

ORIGINAL ARTICLE

Effect of Substrate Topography and Chemistry on Human Mesenchymal Stem Cell Markers: A Transcriptome Study

Bo Zhang^{1,2,*}, Naresh Kasoju^{1,*,†}, Qiongfang Li^{3,*}, Jinmin Ma³,
Aidong Yang², Zhanfeng Cui¹, Hui Wang^{1,3,4}, Hua Ye¹

¹*Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK*

²*Department of Engineering Science, University of Oxford, Oxford, UK*

³*BGI-Shenzhen, Shenzhen 518083, China*

⁴*Oxford Suzhou Centre for Advanced Research, Suzhou Industrial Park, Jiangsu, China*

Background and Objectives: The International Society for Cellular Therapy (ISCT) proposed a set of minimal markers for identifying human mesenchymal stromal cells (hMSCs) in 2007. Since then, with the growing interest of better characterising hMSCs, various additional surface markers have been proposed. However, the impact of how culture conditions, in particular, the culture surface, vary the expression of hMSC markers was overlooked.

Methods and Results: In this study, we utilized the RNA sequencing data on hMSCs cultured on different surfaces to investigate the variation of the proposed hMSC biomarkers. One of the three ISCT proposed positive biomarker, CD90 was found to be significantly down regulated on hMSCs culture on fibrous surfaces when compared to flat surfaces. The detected gene expression values for 177 hMSCs biomarkers compiled from the literature are reported here. Correlation and cluster analysis revealed the existence of different biomarker communities that displayed a similar expression profile. We found a list of hMSCs biomarkers which are the least sensitive to a change in surface properties and another list of biomarkers which are found to have high sensitivity to a change in surface properties.

Conclusions: This study demonstrated that substrate properties have paramount effect on altering the expressions of hMSCs biomarkers and the proposed list of substrate-stable and substrate-sensitive biomarkers would better assist in the population characterisation. However, proteomic level analysis would be essential to confirm the observations noted.

Keywords: Human mesenchymal stromal cells, Surface markers, Cell biomaterial interactions, Next generation sequencing, Quality control, Regenerative medicine

Received: October 29, 2018, Revised: January 25, 2019, Accepted: February 4, 2019, Published online: February 28, 2019

Correspondence to **Hui Wang**

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, UK

Tel: +44-1865-617914, Fax: +44-1865-617701, E-mail: hui.wang@eng.ox.ac.uk

Co-Correspondence to **Hua Ye**

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, UK

Tel: +44-1865-617689, Fax: +44-1865-617701, E-mail: hua.ye@eng.ox.ac.uk

*These authors contributed equally to this work.

†Current affiliation: Division of Tissue Culture, Department of Applied Biology, Biomedical Technology Wing, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Thiruvananthapuram 695012, Kerala, India

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2019 by the Korean Society for Stem Cell Research

Introduction

Mesenchymal stromal cells (MSCs), the clonogenic fibroblastic colony-forming units, hold immense potential as cells of choice in tissue engineering and regenerative medicine applications. Though bone marrow was the first and widely recognized source tissue, the latest investigations have led to the discovery of a wide variety of MSCs-harboring placenta, umbilical cord and virtually any vascularized tissue (1). MSCs can not only be differentiated into tissue-specific lineages but can also be trans-differentiated across lineages (2). Apart from such multipotency, the immunomodulatory property of MSCs is attracting attention (3). While these developments highlight the potential of MSCs, they also lead to some confusion in the field, majorly due to the identity paradox. To this effect, the International Society for Cellular Therapy (ISCT) proposed (a) plastic-adherence nature, (b) CD105⁺, CD73⁺, CD90⁺, and CD45⁻, CD34⁻, CD14⁻ or CD11b⁻, CD79 α ⁻ or CD19⁻ and HLA-DR⁻ surface markers, and (c) tri-potency ability as the defining criteria to identify MSCs (4). Recently, a few others such as Stro-1, CD146 and CD49f were also proposed as candidate markers. However, source dependent differences left MSC identity still obscure. Besides, the discovery of culture conditions dependent effects on surface marker expression, such as induction of SSEA-4 by fetal calf serum (5), and suppression of CD49f by confluency (6), further enhanced the ambiguity over MSCs identity. Pham et al. showed that cryopreservation reduced the expression of CD73 (7). Lv et al. articulated these issues in great detail in their recent review (8).

Along with conventional culture conditions optimization studies, investigations on the use of biomaterials to help translate MSC therapy from bench to bedside are on the rise (9). It is an indisputable fact that the use of biomaterials in facilitating isolation (10), in enhancing *in vitro* expansion (11), in long-term storage (12), in modulating differentiation (13), and in aiding site specific delivery (14), has expanded the horizons of MSCs to the next level. In this regard, it is important to understand the underlying cell – biomaterial interactions since subtle alterations in substrate topography and chemistry influence the cell fate to a significant extent (15-17). For instance, McMurray et al. discovered that a nano structured surface retains stem-cell phenotype and maintains stem-cell growth over eight weeks (18), while, Dalby et al. demonstrated the use of nano scale disorder to stimulate MSCs to produce bone mineral *in vitro* (19). Similarly, Benoit et al. reported

that substrates functionalized with small molecules can control MSCs differentiation encapsulated in hydrogels (20), while, induction of osteogenic and adipogenic differentiation on amine functionalized surfaces in contrast to pristine samples shows the effect of surface chemistry on cell response (21). Whereas there is evidence on how biomaterials act in terms of controlling stem cell differentiation, there is little information on whether or not these biomaterials have any influence on surface marker expression in the context of MSCs expansion.

Such information is needed not just to ensure quality control of the expanded MSCs but to design next-generation biomaterials-based substrates for MSCs expansion. Inspired by observations of Zamparelli et al. (22) and Duffy et al. (23) on the effect of biomaterials on surface antigen expression in MSCs, here, we designed a study to unravel the effects of substrate topography and substrate chemistry on expression levels of surface markers in human bone marrow-derived MSCs (hbm-MSCs). However, unlike conventional flow cytometry or qRT-PCR methods, here we tested the gene expression levels by following the whole transcriptome shotgun sequencing, also known as RNA-sequencing (24, 25). This next-generation sequencing approach helps to (a) evaluate the stability and sensitivity of ISCT and few other known hMSCs markers, (b) identify potential gene networks or molecular pathways associated with these markers, and (c) screen for new substrate-stable and substrate-sensitive candidate markers. For this, we have prepared poly (L-lactide) (PLLA) based substrates with two variables of topography viz. flat (Fl) and fibrous (Fs) and two variables of chemistry viz. pristine (Pr) and aminated (Am). The flat PLLA surface (Fl-PLLA) represents a conventional two-dimensional culture substrate, whereas, the fibrous PLLA (Fs-PLLA) surface represents an advanced three-dimensional culture substrate. The pristine PLLA (Pr-PLLA) surface represents a hydrophobic surface that a majority of synthetic polymers exhibit, whereas, the aminated PLLA (Am-PLLA) surface represents a hydrophilic surface that a majority of natural polymers exhibit. The cell response on the test materials (Fl-Pr-PLLA, Fl-Am-PLLA, Fs-Pr-PLLA and Fs-Am-PLLA) was compared with that on a control plate (TCPS).

Materials and Methods

Surface material, cell culture and RNA extraction

The detailed information on surface preparation, cell culture and RNA extraction has been reported previously

(26). A total of four synthesized surfaces and tissue culture plate were used for cell culture. The four surfaces were synthesized with poly (L) Lactic acid (PLLA), including two types of surface topography (flat – Fl or fibrous -Fs) and two types of surface chemistry (aminated – Am or non-aminated – Pr). The standard tissue culture polystyrene substrate was used as the control surface.

Human bone marrow-derived mesenchymal stem cells (hbm-MSCs, PoieticsTM, cat. no. PT-2501, Lonza UK), were cultured in a proprietary medium (cat. no. PT-3001, Lonza UK), supplemented with antibiotics (10 U/mL penicillin and 10 μ g/mL streptomycin, Thermo Scientific, UK). Cells were routinely incubated in a CO₂ (5%) incubator maintained at 37°C temperature, 95% relative humidity. hbm-MSCs (passage 3) were seeded onto various test and control substrates at a rate of 50,000 cells/well and cultured for 7 days with viability measured at day 3, 5 and 7 using the Alamar blue assay (Thermo Scientific, UK).

RNA extraction was done using the RNeasy Mini Kit and followed the supplier's instructions (Qiagen, UK). RNA agarose gel electrophoresis and Nanodrop spectrophotometer (1000, Thermo Scientific) were used to check the quality and quantity of the extracted RNA (26). RNA was isolated at each time point; however, RNA *Seq* was done for day 3 sample to capture the substrate-induced early commitment of MSCs.

RNA sequencing

Total RNAs were extracted from all samples using a commercial extraction kit (RNeasy Mini kit, Cat No. 74106). Sequencing platform of BGI-500 (BGI, Shenzhen, China) was used to obtain gene expression profiles. SOAP-nuke was used to remove low quality reads. Quality control checks were performed to confirm sequencing saturation and gene mapping distribution. Fragments per Kilobase of Transcript per Million mapped reads (FPKM) value were used to express relative gene abundance. Genes with expression levels less than 1 across all samples were neglected in order to increase data set confidence.

Data analysis

Previously reported 177 hMSCs biomarkers were compiled from the literature (8, 27, 28), including proteoglycans, adhesion molecules, receptors, etc. The full list of genes is included in Supplementary Table S1. The expression levels for these biomarkers were checked against the samples cultured on different surfaces in this study. Hierarchical cluster analysis was conducted on the compiled biomarkers and the total sequenced genes, samples

were grouped based on their pair wise differences.

Gene-wide correlation analysis was performed with $p < 0.01$ and the Pearson coefficient (r) > 0.9 or < -0.9 . Normalized standard deviation, also known as coefficient of variation, was used to quantify the stability or sensitivity of gene expression across samples. The stability and sensitivity of the biomarkers was examined by evaluating the consistency of expression in samples cultured on the different types of surfaces. Differentially expressed genes were determined with greater than one-fold change in pair wise comparison and a false discovery rate less than 0.001.

The biomarker expressions and correlations were used to develop a network structure and was graphically represented using Cytoscape (29). Positive and negative correlations of each gene expression across samples was determined. The average expression levels for each gene and their ranked percentile determines the size of the node. Cluster analysis was conducted using Newman's method, edge between-ness and modularity were determined as the method described in (30). The Octave-networks-toolbox was utilized (31). The modularity for the different number of communities were screened, the number of communities with the highest modularity was selected as the optimal clustering configuration. The optimal clustering configuration was processed using the Newman-Girvan algorithm to generate the distribution and the specific genes for each cluster. The clusters with more than five nodes/genes were considered as true clusters and further analysed.

Results and Discussion

General characteristics of materials and cell response

Details of the material characteristics such as morphological features of Fl-PLLA and Fs-PLLA before and after amination by SEM, confirmation of amine functionalization by FITC staining and ATR-FTIR was previously reported by our group (26). The diameter of electrospun fibres of Fs-PLLA scaffold used in the study was 603 ± 197 nm (Mean \pm SD). There was no significant change in the fibre diameter after amine functionalization, as confirmed by SEM. Besides, qualitative and quantitative analysis of cell adhesion and proliferation on various substrates was also previously reported by our group (26). Briefly, the cellular adhesion and proliferation was relatively high on control TCP substrate compared to test PLLA substrates. This was apparently due to lack of RGD moieties on PLLA. Amongst the test substrates, significant changes in the overall cell response was noted amongst substrates varying in surface topography (Fs-PLLA and Fl-PLLA)

than the surface chemistry (Pr-PLLA and Am-PLLA).

Analysis from the perspective of ISCT recommended markers

The set of surface markers proposed by ISCT has not only streamlined the scientific investigations on hMSCs, but has also assisted commercial organizations in characterising hMSCs-based products for clinical trials (32). Although there is variability in the marker set reported in the literature, in many cases, the ISCT-proposed set of markers is considered the gold standard (8). Following this trend, in the current study, we started off by investigating the expression profiles of ISCT-proposed markers. With no surprises, irrespective of the culture substrate properties, hMSCs expressed *CD105*, *CD73* and *CD90*, and were negative for expression of *CD45*, *CD34*, *CD14*, *CD19* and *HLA-DR* (Fig. 1; the expression levels of the negative markers were undetected or found to be negligible, thus are ignored). However, a close look at the relative expression levels revealed two compelling observations: firstly, the expression level of a given marker varied depending on the culture substrate. Typically, hMSCs cultured on flat substrates (FI-PLLA) showed up regulated surface marker expression as compared to those cultured on fibrous substrates (Fs-PLLA). There was no visible distinction in the surface marker expression in hMSCs cultured on pristine (Pr-PLLA) and aminated (Am-PLLA) substrates. Although the substrate properties were known to modulate the hMSCs fate, little was known of their effects on the surface marker expression levels. Previously,

Skardal et al. described that immune-histochemical (IHC) staining for surface markers such as *CD90* and *CD105* was strong for cells cultured on synthetic substrates with 2~5 kPa stiffness as compared to the cells cultured on stiffer (15, 50 kPa) substrates. Our observations not only support the findings of Skardal et al. but also provide quantitative evidence towards substrate-dependant effects on surface marker expression (33).

The second curious finding from Fig. 1 was that the overall expression levels of *CD105* and *CD90* were positively correlated ($p=0.0186$, $r=0.916$, Fig. 1) whereas correlation between *CD73* and *CD90* was not significant ($p=0.305$, $r=-0.276$, Fig. 1). Intrigued by this observation, we performed a correlation analysis. Out of 17,748 genes, identified by RNA-seq, across hMSCs samples cultured on five different surfaces, a total of 871 genes were found to have a positive or negative correlation. Of these, 464 genes were correlated with *CD90*, 467 genes with *CD105* and 334 genes with *CD73* (Fig. 2). It was again apparent that expression levels of *CD105* and *CD90* were interactive whereas the *CD73* level was independent. Our results comply with those reported by Zamparelli et al. where rat bone marrow-derived MSCs cultured on PLLA substrates showed, using flow cytometry, an $82\pm 8\%$ *CD90* positive population but only a $35\pm 19\%$ *CD73* positive population (22), and by Skardal et al. where human amniotic fluid-derived MSCs cultured on synthetic substrates showed comparable IHC staining for *CD105* and *CD90* markers (33). Furthermore, in the current study, *CD90* and *CD105* shared over 84% ($n=394$) of their in-

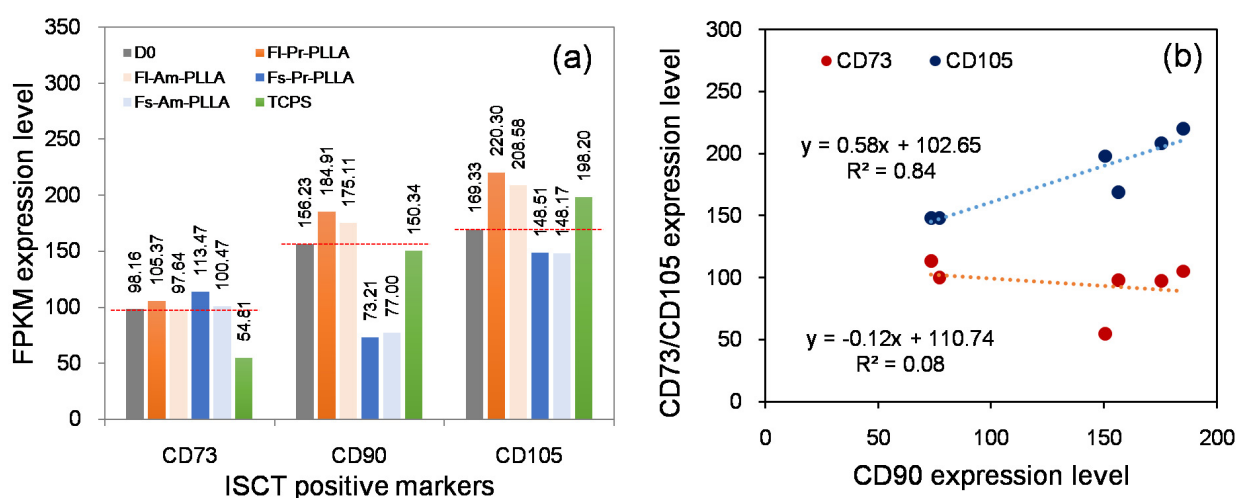


Fig. 1. Expression levels of ISCT recommended hMSCs surface markers: (a) *CD90* and *CD105* showed substrate-sensitive response whereas *CD73* showed substrate-stable response. *CD90* and *CD105* showed variations on expression levels that were higher than that of *CD73*. (b) The FPKM values for *CD90* were plotted against the FPKM values of *CD73* and *CD105*, line of best fit and R-square is shown in the figure.

dividually correlated genes, whereas *CD73* showed no commonly correlated genes with either of the other two markers (Fig. 2). As evident from the NCBI Gene Database, although the genomic context is different, *CD90* (*THY1*) and *CD105* (*ENG*) encode proteins that are involved in cell communication, whereas *CD73* (*NT5E*) encodes a protein that catalyses the conversion of extracellular nucleotides to membrane-permeable nucleosides. Such functional relations may perhaps explain why *CD90* and *CD105* were expressed correlatively. Additionally, based on a report by Colgan et al. suggesting a major involvement of *CD73* in nucleotide metabolism during hypoxia and ischemia, it was most likely that the lower *CD73* expression level in the current study may be attributed to the normoxic culture conditions (34).

Analysis from the perspective of a comprehensive list of markers reported in the literature

Of the three positive ISCT markers, *CD90* and *CD105* might be proposed as a substrate-sensitive marker, while *CD73* appeared as a substrate-stable marker. However, we believe that the concept of substrate-sensitive/stable markers should be consolidated further, thus incorporation of additional markers is needed to propose a comprehensive set of biomaterials-specific markers for potential use in the field of MSCs-based regenerative medicine. The list of 177 hMSCs markers that were reported in various contexts from the literature is examined (Supplementary Table S1). Overall, the expression profiles between them vary to a significant extent. Preferential expressions of certain markers were observed, where, some markers exhibited stable expression and some showed sensitiveness

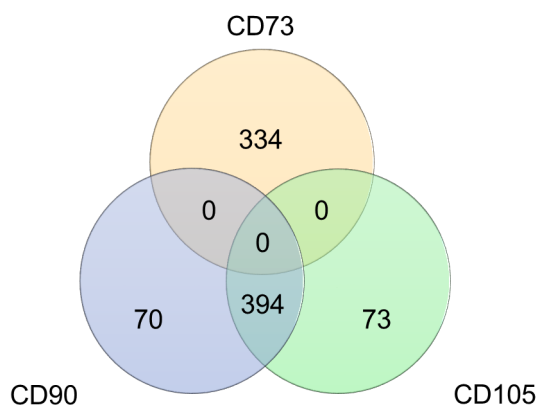


Fig. 2. Genes correlated with the three positive ISCT recommended hMSCs markers: The total number of correlated genes with *CD73* (334) was less than that with *CD90* (464) and *CD105* (467). *CD90* and *CD105* share over 84% of their individually correlated genes, whereas, *CD73* has no common genes with the other two markers.

to a change in culture surface conditions. Expression levels of these 177 biomarkers were subjected to principal component analysis (PCA) analysis and dendrogram analysis. As shown in Fig. 3, the effect of chemistry treatments on the gene expression profile was relatively smaller than that of surface topography induced changes, and the samples cultured on flat surfaces exhibited significantly smaller changes to the control cluster than those in the fibrous samples. This feature conformed to the gene expression patterns reported for all genes (26), strongly suggesting that the marker set genuinely represented the overall gene expression. To our knowledge, this study is the first of its kind reporting the effect of substrate topography and chemistry on the expression of 177 biomarkers reported in the literature.

Proceeding further, we have performed a correlation analysis within these 177 biomarker gene expression levels to seek more insights (Fig. 4). Out of 177 marker genes, 22 genes were found without any significantly correlated pairs and therefore were excluded from the figure. Amongst the rest, both positive (represented by the green line) and negative (represented by the red line) correlations were discovered. The statistical relationships were not straight forward and intertwined for most of the genes. However, certain genes were found to correlate to only one or a few other genes. For easy reference, the results are interpreted with reference to the ISCT markers. As shown in Fig. 4a, the three ISCT markers have been rearranged to the right-hand-side of the figure to highlight their high degrees of connectivity within the network. Overall, the per-

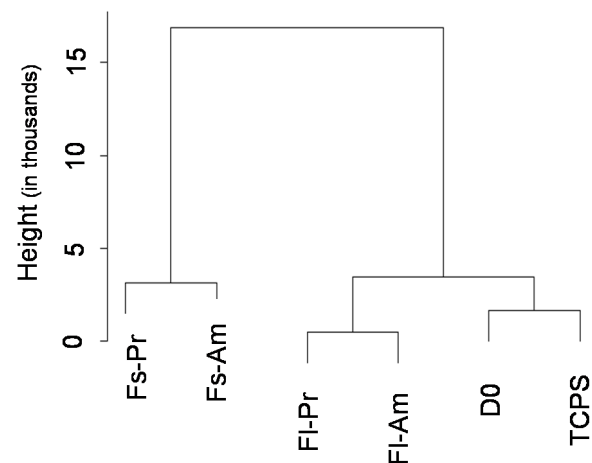


Fig. 3. Substrate induced changes in expression of 177 hMSCs biomarkers compiled from the literature: Dendrogram shows that the expression levels of the biomarkers were significantly altered with respect to changes in topography in comparison with chemistry induced changes.

centiles of the average expression levels across samples are represented by the size of the nodes. The three ISCT markers exhibited considerably significant levels of correlation. A positive correlation was identified between

CD90 (*THY1*) and *CD105* (*ENG*), whereas there was no correlation between *CD73* (*NT5E*) and either of the other two. Further, the networks including these three markers and their first-degree-neighbours are isolated and pre-

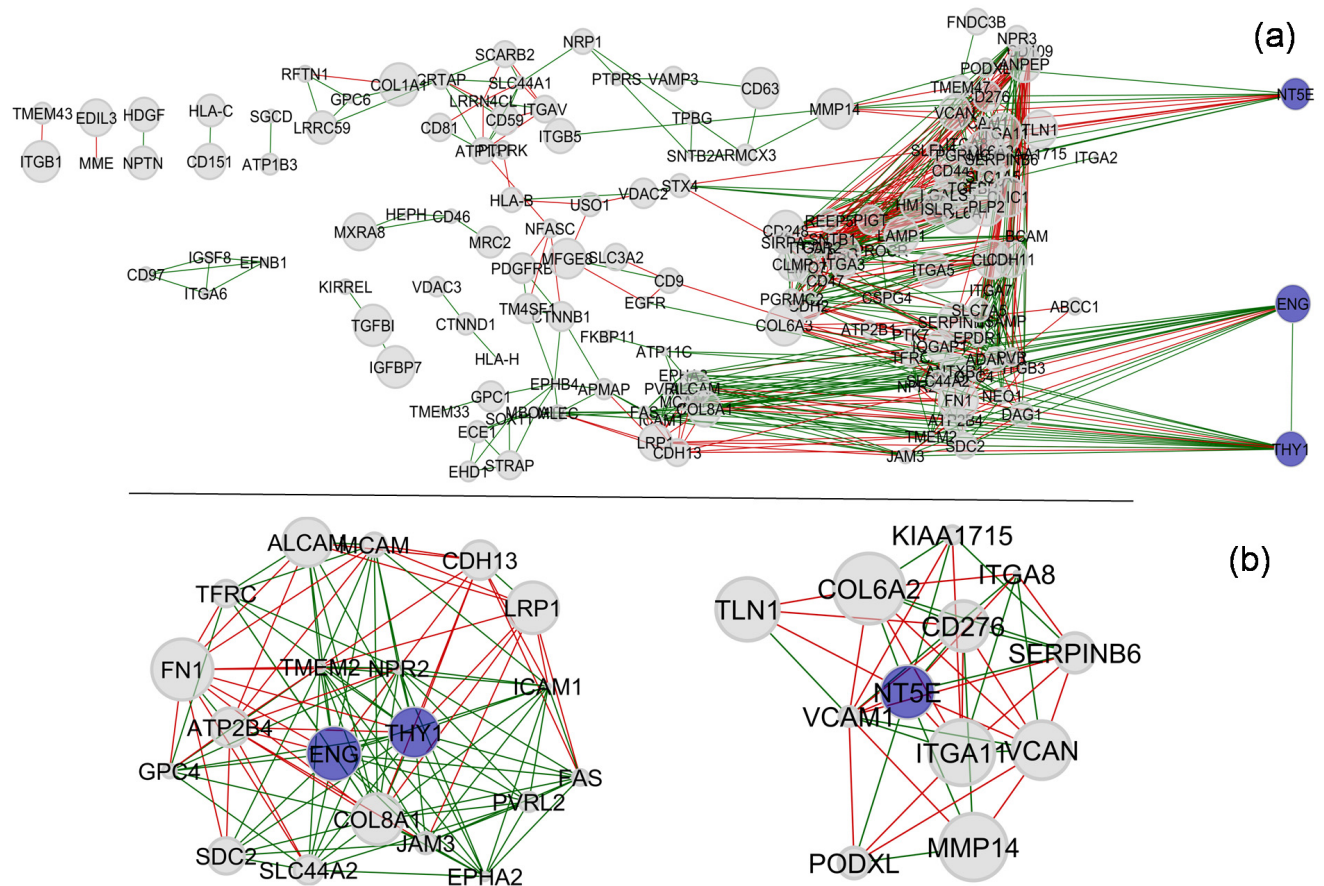


Fig. 4. Correlation analysis of gene expression levels of 177 hMSCs biomarkers compiled from the literature: (a) overall observation with reference to three ISCT markers highlighted and (b) first-degree-connections of ISCT markers suggested correlation between *CD90* and *CD105* including its neighbors but no correlation with that of *CD73* and its neighbors. A negative correlation between two genes is shown by a red line and a positive correlation is shown by a green line. The size of the node represents the mean FPKM percentile for that particular gene.

Table 1. List of biomaterials-stable hMSCs markers identified from the 177 biomarkers reported in the literature

Gene ID	Gene Name	D0	FI-Pr	FI-Am	Fs-Pr	Fs-Am	TCP	CoV
1495	CTNNA1	117.4	120.1	124.3	117.8	127.6	119.3	0.03
10085	EDIL3	176.2	160.4	159.7	173.2	166.4	172.0	0.04
3916	LAMP1	195.8	187.7	193.8	203.4	194.5	177.4	0.05
6717	SRI	26.0	28.9	25.2	25.8	28.0	26.8	0.05
3490	IGFBP7	1375.9	1304.5	1208.1	1211.9	1166.4	1273.2	0.06
781	CACNA2D1	13.1	12.3	11.8	13.2	11.6	13.6	0.06
6443	SGCB	30.1	31.0	32.3	28.1	27.6	27.8	0.06
8910	SGCE	10.9	10.9	10.9	9.3	11.2	10.5	0.07
3688	ITGB1	810.3	787.9	801.8	942.6	834.5	819.7	0.07
9217	VAPB	8.9	9.0	9.8	7.9	8.7	9.0	0.07

sented in Fig. 4b. There are quite a few first-degree-neighbours of *CD90* or *CD105* which showed correlation within this network (Supplementary Table S2). However,

none of the first-degree-neighbours of *CD73* exhibited a correlation with those of *CD90* or *CD105*. This may perhaps be due to the functional context of these genes as

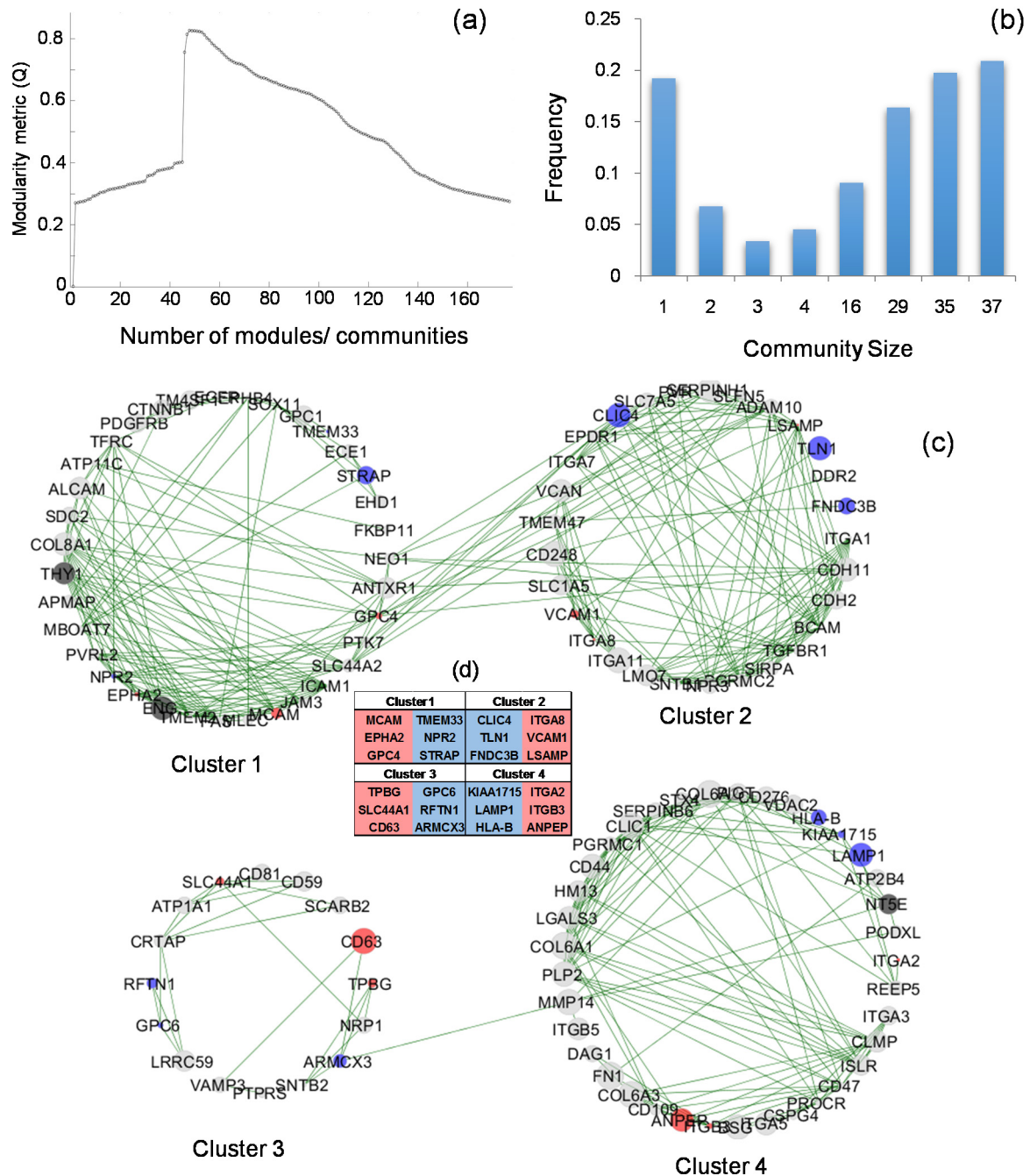


Fig. 5. Cluster analysis of the 177 compiled hMSCs markers: (a) network modularity determination figure showing the configuration of 48 communities resulted in the highest modularity. (b) Gene frequency graph, the frequency (number) of genes that was allocated into the different size of communities, as a result of the optimal cluster configuration. (c) The clustered network consists of the circular subplots; each of the parameter circles represents a community/cluster. Each green line represents the positive correlation of the gene expressions. The interactions within each community are noticeably more than the interactions with other communities. The size of the node represents the average expression level of that gene. The three positive ISCT markers are highlighted in grey. The top three stable or variably expressed genes in each community are highlighted in (c) and the gene names in panel (d), in blue and red, respectively.

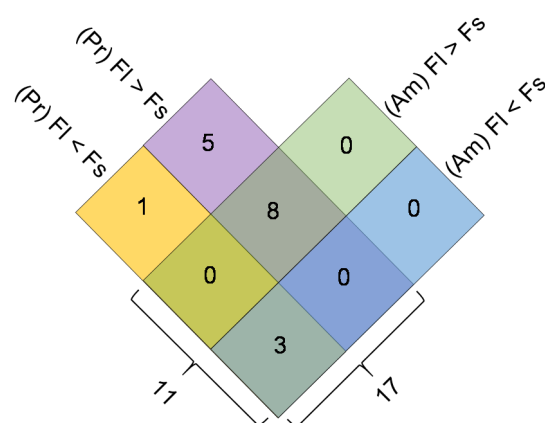
discussed earlier.

Findings on biomaterial- stable and sensitive markers

Besides the conventional ISCT markers, we examined the stability of the expressions for the 177 biomarkers as a result of being cultured on different biomaterials. The stability of the gene was quantified by the coefficient of variance (CoV) across samples. A smaller CoV entails a higher independence of this biomarker gene to the culture substrate. The top 10 genes with the lowest coefficient of variance are shown in Table 1. None of these top 10 genes was found to correlate to the three positive ISCT markers.

Furthermore, we examined the connectedness of the 177 expressed biomarkers. The positive correlation results from the previous analysis (section 3.2) were subjected to cluster analysis to identify the community groupings that maintain the highest level of network integrity. The grouping that yielded the highest network modularity ($Q=0.82$) was taken as the set of the most representative clusters (Fig. 5a). A total of 48 communities were identified in this network arrangement, among which four communities consisted of more than five genes. The frequency plot describing the number of genes in each cluster size is shown in Fig. 5b. The correlation between genes and clusters are shown in Fig. 5c. The list of genes in each of these four communities are closely correlated with each other, in other words, the expression of one marker gene is expected to represent that of the others; the top three most stably (in blue) or variably expressed (in red) genes from each of the communities are shown in Fig. 5d as potential marker genes. Representative genes (stably ex-

pressed genes in Fig. 5d) from each community can be selected to capture the expression pattern for that population of markers. Consistent with earlier findings, *CD90* and *CD105* belong to the same community, while *CD73* belongs to another community, and there is no correlation detected between the *CD73* community and the *CD90/CD105* community.



1: ITGA2; 3: FN1, CD9, ITGB3; 5: SLC44A2, PVRL2, LMO7, BCAM, ITGA1; 8: THY1, TMEM2, COL8A1, NPR3, ANTXR1, GPC4E, PHA2, MCAM

Fig. 6. Identification of hMSCs biomarkers that respond to changes in substrate topography: The number of differentially regulated genes (>for up-regulation and <for down-regulation) are shown in each block, along with the gene names listed. The two chemistry substrates are represented by Pr and Am. The underlined genes were also differentially expressed compared to those from TCP.

Table 2. List of biomaterials-sensitive hMSCs markers identified from the 177 biomarkers reported in the literature

Gene ID	Gene Name	D0	FI-Pr	FI-Am	Fs-Pr	Fs-Am	TCP	CoV
4162	MCAM	6.5	34.3	36.1	9.8	9.4	26.9	0.66
3690	ITGB3	10.7	5.6	5.4	17.8	13.9	3.3	0.60
1969	EPHA2	6.2	8.9	6.5	1.8	1.9	6.3	0.53
928	CD9	61.6	26.7	29.9	59.7	83.1	25.3	0.50
3672	ITGA1	4.6	2.6	3.2	1.2	1.7	4.3	0.48
3673	ITGA2	5.8	2.9	3.3	6.9	4.0	1.7	0.46
4059	BCAM	2.2	4.5	4.2	1.8	2.5	5.5	0.43
2239	GPC4	8.0	13.1	13.1	5.3	5.7	13.5	0.40
2335	FN1	7292.7	7148.9	7036.7	15578.7	14296.8	8124.2	0.40
84168	ANTXR1	73.6	121.5	135.6	54.5	60.2	129.1	0.39
4883	NPR3	42.6	51.5	54.2	25.5	24.3	67.7	0.38
1295	COL8A1	113.8	205.7	196.0	85.1	89.0	171.8	0.38
4008	LMO7	92.4	114.1	129.1	54.0	71.3	160.7	0.38
23670	TMEM2	5.5	9.1	8.8	3.3	4.1	7.5	0.38
7070	THY1	156.2	184.9	175.1	73.2	77.0	150.3	0.36
5819	PVRL2	17.7	22.6	17.7	9.4	9.1	16.9	0.34
57153	SLC44A2	29.9	39.2	35.2	19.3	20.4	36.5	0.28

To further investigate the surface sensible markers, we looked at the biomarkers that displayed greater dependency to a change in culture surfaces, in particular, a change in surface topography (between flat and fibrous surfaces). The 17 biomarkers that were found to be differentially expressed between the samples cultured on different topographies are shown in Table 2, including one of the ISCT markers, *CD90* (*THY1*). Among these biomarkers, 13 displayed significant down-regulation when cultured on fibrous surfaces, whereas the other four experienced significant up-regulation.

The impact of different chemistry substrates did not significantly alter the expression levels of any of the studied biomarkers. To further illustrate the interactions of surface topography and chemistry, a Venn diagram was prepared to show its differential regulation on both pristine (Pr) and aminated (Am) substrates, as presented in Fig. 6. The numbers of up- and down-regulated genes along with gene names are shown for each chemistry substrate. A total of 17 and 11 biomarkers were found to be differentially regulated by topography, on pristine-based and aminated substrate surfaces, respectively. Among which, nine biomarkers were consistently down-regulated on fibrous surfaces for both substrates, including *CD90*. The three biomarkers that were significantly up-regulated on fibrous surfaces for both substrates were *FNI*, *CD9* and *ITGB3*. These 11 (eight down-regulated and three up-regulated) genes provide strong evidence demonstrating the sensitivity of these biomarkers to a change in surface topography.

Conclusions

To conclude, we hypothesized that the transcriptome analysis of human bone marrow derived MSCs (hbm-MSCs) cultured on standard tissue culture-treated polystyrene (TCP) and poly (L-lactide) (PLLA) based artificial substrates with varying topography (F1: flat and F5: fibrous) and chemistry (Pr: pristine and Am: aminated) could offer novel insights into substrate-specific markers. Apart from ISCT proposed markers, we have identified and compiled 177 markers that are reported in the literature in various contexts. Comparative gene expression level analysis suggested that *CD90* and *CD105* showed comparable substrate-sensitive response whereas *CD73* showed substrate-stable response. A correlative analysis found 871 genes, out of 17748, to be correlated with ISCT markers, of which, *CD90* and *CD105* showed comparable results that are distinct from *CD73*. Subsequently, a dendrogram of 177 markers suggested that chemistry induced

changes on gene expression profile was relatively smaller than that of surface topography induced changes, and the samples cultured on flat surfaces exhibited significantly smaller differences than those in the fibrous samples. From the correlation analysis, the first-degree-connections of ISCT markers and the 177 markers found in the literature showed a correlation between *CD90* and *CD105* including its neighbors but no correlation with that of *CD73* and its neighbors. Positive correlation results yielded a top-10 list of substrate-stable and sensitive markers for hMSCs studies in the context of biomaterials research. Furthermore, among the 48 communities identified from the cluster analysis, the 3 most representative and variable genes were reported for each cluster with greater than 5 genes. The study indicates that an omics approach can help analyse the variations in a comprehensive manner and perhaps give clarity on the MSCs identity in the context of biomaterial research. However, in order to confirm that the changes observed at the gene transcription level are reflected at the protein level, a proteomic analysis may be required.

Acknowledgments

We are grateful to China Regenerative Medicine International Limited, Hong Kong for funding this work.

Potential Conflict of Interest

The authors have no conflicting financial interest.

Author Contributions

NK, HY and HW conceived and designed the study. NK, BZ and JM performed the experiments. NK, BZ, QL, JM and HW performed data analysis. NK, BZ and HW wrote the manuscript. ZC, HW, HY and AY reviewed the manuscripts.

Supplementary Materials

Supplementary data including two table can be found with this article online at <http://pdf.medrang.co.kr/paper/pdf/IJSC/IJSC-12-s18102.pdf>.

References

1. Ullah I, Subbarao RB, Rho GJ. Human mesenchymal stem cells - current trends and future prospective. *Biosci Rep* 2015;35. pii: e00191
2. Song L, Tuan RS. Transdifferentiation potential of human mesenchymal stem cells derived from bone marrow. *FASEB J* 2004;18:980-982
3. Gao F, Chiu SM, Motan DA, Zhang Z, Chen L, Ji HL, Tse

- HF, Fu QL, Lian Q. Mesenchymal stem cells and immunomodulation: current status and future prospects. *Cell Death Dis* 2016;7:e2062
4. Dominici M, Le Blanc K, Mueller I, Slaper-Cortenbach I, Marini F, Krause D, Deans R, Keating A, Prockop DJ, Horwitz E. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* 2006;8:315-317
5. Suila H, Pitkänen V, Hirvonen T, Heiskanen A, Anderson H, Laitinen A, Natunen S, Miller-Podraza H, Satomaa T, Natunen J, Laitinen S, Valmu L. Are globoseries glycosphingolipids SSEA-3 and -4 markers for stem cells derived from human umbilical cord blood? *J Mol Cell Biol* 2011;3:99-107
6. Lee RH, Seo MJ, Pulin AA, Gregory CA, Ylostalo J, Prockop DJ. The CD34-like protein PODXL and alpha6-integrin (CD49f) identify early progenitor MSCs with increased clonogenicity and migration to infarcted heart in mice. *Blood* 2009;113:816-826
7. Pham H, Tonai R, Wu M, Birtolo C, Chen M. CD73, CD90, CD105 and Cadherin-11 RT-PCR screening for mesenchymal stem cells from cryopreserved human cord tissue. *Int J Stem Cells* 2018;11:26-38
8. Lv FJ, Tuan RS, Cheung KM, Leung VY. Concise review: the surface markers and identity of human mesenchymal stem cells. *Stem Cells* 2014;32:1408-1419
9. Zhang Z, Gupte MJ, Ma PX. Biomaterials and stem cells for tissue engineering. *Expert Opin Biol Ther* 2013;13:527-540
10. Mahara A, Yamaoka T. Continuous separation of cells of high osteoblastic differentiation potential from mesenchymal stem cells on an antibody-immobilized column. *Biomaterials* 2010;31:4231-4237
11. Chen AK, Reuveny S, Oh SK. Application of human mesenchymal and pluripotent stem cell microcarrier cultures in cellular therapy: achievements and future direction. *Biotechnol Adv* 2013;31:1032-1046
12. Sambu S, Xu X, Schiffer HA, Cui ZF, Ye H. RGDS-fuctionalized alginates improve the survival rate of encapsulated embryonic stem cells during cryopreservation. *Cryo Letters* 2011;32:389-401
13. Lee J, Abdeen AA, Kilian KA. Rewiring mesenchymal stem cell lineage specification by switching the biophysical microenvironment. *Sci Rep* 2014;4:5188
14. Choi YS, Park SN, Suh H. Adipose tissue engineering using mesenchymal stem cells attached to injectable PLGA spheres. *Biomaterials* 2005;26:5855-5863
15. Dalby MJ, Gadegaard N, Oreffo RO. Harnessing nanotopography and integrin-matrix interactions to influence stem cell fate. *Nat Mater* 2014;13:558-569
16. Walters NJ, Gentleman E. Evolving insights in cell-matrix interactions: elucidating how non-soluble properties of the extracellular niche direct stem cell fate. *Acta Biomater* 2015;11:3-16
17. Jang HK, Kim BS. Modulation of stem cell differentiation with biomaterials. *Int J Stem Cells* 2010;3:80-84
18. McMurray RJ, Gadegaard N, Tsimbouri PM, Burgess KV, McNamara LE, Tare R, Murawski K, Kingham E, Oreffo RO, Dalby MJ. Nanoscale surfaces for the long-term maintenance of mesenchymal stem cell phenotype and multipotency. *Nat Mater* 2011;10:637-644
19. Dalby MJ, Gadegaard N, Tare R, Andar A, Riehle MO, Herzyk P, Wilkinson CD, Oreffo RO. The control of human mesenchymal cell differentiation using nanoscale symmetry and disorder. *Nat Mater* 2007;6:997-1003
20. Benoit DS, Schwartz MP, Durney AR, Anseth KS. Small functional groups for controlled differentiation of hydrogel-encapsulated human mesenchymal stem cells. *Nat Mater* 2008;7:816-823
21. Phillips JE, Petrie TA, Creighton FP, García AJ. Human mesenchymal stem cell differentiation on self-assembled monolayers presenting different surface chemistries. *Acta Biomater* 2010;6:12-20
22. Zamparelli A, Zini N, Cattini L, Spaletta G, Dallatana D, Bassi E, Barbaro F, Iafisco M, Mosca S, Parrilli A, Fini M, Giardino R, Sandri M, Sprio S, Tampieri A, Maraldi NM, Toni R. Growth on poly(L-lactic acid) porous scaffold preserves CD73 and CD90 immunophenotype markers of rat bone marrow mesenchymal stromal cells. *J Mater Sci Mater Med* 2014;25:2421-2436
23. Duffy CR, Zhang R, How SE, Lilienkamp A, De Sousa PA, Bradley M. Long term mesenchymal stem cell culture on a defined synthetic substrate with enzyme free passaging. *Biomaterials* 2014;35:5998-6005
24. Kasoju N, Wang H, Zhang B, George J, Gao S, Triffitt JT, Cui Z, Ye H. Transcriptomics of human multipotent mesenchymal stromal cells: retrospective analysis and future prospects. *Biotechnol Adv* 2017;35:407-418
25. Sargent A, Shano G, Karl M, Garrison E, Miller C, Miller RH. Transcriptional profiling of mesenchymal stem cells identifies distinct neuroimmune pathways altered by CNS disease. *Int J Stem Cells* 2018;11:48-60
26. Li Q, Zhang B, Kasoju N, Ma J, Yang A, Cui Z, Wang H, Ye H. Differential and interactive effects of substrate topography and chemistry on human mesenchymal stem cell gene expression. *Int J Mol Sci* 2018;19. pii: E2344
27. Niehage C, Steenblock C, Pursche T, Bornhäuser M, Corbeil D, Hoflack B. The cell surface proteome of human mesenchymal stromal cells. *PLoS One* 2011;6:e20399
28. Cho KA, Park M, Kim YH, Woo SY, Ryu KH. RNA sequencing reveals a transcriptomic portrait of human mesenchymal stem cells from bone marrow, adipose tissue, and palatine tonsils. *Sci Rep* 2017;7:17114
29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-2504
30. Girvan M, Newman ME. Community structure in social and biological networks. *Proc Natl Acad Sci U S A* 2002;99:7821-7826
31. Bounova G. Octave networks toolbox. *Zendo* 2015 doi:10.5281/zenodo.22398.

32. Mendicino M, Bailey AM, Wonnacott K, Puri RK, Bauer SR. MSC-based product characterization for clinical trials: an FDA perspective. *Cell Stem Cell* 2014;14:141-145
33. Skardal A, Mack D, Atala A, Soker S. Substrate elasticity controls cell proliferation, surface marker expression and motile phenotype in amniotic fluid-derived stem cells. *J Mech Behav Biomed Mater* 2013;17:307-316
34. Colgan SP, Eltzschig HK, Eckle T, Thompson LF. Physiological roles for ecto-5'-nucleotidase (CD73). *Purinergic Signal* 2006;2:351-360

Table S1. The expression for the compiled 177 hMSCs biomarkers identified in the literature

Gene ID	Gene Name	D0	Fl-Pr	Fl-Am	Fs-Pr	Fs-Am	TCP	Average	Median	Stdev	normalized stdev/coef- ficient of variation
1495	CTNNA1	117.4	120.1	124.3	117.8	127.6	119.3	121.8	120.1	4.0	0.03
10085	EDIL3	176.2	160.4	159.7	173.2	166.4	172.0	166.3	166.4	6.3	0.04
3490	IGFBP7	1375.9	1304.5	1208.1	1211.9	1166.4	1273.2	1232.8	1211.9	55.3	0.04
25923	ATL3	27.4	23.0	24.0	23.7	22.0	21.6	22.9	23.0	1.0	0.05
3916	LAMP1	195.8	187.7	193.8	203.4	194.5	177.4	191.4	193.8	9.6	0.05
5796	PTPRK	29.7	25.7	26.3	25.6	23.2	26.7	25.5	25.7	1.3	0.05
6717	SRI	26.0	28.9	25.2	25.8	28.0	26.8	26.9	26.8	1.5	0.06
1500	CTNND1	24.4	33.4	32.6	30.1	35.1	31.4	32.5	32.6	1.9	0.06
10082	GPC6	14.7	9.2	9.1	8.8	8.2	8.0	8.7	8.8	0.5	0.06
7045	TGFB1	1635.7	1003.3	939.6	930.3	857.7	992.1	944.6	939.6	58.0	0.06
23114	NFASC	21.7	24.3	25.2	27.0	27.1	28.3	26.4	27.0	1.6	0.06
51566	ARMCX3	24.5	32.5	31.2	32.1	30.5	27.6	30.8	31.2	1.9	0.06
4179	CD46	20.9	16.8	14.8	15.6	15.9	14.3	15.5	15.6	1.0	0.06
64778	FNDC3B	61.6	47.6	51.1	48.6	46.4	54.8	49.7	48.6	3.3	0.07
11171	STRAP	72.3	70.4	65.5	62.4	58.3	61.9	63.7	62.4	4.5	0.07
8910	SGCE	10.9	10.9	10.9	9.3	11.2	10.5	10.5	10.9	0.8	0.07
781	CACNA2D1	13.1	12.3	11.8	13.2	11.6	13.6	12.5	12.3	0.9	0.07
6443	SGCB	30.1	31.0	32.3	28.1	27.6	27.8	29.4	28.1	2.1	0.07
3688	ITGB1	810.3	787.9	801.8	942.6	834.5	819.7	837.3	819.7	61.5	0.07
3105	HLA-A	260.6	238.8	204.9	219.7	199.1	204.9	213.5	204.9	16.1	0.08
9217	VAPB	8.9	9.0	9.8	7.9	8.7	9.0	8.9	9.0	0.7	0.08
7416	VDAC1	102.3	111.1	118.7	120.0	107.3	98.7	111.1	111.1	8.7	0.08
23180	RFTN1	19.4	23.8	23.7	23.1	20.1	20.6	22.3	23.1	1.8	0.08
118429	ANTXR2	15.9	12.8	14.4	15.3	13.8	12.5	13.8	13.8	1.1	0.08
55161	TMEM33	5.3	4.5	4.5	4.1	3.7	3.9	4.1	4.1	0.3	0.08
27020	NPTN	99.9	97.8	109.1	96.8	112.7	92.1	101.7	97.8	8.7	0.09
9218	VAPA	17.2	15.8	16.7	18.9	15.5	15.7	16.5	15.8	1.4	0.09
79188	TMEM43	45.1	33.9	33.1	27.0	31.2	32.7	31.6	32.7	2.7	0.09
481	ATP1B1	9.2	9.8	11.3	10.9	12.5	10.7	11.1	10.9	1.0	0.09
3728	JUP	3.0	5.4	4.9	4.2	4.6	4.6	4.8	4.6	0.4	0.09
950	SCARB2	58.7	54.3	54.0	50.1	48.6	43.3	50.1	50.1	4.5	0.09
80856	KIAA1715	12.6	11.6	11.4	12.7	11.5	9.7	11.4	11.5	1.1	0.09
3068	HDGF	84.2	112.2	124.2	111.6	129.6	102.8	116.1	112.2	10.8	0.09
10134	BCAP31	59.7	70.3	63.9	59.8	68.5	55.5	63.6	63.9	6.1	0.10
6645	SNTB2	5.2	5.3	5.3	5.5	4.9	4.3	5.0	5.3	0.5	0.10
8826	IQGAP1	80.3	73.6	81.0	92.1	89.4	76.5	82.5	81.0	8.0	0.10

Table S1. Continued

Gene ID	Gene Name	D0	FI-Pr	FI-Am	Fs-Pr	Fs-Am	TCP	Average	Median	Stdev	normalized stdev/coef- ficient of variation
3106	HLA-B	54.5	44.9	42.4	45.3	51.3	38.7	44.5	44.9	4.6	0.10
4267	CD99	403.0	572.5	708.4	644.0	646.5	544.8	623.2	644.0	65.1	0.10
10491	CRTAP	31.5	33.0	31.8	30.3	27.7	25.4	29.6	30.3	3.1	0.10
10857	PGRMC1	28.2	35.8	34.5	38.9	37.6	29.4	35.2	35.8	3.7	0.10
257194	NEGR1	9.1	10.6	10.6	10.6	8.2	9.5	9.9	10.6	1.1	0.11
9902	MRC2	99.8	108.8	84.3	94.3	95.4	83.7	93.3	94.3	10.2	0.11
4882	NPR2	7.5	8.1	7.6	6.4	6.3	7.6	7.2	7.6	0.8	0.11
3107	HLA-C	120.9	104.1	84.5	106.6	108.3	87.7	98.2	104.1	11.2	0.11
7094	TLN1	125.8	178.4	199.4	169.8	190.6	228.1	193.2	190.6	22.5	0.12
9843	HEPH	4.0	5.4	4.3	4.9	5.1	4.1	4.8	4.9	0.6	0.12
54867	TMEM214	34.3	39.5	33.7	32.0	34.9	28.5	33.7	33.7	4.0	0.12
3685	ITGAV	78.0	47.9	51.9	53.9	55.3	65.7	54.9	53.9	6.6	0.12
25932	CLIC4	209.0	210.3	200.7	162.4	177.7	219.0	194.0	200.7	23.4	0.12
5269	SERPINB6	40.5	39.5	39.1	43.7	41.3	31.2	38.9	39.5	4.7	0.12
7905	REEP5	45.1	34.0	36.0	41.3	36.2	29.4	35.4	36.0	4.3	0.12
8615	USO1	37.8	36.3	39.9	34.0	29.7	40.9	36.2	36.3	4.5	0.13
4240	MFGF8	400.9	328.5	323.3	370.4	418.4	310.0	350.1	328.5	44.4	0.13
221294	NT5DC1	5.0	5.1	5.5	3.8	5.1	5.1	4.9	5.1	0.6	0.13
1889	ECE1	38.0	37.6	33.3	30.1	26.5	31.5	31.8	31.5	4.1	0.13
8829	NRP1	71.3	50.8	50.8	50.2	43.1	37.4	46.5	50.2	6.0	0.13
476	ATP1A1	64.4	57.3	51.3	48.8	46.8	39.8	48.8	48.8	6.4	0.13
3956	LGALS1	1092.8	1156.2	966.1	1051.1	870.7	845.0	977.8	966.1	128.9	0.13
5802	PTPRS	12.7	16.2	13.5	16.4	13.7	11.9	14.3	13.7	1.9	0.13
81502	HM13	104.4	99.9	91.9	109.6	111.7	79.8	98.6	99.9	13.2	0.13
977	CD151	164.0	153.3	120.5	163.5	161.8	129.3	145.7	153.3	19.6	0.13
55379	LRRC59	96.3	116.7	111.2	102.6	86.6	87.3	100.9	102.6	13.6	0.14
1499	CTNNB1	69.5	85.4	80.0	66.3	64.9	64.5	72.2	66.3	9.8	0.14
91663	MYADM	43.9	57.3	50.7	73.2	59.2	64.0	60.9	59.2	8.4	0.14
483	ATP1B3	46.7	32.1	32.3	32.0	35.0	23.7	31.0	32.1	4.3	0.14
682	BSG	248.2	250.6	216.3	273.9	238.5	189.6	233.8	238.5	32.3	0.14
3655	ITGA6	6.2	9.8	8.2	8.0	7.2	7.0	8.0	8.0	1.1	0.14
51604	PIGT	67.3	65.0	66.3	78.9	70.2	53.1	66.7	66.3	9.3	0.14
1605	DAG1	38.5	42.5	42.8	57.1	50.3	41.4	46.8	42.8	6.8	0.14
7417	VDAC2	78.1	75.0	68.1	77.6	88.1	58.6	73.5	75.0	11.0	0.15
4363	ABCC1	19.2	22.1	23.0	25.5	28.3	18.8	23.5	23.0	3.6	0.15
953	ENTPD1	1.5	1.1	0.8	1.2	0.9	1.0	1.0	1.0	0.1	0.15

Table S1. Continued

Gene ID	Gene Name	D0	FI-Pr	FI-Am	Fs-Pr	Fs-Am	TCP	Average	Median	Stdev	normalized stdev/coef- ficient of variation
1956	EGFR	17.0	12.1	12.7	10.1	8.7	12.3	11.2	12.1	1.7	0.15
490	ATP2B1	31.3	25.2	24.4	22.8	19.7	30.1	24.5	24.4	3.8	0.16
7037	TFRC	37.2	32.2	32.9	25.1	23.0	31.5	28.9	31.5	4.6	0.16
966	CD59	139.9	95.0	85.0	83.3	75.5	61.1	80.0	83.3	12.6	0.16
6383	SDC2	88.6	53.8	52.2	36.3	41.6	50.5	46.9	50.5	7.6	0.16
975	CD81	98.7	75.0	62.5	59.1	58.7	47.6	60.6	59.1	9.8	0.16
162394	SILFN5	23.9	18.9	20.2	17.3	16.3	24.4	19.4	18.9	3.2	0.16
1192	CLIC1	295.7	279.7	280.6	339.4	301.7	212.4	282.8	280.6	46.2	0.16
51303	FKBP11	14.3	13.2	15.5	11.1	10.3	13.7	12.8	13.2	2.1	0.16
2817	GPC1	90.9	91.7	89.9	73.6	62.0	70.9	77.6	73.6	12.8	0.16
23385	NCSTN	52.0	48.8	53.0	39.1	54.9	62.3	51.6	53.0	8.5	0.17
7419	VDAC3	27.4	31.8	32.4	23.3	35.6	26.2	29.9	31.8	5.0	0.17
55243	KIRREL	19.1	19.6	17.0	16.1	12.7	19.5	17.0	17.0	2.8	0.17
3693	ITGB5	148.1	144.0	126.3	127.4	128.0	88.0	122.8	127.4	20.7	0.17
3675	ITGA3	67.2	67.7	57.3	81.4	73.6	54.1	66.8	67.7	11.3	0.17
9341	VAMP3	50.4	49.1	40.9	50.1	38.1	33.3	42.3	40.9	7.2	0.17
1012	CDH13	82.5	47.5	47.0	65.5	68.0	56.5	56.9	56.5	9.8	0.17
4756	NEO1	13.9	20.0	20.7	14.0	15.8	20.8	18.3	20.0	3.1	0.17
214	ALCAM	118.6	131.0	138.7	94.5	96.6	122.5	116.7	122.5	20.1	0.17
83604	TMEM47	44.1	31.5	35.0	31.0	30.0	44.5	34.4	31.5	5.9	0.17
967	CD63	694.8	911.0	791.9	888.2	693.7	588.8	774.7	791.9	135.0	0.17
6520	SLC3A2	69.8	61.0	55.8	46.9	38.4	57.0	51.8	55.8	9.1	0.18
4921	DDR2	32.6	22.6	20.3	17.1	20.9	27.5	21.7	20.9	3.8	0.18
3678	ITGA5	238.7	165.0	135.9	192.5	158.7	121.4	154.7	158.7	27.4	0.18
6810	STX4	26.6	37.6	32.5	40.7	42.7	26.8	36.0	37.6	6.5	0.18
4311	MME	9.9	11.0	11.2	7.8	8.8	7.7	9.3	8.8	1.7	0.18
54587	MXRA8	160.1	210.8	153.7	187.9	181.9	129.7	172.8	181.9	31.5	0.18
2022	ENG	169.3	220.3	208.6	148.5	148.2	198.2	184.8	198.2	34.1	0.18
83700	JAM3	20.9	21.9	20.3	13.9	15.4	18.3	18.0	18.3	3.3	0.19
355	FAS	8.9	14.2	12.9	9.6	9.1	11.5	11.5	11.5	2.1	0.19
3136	HLA-H	2.8	2.7	2.4	2.0	3.1	2.1	2.5	2.4	0.5	0.19
5355	PLP2	173.4	220.0	207.1	266.6	240.3	153.6	217.5	220.0	42.2	0.19
1277	COL1A1	7846.3	6205.6	6507.3	6103.8	8942.6	8808.9	7313.6	6507.3	1434.5	0.20
1292	COL6A2	1228.7	1349.4	1239.6	1517.2	1392.6	860.4	1271.8	1349.4	250.6	0.20
1291	COL6A1	1191.9	1215.0	1171.8	1482.5	1429.7	868.2	1233.4	1215.0	244.0	0.20
57136	APMAP	35.2	38.4	35.7	24.7	25.8	28.1	30.5	28.1	6.2	0.20

Table S1. Continued

Gene ID	Gene Name	D0	FI-Pr	FI-Am	Fs-Pr	Fs-Am	TCP	Average	Median	Stdev	normalized stdev/coef- ficient of variation
7162	TPBG	20.8	16.5	15.7	16.7	13.6	9.7	14.4	15.7	2.9	0.20
960	CD44	286.1	185.9	192.6	230.8	226.5	131.7	193.5	192.6	39.9	0.21
976	CD97	18.5	21.7	17.0	15.2	13.6	13.7	16.2	15.2	3.4	0.21
102	ADAM10	42.5	44.2	42.9	28.4	31.5	46.6	38.7	42.9	8.2	0.21
4035	LRP1	134.8	122.4	121.1	190.2	186.2	157.7	155.5	157.7	33.3	0.21
961	CD47	12.1	9.1	8.6	12.3	10.7	7.0	9.5	9.1	2.0	0.21
871	SERPINH1	352.4	247.9	218.4	165.3	157.7	252.9	208.5	218.4	44.9	0.22
10938	EHD1	32.5	39.5	30.7	28.2	21.5	29.8	29.9	29.8	6.4	0.22
79827	CLMP	130.3	103.5	98.1	137.7	128.3	78.8	109.3	103.5	23.7	0.22
9761	MLEC	19.5	24.8	24.6	16.4	15.6	18.9	20.1	18.9	4.4	0.22
6510	SLC1A5	132.7	91.8	87.6	69.0	72.6	117.8	87.7	87.6	19.4	0.22
2050	EPHB4	6.1	7.5	6.9	5.1	4.5	5.2	5.8	5.2	1.3	0.22
93185	IGSF8	10.4	13.6	10.6	10.3	7.8	8.2	10.1	10.3	2.3	0.23
80381	CD276	100.5	87.5	73.5	91.3	87.3	47.5	77.4	87.3	18.0	0.23
6444	SGCD	16.0	15.5	16.5	14.6	16.1	8.2	14.2	15.5	3.4	0.24
4907	NT5E	98.2	105.4	97.6	113.5	100.5	54.8	94.4	100.5	22.9	0.24
1000	CDH2	73.9	69.6	75.1	44.9	52.5	83.7	65.1	69.6	16.1	0.25
54749	EPDR1	30.2	29.4	26.0	17.8	18.0	31.2	24.5	26.0	6.3	0.26
5817	PVR	44.9	28.3	26.2	19.7	17.4	33.2	25.0	26.2	6.4	0.26
6664	SOX11	0.4	1.2	1.0	0.8	0.6	0.8	0.9	0.8	0.2	0.26
286410	ATP11C	9.8	8.4	9.8	5.5	5.4	8.4	7.5	8.4	2.0	0.26
140885	SIRPA	26.5	35.3	39.6	24.9	25.8	46.2	34.3	35.3	9.1	0.26
5754	PTK7	19.3	20.6	16.5	11.3	11.4	18.5	15.7	16.5	4.2	0.27
10544	PROCR	11.4	10.0	8.6	13.6	10.5	6.3	9.8	10.0	2.7	0.27
3958	LGALS3	149.6	140.4	130.9	180.4	173.8	83.0	141.7	140.4	39.0	0.28
1009	CDH11	206.0	173.1	183.6	105.1	116.1	203.6	156.3	173.1	43.3	0.28
23446	SLC44A1	15.5	17.2	16.0	14.5	12.7	7.5	13.6	14.5	3.8	0.28
1464	CSPG4	15.0	14.2	11.0	20.2	15.7	10.1	14.2	14.2	4.1	0.28
1947	EFNB1	6.1	8.6	6.4	5.4	4.8	4.4	5.9	5.4	1.7	0.29
3383	ICAM1	1.2	1.9	1.6	1.0	1.0	1.4	1.4	1.4	0.4	0.30
5420	PODXL	8.6	33.5	31.9	35.5	31.6	13.9	29.3	31.9	8.7	0.30
221091	LRRN4CL	2.6	1.8	2.4	2.7	3.0	4.0	2.8	2.7	0.8	0.30
3679	ITGA7	5.0	6.2	5.0	3.5	3.7	6.9	5.1	5.0	1.5	0.30
4323	MMP14	358.5	578.3	497.1	570.6	523.1	229.7	479.8	523.1	143.8	0.30
79143	MBOAT7	15.0	15.5	12.5	8.6	7.2	10.0	10.8	10.0	3.3	0.30
10424	PGRMC2	39.2	34.5	32.7	18.4	22.4	40.8	29.8	32.7	9.2	0.31

Table S1. Continued

Gene ID	Gene Name	D0	Fl-Pr	Fl-Am	Fs-Pr	Fs-Am	TCP	Average	Median	Stdev	normalized stdev/coef- ficient of variation
57124	CD248	245.9	373.8	370.3	226.7	332.9	544.2	369.6	370.3	114.3	0.31
6641	SNIB1	7.8	6.6	6.4	3.8	4.8	8.8	6.1	6.4	1.9	0.31
57153	SLC44A2	29.9	39.2	35.2	19.3	20.4	36.5	30.1	35.2	9.5	0.32
4071	TM4SF1	106.9	103.4	91.7	62.2	54.0	52.5	72.8	62.2	23.3	0.32
493	ATP2B4	68.6	41.6	43.3	82.0	74.2	48.6	57.9	48.6	18.8	0.32
3671	ISLR	133.9	101.9	76.5	127.9	116.8	50.7	94.8	101.9	31.3	0.33
8140	SLC7A5	84.2	43.8	38.1	23.1	22.3	50.9	35.6	38.1	12.7	0.36
1462	VCAN	113.4	109.1	138.3	96.4	104.8	213.6	132.4	109.1	48.1	0.36
7046	TGFBR1	17.4	13.1	13.4	8.1	9.6	21.0	13.0	13.1	5.0	0.38
5819	PVRL2	17.7	22.6	17.7	9.4	9.1	16.9	15.1	16.9	5.8	0.39
1295	COL8A1	113.8	205.7	196.0	85.1	89.0	171.8	149.5	171.8	58.4	0.39
1293	COL6A3	789.2	733.9	713.9	1388.3	1183.0	539.0	911.6	733.9	357.2	0.39
84168	ANTXR1	73.6	121.5	135.6	54.5	60.2	129.1	100.2	121.5	39.5	0.39
2335	FN1	7292.7	7148.9	7036.7	15578.7	14296.8	8124.2	10437.1	8124.2	4155.0	0.40
290	ANPEP	141.1	126.4	130.1	225.1	216.4	80.2	155.6	130.1	62.7	0.40
135228	CD109	21.1	20.4	20.5	36.6	33.9	12.9	24.9	20.5	10.0	0.40
5159	PDGFRB	64.4	116.7	98.8	59.7	63.0	41.8	76.0	63.0	30.7	0.40
7070	THY1	156.2	184.9	175.1	73.2	77.0	150.3	132.1	150.3	53.6	0.41
4059	BCAM	2.2	4.5	4.2	1.8	2.5	5.5	3.7	4.2	1.5	0.41
4008	LMO7	92.4	114.1	129.1	54.0	71.3	160.7	105.8	114.1	43.3	0.41
23670	TMEM2	5.5	9.1	8.8	3.3	4.1	7.5	6.6	7.5	2.7	0.41
2239	GPC4	8.0	13.1	13.1	5.3	5.7	13.5	10.1	13.1	4.2	0.42
4883	NPR3	42.6	51.5	54.2	25.5	24.3	67.7	44.6	51.5	19.0	0.43
22801	ITGA11	299.8	243.7	257.3	161.2	211.8	464.0	267.6	243.7	115.8	0.43
3672	ITGA1	4.6	2.6	3.2	1.2	1.7	4.3	2.6	2.6	1.2	0.47
3673	ITGA2	5.8	2.9	3.3	6.9	4.0	1.7	3.8	3.3	1.9	0.52
7412	VCAM1	16.0	10.1	13.2	7.0	10.0	24.2	12.9	10.1	6.7	0.52
4162	MCAM	6.5	34.3	36.1	9.8	9.4	26.9	23.3	26.9	13.0	0.56

Table S1. Continued

Gene ID	Gene Name	D0	FI-Pr	FI-Am	Fs-Pr	Fs-Am	TCP	Average	Median	Stdev	normalized stdev/coef- ficient of variation
928	CD9	-	26.7	29.9	59.7	83.1	25.3	44.9	29.9	25.6	0.57
1969	EPHA2	Cluster1	8.9	6.5	1.8	1.9	6.3	5.1	6.3	3.1	0.61
3690	ITGB3	Cluster4	5.6	5.4	17.8	13.9	3.3	9.2	5.6	6.3	0.68
4045	LSAMP	Cluster2	1.6	1.5	0.3	0.2	2.1	1.1	1.5	0.8	0.72
8516	ITGA8	Cluster2	1.2	1.2	0.5	0.7	3.0	1.3	1.2	1.0	0.73

References:

Niehage, Christian, et al. "The cell surface proteome of human mesenchymal stromal cells." *PloS one* (2011): e20399.
Cho, Kyung-Ah, et al. "RNA sequencing reveals a transcriptomic portrait of human mesenchymal stem cells from bone marrow, adipose tissue, and palatine tonsils." *Scientific reports* (2017): 17114.
Lv, Feng-luan, et al. "Concise review: the surface markers and identity of human mesenchymal stem cells." *Stem cells* (2014): 1408-1419.

Table S2. The 20 genes with high correlation with CD90 and/or CD105

gene_id	gene_name	D0	Fl-Pr	Fl-Am	Fs-Pr	Fs-Am	TCP	Pearson coefficient	
								with THY1	with ENG
8826	IQGAP1	80.3	73.6	81.0	92.1	89.4	76.5	−0.92	−0.89
4882	NPR2	7.5	8.1	7.6	6.4	6.3	7.6	0.98	0.93
214	ALCAM	118.6	131.0	138.7	94.5	96.6	122.5	0.97	0.94
2022	ENG	169.3	220.3	208.6	148.5	148.2	198.2	0.92	1.00
83700	JAM3	20.9	21.9	20.3	13.9	15.4	18.3	0.96	0.81
57136	APMAP	35.2	38.4	35.7	24.7	25.8	28.1	0.91	0.77
102	ADAM10	42.5	44.2	42.9	28.4	31.5	46.6	0.93	0.85
4035	LRP1	134.8	122.4	121.1	190.2	186.2	157.7	−0.97	−0.85
9761	MLEC	19.5	24.8	24.6	16.4	15.6	18.9	0.90	0.91
286410	ATP11C	9.8	8.4	9.8	5.5	5.4	8.4	0.91	0.70
5754	PTK7	19.3	20.6	16.5	11.3	11.4	18.5	0.93	0.80
3383	ICAM1	1.2	1.9	1.6	1.0	1.0	1.4	0.90	0.98
57153	SLC44A2	29.9	39.2	35.2	19.3	20.4	36.5	0.96	0.97
493	ATP2B4	68.6	41.6	43.3	82.0	74.2	48.6	−0.89	−0.98
5819	PVRL2	17.7	22.6	17.7	9.4	9.1	16.9	0.97	0.90
2335	FN1	7292.7	7148.9	7036.7	15578.7	14296.8	8124.2	−0.98	−0.84
23670	TMEM2	5.5	9.1	8.8	3.3	4.1	7.5	0.92	0.99
928	CD9	61.6	26.7	29.9	59.7	83.1	25.3	−0.79	−0.92
1969	EPHA2	6.2	8.9	6.5	1.8	1.9	6.3	0.98	0.92
3690	ITGB3	10.7	5.6	5.4	17.8	13.9	3.3	−0.86	−0.91
7070	THY1	156.2	184.9	175.1	73.2	77.0	150.3		
2022	ENG	169.3	220.3	208.6	148.5	148.2	198.2		