Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/scr

Global proteomic signature of undifferentiated human bone marrow stromal cells: Evidence for donor-to-donor proteome heterogeneity

Samuel T. Mindaye [a], Moonjin Ra [a], Jessica L. Lo Surdo [b], Steven R. Bauer [b], Michail A. Alterman [a,*]

[a] Tumor Vaccines and Biotechnology Branch, Division of Cellular and Gene Therapies, Center for Biologics Evaluation and Research, US Food and Drug Administration, Bethesda, MD, USA
[b] Cellular and Tissue Therapies Branch, Division of Cellular and Gene Therapies, Center for Biologics Evaluation and Research, US Food and Drug Administration, Bethesda, MD, USA

Abstract   The clinical application of human bone marrow stromal cells (hBMSCs) largely depends on their capacity to expand *in vitro*. We have conducted a comprehensive comparative proteomic analysis of culture-expanded hBMSCs obtained from different human donors. The data reveal extensive donor-to-donor proteomic heterogeneity. Processing and database-searching of the tandem MS data resulted in a most comprehensive to date proteomic dataset for hBMSC. A total of 7753 proteins including 712 transcription and translation regulators, 384 kinases, 248 receptor proteins, and 29 cytokines were confidently identified. The proteins identified are mainly nuclear (43.2%) and the share of proteins assigned to more than one subcellular location constitutes 10% of the identified proteome. Bioinformatics tools (IPA, DAVID, and PANTHER) were used to annotate proteins with respect to cellular locations, functions, and other physicochemical characteristics. We also compared the proteomic profile of hBMSCs to recently compiled datasets for human and mouse pluripotent stem cells. The result shows the extent of similarity between the three cell populations and also identified 253 proteins expressed uniformly by all lines of hBMSCs but not reported in the proteomic datasets of the two pluripotent stem cells. Overall, the proteomic database reported in this paper can serve as a reference map for extensive evaluation of hBMSC to explain their biology as well as identify possible marker candidates for further evaluation.
Published by Elsevier B.V.

Introduction

Mesenchymal stromal cells were originally described as stromal cells from bone marrow in the hematopoietic microenvironment that formed adherent colonies when cultured *ex vivo* and demonstrate osteogenic potential (Friedenstein et al., 1968, 1970; Sensebe et al., 2010). Since their first description, cells with similar characteristics have been derived from numerous tissues including cord blood, adipose tissue, cartilage, dental pulp, and muscle (Kuhn and Tuan, 2010). The cells obtained from bone marrow were named mesenchymal stem cells in 1991 by Caplan (Caplan, 1991). In 2005, the International Society for Cellular Therapy (ISCT) recommended the term multipotent

---

 * Corresponding author. Fax: +1 301 827 0449.
    E-mail address: Michail.Alterman@fda.hhs.gov (M.A. Alterman).

mesenchymal stromal cells to be used to refer to fibroblast-like cells with a set of properties including plastic-adherence, *in vitro* trilineage differentiation capacity, and expression of a defined set of cell-surface antigens (Dominici et al., 2006; Horwitz et al., 2005). The ISCT's definition has been widely adopted although recent evidence has shown that the characteristics of stromal cells vary depending on their tissue sources. Moreover, the true multipotency and self-renewing capacity of stromal cells from various tissues have not been confirmed with rigorous bioassays (Bianco et al., 2013). A particular challenge to the field has been the absence of the unique set of markers that can be used to enrich MSCs from other connective tissue cell populations and define them functionally. There is much discussion of the functional definition, nomenclature, and experimental handling of multipotent stem cells as can be observed in recent reviews (Bianco et al., 2010, 2013; Keating, 2012). In this paper the term bone marrow stromal cells (BMSCs) is used to refer to plastic adhering bone marrow-derived colonies of stromal progenitors that express a set of cell-surface phenotypes defined by ISCT. Such cells have been referred to by various names in the literature including mesenchymal stem/stromal cells (MSCs).

While stem cell-based therapies hold great potential for the treatment of a wide array of medical conditions they are so novel that product characterization is particularly challenging. Despite considerable progress, the molecular regulatory mechanisms of self-renewal and lineage specification in these cell types are largely unexplored. In recent years a number of "omics" technologies were applied to investigate MSCs (Jansen et al., 2010; Kulterer et al., 2007; Ng et al., 2008; Ren et al., 2011). The majority of earlier proteomic studies were performed by a combination of two dimensional electrophoresis (2DE) and matrix assisted laser desorption ionization mass spectrometry (MALDI MS). The first proteomic investigation reported by Colter et al. resulted in the identification of 40 differentially regulated proteins between rapidly self-renewing and mature human BMSCs (hBMSCs) (Colter et al., 2001). Similar techniques were used to study the effect of transforming growth factor beta (TGF-β) (Wang et al., 2004), shear stress (Yi et al., 2010), and mechanical strain and TGF-β (Kurpinski et al., 2009), or disease conditions such as rheumatoid arthritis (Kastrinaki et al., 2008), osteoarthritis (Rollin et al., 2008), and idiopathic scoliosis (Zhuang et al., 2011) on hBMSCs. Other studies used combinations of 2DE and MALDI MS to compare the proteomic variability between MSCs isolated from various sources such as amniotic fluid, bone marrow, umbilical cord, placenta, adipose tissue, and synovial membrane (Roche et al., 2009).

However, the well-documented poor performance of 2DE with regard to membrane, basic, and low abundance proteins limited the exploration of such complex biological samples as MSCs (Chevalier, 2008). Recent trends show that on-line multi-dimensional liquid chromatography (LC) coupled with MS significantly improves proteomic coverage. This approach dramatically increased the number of proteins identified (~900) including hundreds of membrane proteins from hBMSCs (Niehage et al., 2011). On the other hand, off-line 2D-LC fractionation followed by MALDI MS was also applied successfully to study the proteomic architecture of distinct populations of hBMSC (Mareddy et al., 2009).

The number of proteins identified to date in hBMSC (<1000) clearly indicates that we have only scratched the surface of the proteome and detected mainly abundantly and moderately expressed proteins. A deeper molecular analysis of the proteome, transcriptome, and protein interactome of hBMSCs would lead to a better understanding of these cells. An additional difficulty is the absence of a unified analytical approach which makes the comparison of data obtained in different laboratories challenging, particularly in combination with the well-documented heterogeneity of hBMSCs.

In this study we applied a combined proteomic approach that included pressure cycling-based protein harvesting, 3D fractionation, and complementary MS strategies (electrospray ionization (ESI) and MALDI) to improve proteomic characterization of culture-expanded hBMSCs obtained from different human donors. We created the largest proteomic database for hBMSCs reported to date. The results obtained highlight the surprisingly large degree of proteomic variability in hBMSC cell lines obtained from six human donors. Furthermore, results of extensive bioinformatic analyses of relevant molecular events, biological processes, signaling pathways, and protein−protein interaction networks (interactome) that are operating in these cell populations are also presented and discussed.

Materials and methods

Cell cultures

HBMSC lines from six human donors (four females and two males, Table 1) were obtained from commercial sources. Cell lines PCBM1641, PCBM1632, and PCBM1662 were obtained at passage 1 (P1) from All Cells. Lines 167696, 110877, and 8F3560 were purchased from Lonza at P1. All donors fulfilled institutional requirements at the time of cell collection. According to the manufacturers, following informed consent bone marrow aspirates were taken and plastic adherent hBMSCs were harvested. The time of culture prior to harvest was 15, 15, and 14 days for cell lines 167696 (at 90% confluence), 110877 (at 45% confluence), and 8F3560 (at 35% confluence). At this stage cells were designated as P0 by the supplier. Cells were further cultured to P1 for 6, 6, and 7 days, respectively, collected at 95% confluence, and frozen. Cell lines PCBM1641, PCBM1632, and PCBM1662 were

Table 1 HBMSC lines and donor characteristics. Cells from two commercial sources were obtained at P1 and a proteomic comparison was performed at P3, P5, and P7.

Donor #	Cell line	Sex	Age	Cell source	Race
1	PCBM1641	F	23	All Cells	Hispanic
2	167696	F	22	Lonza	Hispanic
3	PCBM1632	M	24	All Cells	Black
4	110877	M	22	Lonza	Black
5	8F3560	F	24	Lonza	Hispanic
6	PCBM1662	F	31	All Cells	Caucasian

in culture to 80% confluence for 14, 15, and 20 days, respectively before collection at P1.

P1 cells from Lonza met commonly used specifications for hBMSC cell surface markers ($\geq$ 90% positive for CD29, CD44, CD105, and CD166; $\leq$ 10% positive for CD14, CD34, and CD45). Extensive characterization and authentication of hBMSC lines from All Cells was performed as described by Lo Surdo and Bauer for PCBM1641 and PCBM1632 (Lo Surdo and Bauer, 2012). In regard to the expression of surface markers the cell lines were at least 90% positive for CD73, CD90, CD105, CD29, CD44, and CD166, and less than 5% positive for CD45, CD34, CD14, CD79α, and HLA-DR. The characterization in terms of colony forming unit assay and adipogenic differentiation potential was performed. All cell lines included in this study met commonly used specifications for hBMSCs (Dominici et al., 2006; Horwitz et al., 2005).

For our experiments, P1 cells were plated at a density of 60 cells/cm^2 density in T175 flasks (Cellstar), culture-expanded in α-MEM (Invitrogen) supplemented with 16.5% fetal bovine serum (FBS) (JM Bioscience, San Diego, CA), L-glutamine and penicillin−streptomycin (Pen/strep) (Invitrogen), and cultured at 37 °C and 5% CO$_2$ as described by Lo Surdo and Bauer (2012). A passage number is the number of times cells were trypsinized prior to freezing. The time to P3, P5, and P7 was calculated: PCBM1641 (7, 8, and 8 days, respectively), 167696 (8, 10, and 12 days), PCBM1632 (7, 9, and 17 days), 110877 (8, 9, and 10 days), 8F3560 (9, 10, 9 days), and PCBM1662 (10, 9, and 12 days). The cell lines vary in their proliferation capacity, which generally decreases with passaging. The full proliferation kinetic characterization for PCBM1641 and PCBM1632 was described by Lo Surdo and Bauer (2012), while the growth profile characterization of 167696 and 110877 was given elsewhere (Mindaye et al., 2013). At 80% confluence, cells were harvested using 0.25% Trypsin/EDTA (Invitrogen), washed using phosphate buffered saline three times, snap frozen in liquid nitrogen, and stored at -80 °C at 5×10^6 cells at passages 3 (P3), 5 (P5), and 7 (P7) for later use. The same serum lot was used for the expansion of all cell lines to 3 passages.

Proteome preparation, MS analyses, data processing, and bioinformatic analyses

The detailed protocol for sample preparation, LC/MALDI MS, LC-ESI MS, and LC-ESI (MS^E), analyses, data processing and database searching, and additional bioinformatic analyses are provided in Supplement 1. For MALDI MS analysis lysates from all cell lines were electrophoretically resolved into six fractions using the GELFREE system that allowed continuous separation of proteins through a gel-packed tube. Eluting proteins are trapped by a molecular weight cut off membrane and subsequently collected in solution phase, which was suitable for downstream MS analysis. From 200 μg of total protein, six fractions were collected (Supplement 1). Based on polyacrylamide gel electrophoresis data the first fraction was estimated to contain ~40% of the total protein loaded and all subsequent five fractions each had ~12% (data not shown). Fractionated samples were subjected to trypsin digestion followed by RP HPLC using an offline MALDI MS analysis. Triplicate analysis of samples from each cell line (at P3, P5, and P7) required a total of 54 LC injections. Each injection

was collected into 372 fractions (spots) on a MALDI target plate so that the total fractions collected from the six cell lines equaled 120,528 (54 × 372 × 6). Subsequently, 1,228,000 tandem MS spectra were collected. All fragment spectra from triplicate runs of each line were combined and an *inChorus* database search was launched using PEAKS Studio software (Perkins et al., 1999; Zhang et al., 2012). Mascot search results, namely, peptide and protein lists, were exported in Excel format. Then, all DTA files generated were combined and used for further filtering based on false discovery rate (FDR) using Scaffold software (http://www. proteomesoftware.com/). Confident protein identification considered a set of filter criteria including identification scores, overall probability, and FDR. Furthermore, proteins were identified using multiple peptides (90% of the proteome) and the majority of the identifications were done with at least 1 unique peptide.

For ESI-based MS analysis, cell lysates from PCBM1641, 167696, and PCBM1632 cells were fractionated using the GELFREE system and digests were prepared as described in Supplement 1. The digest from PCBM1641 was analyzed with the LTQ MS after 1D chromatographic peptide separation. The tandem MS experiment results were used for SwissProt database search. The digests from 167696 and PCBM1632 were analyzed using a Synapt G2 QTof MS. Tryptic peptides in this case were separated using 2D nanoAcquity UPLC (Supplement 1). Following MS^E acquisition, the data were processed using PLGS software (Li et al., 2009) and queried against target as well as one-time randomized human database (SwissProt). In contrast, the cell lysates from lines 110877, 8F3560, and PCBM1662 were digested without prior GELFREE protein fractionation steps and analyzed using a Synapt G2 QTof MS after 2D nanoAcquity UPLC peptide separation. In this case, instead of protein fractionation, which significantly adds to the overall analysis time, the number of fractions in the first dimension separation was increased to 13 prior to the second dimension separation (Supplement 1).

Results

Cell culturing and protein identification

Proteomic analyses of hBMSCs obtained from six donors were performed by a comprehensive approach developed in our laboratory (Mindaye et al., 2013). This approach includes pressure-assisted protein extraction, GELFREE protein fractionation, and multidimensional chromatographic peptide separation with subsequent complimentary ionization technique MS analysis for improved protein identification. Proteomic analyses were performed using 5×10^6 cells. Based on the stringent criteria, a total of 6192 proteins at a FDR of 0.8% were identified from six cell lines using MALDI MS data. The protein identification summary from each cell line is given in Table S1 (Supplement 2) and the list of proteins identified using MALDI MS is provided in Table S2. ESI-based proteome characterization was also performed either after GELFREE fractionation of the cell lysate (for PCBM1641, 167696, and PCBM1632) or without intact protein fractionation (cell lines 110877, 8F3560, and PCBM1662). The summary of proteins identified using ESI technique is

provided in Table S1 (Supplement 2) and the list of proteins is provided in Table S3 (Supplement 2).

Overall, data acquired using ESI MS led to the identification of 8828 proteins from six cell lines with 3866 of them being non-redundant (Supplement 2, Table S3). Similarly, from MALDI MS data a sum of 21,013 proteins was recovered, where 6192 were identified as non-redundant (Supplement 2, Table S2). Thus, the total number of proteins identified in this study using the two complimentary MS techniques was 29,841. As shown in Fig. 1A, only a third of the non-redundant proteins (31%) were identified using both ESI MS and MALDI MS datasets. Apparently, MALDI MS identified more proteins than ESI MS (82% of the total proteome as compared to 51% for ESI MS). Similar observations have also been reported from other proteomic studies, where MALDI MS increases the number of additional protein identifications from complex biological samples by an average of 45% (Yang et al., 2007), which is close to the improvement

achieved in this study (52%). Overall, the total number of non-redundant proteins identified from six cell lines was 7753 (Supplement 2, Tables S1).

The physicochemical properties of proteins from ESI MS and MALDI MS datasets were compared and summarized as shown in Figs. 1B–D. The proteins were mapped in relation to their theoretical molecular weight (Mw) and isoelectric point (pI), which were calculated with the help of Compute pI/Mw tool (www.expasy.org). The predicted pI distributions from the two datasets (MALDI vs. ESI) follow similar trends of bimodality centered at around pI 5 and 9, and these peaks are separated by a trough at around pI 7.5. Consistent with data on other cell types, the proteome of hBMSCs shows the characteristic trend and that most proteins have evolved to partition away from the close to neutral cytosolic pH, at which proteins are generally less soluble (Weiller et al., 2004). The trough consists mostly of proteins of nuclear origin. This trimodal distribution is characteristic of the
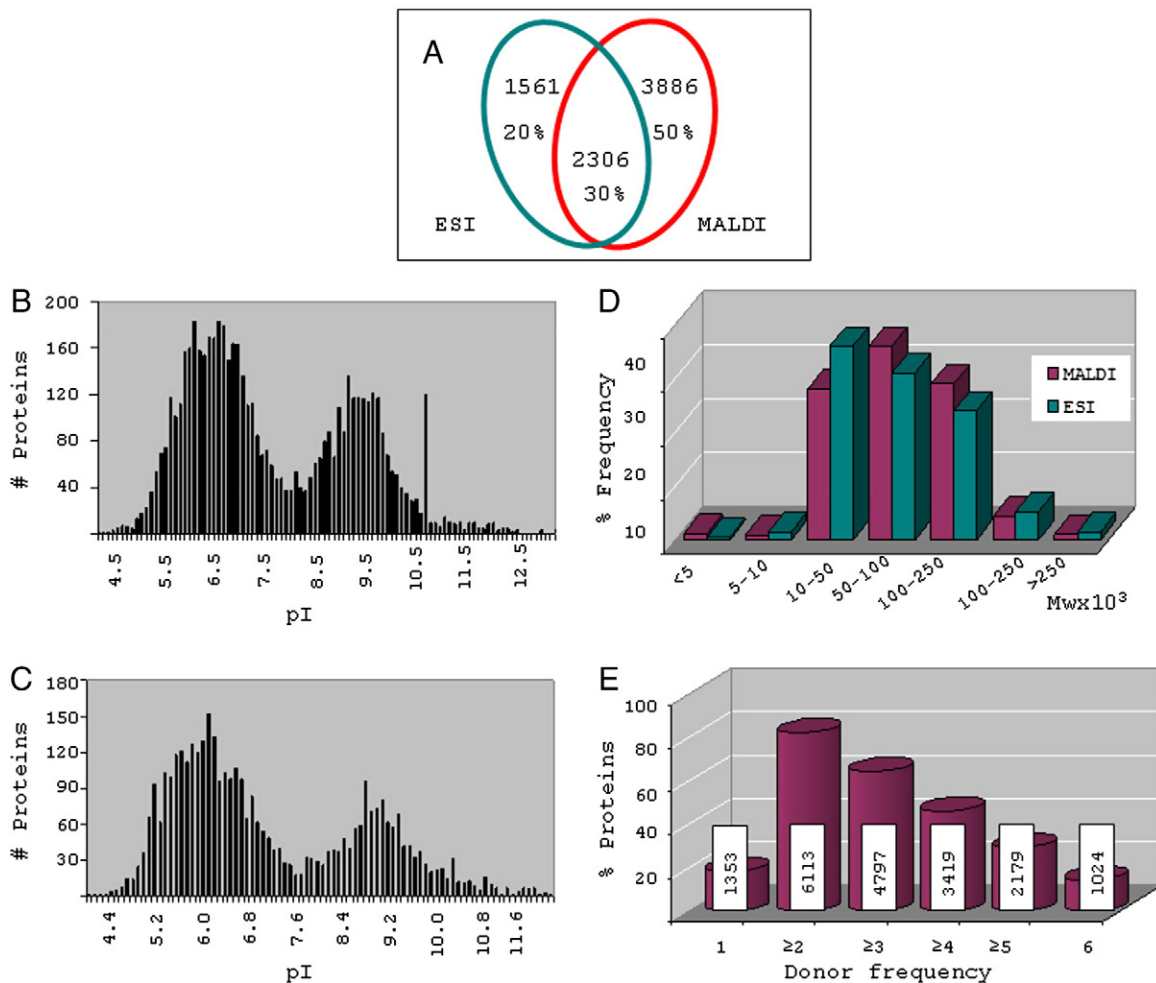

Figure 1   A summary of the proteomic characterization of hBMSCs. Panel A compares proteins identified using ESI MS and MALDI MS. Only 30% of the total proteins identified from hBMSCs were identified using the two MS techniques. Panels B and C, respectively, show the isoelectric point distribution of proteins of hBMSCs identified using LC/MALDI MS and LC-ESI MS. Panel D reveals the molecular weight distribution of proteins identified from hBMSCs using LC-ESI MS and LC/MALDI MS. Revealed in Panel E is the frequency distribution of proteins identified from six hBMSC lines. Shown on the histogram is the cumulative number of proteins expressed by ≥n number of cell lines.

eukaryotic proteome. In both cases, the peak on the acidic side rises sharply and is more abundant (62% and 56% for ESI MS and MALDI MS, respectively) than the one on the alkaline side.

Expression overlap between cell lines: the proteomic basis of donor-to-donor variability

One of the challenges facing clinical use of MSCs is the differences in biological activities between cells obtained from different donors (Pevsner-Fischer et al., 2011; Phinney, 2007, 2012). The qualitative and/or quantitative variability of the cell proteome and the posttranslational modifications are likely the most significant contributor to this functional heterogeneity. We attempted to evaluate the extent of protein expression variability in cell lines obtained from different donors. Fig. 1E summarizes the frequency of protein identification. From the total proteins, approximately one out of six (1353 proteins) was identified from one cell line but not from the others, and only 13% (1024 proteins) were identified from all cell lines (Table S4, Supplement 2). This fraction included proteins critical for major biological functions in MSCs. For example, leukemic inhibitory factor 2 (LIF2) is implicated in the maintenance of stemness in MSCs (Kolf et al., 2007). Beta-catenin, a component of the canonical Wnt signaling pathway, is also important for MSC self-renewal (Kolf et al., 2007; Ling et al., 2009). Also, serine/threonine−protein kinase (DCLK3) regulates a number of transcription factors and determines MSC fate through its positive effectors of the non-canonical Wnt signaling pathway (Ling et al., 2009). Likewise, other important transcription factors such as RAS-responsive element binding protein 1, HMG box transcription factors (e.g., nuclear transcription factor 1), and basic helix−loop−helix transcription factors (e.g., aryl hydrocarbon receptor nuclear transcription factor 2) have also been found to be expressed uniformly between cell lines.

Proteins expressed in at least 50% of the cell lines accounted for 62% of the total proteome identified, while the number increases to 79% (6113 proteins) if expression is evaluated from two or more cell lines (Supplement 2, Table S2 and S3).

Sub-cellular localization of proteins

Proteins perform biological functions within the context of their spatial time-resolved localization in a living cell. Generally, the ultimate localization of proteins can be determined by the amino acid sequences because some of these sequences are recognized by specific location-receptor proteins (Imai and Nakai, 2010). This principle, together with other features such as protein sorting signals, sequence homology with a protein of known localization, and protein−protein interaction data, is used in subcellular prediction algorithms. The subcellular localization of hBMSC proteins was classified using the ngLOC web-based server from the University of Nebraska Medical Center (http://ngloc.unmc.edu/) (King and Guda, 2007). ngLOC is a Bayesian method for predicting sub-cellular localization of proteins at proteome scale. The method relies on the density distribution of fixed length peptide sequences (n-gram) derived from the primary structure of proteins. It can classify proteins into at least 10 distinct subcellular localizations including cytoplasm, endoplasmic reticulum, extracellular/secreted, golgi, lysosomes, mitochondria, nucleus, junction, peroxysomes, and plasma membrane. Furthermore, it can also infer the localization of proteins if they appear at more than one site.

A summary of the subcellular localization of the hBMSC proteome is given in Figs. 2A and B. Fig. 2A represents proteins that are predicted to localize at a unique subcellular location (90% of the total proteome); the majority of them are nuclear (43.2%), followed by cytoplasmic (17.6%), and then plasma membrane proteins (13.6%). Of the total proteome, 10% were predicted to appear at more than one location and Fig. 2B describes their distribution. The existence of proteins at multiple locations is a widely appreciated phenomenon (Zhang et al., 2008). Such proteins, which may account for as much as 35% of the total proteome, may simultaneously be located within or undergoing transport between different cellular compartments (Zhang et al., 2008). Not surprisingly, the majority (47%) of such proteins reside at the cytoplasm/nucleus site. Examples include pergerin, a component of the MOZ/MORF complex, which has a transcription regulation role (Ullah et al., 2008); endothelial differentiation-related factor 1, which is another transcription regulator that can be involved in cell differentiation (Dragoni et al., 1998); and cystatin-B, which is a reversible inhibitor of cathepsins L, H, and B. Moreover, proteasome activator complex subunit 3, a subunit of the 11S REG-gamma proteosome regulator, resides in the cytoplasm during mitosis following nuclear envelope breakdown (Brooks et al., 2000). Proteins residing at cytoplasm/plasma membrane site are the second most abundant in the class (10%) followed by those at cytoplasm/cytoskeleton (8%).

Major biological processes, pathways, and protein−protein interaction networks in hBMSCs

Most large-scale functional studies rely on resources provided by the Gene Ontology (GO) Consortium. GO annotations can be accessed from major biological databases such as UniProt, Ensembl, EntrezGene, and others (Barrell et al., 2009). We used the publicly available PANTHER server to do functional classification of the hBMSC proteome (Thomas et al., 2003). PANTHER (Protein ANalysis THrough Evolutionary Relationships) allows browsing of a manually curated database by functions, where the database was established using published evidence as well as evolutionary relationships. The accuracy of protein function prediction was estimated using statistical models (Hidden Markov Models, HMMs) constructed for each functional protein group (Thomas et al., 2003).

The UniProt accession numbers of hBMSC proteins were used to search in the PANTHER database. The search results were displayed as summaries of molecular function, biological process, cellular compartment, protein class, and pathway enrichment using the input proteomic dataset (Fig. 3). The y-axis shows the number of genes related to a particular biological process shown on the x-axis. Evidently, culture-expanded hBMSCs were metabolically active. In addition to the primary metabolic processes (carbohydrate, protein,
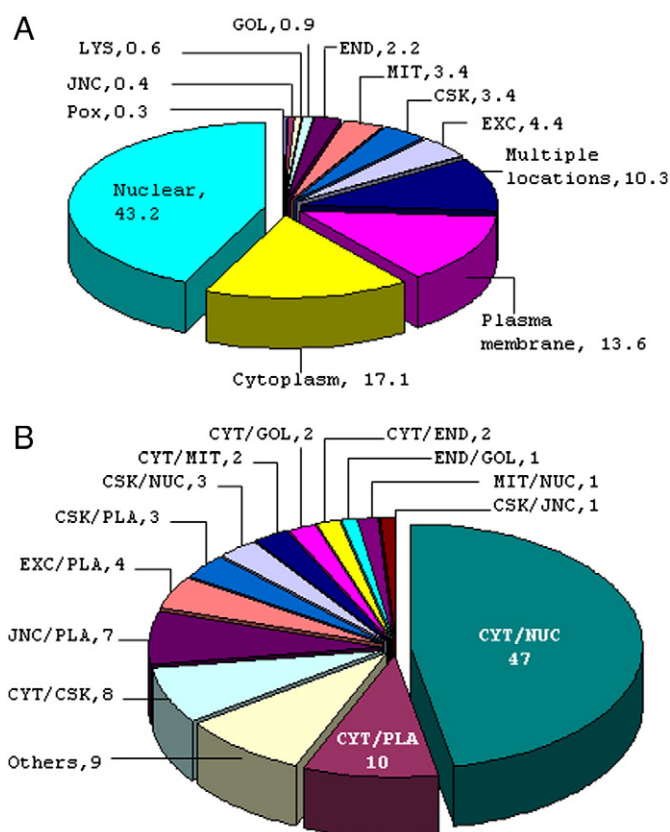
Figure 2   Predicted sub-cellular localization of the proteome of hBMSCs (n = 7753 proteins). Shown in A is the profile of proteins predicted to reside at one cellular location, while shown in B is the profile for proteins expected to be present at two locations. The notations are as follows: CYT (cytoplasm), EXC (extracellular), NUC (nuclear), PLA (plasma membrane), GOL (golgi), END (endoplasmatic), LYS (lysosome), CSK (cytoskeleton), JNC (junction), MIT (mitochondria), POX (peroxisome). Annotation was performed using ngLOC program.

lipid, amino acid, and nucleotide metabolism), specialized processes including coenzyme metabolism, ferredoxin and porphyrin metabolism, and vitamin metabolism have been enriched. Proteins that are involved in various cellular processes including cell adhesion, chromosome segregation, cell communication, cell motion, and cytokinesis were also identified. Proteins involved in mitotic proliferation of hBMSCs are represented by 112 proteins. Members of the caspase family (e.g., caspase 2), cyclins and cyclin-dependent kinases (CDKs) have been identified and these proteins are known to influence cell number through their involvement in mitosis and apoptotic events (Malumbres, 2011; Zhu et al., 2006). The immunomodulatory role of MSCs has been one of the capacities explored for potential application in cell therapy (Chen et al., 2011). HBMSCs express both pro- and anti-inflammatory proteins and the proteome is enriched with proteins involved in immune response through antigen processing and presentation (e.g., cathepsins), cellular activation (e.g., retinoic acid and tumor necrosis factors) and various other chemokines and cytokines. In contrast to the marked molecular variability described above (only 13% of the proteome overlaps), the overall profile of biological processes demonstrated remarkable similarity (Fig. 3). The same pattern was observed when the proteomes were screened based on signaling pathways (data not shown). This

phenomenon may be partly explained by the fact that cellular functions could be carried out by different but functionally similar proteins (Omelchenko et al., 2010).

For lack of a reference proteome for culture-expanded hBMSCs we assembled one by using proteins expressed from at least 50% of cell lines (three out of six). The reference proteomic map (4797 proteins, Supplement 2, Table S5) was used to enrich molecular events representative of hBMSCs. Accordingly, the reference proteome, which was mapped into 4645 genes in the PANTHER database, helped to enrich 5098 molecular functions, 8309 biological processes, and 2252 pathways potentially operating in hBMSCs. As an example, the classification based on signaling pathways is summarized in Fig. 4. As shown, integrin signaling is the most enriched pathway with 120 genes associated with it. The integrins are transmembrane receptors and are involved primarily in sensing the extracellular matrix environment. Their binding with ligands from the extracellular environment leads to the activation of intracellular signaling events that are essential in cell migration, growth and survival (Danmark et al., 2012; Martin et al., 2002; Parsons et al., 2012). Chemokine and cytokine signaling pathways related to inflammation are enriched by 105 genes. Also, hBMSCs express a wide array of chemokines and cytokines, which together with their receptors, play important roles in
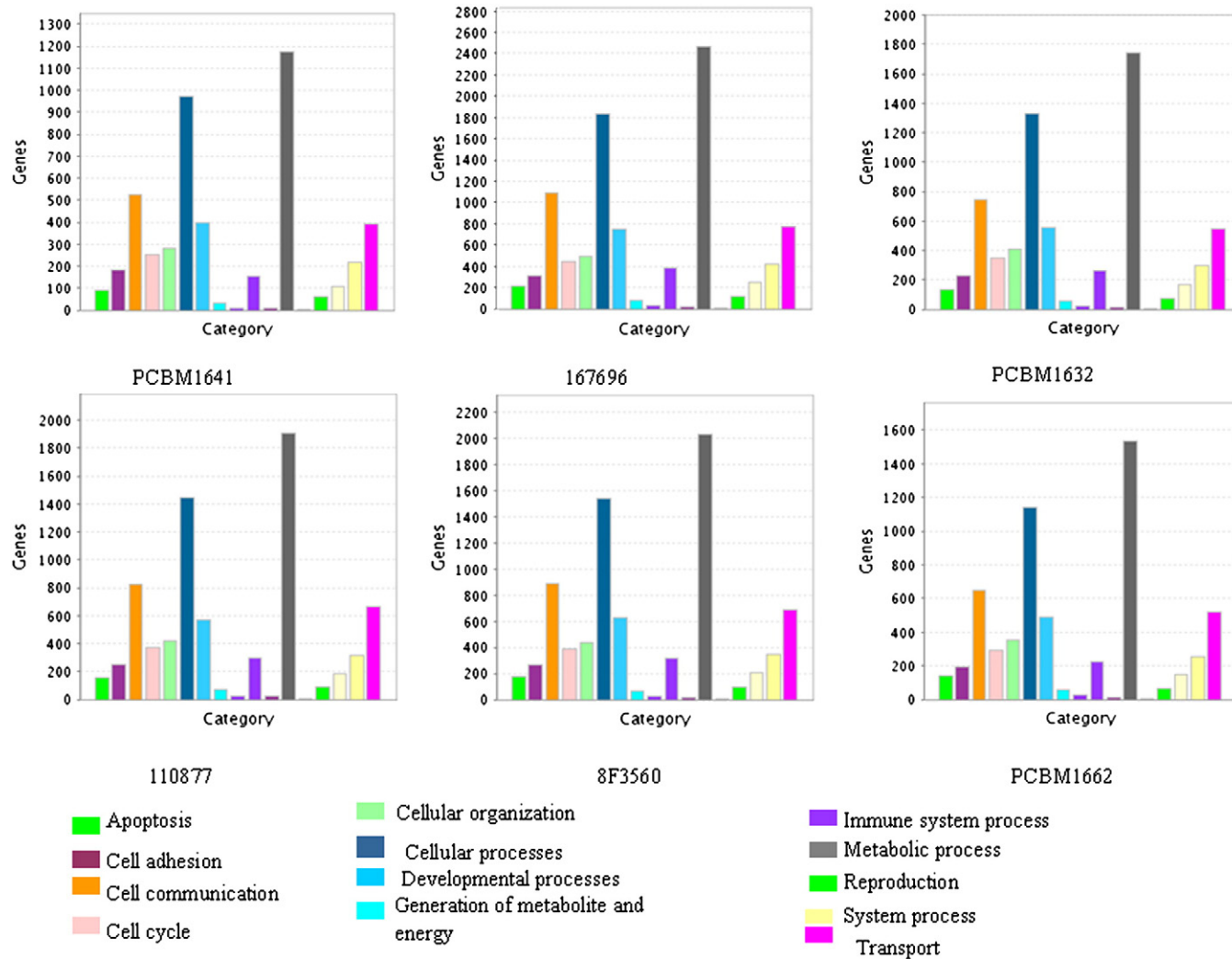
Figure 3 Biological processes enriched using proteins identified from six hBMSC lines. Enrichment was carried out using the PANTHER program as described in Supplement 1.
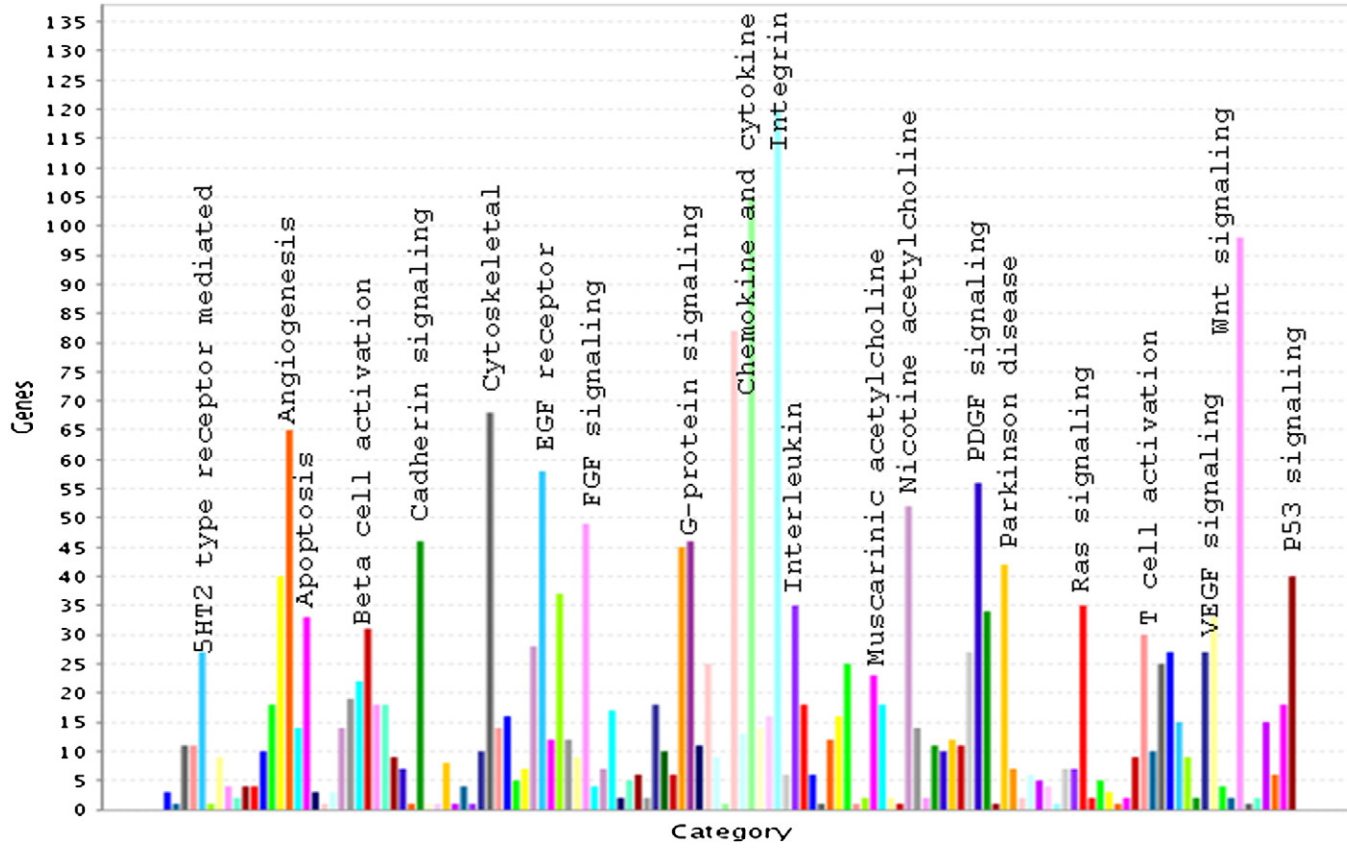
Figure 4 Representative signaling events for hBMSCs. Enrichment was performed with the PANTHER program using proteins expressed in at least 50% of cell lines (n = 4797 proteins).

homing, migration, and engraftment of cells to the inflammation sites (Docheva et al., 2008). The Wnt signaling pathway, which has a vital role in developmental regulation and cell fate determination (Etheridge et al., 2004; Ling et al., 2009), was enriched with 98 genes. Excessive activation of beta-catenin-dependent Wnt signaling leads to activated DNA damage followed by increased senescence and reduces proliferation of MSCs (Zhang et al., 2011). Wnt signaling molecules have also been shown to inhibit early stages of adipogenesis (Laudes, 2011). Represented by 40 genes is the p53 pathway, which plays a role in the differentiation (Molchadsky et al., 2008) and aging of MSCs (Zhang et al., 2011). Other major signaling pathways include cytoskeletal regulation by Rho GTPase (68 genes), angiogenesis (65 genes), and cadherin signaling (46 genes). Growth factor signaling pathways have also been enriched: EGF receptor-mediated (58 genes), FGF (49 genes), PDGF (56 genes), VEGF (33 genes), and TGF-β (25 genes). These have been reported to be involved in developmental regulation of MSCs through differentiation (Ng et al., 2008), migration (Veevers-Lowe et al., 2011), and self-renewal (Coutu and Galipeau, 2011; Rodrigues et al., 2010).

Proteins rarely act alone; to exert their functions efficiently they interact with each other. To see major relationships between proteins of hBMSCs, we performed protein−protein interaction network analysis using the Ingenuity Pathway Analysis (IPA) software. The accession numbers in the reference proteome (4797 proteins, Supplement 2, Table S5) were uploaded and 4713 of the proteins (98%) found matches in the IPA knowledge base (IPAKB). A Core Analysis was run to enrich interaction networks. The analysis considered only direct interaction relationships, which have been confirmed experimentally or predicted with high confidence. According to the IPA, direct interaction exists when there is evidence that proteins interact physically with one another. Networks were ranked based on their score, which is the negative log of their p-value. The scoring takes into account the number of eligible molecules, the size of the network, and the total number of molecules known to be associated with that network in IPAKB. The information in the IPAKB is curated manually and is composed from published literature and major external databases. The list of top 25 high scoring interaction networks according to our screening is presented in Supplement 3.

A general interactome scheme between molecules in the top-ranking networks was built by merging 7 relevant networks (Supplement 4, Fig. S1). The interactome map is represented graphically by 648 nodes (molecules) and 3127 edges (biological relationships between nodes) with an edge-to-node ratio exceeding 4.8, which means that the interactions in the network are highly unlikely to be random events. As the general interactome illustrates there are a number of biologically relevant interaction networks that are good candidates for further examination to generate a testable hypothesis or for an in-depth scientific scrutiny in regard to their importance in hBMSC biology. Nevertheless, even though the direct interaction relationships revealed in the general interactome model are supported by literature evidence from IPAKB, it needs to be reminded that not all of them have been proven experimentally, and interpretation of these modeled interactions needs to be treated with caution.

Comparison between proteomes of hBMSCs and pluripotent stem cells (PSCs)

The proteomes of human and mouse pluripotent stem cells (hPSCs and mPSCs) that were recently compiled accounted for a total of 7487 and 7295 proteins, respectively (Gundry et al., 2011). We compared the proteome set compiled in this study for hBMSCs against the datasets for hPSCs and mPSCs. For the analysis, the accession numbers of proteins were uploaded onto the IPA software and parsed using the "Compare data" option of the software to output proteins shared between the three cell types. The screening shows that the proteomes of the three distinct cell populations share at least 2479 proteins between them (Supplement 4, Fig. S3). The overlapping proteins encompass a network of at least 152 transcription regulators, 971 binding proteins, 911 proteins with catalytic properties, 89 receptors, ion channels, and transporters. The unique part of the hBMSCs proteome accounted for 28% (2789 proteins). Furthermore, the comparative map of CDs expressed on the three cell types is presented in Table S7 (Supplement 2). The comparison involves 116 hBMSC CDs, 102 hPSCs CDs, and 111 mPSCs CDs and revealed that the three cell types expressed at least 41 CDs in common.

Discussion

The current consensus is that not many proteins are expressed in a cell-specific manner. In fact, it has been reported that less than 1% of the human proteome is expressed in either single or a few cell types (Gry et al., 2010). This epitomizes the huge challenge that has to be overcome during the identification of protein markers that correlate with specific biological or clinical outcomes. Human BMSCs are poorly characterized in part because of the lack of comprehensive molecular evaluation. Therefore, the data presented in this report, in addition to its correlation with complex biological and molecular networks operating in hBMSCs, serves as a reference proteomic database for comparative work to identify and evaluate possible marker candidates for hBMSCs. The data obtained in this study also provide insight into donor-to-donor MSC proteome heterogeneity.

Previous proteomic studies on hBMSCs had limited proteomic coverage and reported only the most abundant proteins. In contrast, this report is based on a total of 7753 proteins, which encompasses 712 transcription and translation regulators including SOX-9, -11, -13, -15, -18, and -30. The expression of SOX-2 and OCT-4, which are factors for pluripotency maintenance, was not detected (Pierantozzi et al., 2011). The expression of Stro-1 could not be detected, which is consistent with previous reports showing that the expression of Stro-1 ceases with cell culture expansion (Kolf et al., 2007). Many members of the STAT family (signal transducer and activator transcription) including STAT-1, -3, -4, 5A, -5B, and -6 have also been identified in this study. STAT proteins are responsible for multiple cellular activities including the regulation of growth, survival, differentiation, motility, and immune response (Akira, 1999). The proteomic data also revealed the expression of 384 kinases (e.g., MAP kinases, PAKs, and CDKs), 248 receptors (e.g., TLR-2, -4, -6,

-7, and -9), and 29 cytokines (*e.g.*, chemokines, interleukins, and tumor necrosis factors).

The donor-to-donor proteomic variability documented in this paper can help to explain previous reports that show functional discordance between cell populations from different donors (Phinney, 2007, 2012). As per our data, only 13% of the total proteins were found universally expressed by all cell lines. Such a high degree of molecular diversity is a little surprising given the fact that samples from all cell lines were treated under similar conditions. The cell lines used in this study were obtained from two commercial sources (Materials and Methods section) and the cell-handling protocols during isolation from bone marrow aspirates could have contributed to the proteomic heterogeneity. To examine the existence of such affect, we compared the proteomic datasets compiled using cell lines from All Cells (n = 3) and Lonza (n = 3). In total, 6327 and 7015 proteins were identified from All Cells and Lonza, respectively. The two datasets shared at least 5596 proteins between them (72% similarity) and evidently the variability between the two group-proteomic datasets is not typically higher than the variability calculated for proteomes from individual donors (Table 2). The reference proteome we constructed based on this study (Supplement 2, Table S5) is not differentially represented by any proteomes of the two commercial cell sources (respectively, 97% and 99% of the proteins in the reference proteome were identified from hBMSCs obtained from All Cells and Lonza). Furthermore, hBMSCs from the two sources have also been compared in terms of growth kinetics, differentiation capacity, and cell surface markers. Although functional differences can be observed, none of these can be associated with the cell source or can be observed in different patterns of cell surface marker expression. Overall, functional and proteomic variabilities which have been observed between cell lines in this study cannot be linked directly to commercial cell source.

Molecular and functional heterogeneities in cell populations is not unique to hBMSCs; it is a widespread phenomenon among stem cells including embryonic (Canham et al., 2010), hematopoietic (Schroeder, 2010), neural (Suslov et al., 2002), and cancer stem cells (Wong et al., 2012). The analysis of major proteomic studies (n = 34) of human and mouse pluripotent stem cells (hPSCs and mPSCs) by Gundry et al. highlighted the degree of heterogeneity among pluripotent stem cell populations (Gundry et al., 2011).

According to their analysis, the comparison of nine major human proteomic studies, which together indexed 6966 proteins, revealed that less than 20% of the proteins were identified from at least 50% of the studies. By comparison, our proteomic screening from six cell lines found 62% of the proteins to be shared by at least 50% of the study subjects. The better proteomic overlap achieved in our study does not necessarily mean that hBMSCs are less heterogeneous than hPSCs. Rather, it reflects uniformity in the analytical proteomic approach.

The comparison between the hBMSC proteomic dataset with pluripotent stem cells revealed that numerous proteins are commonly expressed between the three cell types. It also has revealed that 28% (2879) of the hBMSC proteins were not identified from the PSC types. However, not all of these proteins were uniformly expressed between cell lines, only 253 have been identified in all six lines (Supplement 2, Table S6). While hBMSCs, hPSCs, and mPSCs expressed at least 41 CDs in common, 39 CDs have been identified exclusively from the hBMSC dataset. Only 15 CDs out of these 39 were identified from at least three of the six hBMSC lines, and only CD351 was identified from all lines. The 15 CDs are mostly receptors (CD5L, CD11D, CD59, CD23, CD85D, CD108, CD123, CD206, CD204, CD247, CD264, and CD351,) and some of them are involved in calcium ion binding (CD437), or are kinases (*e.g.*, CD167, CDw293).

The reference proteomic map for hBMSCs, which was constructed by combining proteins identified from at least 50% of cell lines, was used as a tool to get a quantitative sense of the proteomic variability between cell lines. For example, cell line 8F3560 has a proteome set of 4716 proteins, of which 3961 were also found in the reference proteome and the remaining proteins (834) were not part of the 8F3560 proteome. To account for the variability caused by the size of a proteome we used a correction factor, which is the ratio between the proteome size of a given cell line and the reference proteome (in this case 4716/4797 = 0.98). The reference proteome consists of 4797 proteins (Supplement 2, Table S5). Thus, the percent variability will be 17.1% (834/4795 * 0.98). Overall, an average variability between lines was 20.9% (Table 2). We also screened the data to determine if highly variably-expressed proteins are uniformly distributed among cellular organelles. For this purpose, the subcellular localizations of the most variably expressed proteins (those expressed in ≤ 2 cell lines) were compared against the profile obtained using the reference proteome (Supplement 4, Fig. S2). There was no apparent bias towards certain organelles in molecular variability between cell lines, although the plasma membrane and mitochondria contain slightly higher proportions of variably-expressed proteins.

Previously, the donor-to-donor heterogeneity in MSCs has been revealed through significant differences in growth rate and clonogenic potential (Phinney et al., 1999). Such inter-subject variabilities (donor-to-donor) have been considered to be caused for the most part by factors imposed by long-term culture conditions (Bonab et al., 2006; Briquet et al., 2010; Pevsner-Fischer et al., 2011; Schallmoser et al., 2010; Lo Surdo and Bauer, 2012). However, intra-subject heterogeneity (hBMSC population from the same donor) has also been recognized and in fact it was proposed that hBMSCs may actually exist *in vivo* as heterogeneous

Table 2 Summary of protein expression variability between six cell lines of hBMSCs.

Cell line	Total ID	# ID matched with ref proteome	% variability
PCBM1641	2768	2434	28.4
167696	5667	4444	8.7
PCBM1632	4225	3615	21.8
110877	4404	3639	22.2
8F3650	4716	3961	17.1
PCBM1662	4395	3431	26.0

Ref proteome is a reference proteome for hBMSCs constructed arbitrarily by taking proteins expressed in at least three of the six human cell lines.

populations (Pevsner-Fischer et al., 2011; Phinney, 2007). For example, only a restricted population of MSCs has been shown to express neuroregulatory proteins (Crigler et al., 2006), or demonstrate selective *in vivo* tumor-homing properties (Bolontrade et al., 2012), or possess the capacity to express interleukin 1 receptor antagonist (Ortiz et al., 2007). Their plastic nature, the capacity to differentiate, and the complex stroma in the bone marrow favor hBMSCs to exist as heterogeneous subpopulations. In fact, such heterogeneity has a biological advantage in tissues, as it allows the selection of the appropriate cell type for various demanding conditions more so than a rigid and homogenous population (Pevsner-Fischer et al., 2011).

Summary

We performed the largest proteomic analysis to date of culture-expanded hBMSC obtained from different human donors. Processing and database-searching of the tandem MS data collected revealed a highly enlarged proteomic dataset consisting of 7753 proteins. Bioinformatics tools including IPA, DAVID, and PANTHER were used to annotate proteins with respect to cellular locations, to analyze relevant molecular events, signaling pathways, and protein–protein interaction networks (interactome) that potentially influence MSC biology. Moreover, comparative analysis of the data highlighted the surprising amount of proteomic heterogeneity between hBMSCs from 6 different donors. The proteome of hBMSCs was compared with recently compiled proteomes of mouse and human pluripotent stem cells. The unique part of the hBMSC proteome has 253 proteins identified in all cell lines. The three cell types expressed 41 CDs in common and 39 CDs were identified exclusively in hBMSCs. Out of these 39 only CD351 was identified from all cell lines. Overall, the results reported and the database compiled in this study can serve as valuable resources in further attempts to understand the complex biological and molecular networks operating in hBMSCs and to identify possible marker candidates for further evaluation.

Author contributions

STM — acquisition, analysis and interpretation of data, drafting the article and final approval of the version to be published; MR — acquisition, analysis and interpretation of data, revising the article and final approval of the version to be published; JLL — acquisition of data, revising the article and final approval of the version to be published; SRB — conception and design of the study, revising the article and final approval of the version to be published; MAA — conception and design of the study, analysis and interpretation of data, drafting and revising the article and final approval of the version to be published.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.scr.2013.05.006.

References

Akira, S., 1999. Functional roles of STAT family proteins: lessons from knockout mice. Stem Cells 17, 138–146.

Barrell, D., Dimmer, E., Huntley, R.P., Binns, D., O'Donovan, C., Apweiler, R., 2009. The GOA database in 2009 — an integrated Gene Ontology Annotation resource. Nucleic Acids Res. 37, D396–D403.

Bianco, P., Robey, P.G., Saggio, I., Riminucci, M., 2010. "Mesenchymal" stem cells in human bone marrow (skeletal stem cells): a critical discussion of their nature, identity, and significance in incurable skeletal disease. Hum. Gene Ther. 21, 1057–1066.

Bianco, P., Cao, X., Frenette, P.S., Mao, J.J., Robey, P.G., Simmons, P.J., Wang, C.Y., 2013. The meaning, the sense and the significance: translating the science of mesenchymal stem cells into medicine. Nat. Med. 19, 35–42.

Bolontrade, M.F., Sganga, L., Piaggio, E., Viale, D.L., Sorrentino, M.A., Robinson, A., Sevlever, G., Garcia, M.G., Mazzolini, G., Podhajcer, O.L., 2012. A specific subpopulation of mesenchymal stromal cell carriers overrides melanoma resistance to an oncolytic adenovirus. Stem Cells Dev. 21, 2689–2702.

Bonab, M.M., Alimoghaddam, K., Talebian, F., Ghaffari, S.H., Ghavamzadeh, A., Nikbin, B., 2006. Aging of mesenchymal stem cell in vitro. BMC Cell Biol. 7, 14.

Briquet, A., Dubois, S., Bekaert, S., Dolhet, M., Beguin, Y., Gothot, A., 2010. Prolonged ex vivo culture of human bone marrow mesenchymal stem cells influences their supportive activity toward NOD/SCID-repopulating cells and committed progenitor cells of B lymphoid and myeloid lineages. Haematologica 95, 47–56.

Brooks, P., Fuertes, G., Murray, R.Z., Bose, S., Knecht, E., Rechsteiner, M.C., Hendil, K.B., Tanaka, K., Dyson, J., Rivett, J., 2000. Subcellular localization of proteasomes and their regulatory complexes in mammalian cells. Biochem. J. 346 (Pt 1), 155–161.

Canham, M.A., Sharov, A.A., Ko, M.S., Brickman, J.M., 2010. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. PLoS Biol. 8, e1000379.

Caplan, A.I., 1991. Mesenchymal stem cells. J. Orthop. Res. 9, 641–650.

Chen, P.M., Yen, M.L., Liu, K.J., Sytwu, H.K., Yen, B.L., 2011. Immunomodulatory properties of human adult and fetal multipotent mesenchymal stem cells. J. Biomed. Sci. 18, 49.

Chevalier, F., 2008. Highlights on the capacities of "Gel-based" proteomics. Proteome Sci. 8, 23.

Colter, D.C., Sekiya, I., Prockop, D.J., 2001. Identification of a subpopulation of rapidly self-renewing and multipotential adult stem cells in colonies of human marrow stromal cells. Proc. Natl. Acad. Sci. U. S. A. 98, 7841–7845.

Coutu, D.L., Galipeau, J., 2011. Roles of FGF signaling in stem cell self-renewal, senescence and aging. Aging (Albany NY) 3, 920–933.

Crigler, L., Robey, R.C., Asawachaicharn, A., Gaupp, D., Phinney, D.G., 2006. Human mesenchymal stem cell subpopulations express a variety of neuro-regulatory molecules and promote neuronal cell survival and neuritogenesis. Exp. Neurol. 198, 54–64.

Danmark, S., Finne-Wistrand, A., Albertsson, A.C., Patarroyo, M., Mustafa, K., 2012. Integrin-mediated adhesion of human mesenchymal stem cells to extracellular matrix proteins adsorbed to polymer surfaces. Biomed. Mater. 7, 035011.

Docheva, D., Haasters, F., Schieker, M., 2008. Mesenchymal stem cells and their cell surface receptors. Curr. Rheumatol. Rev. 4, 155–160.

Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., Marini, F., Krause, D., Deans, R., Keating, A., Prockop, D., Horwitz, E., 2006. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. Cytotherapy 8, 315–317.

Dragoni, I., Mariotti, M., Consalez, G.G., Soria, M.R., Maier, J.A., 1998. EDF-1, a novel gene product down-regulated in human endothelial cell differentiation. J. Biol. Chem. 273, 31119–31124.

Etheridge, S.L., Spencer, G.J., Heath, D.J., Genever, P.G., 2004. Expression profiling and functional analysis of wnt signaling mechanisms in mesenchymal stem cells. Stem Cells 22, 849–860.

Friedenstein, A.J., Petrakova, K.V., Kurolesova, A.I., Frolova, G.P., 1968. Heterotopic of bone marrow. Analysis of precursor cells for osteogenic and hematopoietic tissues. Transplantation 6, 230–247.

Friedenstein, A.J., Chailakhjan, R.K., Lalykina, K.S., 1970. The development of fibroblast colonies in monolayer cultures of guinea-pig bone marrow and spleen cells. Cell Tissue Kinet. 3, 393–403.

Gry, M., Oksvold, P., Pontén, F., Uhlén, M., 2010. Tissue-specific protein expression in human cells, tissues and organs. J. Proteome Bioinform. 3, 286–293.

Gundry, R.L., Burridge, P.W., Boheler, K.R., 2011. Pluripotent stem cell heterogeneity and the evolving role of proteomic technologies in stem cell biology. Proteomics 11, 3947–3961.

Horwitz, E.M., Le Blanc, K., Dominici, M., Mueller, I., Slaper-Cortenbach, I., Marini, F.C., Deans, R.J., Krause, D.S., Keating, A., 2005. Clarification of the nomenclature for MSC: the International Society For Cellular Therapy position statement. Cytotherapy 7, 393–395.

Imai, K., Nakai, K., 2010. Prediction of subcellular locations of proteins: where to proceed? Proteomics 10, 3970–3983.

Jansen, B.J., Gilissen, C., Roelofs, H., Schaap-Oziemlak, A., Veltman, J.A., Raymakers, R.A., Jansen, J.H., Kogler, G., Figdor, C.G., Torensma, R., Adema, G.J., 2010. Functional differences between mesenchymal stem cell populations are reflected by their transcriptome. Stem Cells Dev. 19, 481–490.

Kastrinaki, M.-C., Sidiropoulos, P., Roche, S., Ringe, J., Lehmann, S., Kritikos, H., Vlahava, V.M., Delorme, B., Eliopoulos, G.D., Jorgensen, C., Charbord, P., Haupl, T., Boumpas, D.T., Papadaki, H.A., 2008. Functional, molecular and proteomic characterisation of bone marrow mesenchymal stem cells in rheumatoid arthritis. Ann. Rheum. Dis. 67, 741–749.

Keating, A., 2012. Mesenchymal stromal cells: new directions. Cell Stem Cell 10, 709–716.

King, B.R., Guda, C., 2007. ngLOC: an n-gram-based Bayesian method for estimating the subcellular proteomes of eukaryotes. Genome Biol. 8, R68.

Kolf, C.M., Cho, E., Tuan, R.S., 2007. Mesenchymal stromal cells. Biology of adult mesenchymal stem cells: regulation of niche, self-renewal and differentiation. Arthritis Res. Ther. 9, 204.

Kuhn, N.Z., Tuan, R.S., 2010. Regulation of stemness and stem cell niche of mesenchymal stem cells: implications in tumorigenesis and metastasis. J. Cell. Physiol. 222, 268–277.

Kulterer, B., Friedl, G., Jandrositz, A., Sanchez-Cabo, F., Prokesch, A., Paar, C., Scheideler, M., Windhager, R., Preisegger, K.H., Trajanoski, Z., 2007. Gene expression profiling of human mesenchymal stem cells derived from bone marrow during expansion and osteoblast differentiation. BMC Genomics 8, 70.

Kurpinski, K., Chu, J., Wang, D., Li, S., 2009. Proteomic profiling of mesenchymal stem cell responses to mechanical strain and TGF-beta1. Cell. Mol. Bioeng. 2, 606–614.

Laudes, M., 2011. Role of WNT signalling in the determination of human mesenchymal stem cells into preadipocytes. J. Mol. Endocrinol. 46, R65–R72.

Li, G.Z., Vissers, J.P., Silva, J.C., Golick, D., Gorenstein, M.V., Geromanos, S.J., 2009. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. Proteomics 9, 1696–1719.

Ling, L., Nurcombe, V., Cool, S.M., 2009. Wnt signaling controls the fate of mesenchymal stem cells. Gene 433, 1–7.

Lo Surdo, J.L., Bauer, S.R., 2012. Quantitative approaches to detect donor and passage differences in adipogenic potential and clonogenicity in human bone marrow-derived mesenchymal stem cells. Tissue Eng. Part C Methods 18, 877–889.

Malumbres, M., 2011. Physiological relevance of cell cycle kinases. Physiol. Rev. 91, 973–1007.

Mareddy, S., Broadbent, J., Crawford, R., Xiao, Y., 2009. Proteomic profiling of distinct clonal populations of bone marrow mesenchymal stem cells. J. Cell. Biochem. 106, 776–786.

Martin, K.H., Slack, J.K., Boerner, S.A., Martin, C.C., Parsons, J.T., 2002. Integrin connections map: to infinity and beyond. Science 296, 1652–1653.

Mindaye, S.T., Ra, M., Lo Surdo, J., Bauer, S.R., Alterman, M.A., 2013. Improved proteomic profiling of the cell surface of culture-expanded human bone marrow multipotent stromal cells. J. Proteome 78C, 1–14.

Molchadsky, A., Shats, I., Goldfinger, N., Pevsner-Fischer, M., Olson, M., Rinon, A., Tzahor, E., Lozano, G., Zipori, D., Sarig, R., Rotter, V., 2008. p53 plays a role in mesenchymal differentiation programs, in a cell fate dependent manner. PLoS One 3, e3707.

Ng, F., Boucher, S., Koh, S., Sastry, K.S., Chase, L., Lakshmipathy, U., Choong, C., Yang, Z., Vemuri, M.C., Rao, M.S., Tanavde, V., 2008. PDGF, TGF-beta, and FGF signaling is important for differentiation and growth of mesenchymal stem cells (MSCs): transcriptional profiling can identify markers and signaling pathways important in differentiation of MSCs into adipogenic, chondrogenic, and osteogenic lineages. Blood 112, 295–307.

Niehage, C., Steenblock, C., Pursche, T., Bornhauser, M., Corbeil, D., Hoflack, B., 2011. The cell surface proteome of human mesenchymal stromal cells. PLoS One 6, e20399.

Omelchenko, M.V., Galperin, M.Y., Wolf, Y.I., Koonin, E.V., 2010. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. Biol. Direct 5, 31.

Ortiz, L.A., Dutreil, M., Fattman, C., Pandey, A.C., Torres, G., Go, K., Phinney, D.G., 2007. Interleukin 1 receptor antagonist mediates the antiinflammatory and antifibrotic effect of mesenchymal stem cells during lung injury. Proc. Natl. Acad. Sci. U. S. A. 104, 11002–11007.

Parsons, J.T., Martin, H.K., Slack, J.K., Boerner, S.A., Martin, C.C., 2012. Integrin signaling pathway. Sci. Signal. (Database of Cell Signaling as seen on 12 June 2013, http://stke.sciencemag.org/cgi/cm/stkecm;CMP_6880).

Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551–3567.

Pevsner-Fischer, M., Levin, S., Zipori, D., 2011. The origins of mesenchymal stromal cell heterogeneity. Stem Cell Rev. 7, 560–568.

Phinney, D.G., 2007. Biochemical heterogeneity of mesenchymal stem cell populations: clues to their therapeutic efficacy. Cell Cycle 6, 2884–2889.

Phinney, D.G., 2012. Functional heterogeneity of mesenchymal stem cells: implications for cell therapy. J. Cell. Biochem. 113, 2806–2812.

Phinney, D.G., Kopen, G., Righter, W., Webster, S., Tremain, N., Prockop, D.J., 1999. Donor variation in the growth properties and osteogenic potential of human marrow stromal cells. J. Cell. Biochem. 75, 424–436.

Pierantozzi, E., Gava, B., Manini, I., Roviello, F., Marotta, G., Chiavarelli, M., Sorrentino, V., 2011. Pluripotency regulators in

human mesenchymal stem cells: expression of NANOG but not of OCT-4 and SOX-2. Stem Cells Dev. 20, 915–923.

Ren, J., Jin, P., Sabatino, M., Balakumaran, A., Feng, J., Kuznetsov, S.A., Klein, H.G., Robey, P.G., Stroncek, D.F., 2011. Global transcriptome analysis of human bone marrow stromal cells (BMSC) reveals proliferative, mobile and interactive cells that produce abundant extracellular matrix proteins, some of which may affect BMSC potency. Cytotherapy 13, 661–674.

Roche, S., Delorme, B., Oostendorp, R.A., Barbet, R., Caton, D., Noel, D., Boumediene, K., Papadaki, H.A., Cousin, B., Crozet, C., Milhavet, O., Casteilla, L., Hatzfeld, J., Jorgensen, C., Charbord, P., Lehmann, S., 2009. Comparative proteomic analysis of human mesenchymal and embryonic stem cells: towards the definition of a mesenchymal stem cell proteomic signature. Proteomics 9, 223–232.

Rodrigues, M., Griffith, L.G., Wells, A., 2010. Growth factor regulation of proliferation and survival of multipotential stromal cells. Stem Cell Res. Ther. 1, 32.

Rollin, R., Marco, F., Camafeita, E., Calvo, E., Lopez-Duran, L., Jover, J.A., Lopez, J.A., Fernandez-Gutierrez, B., 2008. Differential proteome of bone marrow mesenchymal stem cells from osteoarthritis patients. Osteoarthr. Cartil. 16, 929–935.

Schallmoser, K., Bartmann, C., Rohde, E., Bork, S., Guelly, C., Obenauf, A.C., Reinisch, A., Horn, P., Ho, A.D., Strunk, D., Wagner, W., 2010. Replicative senescence-associated gene expression changes in mesenchymal stromal cells are similar under different culture conditions. Haematologica 95, 867–874.

Schroeder, T., 2010. Hematopoietic stem cell heterogeneity: subtypes, not unpredictable behavior. Cell Stem Cell 6, 203–207.

Sensebe, L., Krampera, M., Schrezenmeier, H., Bourin, P., Giordano, R., 2010. Mesenchymal stem cells for clinical application. Vox Sang. 98, 93–107.

Suslov, O.N., Kukekov, V.G., Ignatova, T.N., Steindler, D.A., 2002. Neural stem cell heterogeneity demonstrated by molecular phenotyping of clonal neurospheres. Proc. Natl. Acad. Sci. U. S. A. 99, 14506–14511.

Thomas, P.D., Kejariwal, A., Campbell, M.J., Mi, H., Diemer, K., Guo, N., Ladunga, I., Ulitsky-Lazareva, B., Muruganujan, A., Rabkin, S., Vandergriff, J.A., Doremieux, O., 2003. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res. 31, 334–341.

Ullah, M., Pelletier, N., Xiao, L., Zhao, S.P., Wang, K., Degerny, C., Tahmasebi, S., Cayrou, C., Doyon, Y., Goh, S.L., Champagne, N., Cote, J., Yang, X.J., 2008. Molecular architecture of quartet MOZ/MORF histone acetyltransferase complexes. Mol. Cell. Biol. 28, 6828–6843.

Veevers-Lowe, J., Ball, S.G., Shuttleworth, A., Kielty, C.M., 2011. Mesenchymal stem cell migration is regulated by fibronectin through alpha5beta1-integrin-mediated activation of PDGFR-beta and potentiation of growth factor signals. J. Cell Sci. 124, 1288–1300.

Wang, D., Park, J.S., Chu, J.S., Krakowski, A., Luo, K., Chen, D.J., Li, S., 2004. Proteomic profiling of bone marrow mesenchymal stem cells upon transforming growth factor beta1 stimulation. J. Biol. Chem. 279, 43725–43734.

Weiller, G.F., Caraux, G., Sylvester, N., 2004. The modal distribution of protein isoelectric points reflects amino acid properties rather than sequence evolution. Proteomics 4, 943–949.

Wong, N.K.Y., Fuller, M., Sung, S., Wong, F., Karsan, A., 2012. Heterogeneity of breast cancer stem cells as evidenced with Notch-dependent and Notch-independent populations. Cancer Med. http://dx.doi.org/10.1002/cam4.18.

Yang, Y., Zhang, S., Howe, K., Wilson, D.B., Moser, F., Irwin, D., Thannhauser, T.W., 2007. A comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-based protein identification and iTRAQ-based shotgun quantitative proteomics. J. Biomol. Tech. 18, 226–237.

Yi, W., Sun, Y., Wei, X., Gu, C., Dong, X., Kang, X., Guo, S., Dou, K., 2010. Proteomic profiling of human bone marrow mesenchymal stem cells under shear stress. Mol. Cell. Biochem. 341, 9–16.

Zhang, S., Xia, X., Shen, J., Zhou, Y., Sun, Z., 2008. DBMLoc: a database of proteins with multiple subcellular localizations. BMC Bioinforma. 9, 127.

Zhang, D.Y., Wang, H.J., Tan, Y.Z., 2011. Wnt/beta-catenin signaling induces the aging of mesenchymal stem cells through the DNA damage response and the p53/p21 pathway. PLoS One 6, e21397.

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., Ma, B., 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol. Cell. Proteomics 1–8. http://dx.doi.org/10.1074/mcp.M111.010587.

Zhu, W., Chen, J., Cong, X., Hu, S., Chen, X., 2006. Hypoxia and serum deprivation-induced apoptosis in mesenchymal stem cells. Stem Cells 24, 416–425.

Zhuang, Q., Li, J., Wu, Z., Zhang, J., Sun, W., Li, T., Yan, Y., Jiang, Y., Zhao, R.C., Qiu, G., 2011. Differential proteome analysis of bone marrow mesenchymal stem cells from adolescent idiopathic scoliosis patients. PLoS One 6, e18834.