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 06 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.