from the tissue from which MSCs were derived. We processed triplicates of each cell type except for h\_AM\_TSCs, h\_AM\_MSCs, and h\_UC\_MSCs, for which we managed to obtain quadruplicates, bringing our total number of samples to 27 human samples and 12 mouse samples (Supplementary Table S2). We used publicly available data; every triplicate was retrieved from a different experiment and a different lab. However, we made sure that the culturing conditions of the cells were similar across all the samples. In addition, all the MSCs samples were primary cultures and sequenced at passage 3. Data for all cell types were converted from the Sequence Read Archive (SRA) format into the FASTQ format using the SRA Toolkit version 2.10.8 for downstream analysis (Leinonen et al., 2010). Moreover, data were filtered; any read with a length less than 50 bp was excluded. Adapter sequences were detected and trimmed using fastp version 0.19.5 (Chen et al., 2018).

## RNA-seq data analysis to find DEGs

We used Kallisto version 0.46.1 for pseudo-alignment and the quantification of abundances of transcripts from the RNA-Seq data (Bray et al., 2016). Human data were pseudo-aligned to the human reference transcriptome GCA\_000001405.15\_GRCh38, while mouse data were pseudo-aligned to the mouse reference transcriptome GCA\_000001635.9\_GRCm39 provided by the Genome Reference Consortium (O'Leary et al., 2016). Pseudomapping was performed using Kallisto (Bray et al., 2016) through the identification of transcripts that a read is compatible with and assigning it a target ID. Each target ID has a corresponding accession number in the index file. Then the abundances of the transcripts are quantified and output files of abundances containing the transcript per million (TPMs) of each target ID and their corresponding accession numbers are produced. After quantification, Sleuth version 0.29.0 was used for the differential expression analysis of the transcript quantifications between mesenchymal stem cells and their tissue specific counterparts.

Sleuth loaded the Kallisto processed data, estimated the parameters for its response error measurement “full” model followed by the estimation of the parameters for its reduced model, and performed differential analysis using the likelihood ratio test. Sleuth normalizes the data by its ability to distinguish between technical and biological variance and performs shrinkage to the model only on the biological component of variance. Sleuth accounts for technical variability in the abundance estimates and models the true abundance using a general linear model, while including the technical variance as error in the response variable. Thereby, distinguishing between technical or biological sources of variance when determining differentially expressed transcripts.

Accordingly, Sleuth produced a table of significant differentially expressed genes (DEGs) with a  $q$  value less than 0.05. This step generated lists of significant DEGs for each tissue type and species. The lists included the gene symbols of significant DEGs and their corresponding TPMs (Pimentel et al., 2017). The gene symbols of each list in the human data were compared and the common DEGs retrieved with their equivalent expression. This step was repeated for the mouse data and the common DEGs retrieved. Subsequently, the list of common DEGs identified between the human MSCs samples were cross referenced and compared to the gene symbols of the common DEGs between the mouse MSCs samples to produce a list of common DEGs between the two species with their equivalent expression. Gene symbols that weren't common between the two species were checked for homology using Homologene (<http://www.ncbi.nlm.nih.gov/homologene/>) and the analysis repeated. Venn diagrams of the common DEGs were constructed using Venny version 2.0 (Oliveros, 2007). Finally, the expression of the common DEGs was compared and visualized using R Studio 3.6.1 that generated heatmaps and t-distributed stochastic neighbor-embedding clustering ([www.r-project.org](http://www.r-project.org)) (Team RStudio, 2019).

## DEGs ontology and enrichment analysis

Biological processes encompassing the DEGs were identified based on GO enrichment analysis using the GOrilla database (Eden et al., 2009). The  $p$ -value threshold was set at  $10^{-3}$ . Afterward, we visualized the enriched GO terms using ReViGO, and a scatter plot was produced showing the  $\log_{10}$   $p$ -value and  $\log$  size of each GO term (Supek et al., 2011).

## Generation of gene interaction network

To further investigate the interactions between the DEGs, we constructed a gene interaction network by mining interaction networks from the GEO, BioGRID, IRefIndex, and I2D using the GeneMANIA Cytoscape plugin (Montejo et al., 2010). This step produced an annotated Cytoscape network of functional interactions between the DEGs.

## Predictive model

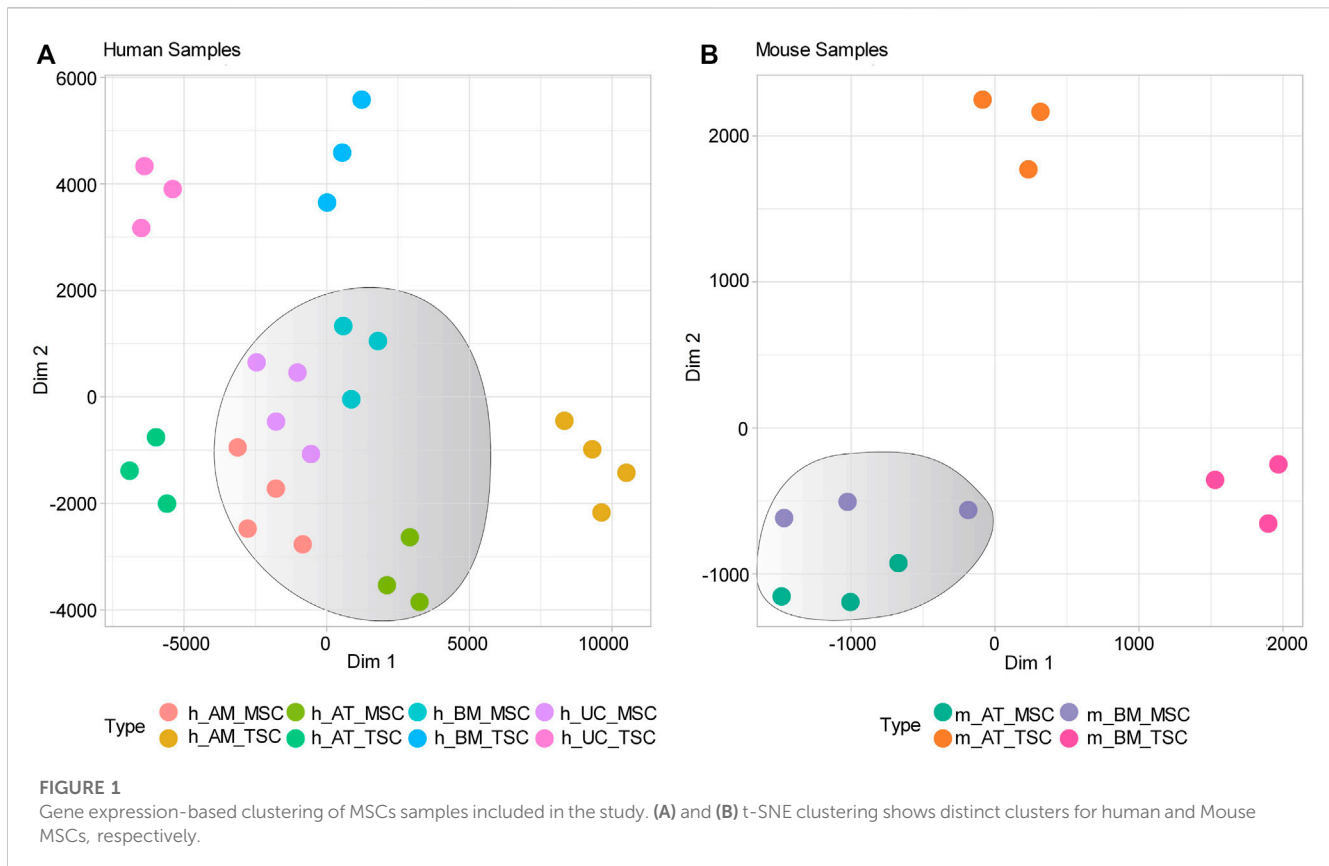
Finally, to assess the ability of the selected DEGs' to identify MSCs, we built a predictive model using the Waikato Environment for Knowledge Analysis (WEKA) software version 3.8.4 (Witten et al., 2016). The gene expression values were converted into the ARFF file format, where our genes of interest were used as attributes in the training dataset. The dataset used for training and testing the model consisted of the expression levels of the 22 proteasome genes across six TSCs, six MSCs mouse samples, thirteen TSCs and fourteen MSCs human samples. We used the AutoWEKAClassifier package (<https://github.com/autowl/autoweka>) to automatically find the best classification model for our provided dataset. AutoWEKAClassifier performed 486 evaluations of its available classifiers and found random forest to be the best classifier with the best error rate for this dataset. The random forest tree classifier was used to train the model with 10-fold cross-validation method. Briefly, this method randomly divided the dataset into 10 parts; it used nine for training and reserved one for testing. This procedure is repeated multiple instances each time reserving a different part for testing. After training the Random Forest model, we created the testing dataset from the training dataset by hiding the type class in order to test its performance in predicting tissue type in both human and mouse samples. WEKA was also used for attribute selection.

# Results

## RNA-seq analysis

We constructed clustering maps using the t-distributed stochastic neighbor embedding (t-SNE) statistical method to visualize the transcriptomic similarities and differences between the samples. The clustering maps showed that all human MSC samples clustered together in a distinct cluster apart from tissue-specific cell (TSCs) samples (Figure 1A). Likewise, the mouse MSC samples clustered together, while TSC samples clustered independently (Figure 1B). Next, we compared the gene expression of human-derived MSCs against their tissue-specific counterparts and identified 20,973, 24,365, 8,296, and 29,197 DEGs for h\_AM\_MSCs, h\_BM\_MSCs, h\_AT\_MSCs, and h\_UC\_MSCs, respectively. We constructed a Venn diagram to visualize the common DEGs shared by the MSCs derived from the four different tissue sources. The common DEGs made up 4.6% of the examined DEGs, equivalent to 2,181 common DEGs (Figure 2A) (Supplementary Table S3).

Similarly, for the mouse datasets, we compared the gene expression of mouse-derived MSCs with their tissue-specific counterparts and identified 14,843 DEGs for m\_AT\_MSCs and 12,783 DEGs for the m\_BM\_MSCs. We found that the common DEGs made up 78.8% of the examined DEGs, equivalent to 12,178 common DEGs (Figure 2B; Supplementary Table S4).



Finally, we compared both lists of common DEGs to determine whether the human and mouse MSCs shared any common DEGs. The Venn diagram showed that the two species had 1,583 (13.3%) DEGs in common (Figure 2C, Supplementary Table S5). Interestingly, the heatmap of the 1,583 common DEGs showed three main clusters: a cluster that included all the human MSC samples, another cluster that included all the mouse-derived MSC samples plus m\_BM\_TSCs, and, finally, the last cluster included the rest of the TSC samples. Each of these clusters included subclusters that grouped the triplicate of each tissue type (Figure 3).

### DEGs ontology and enrichment analysis

In the gene ontology analysis, we produced a list of 157 enriched gene ontology (GO) terms with a threshold  $p$ -value of  $10^{-3}$  (Supplementary Table S6). ReViGO generated a scatter plot of the enriched GO terms organized according to their significance ( $p$ -value) and uniqueness. Following visualization, we identified a unique GO term (GO:2000736) to regulate stem cell differentiation with a significant  $p$ -value of  $8.69E-4$  and a  $q$  value of  $4.64E-2$ . Moreover, this GO term had a frequency of 0.010% and a uniqueness score of 0.70 (Supplementary Figure S1). Other GO terms were more general, less unique, and not explicitly specific to stem cell function. The GO term GO:2000736 included 23 genes that belonged to the proteasomal degradation pathway (Table 1).

### Gene expression analysis

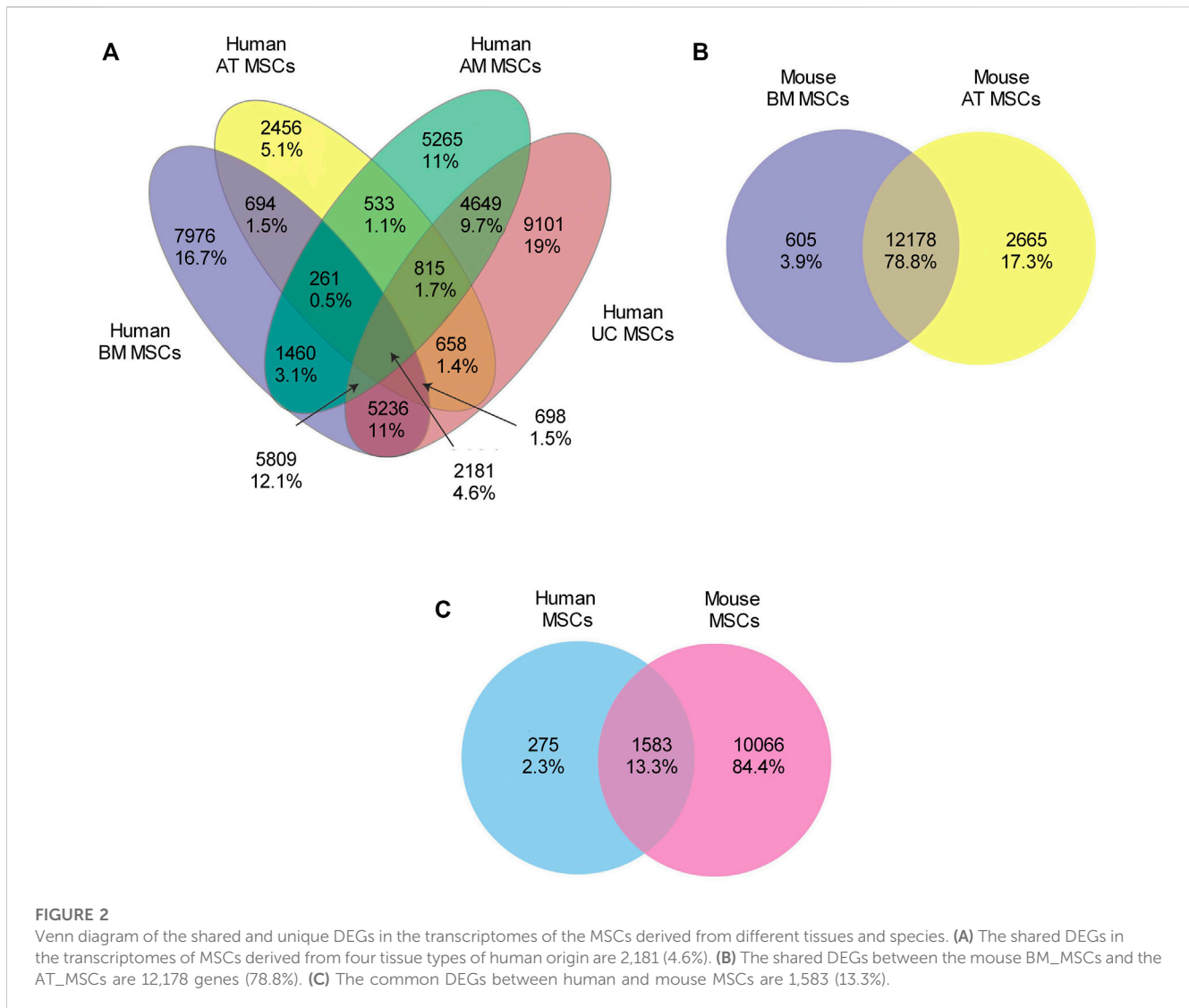
Now that our attention was drawn to these 23 genes, we wanted to take a closer look at their behavior. First, we inspected their expression

patterns in both species. We found that the 23 genes appeared to be upregulated in both the human and mouse MSC samples, except for PSMD4, which was downregulated in the mouse MSC samples compared to the TSC samples (Figures 4A, B). PSMD4 had a  $p$ -value of  $6.27E-03$  and a fold change of 0.136934557 in m\_AT\_MSCs, while in m\_BM\_MSCs, it had a  $p$ -value of  $9.79E-03$  and a fold change of  $-0.224675507$ . Since most genes were upregulated, we attempted to further explore the interplay between these genes and other systems that assist the proteasome in antioxidant defense, namely, antioxidant enzymes.

Consequently, members of the superoxide dismutase, glutathione peroxidase, peroxiredoxin, thioredoxin, and peroxidase families were cross-referenced against the 1,583 common DEGs identified earlier. We found SOD3, GPX7, GPX8, PRDX2, PRDX4, TXN2, and PXDN were present in our list of common DEGs. The majority of the transcripts of these genes were upregulated but not all. Out of the seven genes, four (GPX7, GPX8, PXDN, TXN2) were upregulated in all analyzed MSC samples. Each of the other three (SOD3, PRDX2, and PRDX4) was upregulated in all the MSC samples except human AT-MSCs and BM-MSCs and mouse BM-MSCs, respectively (Figures 4C, D).

### Gene interaction network

Next, we wanted to shed more light on the interaction between these antioxidant enzymes and the 23 members of the proteasomal degradation system identified earlier. We employed the help of Cytoscape and the Genemania database to understand the interplay between these antioxidant enzymes and the proteasomal genes. A



gene interaction network was generated, and it showed all 23 genes of the proteasomal degradation pathway were co-expressed together and co-expressed with the antioxidant genes. Specifically, it showed PSMA7 to be co-expressed with GPX7, which in turn was co-expressed with GPX8, SOD3, and PXDN. Additionally, PRDX2 was co-expressed with PSMA7, PSMB3, PSMB6, PSMB7, PSMC4, PSMD3, and PSMD8. Furthermore, TXN2 was co-expressed with PSMA1, PSMA7, PSMB3, and PSMB6. Finally, PRDX4 was co-expressed with PSMA5, PSMB1, PSMB2, PSMB5, PSMB6, PSMC1, PSMC2, PSMC5, PSMD1, PSMD8, and PSMD14 (Figure 5).

### Predictive model

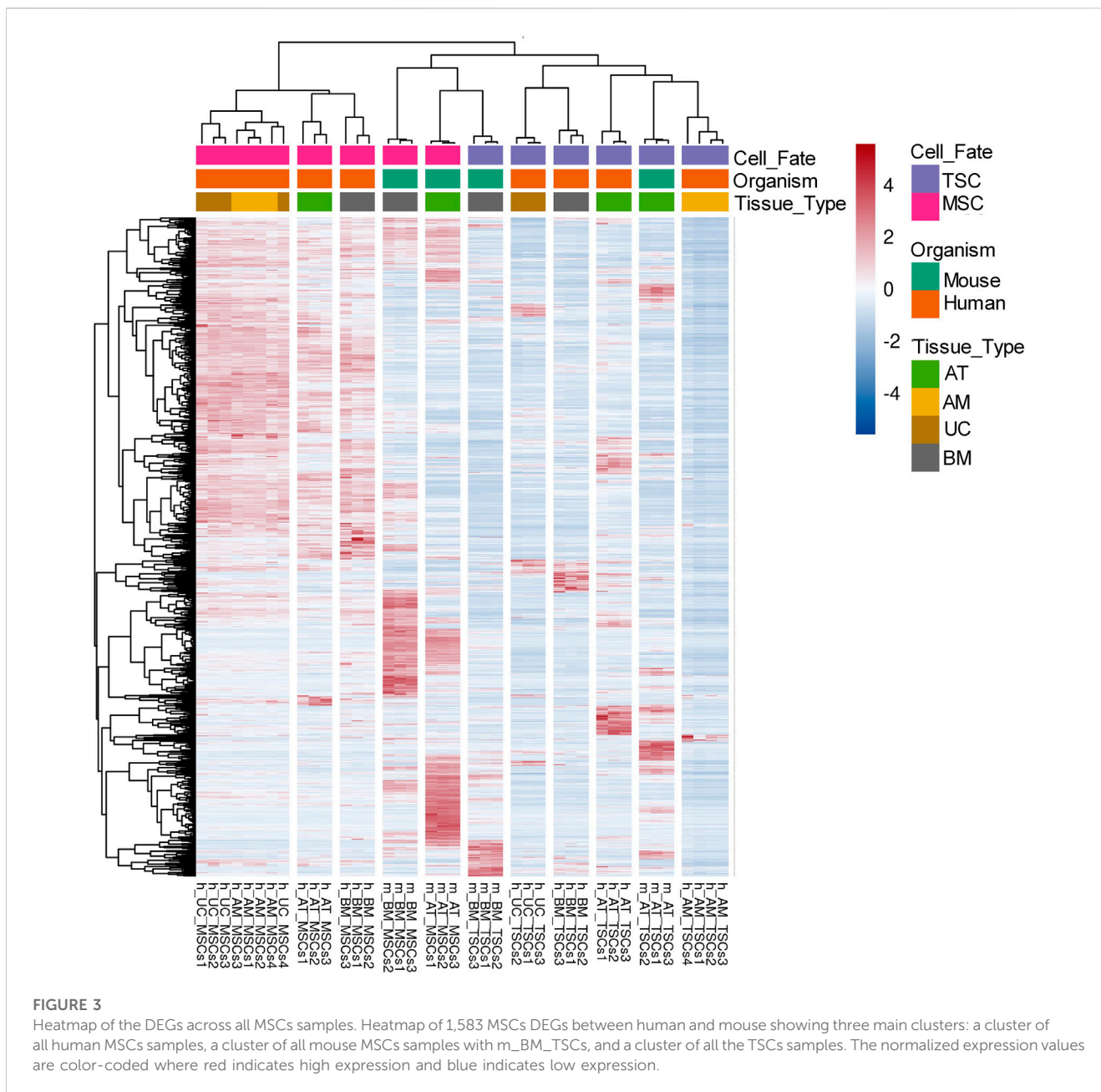
Finally, we wanted to test the genes' efficiency in predicting the identity of MSCs across the different tissue sources in both human and mouse species. To test this hypothesis, AutoWEKAClassifier performed 486 evaluations of available classifiers and found random forest to be the best classifier with the best error rate. The random forest tree classifier was used to train the model with 10-fold cross-validation, and the trained model was finalized. The final model was

loaded to test its performance in predicting stem cell type on the testing data. We used the upregulated proteasomal genes as attributes, and we removed PSMD4 from the list since it had inconsistent expression across both species. We proceeded with the other 22 genes and ran the random forest model. The model tested the data 40 times and showed that MSCs were correctly classified in all 40 instances. To test whether all 22 genes contributed equally to the classification process, we ran the gain ratio attribute selection evaluator in WEKA. We found six genes to be the top contributors in the classification: PSMB5, PSMB1, PSMD14, PSMC4, PSMA1, and PSMD8 (Supplementary Table S7). We repeated the random forest model using these six genes and these six genes were enough to correctly classify the MSCs all 40 instances (Supplementary Table S8).

### Discussion

Ever since the discovery of MSCs by Friedenstein et al. (1968), researchers have debated their identity; however, the criteria





proposed by the ISCT still fail to adequately describe MSCs and shows discrepancies across species and tissues of origin. As the focus shifted to stemness and stemness-related gene expression to aid in identifying MSCs, the search for adequate markers has intensified. Here, we show that members of the proteasome degradation system can be used as potential stemness-related markers to validate the identity of MSCs.

In this study, we integrated the RNA-seq data of MSCs derived from four different human tissues (AM, BM, AT, and UC) and two different mouse tissues (AT, BM). Differential expression analysis presented us with a list of 1,583 DEGs common to MSCs and TSCs across all tissue types in humans and mice. Further gene ontology enrichment analysis categorized these genes into GO terms, one of which was the GO term for regulating stem cell differentiation.

GO terms such as (GO:0055114) involved in oxidation-reduction process and GO term (GO:0006123) involved in mitochondrial electron transport, cytochrome c to oxygen were present in our results, however, they had higher frequencies than the GO term (GO:2000736) for stem cell regulation, which were 0.172%, 0.044% and 0.010% respectively. Since a higher frequency denotes a more general term, we focused on the GO term (GO: 2000736) for stem cell regulation due to its uniqueness, high significance, and its specificity to stem cell processes. It included 23 members of the proteasome degradation system. Compelling evidence suggests a pivotal role for the proteasome in maintaining the pluripotency of mouse and human embryonic stem cells by supporting the clean-up of proteins oxidatively damaged during differentiation (Schröter and Adjaye, 2014).

TABLE 1 List of DEGs belonging to the proteasomal degradation pathway identified in the GO enrichment analysis (GO:2000736).

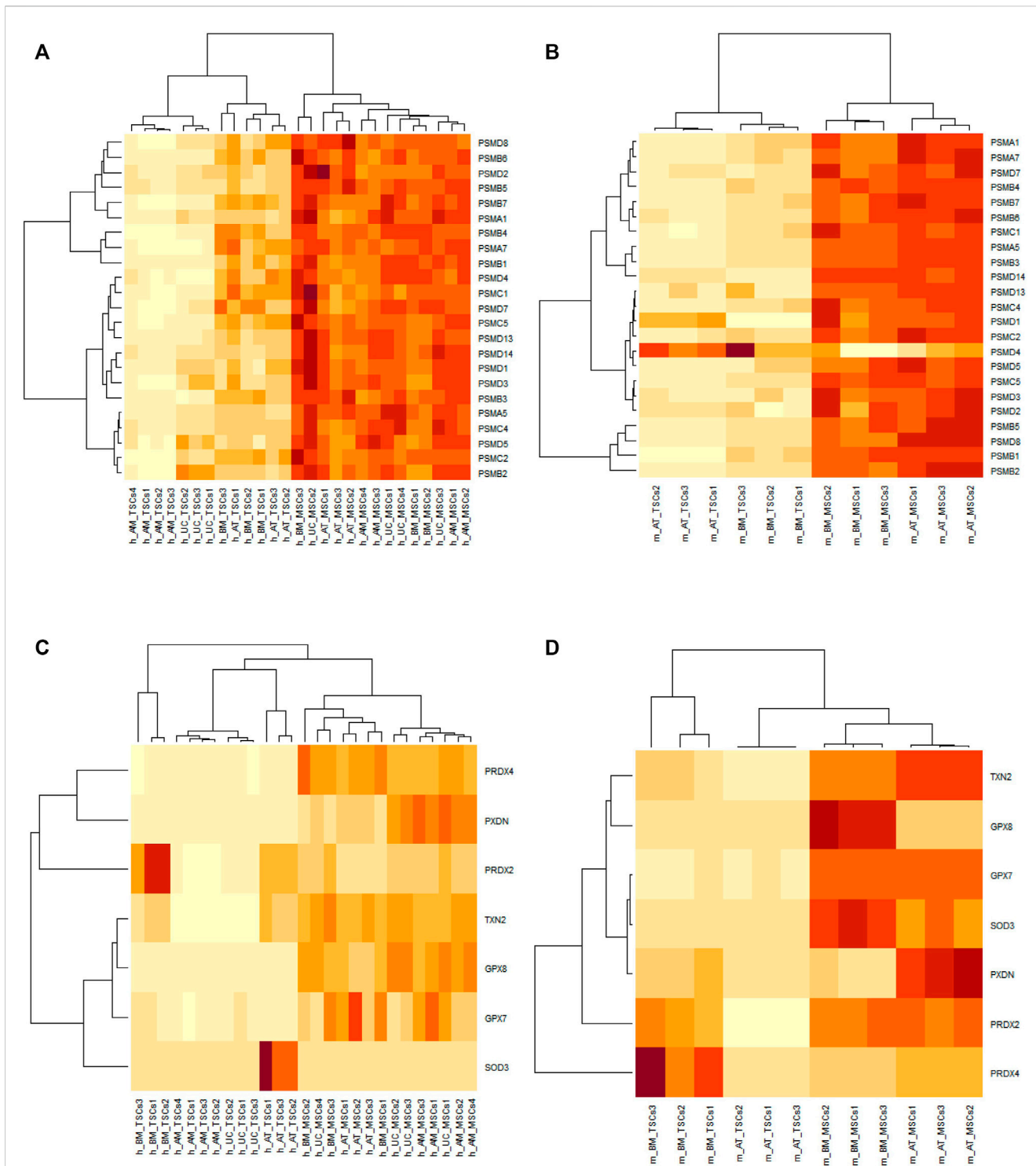
HUGO ID	Systemic ID	Gene
PSMA1	$\alpha 6$	proteasome (prosome, macropain) subunit, alpha type, 1
PSMA5	$\alpha 1$	proteasome (prosome, macropain) subunit, alpha type, 5
PSMA7	$\alpha 4$	proteasome (prosome, macropain) subunit, alpha type, 7
PSMB1	$\beta 6$	proteasome (prosome, macropain) subunit, beta type, 1
PSMB2	$\beta 4$	proteasome (prosome, macropain) subunit, beta type, 2
PSMB3	$\beta 3$	proteasome (prosome, macropain) subunit, beta type, 3
PSMB4	$\beta 7$	proteasome (prosome, macropain) subunit, beta type, 4
PSMB5	$\beta 5$	proteasome (prosome, macropain) subunit, beta type, 5
PSMB6	$\beta 1$	proteasome (prosome, macropain) subunit, beta type, 6
PSMB7	$\beta 2$	proteasome (prosome, macropain) subunit, beta type, 7
PSMC1	Rpt2	proteasome (prosome, macropain) 26s subunit, atpase, 1
PSMC2	Rpt1	proteasome (prosome, macropain) 26s subunit, atpase, 2
PSMC4	Rpt3	proteasome (prosome, macropain) 26s subunit, atpase, 4
PSMC5	Rpt6	proteasome (prosome, macropain) 26s subunit, atpase, 5
PSMD1	Rpn2	proteasome (prosome, macropain) 26s subunit, non-atpase, 1
PSMD2	Rpn1	proteasome (prosome, macropain) 26s subunit, non-atpase, 2
PSMD3	Rpn3	proteasome (prosome, macropain) 26s subunit, non-atpase, 3
PSMD4	Rpn10	proteasome (prosome, macropain) 26s subunit, non-atpase, 4
PSMD5	Rpn4	proteasome (prosome, macropain) 26s subunit, non-atpase, 5
PSMD7	Rpn8	proteasome (prosome, macropain) 26s subunit, non-atpase, 7
PSMD8	Rpn12	proteasome (prosome, macropain) 26s subunit, non-atpase, 8
PSMD13	Rpn9	proteasome (prosome, macropain) 26s subunit, non-atpase, 13
PSMD14	Rpn11	proteasome (prosome, macropain) 26s subunit, non-atpase, 14

The proteasome is an essential component of protein quality-control systems and plays a critical role in cellular homeostasis. It is involved in the degradation of abnormal, oxidized, or otherwise damaged proteins (Raynes et al., 2016). The accumulation of oxidized proteins in cells leads to their decreased life span (Reeg and Grune, 2015). Moreover, it has been demonstrated that dysfunction of the proteasome is heavily implicated in cell ageing (Chondrogianni et al., 2003). A recent study revealed that impairment of proteasome function resulted in an accumulation of oxidatively modified proteins in senescent Wharton's jelly (WJ) and adipose-derived human adult mesenchymal stromal/stem cells. More importantly, this study showed that senescence of these cells' is accompanied by a decline in proteasome content and activities, coupled with the concurrent loss of their stemness (Kapetanou et al., 2017). Although the degradation of oxidized proteins can occur by ubiquitin-dependent (26S-proteasome) and ubiquitin-independent (20S-proteasome) mechanisms (Shang et al., 2001; Goldberg, 2003; Bader et al., 2007), various studies have shown that the 20S proteasome might be the major machinery involved in this process (Silva et al., 2012; Jung et al., 2014). Here, we show that members of the 20S proteasome (PSMA1, PSMA5, and PSMA7) of

the alpha subunits and all members of the beta subunit (PSMB1, PSMB2, PSMB3, PSMB4, PSMB5, PSMB6, and PSMB7) are not only differentially expressed in MSCs but are also upregulated.

Furthermore, we show that members of the 19S proteasome base (PSMC1, PSMC2, PSMC4, PSMC5, PSMD1, and PSMD2) and lid (PSMD3, PSMD5, PSMD7, PSMD8, PSMD13, and PSMD14) are also differentially expressed and upregulated in MSCs in comparison with TSCs. However, our results also showed that PSMD4 expression in mouse MSC samples was downregulated. PSMD4's central role in the 19S lid is to recognize polyubiquitinated protein substrates and detach the ubiquitin molecules from them for their subsequent degradation through the 26S proteasomal pathway (da Fonseca et al., 2012). PSMD4 is not the only ubiquitin receptor in the 19S lid. PSMD2 is another ubiquitin receptor that recognizes and binds both ubiquitin and ubiquitin-like proteins (Chojnacki et al., 2017). We found PSMD2 to be upregulated in our mouse MSCs. It could be that mouse MSCs rely mainly on PSMD2 to recognize polyubiquitinated protein substrates, thereby rendering PSMD4 dispensable.

Studies have shown that during MSC proliferation, ROS are produced as byproducts of oxidative metabolism. However,

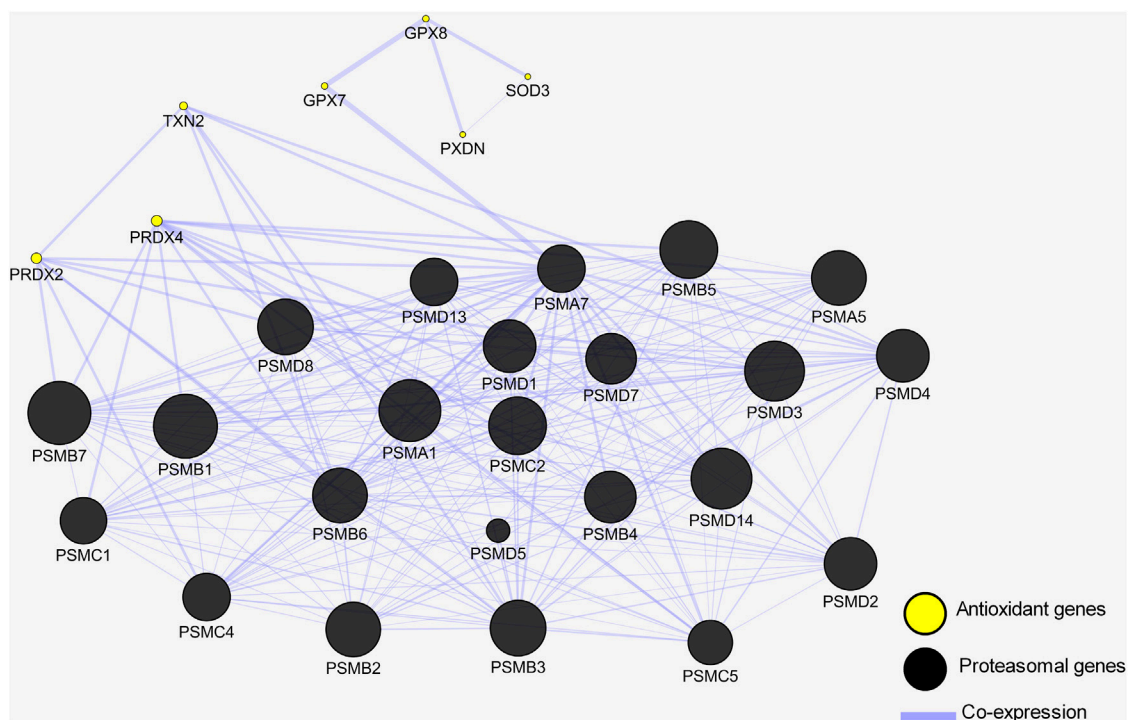


**FIGURE 4**  
Heatmaps of DEGs between MSCs and their TSCs generated by RStudio. Orange indicates upregulation and yellow indicates downregulation. **(A)** and **(B)** heatmaps of the proteasomal genes' expression in human and mouse samples, respectively, showing all genes to be upregulated in MSCs samples except PSMD4 in mouse MSCs. **(C)** and **(D)** heatmaps showing the expression of the antioxidant genes is upregulated in human and mouse MSCs samples, respectively, with the exception of SOD3 and PRDX2 in human AT\_MSCs and BM\_MSCs, and PRDX4 in mouse BM\_MSCs.

increased ROS levels may lead to a decrease in cell survival and have also been implicated in cell senescence (Ko et al., 2012). To counteract these detrimental effects, the cell has antioxidant

defense systems activated by high ROS levels. Increased ROS concentration causes Nrf2 (a stress-responsive transcription factor) to dissociate from its inhibitory complex with Keap1. This





**FIGURE 5**

Gene interaction network shows antioxidant genes are co-expressed with proteasomal genes. Genes are depicted by nodes and the types of interaction are depicted by edges. Black circles represent proteasomal genes co-expressed with the antioxidant genes represented by yellow circles.

enables Nrf2 to accumulate and translocate to the nucleus, where it binds to antioxidant-response elements (ARE), thus promoting the expression of several antioxidant (Dai et al., 2020) and proteasomal genes (Kwak et al., 2003). These antioxidant enzymes and the proteasome degradation machinery work together as a defense against damaging high ROS levels. We demonstrated that members of the proteasome degradation machinery were upregulated across the MSCs samples tested. We also demonstrated that GPX7, GPX8, TXN2, and PXDN antioxidant genes were differentially expressed and upregulated in MSCs compared to TSCs. Additionally, we provided evidence that these genes are co-expressed with the proteasome degradation machinery members by data mining gene interaction databases. Taken together, these results point to the efficiency of MSCs in counteracting oxidative stress, in which the proteasome is integral.

Finally, to show the competence of these proteasome genes in validating the identity of MSCs, we employed the aid of predictive models. Predictive models have been used robustly to identify a general MSC phenotype that could distinguish MSCs from other cell types. A recent study showed that gene expression levels in prediction models increase the classification accuracy of the combined set of traditional MSC cell surface markers (Rohart et al., 2016). Using the random forest model, we showed that the expression of six proteasome genes could accurately distinguish MSCs from their tissue-specific counterparts. Of these six genes, PSMB5, PSMB1, and PSMD14 have been linked to stem cell function. As previously mentioned, PSMB1 and PSMB5 are catalytic subunits of the 20S proteasome, and reducing their

expression leads to a decrease in cell proliferation and an increase in replicative senescence in hBMSCs (Yu et al., 2015). Likewise, Kapetanou and others reported a similar decline in the expression of these two genes in senescent WJ-MSCs. They also showed that PSMB5 overexpression rescues these senescent cells from age-related reductions in proteasome expression and function, improving their stemness and extending their lifespan (Kapetanou et al., 2017). Finally, PSMD14 is essential for proper 26S assembly (Quinet et al., 2020); it also plays a role in cleaving polyubiquitin chains at a proximal site and recycling ubiquitin chains (Lu et al., 2012). PSMD14 is a crucial regulator of stem cell maintenance. A reduction in its levels leads to a marked decrease in Oct4 protein expression, accompanied by abnormal morphology in embryonic stem cells (Buckley Shannon et al., 2012). However, no data currently exists on the three remaining genes (PSMC4, PSMA1, and PSMD8) that link them to any stem cell function.

Our study carried out a comprehensive comparative analysis of MSCs RNA-seq data across two species and six different tissue types to ascertain potential identity markers. Our results showed that six members of the proteasomal machinery are promising candidates for validating the identity of MSCs. Moreover, we shed a light on their association with antioxidant enzymes in defending MSCs against high ROS levels, thereby maintaining their proliferation and self-renewal. These six genes can be used as additional stemness-related markers to refine and enhance the accuracy of MSC identification, which is a critical step in ensuring the yield of a pure population for consequent applications in regenerative therapies.

Our prediction model is based only on MSCs and TSCs data for its training and testing sets; experimental validation of the expression levels of the six proposed stemness-related markers in MSCs from different sources, both on the RNA and protein levels, is crucial to confirm their efficiency in identifying MSCs. Experimental validation is beyond the scope of this study; however, it should be the focus of future confirmatory studies.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding authors.

## Author contributions

Conceived the study: AA and MG. Designed the study: AA, AM, and MG. Collected the data: MG and AE. Analyzed the data: MG, AE, AM, and AA. Wrote the paper: MG and AE. Review and Editing the paper: AM and AA.

## References

- Alviano, F., Fossati, V., Marchionni, C., Arpinati, M., Bonsi, L., Franchina, M., et al. (2007). Term amniotic membrane is a high throughput source for multipotent mesenchymal stem cells with the ability to differentiate into endothelial cells *in vitro*. *BMC Dev. Biol.* 7 (1), 11. doi:10.1186/1471-213X-7-11
- Athar, A., Füllgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., et al. (2018). ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Res.* 47 (D1), D711–D715. doi:10.1093/nar/gky964
- Bader, N., Jung, T., and Grune, T. (2007). The proteasome and its role in nuclear protein maintenance. *Exp. Gerontol.* 42 (9), 864–870. doi:10.1016/j.exger.2007.03.010
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2012). NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res.* 41 (D1), D991–D995. doi:10.1093/nar/gks1193
- Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34 (5), 525–527. doi:10.1038/nbt.3519
- Buckley Shannon, M., Aranda-Orgilles, B., Strikoudis, A., Apostolou, E., Loizou, E., Moran-Crusio, K., et al. (2012). Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. *Cell stem cell* 11 (6), 783–798. doi:10.1016/j.stem.2012.09.011
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), i884–i890. doi:10.1093/bioinformatics/bty560
- Chojnacki, M., Mansour, W., Hameed, D. S., Singh, R. K., El Oualid, F., Rosenzweig, R., et al. (2017). Polyubiquitin-photoactivatable crosslinking reagents for mapping ubiquitin interactome identify Rpn1 as a proteasome ubiquitin-associating subunit. *Cell Chem. Biol.* 24 (4), 443–457.e6. doi:10.1016/j.chembiol.2017.02.013
- Chondrogianni, N., Petropoulos, I., Grimm, S., Georgila, K., Catalgol, B., Friguet, B., et al. (2014). Protein damage, repair and proteolysis. *Mol. Asp. Med.* 35, 1–71. doi:10.1016/j.mam.2012.09.001
- Chondrogianni, N., Stratford, F. L. L., Trougakos, I. P., Friguet, B., Rivett, A. J., and Gonos, E. S. (2003). Central role of the proteasome in senescence and survival of human fibroblasts: Induction of a senescence-like phenotype upon its inhibition and resistance to stress upon its activation. *J. Biol. Chem.* 278 (30), 28026–28037. doi:10.1074/jbc.M301048200
- Choo, K. B., Tai, L., Hymavathe, K. S., Wong, C. Y., Nguyen, P. N. N., Huang, C.-J., et al. (2014). Oxidative stress-induced premature senescence in Wharton's jelly-derived mesenchymal stem cells. *Int. J. Med. Sci.* 11 (11), 1201–1207. doi:10.7150/ijms.8356
- da Fonseca, P. C., He, J., and Morris, E. P. (2012). Molecular model of the human 26S proteasome. *Mol. cell* 46 (1), 54–66. doi:10.1016/j.molcel.2012.03.026
- Dai, X., Yan, X., Wintergerst, K. A., Cai, L., Keller, B. B., and Tan, Y. (2020). Nrf2: Redox and metabolic regulator of stem cell state and function. *Trends Mol. Med.* 26 (2), 185–200. doi:10.1016/j.molmed.2019.09.007
- Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., Marini, F., Krause, D., et al. (2006). Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* 8 (4), 315–317. doi:10.1080/14653240600855905
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinform.* 10, 48. doi:10.1186/1471-2105-10-48
- Erices, A., Conget, P., and Minguell, J. J. (2000). Mesenchymal progenitor cells in human umbilical cord blood. *Br. J. Haematol.* 109 (1), 235–242. doi:10.1046/j.1365-2141.2000.01986.x
- Friedenstein, A. J., Petrakova, K. V., Kurolesova, A. I., and Frolova, G. P. (1968). Heterotopic of bone marrow. Analysis of precursor cells for osteogenic and hematopoietic tissues. *Transplantation* 6 (2), 230–247. doi:10.1097/00007890-196803000-00009
- Goldberg, A. L. (2003). Protein degradation and protection against misfolded or damaged proteins. *Nature* 426 (6968), 895–899. doi:10.1038/nature02263
- Jung, T., and Grune, T. (2008). The proteasome and its role in the degradation of oxidized proteins. *IUBMB Life* 60 (11), 743–752. doi:10.1002/iub.114
- Jung, T., Höhn, A., and Grune, T. (2014). The proteasome and the degradation of oxidized proteins: Part II – protein oxidation and proteasomal degradation. *Redox Biol.* 2, 99–104. doi:10.1016/j.redox.2013.12.008
- Kapetanou, M., Chondrogianni, N., Petrakis, S., Koliakos, G., and Gonos, E. S. (2017). Proteasome activation enhances stemness and lifespan of human mesenchymal stem cells. *Free Radic. Biol. Med.* 103, 226–235. doi:10.1016/j.freeradbiomed.2016.12.035
- Ko, E., Lee, K. Y., and Hwang, D. S. (2012). Human umbilical cord blood-derived mesenchymal stem cells undergo cellular senescence in response to oxidative stress. *Stem Cells Dev.* 21 (11), 1877–1886. doi:10.1089/scd.2011.0284
- Kwak, M. K., Wakabayashi, N., Greenlaw, J. L., Yamamoto, M., and Kensler, T. W. (2003). Antioxidants enhance mammalian proteasome expression through the Keap1-Nrf2 signaling pathway. *Mol. Cell. Biol.* 23 (23), 8786–8794. doi:10.1128/MCB.23.23.8786-8794.2003
- Leinonen, R., Sugawara, H., Shumway, M., and Collaboration, N. S. D. (2010). The sequence read archive. *Nucleic Acids Res.* 39 (1), D19–D21. doi:10.1093/nar/gkq1019
- Liang, R., and Ghaffari, S. (2014). Stem cells, redox signaling, and stem cell aging. *Antioxidants redox Signal.* 20 (12), 1902–1916. doi:10.1089/ars.2013.5300
- Lu, L., Song, H.-F., Zhang, W.-G., Liu, X.-Q., Zhu, Q., Cheng, X.-L., et al. (2012). Potential role of 20S proteasome in maintaining stem cell integrity of human bone marrow stromal cells in prolonged culture expansion. *Biochem. Biophysical Res. Commun.* 422 (1), 121–127. doi:10.1016/j.bbrc.2012.04.119
- Minguell, J. J., Erices, A., and Conget, P. (2001). Mesenchymal stem cells. *Exp. Biol. Med.* 226 (6), 507–520. doi:10.1177/153537020122600603

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2023.1065050/full#supplementary-material>

- Montejo, J., Zuberi, K., Rodriguez, H., Kazi, F., Wright, G., Donaldson, S. L., et al. (2010). GeneMANIA Cytoscape plugin: Fast gene function predictions on the desktop. *Bioinformatics* 26 (22), 2927–2928. doi:10.1093/bioinformatics/btq562
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciuffo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44 (D1), D733–D745. doi:10.1093/nar/gkv1189
- Oliveros, J. C. (2007). *VENNY. An interactive tool for comparing lists with Venn Diagrams.*
- Peister, A., Mellad, J. A., Larson, B. L., Hall, B. M., Gibson, L. F., and Prockop, D. J. (2004). Adult stem cells from bone marrow (MSCs) isolated from different strains of inbred mice vary in surface epitopes, rates of proliferation, and differentiation potential. *Blood* 103 (5), 1662–1668. doi:10.1182/blood-2003-09-3070
- Pimentel, H., Bray, N. L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. methods* 14 (7), 687–690. doi:10.1038/nmeth.4324
- Quinet, G., Gonzalez-Santamarta, M., Louche, C., and Rodriguez, M. S. (2020). Mechanisms regulating the UPS-als crosstalk: The role of proteaphagy. *Molecules* 25 (10), 2352. doi:10.3390/molecules25102352
- Raynes, R., Pomatto, L. C. D., and Davies, K. J. A. (2016). Degradation of oxidized proteins by the proteasome: Distinguishing between the 20S, 26S, and immunoproteasome proteolytic pathways. *Mol. Aspects Med.* 50, 41–55. doi:10.1016/j.mam.2016.05.001
- Reeg, S., and Grune, T. (2015). Protein oxidation in aging: Does it play a role in aging progression? *Antioxidants Redox Signal.* 23 (3), 239–255. doi:10.1089/ars.2014.6062
- Rohart, F., Mason, E. A., Matigian, N., Mosbergen, R., Korn, O., Chen, T., et al. (2016). A molecular classification of human mesenchymal stromal cells. *PeerJ* 4, e1845. doi:10.7717/peerj.1845
- Sart, S., Song, L., and Li, Y. (2015). Controlling redox status for stem cell survival, expansion, and differentiation. *Oxidative Med. Cell. Longev.* 2015, 105135. doi:10.1155/2015/105135
- Schröter, F., and Adjaye, J. (2014). The proteasome complex and the maintenance of pluripotency: Sustain the fate by mopping up? *Stem Cell Res. Ther.* 5 (1), 24. doi:10.1186/scrt413
- Shang, F., Nowell, T. R., and Taylor, A. (2001). Removal of oxidatively damaged proteins from lens cells by the ubiquitin-proteasome pathway. *Exp. Eye Res.* 73 (2), 229–238. doi:10.1006/exer.2001.1029
- Silva, G. M., Netto, L. E. S., Simões, V., Santos, L. F. A., Gozzo, F. C., Demasi, M. A. A., et al. (2012). Redox control of 20S proteasome gating. *Antioxid. Redox Signal.* 16 (11), 1183–1194. doi:10.1089/ars.2011.4210
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLOS ONE* 6 (7), e21800. doi:10.1371/journal.pone.0021800
- Tanaka, K. (2009). The proteasome: Overview of structure and functions. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 85 (1), 12–36. doi:10.2183/pjab.85.12
- Team RStudio (2019). *Integrated development environment for R.* Boston, MA: RStudio, Inc.
- Valle-Prieto, A., and Conget, P. A. (2010). Human mesenchymal stem cells efficiently manage oxidative stress. *Stem Cells Dev.* 19 (12), 1885–1893. doi:10.1089/scd.2010.0093
- Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). “Data mining,” in *Practical machine learning tools and techniques*. Fourth Edition (Morgan Kaufmann Publishers Inc, United states).
- Yu, H., Lai, H.-J., Lin, T.-W., and Lo Szecheng, J. (2015). Autonomous and non-autonomous roles of DNase II during cell death in *C. elegans* embryos. *Biosci. Rep.* 35 (3), e00203. doi:10.1042/BSR20150055
- Zhao, W., Li, Y., and Zhang, X. (2017). Stemness-related markers in cancer. *Cancer Transl. Med.* 3 (3), 87–95. doi:10.4103/ctm.ctm\_69\_16
- Zuk, P. A., Zhu, M., Ashjian, P., De Ugarte, D. A., Huang, J. I., Mizuno, H., et al. (2002). Human adipose tissue is a source of multipotent stem cells. *Mol. Biol. Cell* 13 (12), 4279–4295. doi:10.1091/mbc.e02-02-0105