Journal of Proteomics 150 (2017) 86–97

Contents lists available at ScienceDirect

Journal of Proteomics

journal homepage: www.elsevier.com/locate/jprot

Complementarity of SOMAscan to LC-MS/MS and RNA-seq for quantitative profiling of human embryonic and mesenchymal stem cells

Anja M. Billing, Hisham Ben Hamidane, Aditya M. Bhagwat, Richard J. Cotton, Shaima S. Dib, Pankaj Kumar, Shahina Hayat, Neha Goswami, Karsten Suhre, Arash Rafii, Johannes Graumann *

Research Division, Weill Cornell Medical College in Qatar, Doha, Qatar

ARTICLE INFO

Article history:
Received 7 June 2016
Received in revised form 1 August 2016
Accepted 29 August 2016
Available online 6 September 2016

Keywords:
Embryonic stem cells
Mesenchymal stem cells
Quantitative LC-MS/MS
RNA-seq
SOMAscan assay

ABSTRACT

Dynamic range limitations are challenging to proteomics, particularly in clinical samples. Affinity proteomics partially overcomes this, yet suffers from dependence on reagent quality. SOMAscan, an aptamer-based platform for over 1000 proteins, avoids that issue using nucleic acid binders. Targets include low expressed proteins not easily accessible by other approaches. Here we report on the potential of SOMAscan for the study of differently sourced mesenchymal stem cells (MSC) in comparison to LC-MS/MS and RNA sequencing. While targeting fewer analytes, SOMAscan displays high precision and dynamic range coverage, allowing quantification of proteins not measured by the other platforms. Expression between cell types (ESC and MSC) was compared across techniques and uncovered the expected large differences. Sourcing was investigated by comparing subtypes: bone marrow-derived, standard in clinical studies, and ESC-derived MSC, thought to hold similar potential but devoid of inter-donor variability and proliferating faster in vitro. We confirmed subtype-equivalency, as well as vesicle and extracellular matrix related processes in MSC. In contrast, the proliferative nature of ESC was captured less by SOMAscan, where nuclear proteins are underrepresented. The complementary of SOMAscan allowed the comprehensive exploration of CD markers and signaling molecules, not readily accessible otherwise and offering unprecedented potential in subtype characterization.

Significance: Mesenchymal stem cells (MSC) represent promising stem cell-derived therapeutics as indicated by their application in >500 clinical trials currently registered with the NIH. Tissue-derived MSC require invasive harvesting and imply donor-to-donor differences, to which embryonic stem cell (ESC)-derived MSC may provide an alternative and thus warrant thorough characterization. In continuation of our previous study where we compared in depth embryonic stem cells (ESC) and MSC from two sources (bone marrow and ESC-derived), we included the aptamer-based SOMAscan assay, complementing LC-MS/MS and RNA-seq data. Furthermore, SOMAscan, a targeted proteomics platform developed for analyzing clinical samples, has been benchmarked against established analytical platforms (LC-MS/MS and RNA-seq) using stem cell comparisons as a model.

1. Introduction

Measuring protein expression level differences is crucial to the fundamental understanding of biological systems. Although the current preferred technique in quantitative proteomics remains liquid chromatography coupled to mass spectrometry (LC-MS/MS), affinity-based techniques have recently gained in analytical power and consequently popularity. The aptamer-based SOMAscan assay [1], a relatively recent addition to the field of targeted affinity proteomics, allows for the simultaneous measurement and quantitation of 1095 proteins by 1129 unique SOMAmers (slow offrate modified aptamers; version 1.1 k). The assay was designed for the analysis of clinical samples, characterized by high complexity and dynamic range. By using only a few microliters of bio fluid (70 μl of plasma or 20 μg of protein sample), and allowing for high throughput measurements (upwards of 84 samples per day in its robotic implementation) it is suitable for clinical studies in the context of human genetic heterogeneity. According to the manufacturer, the dynamic range of the assay covers 8 orders of magnitude and measures molecules with high sensitivity (median lower limit of detection of 40 fM) and specificity superior to antibody-mediated

Abbreviations: ESC, embryonic stem cells; ESC-MSC, embryonic stem cell-derived mesenchymal stem cells; BM-MSC, bone marrow-derived mesenchymal stem cells; SOMA, SOMAscan assay, a trademark by SomaLogic, Boulder CO; RNA, RNA sequencing; PROT, liquid chromatography coupled to mass spectrometry-based proteomics; SDEs, significant differentially expressed.

* Corresponding author.
E-mail addresses: anb2061@qatar-med.cornell.edu (A.M. Billing), hbh2002@qatar-med.cornell.edu (H. Ben Hamidane), adb2018@qatar-med.cornell.edu (A.M. Bhagwat), rjc2003@qatar-med.cornell.edu (R.J. Cotton), ssd2002@qatar-med.cornell.edu (S.S. Dib), pankaj123@gmail.com (P. Kumar), shh2026@qatar-med.cornell.edu (S. Hayat), neg2007@qatar-med.cornell.edu (N. Goswami), kas2049@qatar-med.cornell.edu (K. Suhre), jat2021@qatar-med.cornell.edu (A. Rafii), jog2030@qatar-med.cornell.edu (J. Graumann).

http://dx.doi.org/10.1016/j.jprot.2016.08.023

detection [2], approaching the reported dynamic range expected in human plasma [3]. Aptamers in this assay are short single stranded 40 base DNAs including non-natural nucleotides. In marked difference to affinity proteomics using antibody arrays, this implies that they are easily produced synthetically and in bulk. Additionally they can be quantified through hybridization to complementary probes on a chip or slide, utilizing the mature technology suite stemming from RNA arrays. Sample preparation is simple, consisting only of dilution and not requiring any pre-processing steps. Overall, the SOMAscan assay (SOMA) targets a limited subset of the proteome but overcomes classical dynamic range limitations. Although it was developed specifically for high throughput screening of clinical samples in biological matrices such as serum, plasma and cerebrospinal fluid, it has recently been applied as well to cell extracts [4] and exosomes [5].

Here we used samples from human embryonic and mesenchymal stem cell populations to assess the complementarity of SOMAscan to more commonly used techniques: liquid chromatography coupled to high resolution mass spectrometry (LC-MS/MS) and next generation RNA sequencing (RNA-seq). The inclusion of different stem cell model systems allowed to probe both massive (ESC vs MSC) and subtle differences in protein expression (ESC-derived MSC vs bone marrow-derived MSC) and thus characterize the SOMAscan performances for both cases. In addition to the juxtaposition of the targeted proteomics assay to mass spectrometry-based proteomics and RNA-seq, we aim to provide complementary data on MSC to our previous report [6], particularly in terms of CD markers and quantitative proteomic cell type signature. This aim is particularly relevant for the latter cell type as, due to their reduced risk in tumor formation, ease of access from adult tissues, allogenicity and in vitro differentiation potential into cells of the mesodermal lineages (osteocytes, chondrocytes and adipocytes [7,8]), MSCs are of great interest in cell-based therapies. Functionally, MSC are immunomodulatory and support tissue regeneration through paracrine activity by secreted molecules or extracellular vesicles rather than differentiation into cells of the injured tissue [9]. Given their paracrine activity, we expected the aptamer-based targeted proteomics platform with its focus on secreted proteins to be advantageous in the analysis of MSC, particularly as those targets remain challenging for exploratory techniques.

2. Experimental procedures

2.1. Cell culture and SILAC labeling of embryonic stem cells (ESC) and mesenchymal stem cells (ESC-MSC, BM-MSC)

ESC-derived MSCs were obtained from three independent differentiation experiments using the previously reported protocol and have been phenotyped to show classic MSC behavior [10].

Bone marrow-derived mesenchymal stem cells were purchased with their mesenchymal characteristics verified. Details of the BM-MSC donors were as follows: 40y/m (StemCell, MSC-001F, lot#BM2893), 39y/m (Lonza, PT2505, lot#1F3422), 27y/m (Lonza, PT2505, lot#318,006), 20y/m (Lonza, PT2505, lot#8F3520).

Permission to use the human embryonic stem cell line (hESC) ES04 (WiCell institute) was obtained from the Cornell/Rockefeller/Sloan Kettering tri-institutional ESC research oversight committee. Funding was secured from nonfederal, US-external funding sources.

For mass spectrometry-based analysis, a triplex SILAC experimental design with label swapping was employed as previously described [6]. For SOMAscan-based analysis and RNA sequencing, cells were grown in standard conditions [6].

2.2. Sample preparation for the SOMAscan assay

Cells were harvested by dispase (ESC; WiCell institute) or trypsin (ESC-MSC, BM-MSC). After PBS washes of the cell pellet, total protein was extracted with the MEM-Per buffer (Thermo Fisher Scientific) supplemented with protease inhibitors (Complete, EDTA-free, Roche). Protein concentration was determined by Bradford assay and adjusted to 0.5 μg/μl. Total protein extracts of ESC, ESC-MSC and BM-MSC samples were sent to SomaLogic (Colorado, USA) as part of a fee for service agreement and subjected to SOMAscan analysis according to standard protocol [11].

2.3. Sample preparation for mass spectrometry

Protein samples were prepared as described previously [6]. Briefly, proteins were extracted using the Nuclear Extract Kit (ActiveMotif) to get cytosolic (CYT), nuclear (NUC) and chromatin-bound (CH) proteins. After methanol-chloroform precipitation, samples were resuspended in 8 M urea buffer (6 M urea, 2 M thiourea, 30 mM HEPES, pH 8) and digested with Lys-C for 3 h, followed by overnight trypsination after dilution to 2 M urea. Peptides were separated by in-solution isoelectric focusing into 12 fractions. After fractionation, peptides were desalted on C_{18} STAGETips [12]. Per sample, 36 fractions (12× CYT, 12× NUC, 12× CH) were measured by mass spectrometry.

2.4. Mass spectrometry

Peptides were subsequently analyzed by liquid chromatography (LC) using an EASY nLC-II system coupled to a Q Exactive mass spectrometer (MS) (Thermo Scientific, Bremen, Germany) as previously described [6,13].

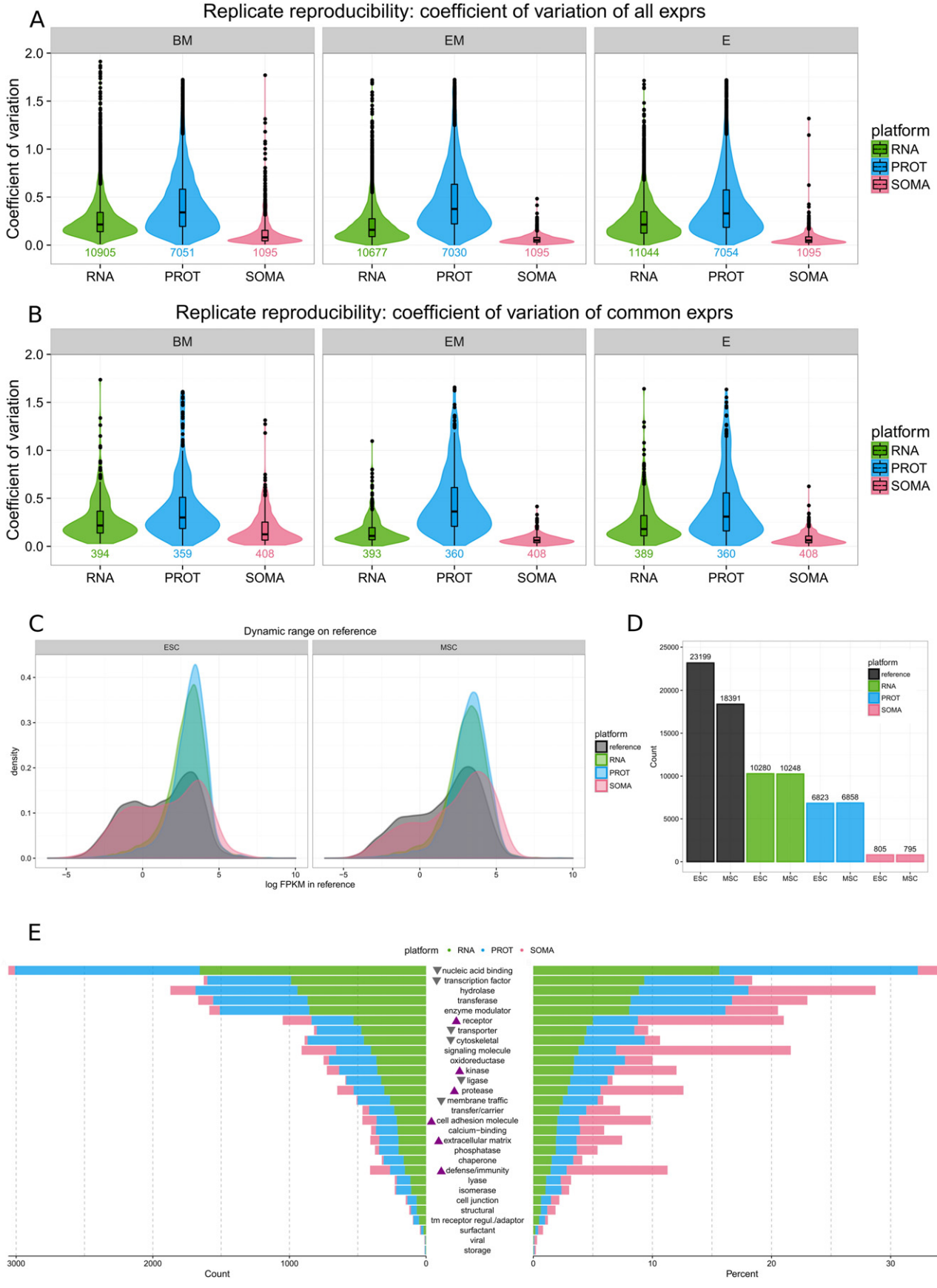
MS data was analyzed by MaxQuant suite of algorithms version 1.4.1.2 [14] using a Homo Sapiens database downloaded from UniprotKB on the 27th of November 2013 and comprising 88,473 protein isoforms entries. MaxQuant analysis was performed on all combined fractions and samples (36 MS runs per sample, grand total of 108 MS acquisitions) with an experimental design template reflecting the triplicated triplex SILAC approach using previously reported search parameter settings [6].

2.5. Next generation RNA sequencing

Next generation RNA sequencing (RNA-seq) was performed as previously described [6]. Briefly, total RNA was extracted with TRIZOL followed by an on-column cleaning step. 100 ng of total RNA was converted to cDNA using the Ovation RNA-seq System V2 (Nugen Technologies, San Carlos, CA). 2 μg of the amplified cDNA was sheared to 150–200 bp size distribution by Adaptive Focused Acoustics using a Covaris E220 instrument (Covaris, Woburn, MA). The sheared cDNA was end-repaired to generate blunt ends, then ligated to Illumina compatible adaptors with indexing tags, followed by 1× AMPure XP beads purification. The final NGS libraries were quantified using Agilent Bioanalyzer DNA Chip 1000 with 11 libraries per pool. Paired-end 100 bp deep sequencing was carried out on HiSeq 2500 (Illumina). RNA-seq analyses were performed using a customized pipeline composed of TopHat [15] Picard, (http://picard.sourceforge.net/index.shtml), Samtools [16] and Cuffdiff [15].

2.6. Statistical analysis

Proteomics data sets (SOMA, PROT) were analyzed with the empirical Bayes moderated *t*-test implemented by the limma bioconductor package [17] in the R environment [18]. *p* values were corrected for multiple hypothesis testing using the Benjamini-Hochberg method (FDR <0.05). Differential expression was calculated on normalized $\log_{10}$ intensities (SOMAscan) or $\log_2$ ratios (nano LC-MS/MS). Differential expression for RNA sequencing data was performed with the Cufflinks [15] package using the Cuffdiff function.


A
Replicate reproducibility: coefficient of variation of all exprs
BM
EM
E
Coefficient of variation
2.0
1.5
1.0
0.5
0.0
10905
7051
1095
10677
7030
1095
11044
7054
1095
RNA
PROT
SOMA
platform
RNA
PROT
SOMA
B
Replicate reproducibility: coefficient of variation of common exprs
394
359
408
393
360
408
389
360
408
C
Dynamic range on reference
ESC
MSC
density
log FPKM in reference
platform
reference
RNA
PROT
SOMA
D
Count
23199
18391
10280
10248
6823
6858
805
795
E
platform • RNA • PROT • SOMA
nucleic acid binding
transcription factor
hydrolase
transferase
enzyme modulator
receptor
transporter
cytoskeletal
signaling molecule
oxidoreductase
kinase
ligase
protease
membrane traffic
transfer/carrier
cell adhesion molecule
calcium−binding
extracellular matrix
phosphatase
chaperone
defense/immunity
lyase
isomerase
cell junction
structural
tm receptor regul./adaptor
surfactant
viral
storage
Count
Percent

2.7. Repeatability and dynamic range coverage comparison

Coefficients of variations (CV) were calculated per feature across replicates on logarithmized FPKM (fragments per kilobase of transcript per million mapped reads) values for RNA sequencing, normalized protein intensities for PROT and relative fluorescence units for SOMA. For PROT, normalized protein intensities were obtained by aggregation of peptide intensities and divided by the total protein intensity per replicate. Distributions of CVs were reported per technique and cell type as violin plots using R. Note that for CV of common gene products, only features quantified in all three replicates per technique are reported.

For dynamic range coverage assessment, reference FPKM values were obtained from ESC [19] (4 concatenated H1 replicates of the BioProject PRJNA269573) and MSC (3 concatenated BM-MSC samples of different donors, available at Galaxy, https://usegalaxy.org/u/cic19/h/mesenchymal-stem-cells-rnaseq) [20] deep transcriptome data. The reference sets were limited to annotated genes with FPKM values >0, retaining only the first isoform. For gene annotations consisting of multiple entries, only the first gene name in the group was used. Gene products identified in RNA, PROT and SOMA were mapped independently onto the reference data sets, considering only products measured in all three replicates. Resulting density distributions for the reference and each technique were overlaid to visualize their respective dynamic range coverage per cell type. Similarly, density plots for features commonly measured in all approaches have been plotted against their respective reference per cell type (Fig. S1).

2.8. Bioinformatics

For classification of proteins or genes, full id sets for each of the three techniques (RNA, PROT, SOMA) were submitted to PANTHER [21]. PCA was performed with the FactoMineR package v.1.29 [22] within the R environment. For cell type comparisons, PCA for PROT was based on non-normalized intensities derived from mass spectrometry. Enrichment and pathway analyses were performed using functionality developed in-house within the R environment [18].

2.9. Data availability

Both RNA and PROT data derive from an earlier publication are accordingly publicly available [6].

3. Results

3.1. Comparative evaluation of the analytical techniques: replicate reproducibility and dynamic range distribution

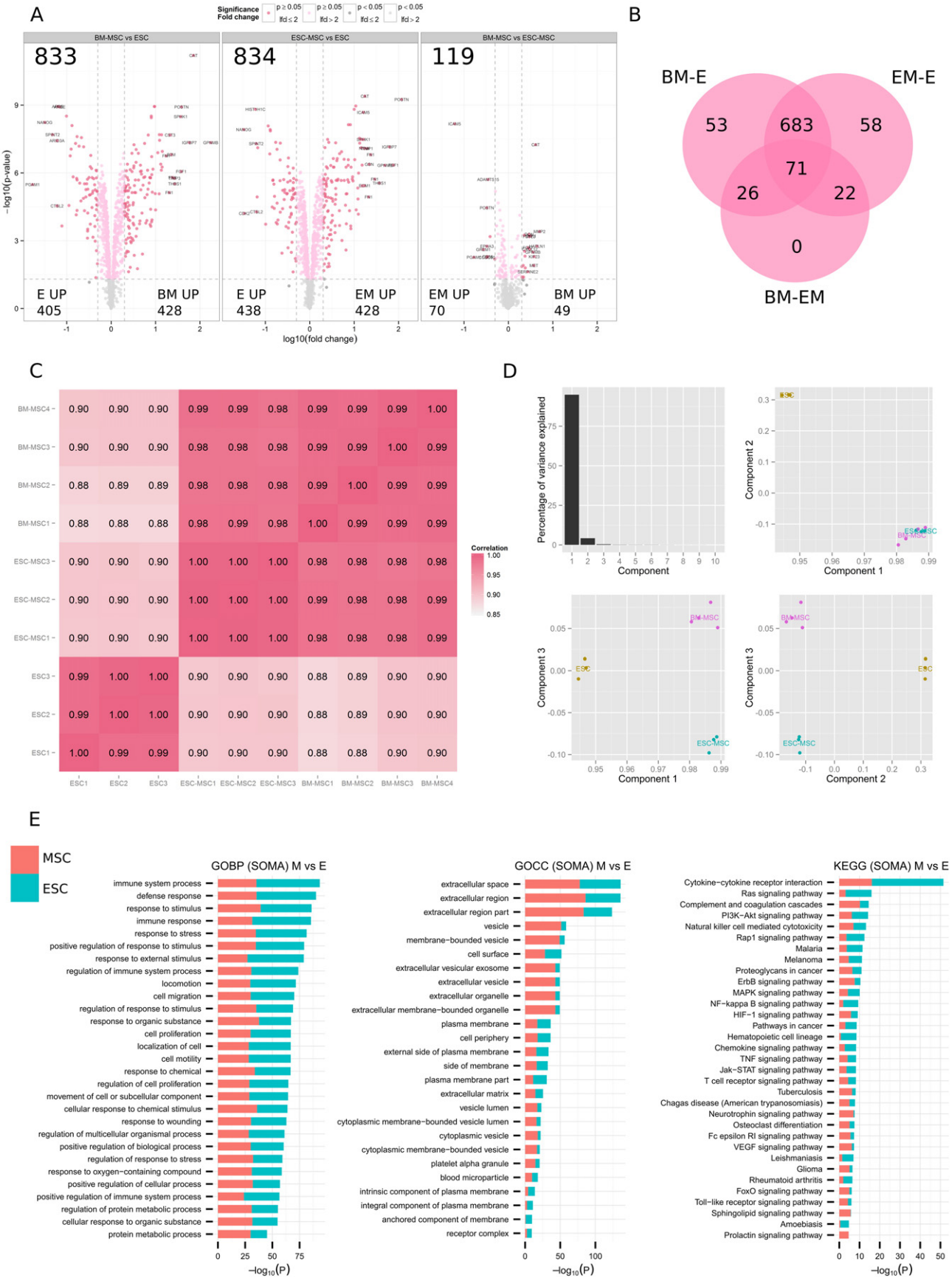
Evaluations of measurement repeatability and dynamic range were considered due to their relative translatability across techniques. Coefficients of variation (CV, also RSD or relative standard deviation) for quantified features were calculated per cell type (BM-MSC: BM, ESC-MSC: EM, ESC: E) between biological replicates for all features measured per technique (Fig. 1A). To increase comparability, data sets were filtered for features commonly measured by all three techniques and quantified in each replicate, which reduced their number to around 400 observations per platform (Fig. 1B). Overall, coefficients of variation (CV) appear comparable across techniques with SOMAscan displaying the narrowest distributions with the lowest median for complete and subset datasets. However, SOMAscan's distributions of CV also show significant kurtosis, particularly apparent for BM-MSC, highlighting disparity in measurement precision between targets. RNA and PROT both display wider CV distribution with technique-based specificities: larger outlier population for RNA and larger standard deviation and median values for PROT. The higher repeatability observed for SOMA can be attributed to its detection specificity (aptamer binding), unsurprisingly exceeding that of the non-targeted RNA and PROT approaches. It is important to note that CV for PROT are based on normalized protein intensities and not on logarithmized ratios (standard for quantitative data analysis) due to the inapplicability of CV to ratio analysis. The propagation of error inherent to a bottom up approach combined with the use of intensities over ratios directly accounts for the lower performance of PROT reported here.

When considering the subset of commonly measured features, the medians and standard deviations of the CV distributions remain similar to those observed in the non-subset data sets, but a bimodal trend for both RNA and PROT becomes apparent, presumably due to a subset of commonly measured features having lower signal to noise values and thus higher absolute variability. This is true in particular for BM-MSC where sample heterogeneity is increased by donor-derived genetic variability.

SomaLogic claims 8 orders of magnitude of dynamic range for their aptamer based assay [2]. In the cell model used, evaluation of that claim and quantification of the obtainable dynamic range are non-trivial, as, to the best of our knowledge, no global protein expression atlas exists for eukaryotes with the exception of yeast [23] or the more recent plasma proteome database [24]. Therefore, the dynamic range of the techniques was compared by mapping the identified gene products onto logarithmized FPKM distributions from common reference transcriptomes of extensive depth for ESC [19] and MSC [20] (Fig. 1C, 23,199 and 18,391 features, respectively). For each analytical platform, distributions of the features mapped to the reference dataset were overlaid per cell type. The overlap between techniques and the reference set provides an estimation of their intrinsic dynamic range for the cell model used. RNA and PROT behave very similar in terms of dynamic range distribution, with a minor sensitivity advantage for RNA. The large depth achieved by PROT derives from the extensive subcellular fractionation approach employed, narrowing the gap with RNA. However, only SOMA provides coverage of lower abundant features from the reference set. Additionally, the SOMA density distribution closely tracks that of the reference, indicating a lower dynamic range bias than for the other approaches used. When considering commonly quantified features on all platforms, the resulting density distributions demonstrate that most of the overlap between techniques occurs for highly abundant targets.

The feature numbers for the reference set for ESC and MSC as well as the quantified features per platform matching the reference are shown in Fig. 1D. Due to the implied analyte amplification, RNA is the most comprehensive technique with over 10,000 features matching the reference, corresponding to >90% of its quantified features, followed by in-depth MS-based proteomic profiling with 6800 features (also >90% of its reported quantified features). With approximately 800 features mapped to the reference set, corresponding to about 70% of the probes, SOMA not only displays the lowest number of mapped features (expected due to the comparatively low number of aptamers), but more importantly the lowest overlap with the reference set. This may stem from SOMAscan quantifying very low abundant proteins that often are at or below the limit of detection of other platforms, including the in-depth RNA sequencing data used as reference.

Fig. 1. Measurement repeatability, dynamic range coverage and protein group distribution across techniques. Violin plots of the coefficients of variation distributions per technique and cell type for (A) all measured features or (B) commonly measured features. (C) Dynamic range coverage of extensive ESC (left) and MSC (right) reference sets are reported for features measured by RNA, PROT and SOMA or commonly measured by all three techniques. (D) Quantified features from reference sets covered by each technique per cell type. (E) PANTHER protein class distributions for each platform (RNA, PROT, SOMA) presented as counts and percentages (left and right, respectively). Over- and underrepresented categories present for the aptamer-based assay (SOMA) are highlighted by purple and grey arrows, respectively.


A
Significance
Fold change
p ≥ 0.05 lfd ≤ 2
p ≥ 0.05 lfd > 2
p < 0.05 lfd ≤ 2
p < 0.05 lfd > 2
BM-MSC vs ESC
ESC-MSC vs ESC
BM-MSC vs ESC-MSC
833
834
119
E UP 405
BM UP 428
E UP 438
EM UP 428
EM UP 70
BM UP 49
−log10(p-value)
log10(fold change)
B
BM-E
EM-E
BM-EM
53
683
58
71
26
22
0
C
Correlation
BM-MSC4
BM-MSC3
BM-MSC2
BM-MSC1
ESC-MSC3
ESC-MSC2
ESC-MSC1
ESC3
ESC2
ESC1
D
Percentage of variance explained
Component
Component 1
Component 2
Component 3
ESC
BM-MSC
ESC-MSC
E
MSC
ESC
GOBP (SOMA) M vs E
immune system process
defense response
response to stimulus
immune response
response to stress
positive regulation of response to stimulus
response to external stimulus
regulation of immune system process
locomotion
cell migration
regulation of response to stimulus
response to organic substance
cell proliferation
localization of cell
cell motility
response to chemical
regulation of cell proliferation
movement of cell or subcellular component
cellular response to chemical stimulus
response to wounding
regulation of multicellular organismal process
positive regulation of biological process
regulation of response to stress
response to oxygen−containing compound
positive regulation of cellular process
positive regulation of immune system process
regulation of protein metabolic process
cellular response to organic substance
protein metabolic process
−log10(P)
GOCC (SOMA) M vs E
extracellular space
extracellular region
extracellular region part
vesicle
membrane−bounded vesicle
cell surface
extracellular vesicular exosome
extracellular vesicle
extracellular organelle
extracellular membrane−bounded organelle
plasma membrane
cell periphery
external side of plasma membrane
side of membrane
plasma membrane part
extracellular matrix
vesicle lumen
cytoplasmic membrane−bounded vesicle lumen
cytoplasmic vesicle
cytoplasmic membrane−bounded vesicle
platelet alpha granule
blood microparticle
intrinsic component of plasma membrane
integral component of plasma membrane
anchored component of membrane
receptor complex
−log10(P)
KEGG (SOMA) M vs E
Cytokine−cytokine receptor interaction
Ras signaling pathway
Complement and coagulation cascades
PI3K−Akt signaling pathway
Natural killer cell mediated cytotoxicity
Rap1 signaling pathway
Malaria
Melanoma
Proteoglycans in cancer
ErbB signaling pathway
MAPK signaling pathway
NF−kappa B signaling pathway
HIF−1 signaling pathway
Pathways in cancer
Hematopoietic cell lineage
Chemokine signaling pathway
TNF signaling pathway
Jak−STAT signaling pathway
T cell receptor signaling pathway
Tuberculosis
Chagas disease (American trypanosomiasis)
Neurotrophin signaling pathway
Osteoclast differentiation
Fc epsilon RI signaling pathway
VEGF signaling pathway
Leishmaniasis
Glioma
Rheumatoid arthritis
FoxO signaling pathway
Toll−like receptor signaling pathway
Sphingolipid signaling pathway
Amoebiasis
Prolactin signaling pathway
−log10(P)

3.2. Targeted proteomics using the SOMAscan assay: exploring pathways and cell signaling of ESC and MSC

3.2.1. Characterization of the subproteome targeted by SOMA

According to the PANTHER database [21], several protein classes are overrepresented in SOMA when compared to RNA and PROT, including signaling molecules (253), receptors (211), defense/immunity proteins (146), proteases (120), cell adhesion molecules (104), kinases (90) and extracellular matrix (66). The high number of signaling molecules stems from the fact that the assay was developed to target specifically proteins present in bio fluids such as plasma, and makes an excellent tool for comprehensive pathway screening, with, among the top KEGG pathways monitored by the assay: cytokine-cytokine receptor interaction (173), MAPK signaling (78), JAK-STAT signaling (65), chemokine signaling (58), cell adhesion (40), axon guidance (39), neurotrophin signaling (38), ErbB signaling (36), Toll-like receptor signaling (32) and TGFβ signaling (25). However, the corollary is the under-representation in the assay of organelle- or nucleus-related proteins, making direct comparison to non-targeted techniques more limited when considering cell line analysis (Fig. 1E).

3.2.2. Intra- and inter sample variability on SDEs

As expected, differences measured between embryonic and mesenchymal stem cells were massive, with 833 and 834 proteins overexpressed in ESC-MSC and BM-MSC versus ESC, respectively. The two MSC populations (ESC-MSC and BM-MSC) were very similar, with only 119 modulated proteins reported between them (Fig. 2A, B, Fig. S1 and Table S1). A correlation matrix shows a high consistency of measurements within each group and further demonstrates the differences between ESC and MSC and the equivalence of the two MSC subpopulations (Fig. 2C). Expectedly, ESC-MSCs derived from a single ES cell line are more homogenous than donor-derived BM-MSC. Differences between ESC and MSC can be explained by the first two components of a principal component analysis (PCA) covering 98% of the observed variance (component 1: 94%, component 2: 4%). Only the third component separates the two MSC types (Fig. 2D).

3.2.3. ESC features

391 SDE proteins were found up-regulated in ESC as compared to MSC, among which the strongest was the only pluripotency-associated transcription factor present in the assay: NANOG (31 fold) (Fig. S2B).

Applying a fold change cutoff of 2 to look for significant changes with higher amplitude, 71 and 66 proteins were found up-regulated in ESC when compared to ESC-MSC and BM-MSC, respectively, with 62 commonly elevated. Many of those (19) were nuclear proteins involved in cell cycle such as checkpoint kinase 1 (CHEK1), aurora kinase B (AURKB), cyclin B1 (CCNB1), cyclin-dependant kinase 2 (CDK2), karyopherin alpha 2 (KPNA2), pescadillo ribosomal biogenesis factor 1 (PES1), and polo-like kinase 1 (PLK1), reflecting the comparatively higher proliferation capacity of ESC. Immune response-related proteins were also found to be highly enriched in ESC and the top KEGG pathways associated to up-regulated proteins in ESC were cytokine-cytokine receptor interaction, Ras signaling, NF-kappa B signaling and hematopoietic cell lineage (Fig. 2E, Fig. S3, DataSet S1). When comparing the ESC profile to both MSC subtypes, very similar results were observed, confirming close relatedness of MSCs despite the sourcing differences.

3.2.4. MSC features

351 SDE proteins were found up-regulated in both MSC subtypes as compared to ESC; upon applying a fold change cut off of 2, that number decreased to 157 up-regulated proteins, with 102 common between ESC-MSC and BM-MSC. These were enriched for extracellular matrix (ECM) and bone development terms, conforming to the mesodermal differentiation potential of MSC, with periostin (POSTN), fibroblast growth factor 1 (FGF1), insulin-like growth factor binding protein 7 (IGFBP7) and transmembrane glycoprotein NMB (GPNMB) as the most up-regulated proteins.

Pathway analysis revealed that vesicle-related proteins are highly prominent in MSC, consistent with other findings reporting the central role of vesicle production in these cells [9,25]. Both MSC populations are enriched with almost identical terms (KEGG pathway) (Fig. 2E, Fig. S3, DataSet S1), among which were complement and coagulation cascade, osteoclast differentiation, ErbB, Hif-1, and TLR signaling. Notably, prolactin, sphingolipid, FoxO, VEGF and neurotrophin signaling were significantly up-regulated in MSC (Fig. 2E). This last finding is particularly relevant as BM-MSC have been shown to have neurotrophic potential and support synapse formation and neurological recovery [26]. Furthermore, neurotrophins regulating neurogenesis were shown recently to induce transdifferentiation of human BM-MSC into neurons [27].

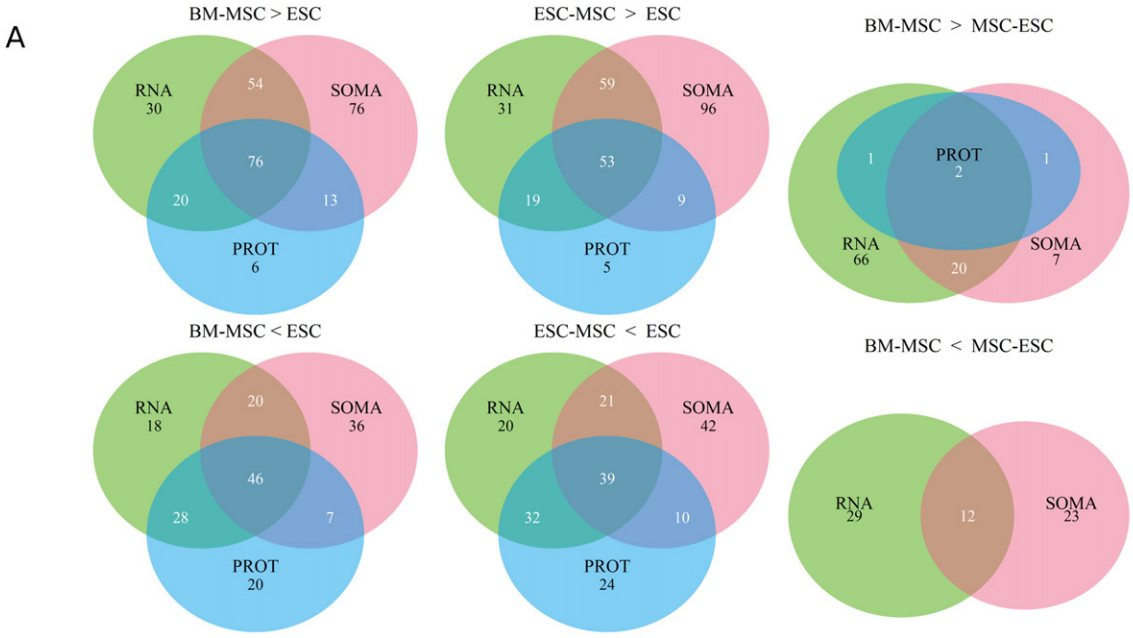
3.2.5. MSC subtype comparison

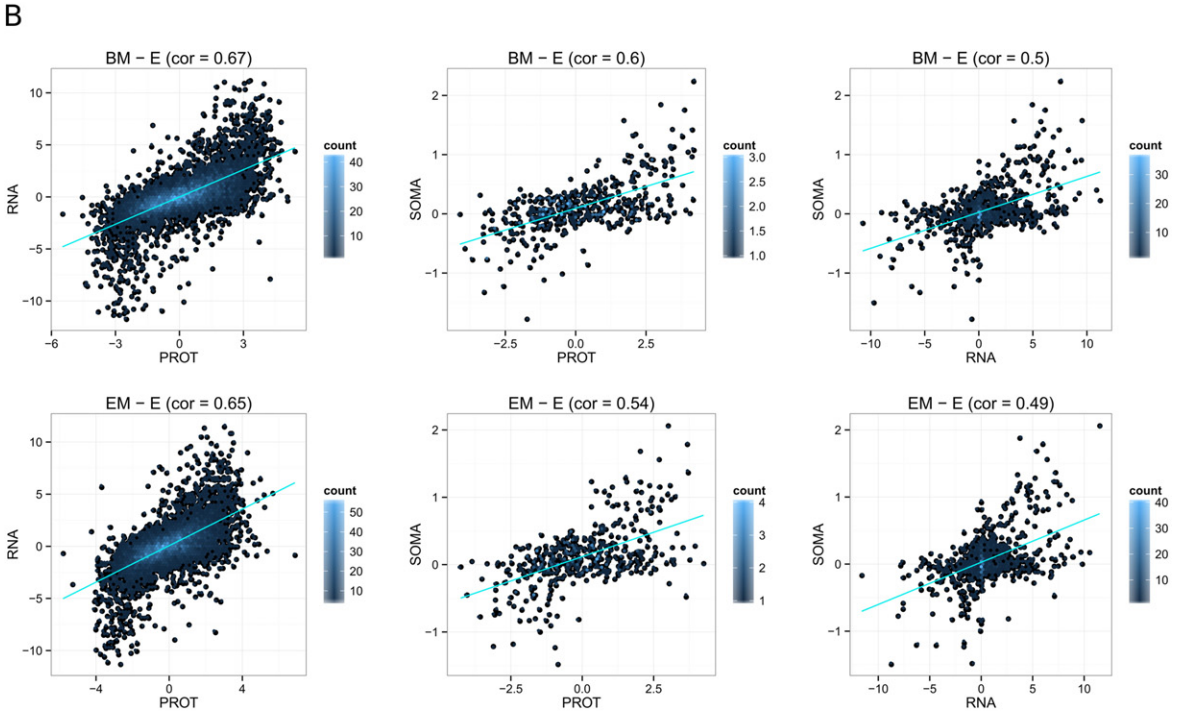
Comparing the two MSC populations revealed that proteins involved in organ and system development were more enriched in BM-MSC, whereas proteins involved in cellular component movement, cell migration, regulation of phosphorylation, extracellular matrix organization and cell communication more in ESC-MSC. The most distinct differences were found for signaling pathways, with cytokine-cytokine receptor interaction, Jak-Stat, neurotrophin and PPAR signaling enriched in ESC-MSC (Fig. S3C). In general, proteins differentially expressed between both MSC populations are associated with terms related to extracellular matrix (ECM) and vesicle proteins. The ECM protein matrix metallopeptidase (MMP) 2 is the strongest up-regulated protein in BM-MSC, whereas MMP1 is the dominant MMP in ESC-MSC. MMPs are implicated in homing [28], one of the characteristic features of MSC. Notably, ICAM5, the protein overexpressed strongest in ESC-MSC when compared to BM-MSC, is a neuron-specific adhesion molecule which stimulates neuron excitability after cleavage by MMPs [29]. Soluble ICAM5 has been shown to strongly induce neurite outgrowth and suppress immune response [30] and although it has not been associated with MSC functionality, it might be involved in the neurogenic potential of MSC.

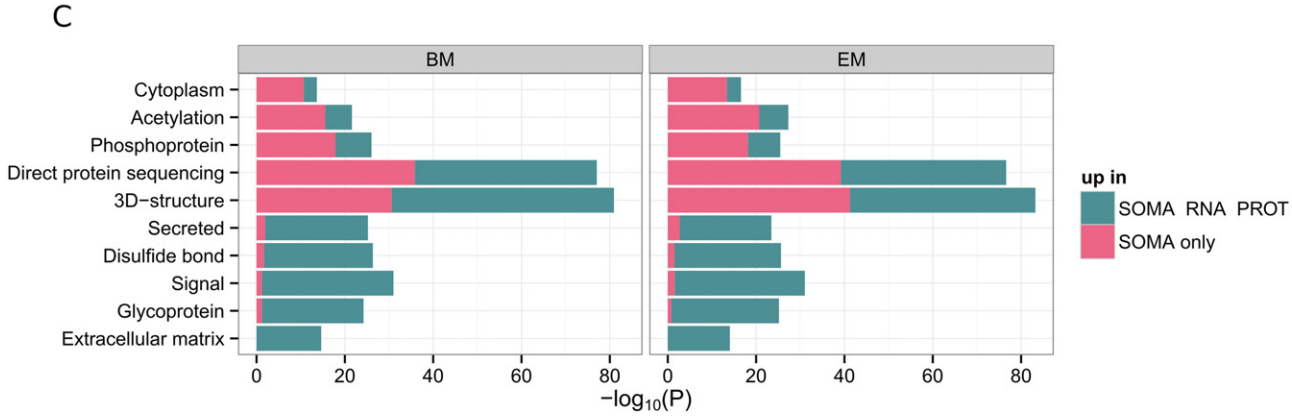
3.3. Comparison of SOMAscan data with LC-MS/MS and RNA sequencing

In total, 599 out of the 1095 proteins targeted by the assay were also measured by nano LC-MS/MS or RNA sequencing, which corresponds to an overlap of approximately 55%. Furthermore, 408 proteins were commonly quantified in all three datasets allowing for analytical platforms comparison (Fig. S4A, top16 in Fig. S6). It is worth noting that the SOMAscan-quantified proteins that were not detected by nano LC-MS/MS or RNA sequencing were low abundant signaling molecules such as cytokines, chemokines and interleukins. Due to this limited measurement overlap between platforms, the validation of the SOMAscan results is rendered difficult, as only 36% of the SDE proteins in SOMA were confirmed by either PROT, RNA or both (Fig. S4B). However, when considering the 408 proteins measured by all techniques, 60% of the SOMA results were validated (Fig. 3A), which is comparable to results obtained when comparing PROT to RNA data [6], however the correlation of expression difference direction is not as substantial as that observed for the latter two techniques (Fig. S4C). Indeed, MSC (both subtypes) to ESC $\log_2$ ratio correlations ranked as follows from best to

Fig. 2. Differential expression analysis for SOMA. (A) Volcano plots for BM-MSC vs ESC (BM-E), ESC-MSC vs ESC (EM-E) and BM-MSC vs ESC-MSC (BM-EM). The total number of differentially expressed proteins and the number of up-regulated proteins per cell type are reported in each panel. (B) Venn diagram of SDEs per comparisons (BM-E, EM-E, BM-EM). (C) Pearson correlation matrix for all replicates. (D) PCA was performed on $\log_{10}$ intensities plotting the percentage of variance per component and reporting 2D plots of the first 3. (E) Enrichment analysis based on GO biological process (GOBP), GO cellular compartment (GOCC), and KEGG pathways for up-regulated proteins in MSC (red) and in ESC (blue). MSC vs ESC comparisons are based on BM-E and EM-E intersections. Bar plots show the 20 most significant enriched terms per cell type sorted by the mean of $-\log_{10}p$ values.


A
BM-MSC > ESC
RNA 30
SOMA 76
PROT 6
54
76
20
13
ESC-MSC > ESC
RNA 31
SOMA 96
PROT 5
59
53
19
9
BM-MSC > MSC-ESC
PROT 2
RNA 66
SOMA 7
1
1
20
BM-MSC < ESC
RNA 18
SOMA 36
PROT 20
20
46
28
7
ESC-MSC < ESC
RNA 20
SOMA 42
PROT 24
21
39
32
10
BM-MSC < MSC-ESC
RNA 29
SOMA 23
12


B
BM − E (cor = 0.67)
BM − E (cor = 0.6)
BM − E (cor = 0.5)
EM − E (cor = 0.65)
EM − E (cor = 0.54)
EM − E (cor = 0.49)
RNA
SOMA
PROT
count


C
BM
EM
Cytoplasm
Acetylation
Phosphoprotein
Direct protein sequencing
3D−structure
Secreted
Disulfide bond
Signal
Glycoprotein
Extracellular matrix
−log10(P)
up in
SOMA RNA PROT
SOMA only

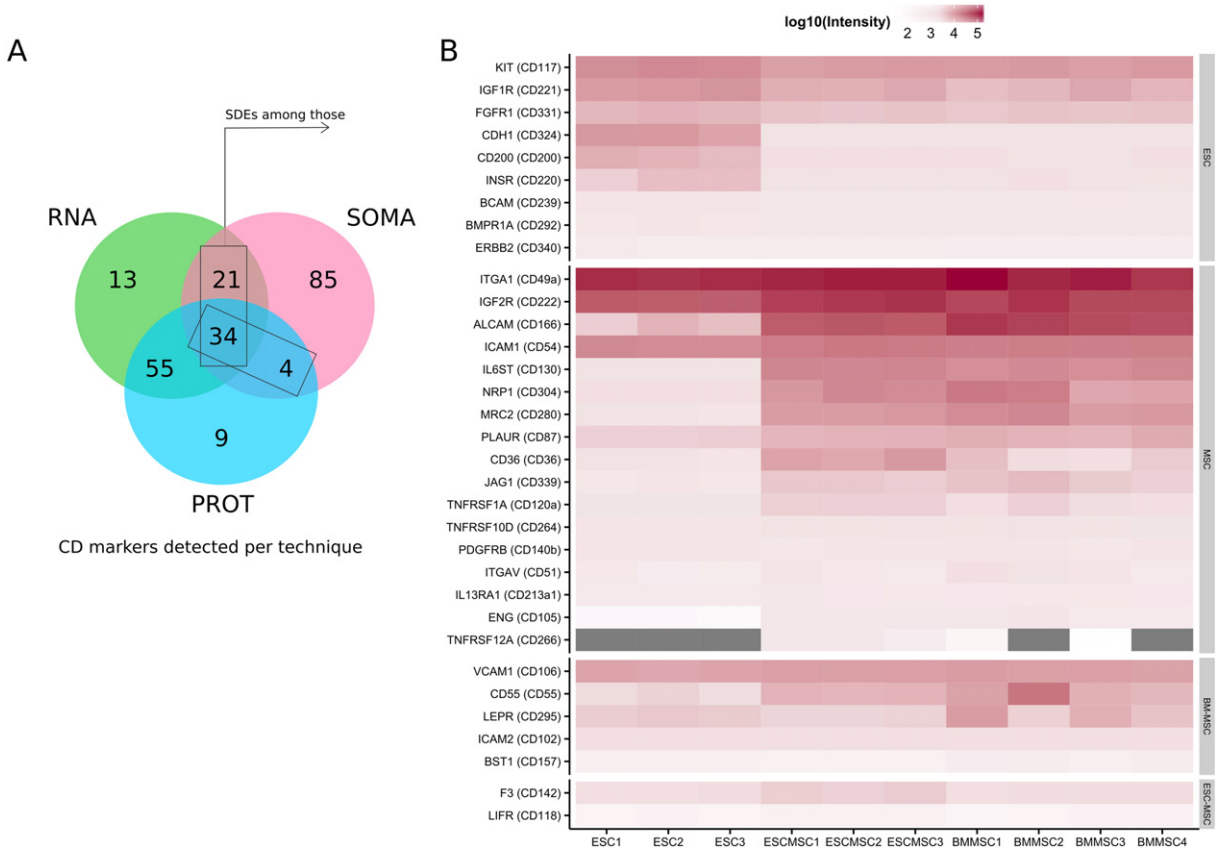


Fig. 4. CD markers. (A) Numbers of CD markers measured by SOMA, RNA and PROT. (B) Differentially expressed CD markers found by SOMA and validated by at least one alternative technique (RNA, PROT).

worst: PROT to RNA (BM-E: 0.67 and EM-E: 0.65), followed by SOMA to PROT (BM-E: 0.6 and EM-E: 0.54) and SOMA to RNA (BM-MSC vs ESC: 0.5 and ESC-MSC vs ESC: 0.49) (Fig. 3B). These results may be partially due to the limited overlap of identifications, but a major analytical difference in the characteristics of the techniques is likely to also play a central role: The underlying principle of the aptamer-based assay is epitope recognition, thus relative quantitation may reflect changes in protein abundance or epitope accessibility, which may occur for example through occlusion by post-translational modification or protein-protein interaction. In fact, enrichment analysis of proteins differentially expressed between subsets by SOMA supports this hypothesis. Among the 408 commonly measured proteins, the SOMA SDEs that were not validated by any of the other two techniques (top16 in Fig. S7) were found to be highly enriched for terms related to post-translational modification ("acetylation" FDR 1×10^{-16}, "phosphorylation" FDR 1×10^{-7}), reflecting the nature of SOMAscan as an assay monitoring epitope availability rather than protein concentration (Fig. 3C).

3.4. Enrichment analysis considering all three techniques: SOMAscan, LC-MS/MS and RNA sequencing

Enrichment analysis was performed for all three data sets and integrated (Fig. S5, DataSet S1). It is apparent that despite the limited overlap between SOMAscan and the other two techniques the targeted proteomics platform picks up very similar functions and pathways as shotgun proteomics and RNA sequencing (Fig. S5A). Top enriched biological processes in MSC include response to wound healing, regulation of cell proliferation, vesicle-mediated transport and cell adhesion. With respect to regeneration, proteins related to developmental processes, blood vessel and skeletal system development were enriched in both MSC populations and were found by all three omics techniques. Top KEGG pathways in MSC include for instance focal adhesion, PI3K-Akt, neurotrophin, and ErbB signaling.

ESCs are exceptional in terms of cellular content and nucleus to cytosol ratio. They are highly enriched in nuclear proteins, which are involved in cell cycle, RNA processing, DNA repair and transcription. Due to the under-representation of those protein classes in the SOMAscan assay, the characteristic nature of this cell type is not well reflected when compared to LC-MS/MS and RNA-seq (Fig. S5B). However, the targeted proteomics approach makes signaling molecules and receptors accessible also in ESC (Fig. S5B, C).

3.5. Phenotyping of ESC and MSC by surface markers: new insights provided by SOMAscan

With MSC being the most prominent candidate in stem cell-based therapy, a comprehensive characterization of surface markers becomes crucial. Previously, we presented a list of 137 CD markers of MSC, which were profiled by a combination of RNA-seq and LC-MS/MS [6]. Remarkably, the SOMAscan assay alone covers 144 CD markers and when

Fig. 3. Differential expression analysis for SOMA, RNA-seq and LC-MS/MS. (A) Cross validation of SOMA by RNA-seq (RNA) and nano LC-MS/MS (PROT). Comparison between significant differentially expressed proteins (SDEs) (FDR < 0.05) found by the three techniques (RNA, PROT, SOMA) for all six cases: BM-MSC > ESC, BM-MSC < ESC, ESC-MSC > ESC, ESC-MSC < ESC, BM-MSC > ESC-MSC, BM-MSC < ESC-MSC. Venn diagrams report only proteins that have been measured by all three techniques (RNA, PROT, SOMA). (B) Pearson correlation between the three large scale data sets (RNA, PROT, SOMA). (C) Enrichment analysis on SwissProt keywords for up-regulated proteins in MSC for BM-E and EM-E comparisons. Based on the 408 common proteins measured by all three techniques, SDEs found by SOMA were subset into proteins cross-validated by RNA and PROT (red) and proteins found solely by SOMA (blue).

A

validated by RNA, PROT, SOMA

MSC

CD markers
ITGAV (CD51)
ICAM1 (CD55)
PLAUR (CD87)
ENG (CD105)
TNFRSF1A (CD120a)
IL6ST (CD130)
PDGFRB (CD140b)
ALCAM (CD166)
IGF2R (CD222)
TNFRSF10D (CD264)
MRC2 (CD280)
NRP1 (CD304)
JAG1 (CD339)

BM-MSC
VCAM1 (CD106)

ESC-MSC
LIFR (CD118)

ESC

CD markers
Kit (CD117)
IGF1R (CD221)
BMPR1A (CD292)
CDH1 (CD324)
FGFR1 (CD331)
ERBB2 (CD340)

B

uniquely measured by SOMA

MSC

CD markers
CD70 (CD70)
LILRB2 (CD85b)
CD86 (CD86)
IL2RG (CD132)
CTLA4 (CD152)
SIGLEC1 (CD169)
FASLG (CD178)
CD207 (CD207)
CD209 (CD209)
TNFSF11 (CD254)
TNFSF13C (CD268)
CD274 (CD274)
ICOSL (CD275)
TLR2 (CD282)
CLEC4M (CD299)
NCR1 (CD335)
TNFRSF18 (CD357)

BM-MSC
IL3RA (CD123)
TNFSF13B (CD257)
EMR2 (CD312)
SIGLEC9 (CD329)
NCR3 (CD337)

ESC-MSC
MPL (CD88)
F3 (CD142)
CDH5 (CD144)

chemokines	interleukins	interleukin receptors
CCL7	IL3	IL2G
CCL14	IL19	IL12RB2
CCL19		IL23R
CCL3L1		

ESC

CD markers
FCGR3B (CD16)
ITGA2B (CD41)
GP1BA (CD42b)
CD48 (CD48)
SELP (CD62)
SELE (CD62E)
SELL (CD62L)
FCGR1A (CD64)
FCAR (CD89)
CSF3R (CD114)
CSF1R (CD115)
IL5RA (CD125)
IL6R (CD126)
TNFRSF4 (CD134)
TNFRSF9 (CD137)
TNFSF8 (CD153)
KIR2DL4 (CD158D)
MRC1 (CD206)
IL10RB (CDW210B)
IL12RB1 (CD212)
IL15RA (CD215)
LAG3 (CD223)
CD226 (CD226)
LY9 (CD229)
PLXNC1 (CD232)
FCER2 (CD23A)
TNFSF14 (CD258)
IL2RA (CD258)
TNFRSF10A (CD261)
TNFRSF13B (CD267)
TNFRSF17 (CD269)
FCRL3 (CD307c)
KLRK1 (CD314)
PECAM1 (CD314)
FCGR2A (CD32)
SIGLEC6 (CD327)
SIGLEC7 (CD328)
NCR2 (CD336)
SLAMF6 (CD352)
CRTAM (CD355)

chemokines	interleukins	interleukin receptors
CCL3	IL2	IL5RA
CCL4L1	IL4	IL6R
CCL13	IL12A	IL10RB
CCL17	IL17B	IL12RB1
CCL20	IL20	IL15RA
CCL21	IL24	IL22RA
CCL22	IL25	IL27RA
CCL23	IL34	
CCL24		
CCL25		
CCL27		
CXCL3		
CXCL10		
CXCL13		

combining it with the other two techniques, 221 CD markers were measured, with 85 uniquely added by SOMA (Fig. 4A, Table S2). Considering SOMA alone, significantly up-regulated in MSC when compared to ESC were 36 common CD markers with 9 specific for BM-MSC and 7 for ESC-MSC. Significantly up-regulated in ESC when compared to MSC were 61 common CD markers with 4 specific when compared to BM-MSC and 5 specific when compared to ESC-MSC. Both MSC populations were differentiated by 9 CD markers (MPL, CD80, F3, DDR1, CD36, TNFRSF12A, IL10RB, MST1R, LIFR) enriched in ESC-MSC and 3 CD markers enriched in BM-MSC (BST1, TEK, EMR2). Some differentially expressed CD markers were validated by at least one other technique (RNA, PROT) (Fig. 4B). CD markers cross validated by all three techniques can be considered a high confidence set and include: for ESC: Kit (CD117), IGF1R (CD221), BCAM (CD239), BMPR1A (CD292), CDH1 (CD324), FGFR1 (CD331), and ERBB2 (CD340); for MSC: ITGAV (CD51), ICAM1 (CD54), PLAUR (CD87), ENG (CD105), TNFRSF1A (CD120a), IL6ST (CD130), PDGFRB (CD140b), ALCAM (CD166), IGF2R (CD222), TNFRSF10D (CD264), MRC2 (CD280), NRP1 (CD304) and JAG1 (CD339) (Fig. 5A).

Some of these high confidence CD markers are well established ESC (CD117, CD324) and MSC markers (CD51, CD54, CD105 and CD166), while others are novel and potentially offer new avenues for cell type definition.

It is believed that surface receptors are involved in homing, immunomodulation and regeneration support of MSC. Among the high confidence MSC marker, Notch ligand Jagged-1 (JAG1, CD339), has recently been shown to be crucial for the MSC-induced expansion of TREG cells [31], further validating the data set beyond more canonical molecules.

High confidence CD marker found for BM-MSC was VCAM1 (CD106), for ESC-MSC LIFR (CD118). In concordance with others, CD106 was found to be low on ESC-MSC in comparison to BM-MSC. Its decrease correlates with diminished lineage differentiation, noted to be less efficient in ESC-MSC than BM-MSC when considering adipogenesis and chondrogenesis [32].

3.6. New findings based on uniquely quantified proteins by SOMAscan

SOMAscan as a targeted proteomics platform is very sensitive and can accurately quantify low expressed proteins, overcoming masking effects by high abundant complex backgrounds. A total of 496 proteins were uniquely measured by SOMA when compared to RNA and PROT (Fig. S4A) with 265 proteins significantly up-regulated in ESC (vs MSC, BM-MSC: 18, ESC-MSC: 23, both MSC: 224) and 131 proteins up-regulated in MSC (vs ESC, BM-MSC: 16, ESC-MSC: 12, both MSC: 103) (top 16 in Fig. S8). Among up-regulated proteins in ESC in comparison to both MSC were for instance chemokines (CCL3, CCL4L1, CCL13, CCL17, CCL20, CCL21, CCL22, CCL23, CCL24, CCL25, CCL27, CXCL3, CXCL10, CXCL13), interleukins (IL2, IL4, IL12A, IL17B, IL20, IL24, IL25, IL34), interleukin receptors (IL5RA, IL6R, IL10RB, IL12RB1, IL15RA, IL22RA, IL27RA), tumor necrosis factors (TNF, TNFSF8, TNFSF14, TNFSF15) and hormones (e.g. parathyroid hormone, pancreatic polypeptide, peptide YY, erythropoietin). Among up-regulated proteins in MSC in comparison to ESC were chemokines (CCL7, CCL14, CCL19, CCL3L1), interleukins (IL3, IL19), interleukin receptors (IL2G, IL12RB2, IL23R), interferons (IFNG, IFNA2), and tumor necrosis factors (TNFSF11). Among the 85 CD markers uniquely measured by SOMA, 17 were up-regulated in MSC (BM and EM) and 40 in ESC (Fig. 5B). Significantly up-regulated in ESC-MSC when compared to ESC and BM-MSC were proteins related to e.g. axon guidance (ICAM5, EPHA3), immune response (IL7, IL12B) and ECM composition (ADAMTS15, SPARCL1). When comparing to the most comprehensive proteomics datasets available on MSC including an antibody-based array of the secretome [33–35], 539 new proteins were measured by SOMA, 118 among those being significantly enriched in MSC. Including our LC-MS/MS dataset in this comparison, 466 proteins were uniquely measured, 97 of which being MSC specific.

4. Discussion

Performing a fair comparison between proteomics technologies is inherently difficult as detection paradigm and overall analytical approaches differ significantly. Although comparing targeted and untargeted approaches is bound to be biased, evaluating quantification repeatability across various platforms may provide useful insights into cross validation strategies. Keeping that in mind, we report here on the complementarity of the aptamer-based SOMAscan assay to other analytical platforms, namely mass spectrometry and RNA sequencing, in the context of mesenchymal stem cell sourcing characterization. The use of a cellular extract as a model system to compare across the analytical platforms allowed us to overcome the classical challenge associated with dynamic range in the analysis of bio fluids by mass spectrometry-based approaches. This comes, however, at the cost of applying SOMAscan to a sample type it is not designed to address. Due to its targeted nature and high sensitivity, the assay nonetheless allows quantification of targets typically at or below limits of detections of standard platforms. Indeed, dynamic range distributions showed SOMAscan to more comprehensively cover the abundance range of reference sets derived from deep transcriptomic analysis from the literature, whereas RNA and PROT gave substantially less congruency. Measurement repeatability assessment revealed comparable distributions between platforms with an expected advantage to the targeted approach. Although a comparison between the SOMAscan assay and a targeted mass spectrometry approach would be more appropriate, establishing targeted MS-assays (using multiple reaction monitoring) covering all SOMAscan targets (1129) was beyond the scope of this work.

Overall, the complementarity of platforms is highlighted by the balance between sensitivity and comprehensiveness and further confirms the benefits of combining approaches for in-depth systems biology. SOMAscan results are largely in accordance with the proteome and transcriptome measurements, based on the 408 proteins commonly quantified with all three techniques. Correlation with RNA sequencing reached 0.5, which is within the expected range between proteome and transcriptome, while comparing to MS-based proteome analysis the correlation score increased to 0.58. Overall, 60% of differentially expressed proteins reported by the SOMAscan assay and quantified by all three techniques were cross validated by either RNA sequencing, MS-based proteomics or both reflecting consistent changes in gene product abundance measured across all techniques. The remaining 40% of differentially expressed proteins which were not cross validated by the other techniques (LC-MS/MS, RNA sequencing) were enriched in terms related to posttranslational modification, suggesting that changes of epitope accessibility, for example by post-translational modification rather than variations in protein abundance, may be at the root of those divergences. Although the overlap of 408 proteins, which were measured across all three techniques, remains modest, enrichment and pathway analysis on the complete SOMAscan data set proved that the targeted proteomics platform gives similar results to MS-based proteomics and RNA sequencing with the notable exception of ESC - likely due to their high content in nuclear proteins, underrepresented in the assay. The proteomic equivalence between the two MSC subtypes has

Fig. 5. Selection of significant up-regulated proteins in MSC and ESC. (A) CD markers for MSC and ESC validated by all three techniques (RNA, PROT, SOMA). (B) CD markers, chemokines, interleukins and interleukin receptors for MSC and ESC measured uniquely in SOMA.

been confirmed by the SOMAscan assay, further validating ESC-MSC as a possible substitution to BM-MSC.

These findings highlight the capabilities and merits of the SOMAscan assay which proves complementary to other omics approaches and allows, due to its combined sensitivity and selectivity, to overcome dynamic range limitations, making it particularly adapted to investigation of signaling pathways.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2016.08.023.

Author contributions

JG, AR and KS directed the study. AMB designed the experiments. AMB and SSD carried out the experiments. AMB, HBH and AB analyzed proteomics data and performed bioinformatics analysis. HBH and NG measured samples by mass spectrometry. SH and PK performed RNA sequencing data processing. AMB, SH, PK and NY performed statistics. AMB, RJC, HBH and AMB created figures. AMB, HBH and JG drafted the manuscript.

Conflict of interest

The authors declare no competing financial interest.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgements

J.G. and the Proteomics Core at WCMC-Q are supported by "Biomedical Research Program" funds at Weill Cornell Medical College in Qatar, a program funded by Qatar Foundation. This work was supported by a grant from Qatar National Research Fund's National Priority Research Program (4-1267-1-194). The statements made herein are solely the responsibility of the authors.

References

[1] L. Gold, J.J. Walker, S.K. Wilcox, S. Williams, Advances in human proteomics at high scale with the SOMAscan proteomics platform, New Biotechnol. 29 (2012) 543–549, http://dx.doi.org/10.1016/j.nbt.2011.11.016.

[2] J.C. Rohloff, A.D. Gelinas, T.C. Jarvis, U.A. Ochsner, D.J. Schneider, L. Gold, N. Janjic, Nucleic acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents, Mol. Ther.—Nucleic Acids. 3 (2014), e201 http://dx.doi.org/10.1038/mtna.2014.49.

[3] S. Surinova, R. Schiess, R. Hüttenhain, F. Cerciello, B. Wollscheid, R. Aebersold, On the development of plasma protein biomarkers, J. Proteome Res. 10 (2011) 5–16, http://dx.doi.org/10.1021/pr1008515.

[4] M.R. Mehan, D. Ayers, D. Thirstrup, W. Xiong, R.M. Ostroff, E.N. Brody, J.J. Walker, L. Gold, T.C. Jarvis, N. Janjic, G.S. Baird, S.K. Wilcox, Protein signature of lung cancer tissues, PLoS One 7 (2012), e35157 http://dx.doi.org/10.1371/journal.pone.0035157.

[5] J. Webber, T.C. Stone, E. Katilius, B.C. Smith, B. Gordon, M.D. Mason, Z. Tabi, I.A. Brewis, A. Clayton, Proteomics analysis of cancer exosomes using a novel modified aptamer-based array (SOMAscanTM) platform, Mol. Cell. Proteomics 13 (2014) 1050–1064, http://dx.doi.org/10.1074/mcp.M113.032136.

[6] A.M. Billing, H. Ben Hamidane, S.S. Dib, R.J. Cotton, A.M. Bhagwat, P. Kumar, S. Hayat, N.A. Yousri, N. Goswami, K. Suhre, A. Rafii, J. Graumann, Comprehensive transcriptomic and proteomic characterization of human mesenchymal stem cells reveals source specific cellular markers, Sci. Rep. 6 (2016) 21507, http://dx.doi.org/10.1038/srep21507.

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

[8] M.F. Pittenger, A.M. Mackay, S.C. Beck, R.K. Jaiswal, R. Douglas, J.D. Mosca, M.A. Moorman, D.W. Simonetti, S. Craig, D.R. Marshak, Multilineage potential of adult human mesenchymal stem cells, Science 284 (1999) 143–147.

[9] S. Rani, A.E. Ryan, M.D. Griffin, T. Ritter, Mesenchymal stem cell-derived extracellular vesicles: toward cell-free therapeutic applications, Mol. Ther. 23 (2015) 812–823, http://dx.doi.org/10.1038/mt.2015.44.

[10] C.M. Raynaud, N. Halabi, D.A. Elliott, J. Pasquier, A.G. Elefanty, E.G. Stanley, A. Rafii, Human embryonic stem cell derived mesenchymal progenitors express cardiac markers but do not form contractile cardiomyocytes, PLoS One 8 (2013), e54524 http://dx.doi.org/10.1371/journal.pone.0054524.

[11] L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E.N. Brody, J. Carter, A.B. Dalby, B.E. Eaton, T. Fitzwater, D. Flather, A. Forbes, T. Foreman, C. Fowler, B. Gawande, M. Goss, M. Gunn, S. Gupta, D. Halladay, J. Heil, J. Heilig, B. Hicke, G. Husar, N. Janjic, T. Jarvis, S. Jennings, E. Katilius, T.R. Keeney, N. Kim, T.H. Koch, S. Kraemer, L. Kroiss, N. Le, D. Levine, W. Lindsey, B. Lollo, W. Mayfield, M. Mehan, R. Mehler, S.K. Nelson, M. Nelson, D. Nieuwlandt, M. Nikrad, U. Ochsner, R.M. Ostroff, M. Otis, T. Parker, S. Pietrasiewicz, D.I. Resnicow, J. Rohloff, G. Sanders, S. Sattin, D. Schneider, B. Singer, M. Stanton, A. Sterkel, A. Stewart, S. Stratford, J.D. Vaught, M. Vrkljan, J.J. Walker, M. Watrobka, S. Waugh, A. Weiss, S.K. Wilcox, A. Wolfson, S.K. Wolk, C. Zhang, D. Zichi, Aptamer-based multiplexed proteomic technology for biomarker discovery, PLoS ONE 5 (2010), e15004 http://dx.doi.org/10.1371/journal.pone.0015004.

[12] J. Rappsilber, Y. Ishihama, M. Mann, Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics, Anal. Chem. 75 (2003) 663–670.

[13] A.R. Liberski, M.N. Al-Noubi, Z.H. Rahman, N.M. Halabi, S.S. Dib, R. Al-Mismar, A.M. Billing, R. Krishnankutty, F.S. Ahmad, C.M. Raynaud, A. Rafii, K. Engholm-Keller, J. Graumann, Adaptation of a commonly used, chemically defined medium for human embryonic stem cells to stable isotope labeling with amino acids in cell culture, J. Proteome Res. 12 (2013) 3233–3245, http://dx.doi.org/10.1021/pr400099j.

[14] J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification, Nat. Biotechnol. 26 (2008) 1367–1372, http://dx.doi.org/10.1038/nbt.1511.

[15] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks, Nat. Protoc. 7 (2012) 562–578, http://dx.doi.org/10.1038/nprot.2012.016.

[16] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools, Bioinforma. Oxf. Engl. 25 (2009) 2078–2079, http://dx.doi.org/10.1093/bioinformatics/btp352.

[17] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (2015), e47 http://dx.doi.org/10.1093/nar/gkv007.

[18] R. Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2015http://www.R-project.org.

[19] M.P. Schwartz, Z. Hou, N.E. Propson, J. Zhang, C.J. Engstrom, V.S. Costa, P. Jiang, B.K. Nguyen, J.M. Bolin, W. Daly, Y. Wang, R. Stewart, C.D. Page, W.L. Murphy, J.A. Thomson, Human pluripotent stem cell-derived neural constructs for predicting neural toxicity, Proc. Natl. Acad. Sci. 112 (2015) 12516–12521, http://dx.doi.org/10.1073/pnas.1516645112.

[20] B. Roson-Burgo, F. Sanchez-Guijo, C.D. Cañizo, J.D.L. Rivas, Transcriptomic portrait of human mesenchymal stromal/stem cells isolated from bone marrow and placenta, BMC Genomics 15 (2014) 910, http://dx.doi.org/10.1186/1471-2164-15-910.

[21] H. Mi, A. Muruganujan, J.T. Casagrande, P.D. Thomas, Large-scale gene function analysis with the PANTHER classification system, Nat. Protoc. 8 (2013) 1551–1566, http://dx.doi.org/10.1038/nprot.2013.092.

[22] F. Husson, J. Josse, S. Le, J, Mazet, FactoMineR, Multivariate Exploratory Data Analysis and Data Mining, 2015 http://cran.r-project.org/web/packages/FactoMineR/index.html accessed February 8, 2015.

[23] S. Ghaemmaghami, W.-K. Huh, K. Bower, R.W. Howson, A. Belle, N. Dephoure, E.K. O'Shea, J.S. Weissman, Global analysis of protein expression in yeast, Nature 425 (2003) 737–741, http://dx.doi.org/10.1038/nature02046.

[24] V. Nanjappa, J.K. Thomas, A. Marimuthu, B. Muthusamy, A. Radhakrishnan, R. Sharma, A.A. Khan, L. Balakrishnan, N.A. Sahasrabuddhe, S. Kumar, B.N. Jhaveri, K.V. Sheth, R.K. Khatana, P.G. Shaw, S.M. Srikanth, P.P. Mathur, S. Shankar, D. Nagaraja, R. Christopher, S. Mathivanan, R. Raju, R. Sirdeshmukh, A. Chatterjee, R.J. Simpson, H.C. Harsha, A. Pandey, T.S.K. Prasad, Plasma proteome database as a resource for proteomics research: 2014 update, Nucleic Acids Res. 42 (2014) D959–D965, http://dx.doi.org/10.1093/nar/gkt1251.

[25] T. Katsuda, N. Kosaka, F. Takeshita, T. Ochiya, The therapeutic potential of mesenchymal stem cell-derived extracellular vesicles, Proteomics 13 (2013) 1637–1653, http://dx.doi.org/10.1002/pmic.201200373.

[26] M. Mauri, D. Lentini, M. Gravati, D. Foudah, G. Biella, B. Costa, M. Toselli, M. Parenti, S. Coco, Mesenchymal stem cells enhance GABAergic transmission in co-cultured hippocampal neurons, Mol. Cell. Neurosci. 49 (2012) 395–405, http://dx.doi.org/10.1016/j.mcn.2012.02.004.

[27] H.-L. Tsai, W.-P. Deng, W.-F.T. Lai, W.-T. Chiu, C.-B. Yang, Y.-H. Tsai, S.-M. Hwang, P.F. Renshaw, Wnts enhance neurotrophin-induced neuronal differentiation in adult bone-marrow-derived mesenchymal stem cells via canonical and noncanonical signaling pathways, PLoS One 9 (2014), e104937 http://dx.doi.org/10.1371/journal.pone.0104937.

[28] I.A.W. Ho, K.Y.W. Chan, W.-H. Ng, C.M. Guo, K.M. Hui, P. Cheang, P.Y.P. Lam, Matrix metalloproteinase 1 is necessary for the migration of human bone marrow-derived mesenchymal stem cells toward human glioma, Stem Cells Dayt. Ohio. 27 (2009) 1366–1375, http://dx.doi.org/10.1002/stem.50.

[29] M. Niedringhaus, X. Chen, R. Dzakpasu, K. Conant, MMPs and soluble ICAM-5 increase neuronal excitability within in vitro networks of hippocampal neurons, PLoS One 7 (2012), e42631 http://dx.doi.org/10.1371/journal.pone.0042631.

[30] C.G. Gahmberg, L. Ning, S. Paetau, ICAM-5: a neuronal dendritic adhesion molecule involved in immune and neuronal functions, Adv. Neurobiol. 8 (2014) 117–132.

[31] E.F. Cahill, L.M. Tobin, F. Carty, B.P. Mahon, K. English, Jagged-1 is required for the expansion of CD4+ CD25+ FoxP3+ regulatory T cells and tolerogenic dendritic cells

by murine mesenchymal stromal cells, Stem Cell Res. Ther. 6 (2015) http://dx.doi.org/10.1186/s13287-015-0021-5.
[32] P.T. Brown, M.W. Squire, W.-J. Li, Characterization and evaluation of mesenchymal stem cells derived from human embryonic stem cells and bone marrow, Cell Tissue Res. 1–16 (2014) http://dx.doi.org/10.1007/s00441-014-1926-5.
[33] R.J. Holley, G. Tai, A.J.K. Williamson, S. Taylor, S.A. Cain, S.M. Richardson, C.L.R. Merry, A.D. Whetton, C.M. Kielty, A.E. Canfield, Comparative quantification of the surfaceome of human multipotent mesenchymal progenitor cells, Stem Cell Rep. 4 (2015) 473–488, http://dx.doi.org/10.1016/j.stemcr.2015.01.007.
[34] S.T. Mindaye, M. Ra, J.L. Lo Surdo, S.R. Bauer, M.A. Alterman, Global proteomic signature of undifferentiated human bone marrow stromal cells: evidence for donor-to-donor proteome heterogeneity, Stem Cell Res. 11 (2013) 793–805, http://dx.doi.org/10.1016/j.scr.2013.05.006.
[35] S.K. Sze, D.P.V. de Kleijn, R.C. Lai, E.K.W. Tan, H. Zhao, K.S. Yeo, T.Y. Low, Q. Lian, C.N. Lee, W. Mitchell, R.M. El Oakley, S.-K. Lim, Elucidating the secretion proteome of human embryonic stem cell-derived mesenchymal stem cells, Mol. Cell. Proteomics MCP. 6 (2007) 1680–1689, http://dx.doi.org/10.1074/mcp.M600393-MCP200.