Intact Cell Mass Spectrometry as a Quality Control Tool for Revealing Minute Phenotypic Changes of Cultured Human Embryonic Stem Cells

Petr Vanhara, Lukáš Kučera, Lubomír Prokeš, Lucie Jurečková, Eladia María Peña-Méndez, Josef Havel, Ales Hampl

Key Words. Embryonic stem cells • Cell culture • Differentiation • Technology • Tissue engineering

ABSTRACT

The stability of in vitro cell cultures is an important issue for any clinical, bio-industrial, or pharmaceutical use. Embryonic stem cells are pluripotent; consequently, they possess the ability to differentiate into all three germ layers and are inherently prone to respond to differentiation stimuli. However, long-term culture inevitably yields clones that are best adapted to the culture conditions, passing regimes, or differentiation sensitivity. This cellular plasticity is a major obstacle in the development of bio-industrial or clinical-grade cultures. At present, the quality control of cell cultures is limited by the lack of reliable (epi)genetic or molecular markers or by the focus on a particular type of instability such as karyotype abnormalities or adverse phenotypic traits. Therefore, there is an ongoing need for robust, feasible, and sensitive methods of determining or confirming cell status and for revealing potential divergences from the optimal state. We modeled both intrinsic and extrinsic changes in human embryonic stem cell (hESC) states using different experimental strategies and addressed the changes in cell status by intact cell mass spectrometry fingerprinting. The analysis of spectral fingerprints by methods routinely used in analytical chemistry clearly distinguished the morphologically and biochemically similar populations of hESCs and provided a biomarker-independent tool for the quality control of cell culture. Stem Cells Translational Medicine 2018;7:109–114

SIGNIFICANCE STATEMENT

Safe biomedical applications of cells require predictable and well-defined cell populations. This study describes intact cell mass spectrometry combined with chemometric and artificial intelligence approaches as a simple, robust, and hypothesis-free quality control tool for determining induced and spontaneous alterations in genetically and morphologically uniform embryonic stem cell populations in routine cell cultures.

INTRODUCTION

Cell-based applications are dependent on rigorous conditions including predictable and stable cell populations. Human embryonic stem cells (hESCs) paved the road for the development of state-of-the-art technologies for regenerative medicine and bio-industry applications. Derived from the pluripotent inner cell mass of a pre-implantation blastocyst, hESCs are inherently unstable and prone to rapid, spontaneous differentiation in vitro. Current culture protocols are focused on maintaining the pluripotent character of hESCs by repressing their differentiation and supporting sustained self-renewal. In culture, hESCs cumulatively acquire various alterations on both the genetic and non-genetic levels [1] and despite sophisticated culture techniques, culture-adapted clones are inevitably selected during long-term in vitro cultures [2]. However, these changes could remain unnoticed until they substantially alter the cell karyotype or cell phenotype, even in case of the stable expression of stemness-associated transcription factors including c-Myc, Sox-2, Klf4, Nanog, or Oct3/4, or their differentiation capacity or typical morphology. Further, molecular, genetic, and/or light-microscopy analyses can fail in the case of the genetically or karyotypically silent changes that arise in cultured cells.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) and its derivatives have diverse applications in classic analytical and structural chemistry and, thanks to their universality, have been adapted for analysis and determination of complex biological samples, including bacteria, yeasts, or mammalian cells [3, 4]. For monitoring of cell cultures, two possible approaches for analyzing mass
spectra are available. The “top-down” approach is based on revealing and identifying individual unique peaks and their use as biomarkers that are correlated with phenotypes or cell states, such as pluripotency, reprogramming events, or differentiation cues [5]. However, despite the precise identification of specific biomarkers that could be statistically associated with cell parameters, their applicability is often based on a random correlation among myriads of other dynamically changing molecules in a cell- or culture-dependent context. Also, we have recently documented that using one or two dominant peaks for quantitative analysis of a cell type reveals non-linear relationships that can be easily influenced by the technical or biological variability in the measurement or the sample itself [6]. The alternative approach assumes that a full spectral fingerprint is sufficiently robust to eliminate any undesired variability but contains enough information to discriminate species, cell types, or even cell states differing in chemical composition, even without the identification of individual molecules. Thus, the analysis without a preceding subjective reduction to several major peaks and their identification can eliminate the bias of the random correlation of individual molecules to a phenotypic change or a cell status and avoid potential misinterpretation. Avoiding the fractionation of cells also helps to reduce the unwanted variability while improving the overall consistency of the mass spectra and their reproducibility.

In addition to MS-based quantitative cell line authentication that we reported recently [6], here we demonstrate that the whole mass spectrum acquired from intact hESCs contains sufficient information to provide an unambiguous fingerprint also of a cell state in an otherwise uniform cell population and offers a simple tool for the monitoring and quality control of pluripotent stem cells in culture.

**Materials and Methods**

**Cell Culture**

Undifferentiated CCTL-14 hESCs (hPSCreg: MUNLe007-A, RRID: CVCL_C860), passage numbers 29, 72, and 269 were maintained in 6-cm Petri dishes (TPP, Trasadingen, Switzerland) on mitotically inactivated mouse embryonic fibroblasts (MEFs) in Dulbecco’s modified Eagle’s medium/F12 supplemented with 15% knockout serum replacement (both from Invitrogen, Life Technologies, Carlsbad, CA), 1-glutamine, minimum essential medium, nonessential amino acids, 0.5% penicillin-streptomycin (both from PAA Laboratories, Pasching, Austria), 2-mercaptoethanol (Sigma-Aldrich, Prague, Czech Republic), and 10 ng/ml fibroblast growth factor-2 (PeproTech, Rocky Hill, NJ). The hESCs were maintained in an incubator at 37°C in a humidified atmosphere containing 5% CO2 and the media was exchanged daily. Retinoic acid was purchased from Sigma-Aldrich (Prague, Czech Republic) and added to the fresh culture medium to the final concentration of 10−8 M.

**Isolation of RNA, Reverse Transcription, and Quantitative PCR**

Total RNA from hESCs was extracted using an RNeasy Mini Kit (Qiagen) and quantified using a NanoDrop (Thermo Scientific). cDNA was synthesized from 1μg DNase I-treated total RNA using a First Strand Reverse Transcription Kit (Roche, Prague, Czech republic). Relative expression was quantified using exon-spanning primers and UPL probes specific for SOX2, NANOG, OCT4, PAX6, SOX1, GATA4, and BRACHYURY. Glyceraldehyde 3-phosphate dehydrogenase was used as the housekeeping gene control (Roche, Prague, Czech Republic). Results were expressed as relative ratio units. All PCR reactions were performed in triplicate from three independent experiments, and reverse transcriptase-negative and template-negative controls were included.

**Immunofluorescent Microscopy**

The hESC colonies were manually harvested under visual control, washed in 1× phosphate-buffered saline (PBS), fixed in 4% formaldehyde for 15 minutes, washed three times in 1× PBS, permeabilized in 0.1% Triton-X100 in PBS for 10 minutes/RT, blocked in 3% BSA in PBS for 1 hour. Then, cells were incubated with anti-Sox2 (AB5603, Merck Millipore, Prague, Czech Republic, dilution 1:500) or anti-Oct3/4 (sc-5279, Santa Cruz Biotechnology, Heidelberg, Germany, dilution 1:100) primary antibodies overnight at 4°C. The cells were then washed three times in 1× PBS and incubated for 60 minutes room temperature with secondary antibody conjugated with AlexaFluor 568 (Sox2) or 488 (Oct3/4) and diluted 1:1,000 (Life Technologies/ThermoFisher Scientific, Prague, Czech Republic, A11004 or A11008, respectively), followed by additional wash in 1× PBS. Hoechst 33342 stain was added to the final concentration 5 μg/ml. Images were acquired using an automated microscope with ×10 objective (Image Xpress MicroXl, Molecular Devices).

**Flow Cytometry**

The hESC colonies were manually harvested under visual control, disaggregated using TryPLE Express (1×) for 2 minutes at 37°C, cells counted and resuspended in FACS buffer (5 mM EDTA, 2% FBS in 1× PBS). Prior staining of intracellular Sox2, cells were fixed using 3.7% formaldehyde in 1× PBS and permeabilized in BD Phosflow Perm Buffer III (BD Biosciences, Prague, Czech Republic). For blocking, 10% mouse serum in 1× PBS was used. Sox2 was stained using anti-Sox2 antibody (Abcam, Cambridge, UK, Ab97959), diluted 1:200 and incubated at 4°C/overnight and anti-rabbit IgG conjugated with Alexa Fluor 488 (Life Technologies/ThermoFisher Scientific, Prague, Czech Republic, A11008).

**Sample Preparation for Mass Spectrometry**

The hESC colonies were manually harvested under visual control, disaggregated using TryPLE Express (1×) for 2 minutes at 37°C, washed in 1× PBS and cells counted. To eliminate traces of PBS, the cells were centrifuged (200g, 2 minutes) and washed three times in 1 ml of 150 mM ammonium bicarbonate buffer. The cell numbers were determined using a Cedex XS cell counter operated with Cedex Control Center software v. 1.0.3. (Innovatis AG, Roche Life Sciences, Prague, Czech Republic). Finally, the cells were resuspended in 150 mM ammonium bicarbonate buffer to the final concentration of 100 × 106 cells per ml. The cell suspension was then mixed with an acidified matrix containing sinapinic acid (30 mg ml−1) and 2,2,2-trifluoroacetic acid (7.5% vol/vol) in a 2:1 sample to matrix ratio and applied to a target plate that was purified in an ultrasonic bath before spotting the samples.

**Mass Spectrometry**

The mass spectra were recorded in linear positive ion mode over the 2,000–12,000 m/z range using an AXIMA-CFR mass spectrometer from Shimadzu Biotech (Kratos, U.K.) equipped with a nitrogen laser (337 nm). The time pulse length of the laser was 3 ns. The maximum laser fluency was 60 mJ per pulse. The full laser power was indicated on the instrument as 180 arbitrary units.
The irradiated spot size was approximately 150 μm in diameter. An external calibration was performed using standard mixtures of peptides and proteins (TofMix, PepMix, LaserBio, France) and standard bacterial extracts (BTS standard) from Bruker BioSpin AG (Fällanden, Switzerland).

Processing of Mass Spectra

The spectral data (Supporting Information Figs. S1, S4) were exported using Launchpad software and imported into MATLAB 2014 Student edition from The MathWorks Inc. (Natick, Massachusetts, USA). The mass spectra in the complete data matrix were reduced to a uniform distribution of the 10,000 data points over the 2,000–12,000 m/z range and aligned using the peak alignment by the fast Fourier transform method. The baseline was subtracted, negative values were set to zero, and an average spectrum was calculated. Afterward, the normalization to the area under the curve or to the total ion count was performed.

Statistical Analysis

Pearson’s correlation and principal component analysis (PCA) were performed in STATISTICA 12 (StatSoft Inc., Tulsa). For the artificial neural network (ANN) analysis, Trojan Neural Network Simulator, Release 3.0 D 1996–1998 from Trojan Software Ltd. (Durham, U.K.) was used. We constructed an ANN containing four neurons in one hidden layer. The intensities of the processed mass spectra served as the input, whereas the sample categories represented the output. The “learning” of the ANN was performed using the back-propagation training algorithm as previously described [7, 8]. The back-propagation was achieved by iteratively adjusting the values of the connection weights in order to minimize the difference between the ANN calculated output value and the experimental value. The optimal ANN architecture was confirmed by plotting the root mean square of the sum of residuals value against the number of neurons in the hidden layer(s) and the number of training cycles (epochs). The ANN input variables (peaks) were selected upon PCA analysis as the peaks covering more than 95% of the variability in the dataset.

RESULTS

First, we investigated if intact cell MALDI-TOF MS, without any previous sample fractionation or identification of dominant peaks, could discriminate between morphologically uniform, but inherently different hESCs. As a well characterized model, we employed hESCs derived from RA-driven differentiation. It has been well documented in vitro that RA induces the differentiation of hESCs to the ectodermal phenotype, including neural lineages and keratinocytes or extra-embryonic tissues [12–14]. Thus, we picked the RA-driven differentiation model to evaluate the subtle changes that occur in the hESC populations during early differentiation. We stimulated the RA-treated hESCs with 10−8 M RA for a period of 12, 24, or 48 hours (Supporting Information Fig. S3A).

We found that the mass spectra from the RA-treated and control cells were highly correlated and lacked obvious specific peaks that would allow a straightforward search for biomarkers (Fig. 2A). Interestingly, the Pearson’s correlation matrix obtained from the average mass spectra revealed a decreased correlation within RA-treated P269-hESCs when compared to RA-treated P29-hESCs and P72-hESCs (Fig. 2B). We then performed the PCA of the control and RA-treated P29-hESCs and compared them with the P269-hESCs. The results revealed that the hESCs without any apparent morphological or molecular alterations were clearly clustered based only on the mass spectral fingerprint (Fig. 2C). Although the cluster analysis only allows for the analysis of an enclosed dataset, we wanted to test whether MS data could be used to predict and correctly categorize an unknown spectral sample. Thus, we constructed and trained a back-propagation ANN [7] based on a reduced dataset containing 23 peak intensities revealed by variable selection approaches (Fig. 2D). Critically, within these time intervals, the hESCs that were induced by RA to differentiate still possessed a typical uniform morphology, maintained high levels of Sox-2, Oct-4, and Nanog, and were completely indistinguishable from the untreated controls. Still, increased level of neuroectodermal determinant Pax6 and downregulation of mesodermal marker Brachyury, as determined by qRT-PCR, clearly spoke for RA-driven entry into differentiation pathway (Fig. 2E, Supporting Information Fig. S3B). The ANN correctly assigned the unknown spectrum to the appropriate sample. On the other hand, using the intensities of only 23 input peaks, the PCA failed to correctly cluster the samples (data not shown). This suggests that non-linear, self-learning approaches are suitable for handling complex data, allowing simple application for the monitoring of routine cultures or for quality control in current good manufacturing practice (CGMP) facilities.

DISCUSSION

The quality of stem cell cultures is an essential prerequisite for applications in which the product variability, lot-to-lot consistency, composition, contamination, or phenotype stability are significant. Here, we present a quality control tool based on the combination of intact cell MS (MALDI-TOF MS) and sophisticated statistical analysis that revealed the hidden variability in hESCs. Current approaches for routine cell identification or characterization use either DNA-based techniques, such as short tandem repeat (STR) profiling or the analysis of the expression of a defined set of
marker genes [15], or molecular and functional characterization of phenotype alterations [16]. However, both approaches focus on a limited set of variables (e.g., gene expression or STR profiling), or in theory do not cover unknown or uncharacterized parameters. The identification of individual biomarkers in large, high-dimensional datasets can also be biased by the formulation of an initial hypothesis, the biological background, or the selection of candidate biomarkers, or can be limited to a particular experimental scenario. For instance, a recent comparison of mouse induced pluripotent cells (miPSCs) and mouse ESCs using high-throughput lectin arrays coupled with surface plasmon resonance imaging revealed specific metabolomic profiles unique to miPSCs and mESCs, indicating differences in the global metabolism of cultured stem cells [17].

Recently, multiple reaction monitoring-mass spectrometry (MRM-MS) was introduced by Baud et al. for the rapid determination and quantitation of pluripotency markers in iPSCs [18]. MRM-MS precisely determined proteotypic peptides generated by

Figure 1. (A): Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry mass spectra obtained from intact hESCs of different passage numbers (P29, P72, and P269). The inset shows the Pearson's correlation of the peak intensities in the average spectra between the P29, P72, and P269 experimental groups. (B): PCA of spectral datasets. Each point in the PCA plot represents a unique biological sample. Abbreviations: hESCs, human embryonic stem cell; PCA, principal component analysis.
Figure 2. (A): Representative matrix-assisted laser desorption/ionization time-of-flight mass spectrometry mass spectra recorded from intact hESCs (P72) treated with RA for 12, 24, or 48 hours. (B): Heat map depicting the Pearson’s correlation matrix of the peak intensities obtained from the average mass spectra. (C): PCA plots of the control and RA-treated P29 or P269 hESCs. Scree plots documenting the contribution of individual factors to the overall variability. (D): Training and leave-one-out verification plots of the RMS versus the number of training cycles (epochs). The first 50,000 iterations are shown. The inset shows the architecture of the artificial neural network used for the prediction of the sample category. (E): Relative normalized expression of stemness (Sox-2, Nanog, Oct-4), neuroectoderm (Pax-6, Sox-1), endoderm (GATA-4), and mesoderm markers (Brachyury) as determined by qRT-PCR. hESCs were cultured with 10^{-8} M RA for 48 hours, or left untreated. The asterisk indicates statistically significant changes in gene expression (p < .05). Abbreviations: hESCs, human embryonic stem cell; PC, principal component; PCA, principal component analysis; RA, retinoic acid; RMS, root mean square.
tryptic digestion of 15 marker proteins associated with self-renewal of reprogrammed cells constituting a pluripotency fingerprint. The intact cell MS approach we propose, employs rather a global mass spectrum as the classification input instead of searching for individual biomarkers, and might represent an independent technique that reflects the rational scale of the intrinsic heterogeneity of the cells and is likewise sensitive enough to reveal small changes that are below the detection limits of other techniques. Interestingly, using this technique we visualized the potential phenotypic drifts that occurred during long-term culture that might contribute to the phenomena of culture adaptation [19]. The intact cell MS, followed by the straightforward multivariate analysis, could therefore discriminate unapparent but critical alterations in stem/progenitor cells that are not detected by other techniques, which has clear application for quality control in routine cell cultures or CGMP.

CONCLUSION

In summary, we introduced a simple, robust, and hypothesis-free quality control tool for determining induced and spontaneous alterations in genetically and morphologically uniform embryonic stem cell populations in routine cell cultures.

ACKNOWLEDGMENTS

This study was supported by the Grant Agency of the Czech Republic (GA15-11707S), the Grant Agency of Masaryk University [MUNI/A/1369/2016], the project COST CZ LD15144 by MEYS CR, the National Program of Sustainability II (project no. LQ1605, MEYS CR), and by funds from the Faculty of Medicine MU to junior researcher (PV).

REFERENCES

1. Amps K, Andrews PW, Anyfantis G et al. Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nat Biotechnol 2011;29:1132–U1113.
2. Canham MA, Sharov AA, Ko MS et al. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. PLoS Biol 2010;8:e1000379.
3. Croxatto A, Prod’hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev 2012;36:380–407.
4. Munteanu B, von Reitzenstein C, Hansch GM et al. Sensitive, robust and automated protein analysis of cell differentiation and of primary human blood cells by intact cell mass spectrometry and artificial neural networks. PLoS One 2016;11:e0147414.
5. Amato F, Lopez A, Pena-Mendez EM et al. Artificial neural networks in medical diagnosis. J Appl Biomed 2013;11:47–58.
6. Houska J, Pena-Mendez EM, Hernandez-Fernaud JR et al. Tissue profiling by nanogold-mediated mass spectrometry and artificial neural networks in the mouse model of human primary hyperoxaluria 1. J Appl Biomed 2014;12:119–125.
7. International Stem Cell Initiative. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat Biotechnol 2007;25:803–816.
8. Harrison NJ, Barnes J, Jones M et al. CD30 expression reveals that culture adaptation of human embryonic stem cells can occur through differing routes. Stem Cells 2009;27:1057–1065.
9. Nguyen HT, Geens M, Spits C. Genetic and epigenetic instability in human pluripotent stem cells. Hum Reprod Update 2013;19:187–205.
10. Baharvand H, Mehrjardi NZ, Hatami M et al. Neural differentiation from human embryonic stem cells in a defined adherent culture condition. Int J Dev Biol 2007;51:371–378.
11. Jagtap S, Meganathan K, Wagh V et al. All-trans retinoic acid and basic fibroblast growth factor synergistically direct pluripotent human embryonic stem cells to extraembryonic lineages. Stem Cell Res 2013;10:228–240.
12. Metallo CM, Ji L, De Pablo JJ et al. Retinoic acid and bone morphogenetic protein signaling synergize to efficiently direct epithelial differentiation of human embryonic stem cells. Stem Cells 2008;26:372–380.
13. Masters JR, Thomson JA, Daly-Burns B et al. Short tandem repeat profiling provides an international reference standard for human cell lines. Proc Natl Acad Sci USA 2001;98:8012–8017.
14. Irelan JT, Wu MJ, Morgan J et al. Rapid and quantitative assessment of cell quality, identity, and functionality for cell-based assays using real-time cellular analysis. J Biomol Screen 2011;16:313–322.
15. Nand A, Singh V, Wang PZ et al. Glycoprotein profiling of stem cells using lectin microarray based on surface plasmon resonance imaging. Anal Biochem 2014;465:114–120.
16. Baud A, Wessely F, Mazzacuva F et al. Multiplex high-throughput targeted proteomic assay to identify induced pluripotent stem cells. Anal Chem 2017;89:2440–2448.
17. Baharvand H, Aghdami N, ed. Advances in Stem Cell Research. 1st ed. Totowa, NJ: Humana Press, 2012:398.