

RESEARCH ARTICLE

# Multivariate Calibration Approach for Quantitative Determination of Cell-Line Cross Contamination by Intact Cell Mass Spectrometry and Artificial Neural Networks

Elisa Valletta<sup>1,2</sup>, Lukáš Kučera<sup>3</sup>, Lubomír Prokeš<sup>1</sup>, Filippo Amato<sup>1</sup>, Tiziana Pivetta<sup>2</sup>, Aleš Hampl<sup>3,4</sup>, Josef Havel<sup>1</sup>, Petr Vaňhara<sup>3\*</sup>

**1** Department of Chemistry, Faculty of Science, Masaryk University, Brno, Czech Republic, **2** Department of Chemical and Geological Sciences, University of Cagliari, Monserrato (CA), Italy, **3** Department of Histology and Embryology, Faculty of Medicine, Masaryk University, Brno, Czech Republic, **4** International Clinical Research Center, St. Anne's University Hospital Brno, Pekarska 53, 656 91, Brno, Czech Republic

These authors contributed equally to this work.

\* [pvanhara@med.muni.cz](mailto:pvanhara@med.muni.cz)



**OPEN ACCESS**

**Citation:** Valletta E, Kučera L, Prokeš L, Amato F, Pivetta T, Hampl A, et al. (2016) Multivariate Calibration Approach for Quantitative Determination of Cell-Line Cross Contamination by Intact Cell Mass Spectrometry and Artificial Neural Networks. PLoS ONE 11(1): e0147414. doi:10.1371/journal.pone.0147414

**Editor:** Patricia Talamas-Rohana, Centro de Investigacion y de Estudios Avanzados del Instituto Politecnico Nacional, MEXICO

**Received:** September 23, 2015

**Accepted:** January 4, 2016

**Published:** January 28, 2016

**Copyright:** © 2016 Valletta et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This study was supported by Grant Agency of Masaryk University (MUNI/M/0041/2013 and MUNI/A/1558/2014), from the European Regional Development Fund (CZ.1.05/1.1.00/02.0123 and CZ.1.07/2.3.00/20.0185) and by funds from the Faculty of Medicine MU to junior researcher (Petr Vaňhara). Elisa Valletta gratefully acknowledges the Masaryk University and its Department of Chemistry,

## Abstract

Cross-contamination of eukaryotic cell lines used in biomedical research represents a highly relevant problem. Analysis of repetitive DNA sequences, such as Short Tandem Repeats (STR), or Simple Sequence Repeats (SSR), is a widely accepted, simple, and commercially available technique to authenticate cell lines. However, it provides only qualitative information that depends on the extent of reference databases for interpretation. In this work, we developed and validated a rapid and routinely applicable method for evaluation of cell culture cross-contamination levels based on mass spectrometric fingerprints of intact mammalian cells coupled with artificial neural networks (ANNs). We used human embryonic stem cells (hESCs) contaminated by either mouse embryonic stem cells (mESCs) or mouse embryonic fibroblasts (MEFs) as a model. We determined the contamination level using a mass spectra database of known calibration mixtures that served as training input for an ANN. The ANN was then capable of correct quantification of the level of contamination of hESCs by mESCs or MEFs. We demonstrate that MS analysis, when linked to proper mathematical instruments, is a tangible tool for unraveling and quantifying heterogeneity in cell cultures. The analysis is applicable in routine scenarios for cell authentication and/or cell phenotyping in general.

## Introduction

In current biomedical research, cells cultured *in vitro* are irreplaceable experimental models and biotechnological tools. However, the research performed on immortalized or tumor-derived cell lines is dependent on proper cell identity and faces continuous significant risk of data misinterpretations due to inadvertent cross-contamination by another cell line [1–4].

for the kind hospitality and for funding her stay in Brno. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

Contamination can easily occur by improper passaging, sharing of culture media for multiple cell lines, or inaccurate labeling and storage. Indeed, sophisticated techniques of cell culture and tissue engineering, such as high-throughput reactors, microfluidics, and stem cell or tissue cultures, require stringent monitoring of cell identity and phenotype stability [5–7]. The current gold standard for authentication of individual cell lines is analysis of Short Tandem Repeat sequences (STR) or Simple Sequence Repeats (SSR) widespread throughout the genome, since each cell line theoretically has a unique STR-profile [1,2]. In specific scenarios, such as co-culture setups of two cell lines, use of various cell populations derived from a single individual or from inbred strains, or the occurrence of phenotypic changes within STR- or otherwise stable cell lines, STR analysis cannot provide sufficient discrimination. Furthermore, in cases of intrinsic heterogeneity or impurity of cell cultures, co-culture setups, microsatellite instability, phenotype shifts, or viral or mycoplasma infections, STR analysis can provide ambiguous results, or, in the best case, allow only a qualitative assessment of cell identity without any information on the extent of contamination or heterogeneity in cell populations. Techniques that can complement STR authentication, such as identification of phenotype- or genotype-related markers, e.g. karyotype, isoenzymes, surface markers, or single nucleotide polymorphisms (SNPs) [8,9], are nevertheless dependent on preceding knowledge of the biological background of the model system used. Optimization of high-resolution methods common in physical and analytical chemistry and advanced mathematical modeling can circumvent the need for specific markers by analysis of global cellular or tissue patterns. Spectral techniques, such as Raman near-infrared or mass spectrometry were recently demonstrated to provide global fingerprints with sufficient capacity to distinguish diseased and normal tissues in models of metabolic disorders, or even individual states of cell differentiation or metabolism [10–13].

Matrix-assisted laser desorption/ionization–time of flight mass spectrometry (MALDI-TOF MS) has recently been used in fields beyond analytical and structural chemistry, such as biomedical research or clinical practice, and has been adapted for characterization of complex biological samples by peptide-mass-fingerprinting or peptide sequence tagging. Biotyping of non-fractionated intact microorganisms by MS is now a fast, routine, and cheap technique in clinical microbiology. Similarly, the concept of biotyping eukaryotic cells by intact cell (IC) MALDI-TOF MS has been suggested to allow identification of cell lines [13–18] or to characterize of physiological events occurring in the cells, such as terminal differentiation or programmed cell death [19]. The IC MALDI-TOF MS technique allows for recording the characteristic profiles of eukaryotic cells in quality sufficient for detailed analyses ranging from ultrastructural molecular cytology, to deep cell phenotyping and tissue analyses [13,15,18–22], to species recognition [23] and ecotoxicology [24,25].

However, a simple visual inspection of mass spectra is often not sufficient to establish an unambiguous cell line-specific set of biomarker peaks. Moreover, specific signal intensity and analytic concentration are not linear except in a narrow concentration range due to various stochastic “MALDI effects,” such as variability in matrix/analyte interactions and energy dissipation or quenching or enhancing of ionization [26]. Mathematical methods that are successfully used in chemometrics, such as bivariate regression, polynomial fitting, multiple linear regression, partial least squares, and artificial intelligence, must therefore be applied in MS analysis of complex biological samples [27–29]. Artificial neural networks (ANNs) represent a robust and versatile mathematical tool for many applications in various fields [30]. ANNs mimic the “learning” and “generalization” abilities of human neural structures. ANNs are able to model highly complex non-linear systems and are used for classification, pattern recognition, modeling, and multivariate data analysis [31]. The basic units of ANNs are “nodes” or “neurons.” They are organized in one “input” layer, in one or more “hidden” layers, and in one “output” layer. Each of the  $i$ -th neurons in a layer is linked to all the  $j$ -th neurons in the next layer. Each

connection is weighted with a weight  $w_{ij}$ . The role of the neurons in the input layer is to receive input data and transfer it to the neurons in the hidden layer through the weighted connections. The neurons in the hidden layer(s) perform mathematical operations on the incoming data (summation, addition of a “*bias*” term, and transformation by a suitable mathematical function). The result is then transferred to the neurons in the output layer where the ANN output is calculated.

Here we study the possibility of using ANNs to determine a quantity of cells of a particular cell line and/or type in two-component mixtures, mimicking a scenario of cell line cross-contamination. To create such situations, we used: a) line CCTL14 of human embryonic stem cells (hESCs), b) mouse embryonic fibroblasts (MEFs), and c) line R1 of mouse embryonic stem cell (mESCs). The cells were arranged into two-component calibration mixtures of hESCs + MEFs and hESCs + mESCs in various ratios. Line CCTL14 of hESCs has previously been thoroughly characterized [32]. MEFs freshly isolated from connective tissue of 11.5 days old mouse embryos are commonly used as a supportive feeder layer for hESCs in a routine co-culture mode. Mouse ESCs [33] represent pluripotent and self-renewing cells that are developmentally and functionally similar to hESCs.

The two-component cell suspensions were analyzed by ANN-coupled IC MALDI-TOF MS in a multivariate calibration approach. We demonstrate that mass spectra contain sufficient information to identify the presence of individual cell types in mixtures, and we report for the first time that ANN analysis of mass spectra from two-component mixtures can correctly predict the level of cell cross-contamination in very complex microenvironment.

## Material and Methods

### Chemicals

Knockout Dulbecco’s modified Eagle’s medium (DMEM), DMEM/F12, knockout serum replacement, fetal bovine serum, L-glutamine, minimum essential medium non-essential amino acids, penicillin-streptomycin, and TrypLE™ Select were purchased from Gibco, Life Technologies Czech Republic Ltd. (Prague, Czech Republic). Ammonium bicarbonate, sinapinic acid, trifluoroacetic acid, DMEM, 2-mercaptoethanol, phosphate buffered saline (PBS), and gelatin were purchased from Sigma-Aldrich Ltd. (Prague, Czech Republic). Matrigel™ was purchased from BD Bioscience, I.T.A.-Intertact Ltd. (Prague, Czech Republic). Fibroblast growth factor-2 was purchased from PeproTech, Baria Ltd. (Prague, Czech Republic). Leukemia Inhibitory Factor Protein was purchased from Chemicon, Merck Millipore (Prague, Czech Republic). Acetonitrile was purchased from J.T. Baker, VWR International Ltd. (Prague, Czech Republic). PepMix standard was purchased from Laser BioLabs (Sophia-Antipolis, France). Tissue culture dishes were purchased from TPP (Trasadingen, Switzerland).

### Cell cultures

Mouse embryonic fibroblasts derived from CF1-mouse embryos were cultured in tissue culture dishes in medium consisting of Knockout DMEM supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 1% minimum essential medium non-essential amino acids, 1% penicillin-streptomycin, and 0.1 mM 2-mercaptoethanol as described previously [34,35]. Human embryonic stem cells [36,37] were cultured in the undifferentiated state in tissue culture dishes coated with Matrigel™ in culture media conditioned by MEFs consisting of DMEM/F12 supplemented with 15% knockout serum replacement, 2 mM L-glutamine, 1% minimum essential medium non-essential amino acids, 0.5% penicillin-streptomycin, 0.1 mM 2-mercaptoethanol, and 4 ng/ml fibroblast growth factor-2. Mouse embryonic stem cells [33] were cultured in tissue culture dishes coated with 0.1% gelatin in medium consisting of DMEM supplemented

with 20% fetal bovine serum, 1% minimum essential medium non-essential amino acids, 1% penicillin-streptomycin, 1 mM 2-mercaptoethanol and 5.5 µg/ml leukemia inhibitory factor (LIF). All cell lines were maintained in an incubator at 37°C with a humidified atmosphere containing 5% CO<sub>2</sub>, with daily media exchange.

## Sample preparation

Cultured cells were washed with 1×PBS and enzymatically disaggregated to single cell suspension using TrypLE™ Select. After 2 min, the enzymatic activity was stopped by the respective culture medium. Detached cells were pelleted by centrifugation at 200 g for 5 min and washed once again with 1×PBS. Cell number was determined by CEDEX XS cell counter operated with CEDEX Control Center software v. 1.0.3. from Innovatis AG, Roche Life Sciences (Prague, Czech Republic).

## Preparation of cell mixtures for MS analysis

**hESCs + MEFs mixture.** Cell suspensions containing a total of  $1 \times 10^6$  hESCs and MEFs in 1×PBS in defined ratios were pelleted by centrifugation at 200 g for 5 min at 4°C and washed three times with an aqueous solution of 150 mM ammonium bicarbonate. Then, the cell pellets were resuspended in 10 µl of 150 mM ammonium bicarbonate and mixed with 5 µl of freshly prepared sinapinic acid matrix (30 mg/ml in 70% acetonitrile and 7.5% trifluoroacetic acid). Two microliters of sample/matrix mixture were immediately spotted in pentaplicates onto the MALDI target and dried at room temperature.

**hESCs + mESCs mixture.** Aliquots of  $1 \times 10^6$  mESCs or hESCs in 1×PBS were pelleted by centrifugation at 200 g for 5 min at 4°C. Supernatant was discarded and pelleted cells were washed three times with 150 mM ammonium bicarbonate solution. Resulting cell pellets were then snap-frozen and stored until further processing. At the time of analysis, both mESCs and hESCs aliquots were quickly thawed and reconstituted in 20 µl of 150 mM ammonium bicarbonate solution and sonicated briefly in a water ultrasound bath. Then, MEFs and hESCs were mixed in given ratios to a total of  $0.5 \times 10^6$  cells per sample and total volume was adjusted to 15 µl with 150 mM ammonium bicarbonate. Each cell suspension was mixed with 7.5 µl of freshly prepared sinapinic acid matrix solution. Two microliters of sample/matrix mixture were immediately spotted onto the MALDI target and dried at room temperature. Each sample was spotted in five technical replicates.

## Mass spectrometry

Mass spectra were recorded on an AXIMA CFR mass spectrometer from Kratos Analytical (Manchester, UK) in linear positive ion mode. The instrument was equipped with a nitrogen laser (337 nm) and delayed extraction was used. The laser energy was expressed in arbitrary units from 0 to 180 a.u. The power of the laser at 180 a.u. was 6 mW, while the irradiated spot size was approximately 150 µm in diameter. External mass calibration was done using the Pep-Mix4 standard. The laser repetition rate was 5 Hz with a pulse time width of 3 ns. Each mass spectrum was obtained by the accumulation of at least 5000 shots. In order to decrease the contribution of chemical noise and possible errors on the baseline and in calibration, the raw mass spectra were pre-processed, cleaned, transformed, and reduced in dimensionality before the data analysis, as described elsewhere [38].

Mass spectra were analyzed using Launchpad Software (Kompact version 2.9.3, 2011) from Kratos Analytical Ltd. Pre-processing of mass spectra and ANN computation were performed using MATLAB 8.6. 2015 from The MathWorks Inc. (Natick, Massachusetts, USA) and Trajan Neural Network Simulator, Release 3.0 D 1996–1998, from Trajan Software Ltd. (Durham, U.

K.). Partial least squares projection to latent structures regression was performed with “leave-one-out” cross-validated prediction in program R ([www.r-project.org](http://www.r-project.org)) using the external *pls* library [39].

### Artificial neural networks

We constructed an artificial neural network containing four neurons in one hidden layer. The intensities of processed mass spectra served as the input, while the number of contaminating cells in the two-component mixtures was the output. The “learning” of the ANN was performed using the back-propagation training algorithm as described elsewhere [40,41]. The back-propagation was achieved by iteratively adjusting the values of connection weights in order to minimize the difference between the ANN calculated output value ( $o^*_{pk}$ ) and the experimental one ( $o_{pk}$ ). After each iteration, the root mean square of the sum of  $(o_{pk} - o^*_{pk})^2$  residuals (RMS) was calculated according to Eq 1:

$$\text{RMS} = \sqrt{\frac{\sum_{p=1}^N \sum_{k=1}^M (o_{pk} - o^*_{pk})^2}{N \times M}} \quad (1)$$

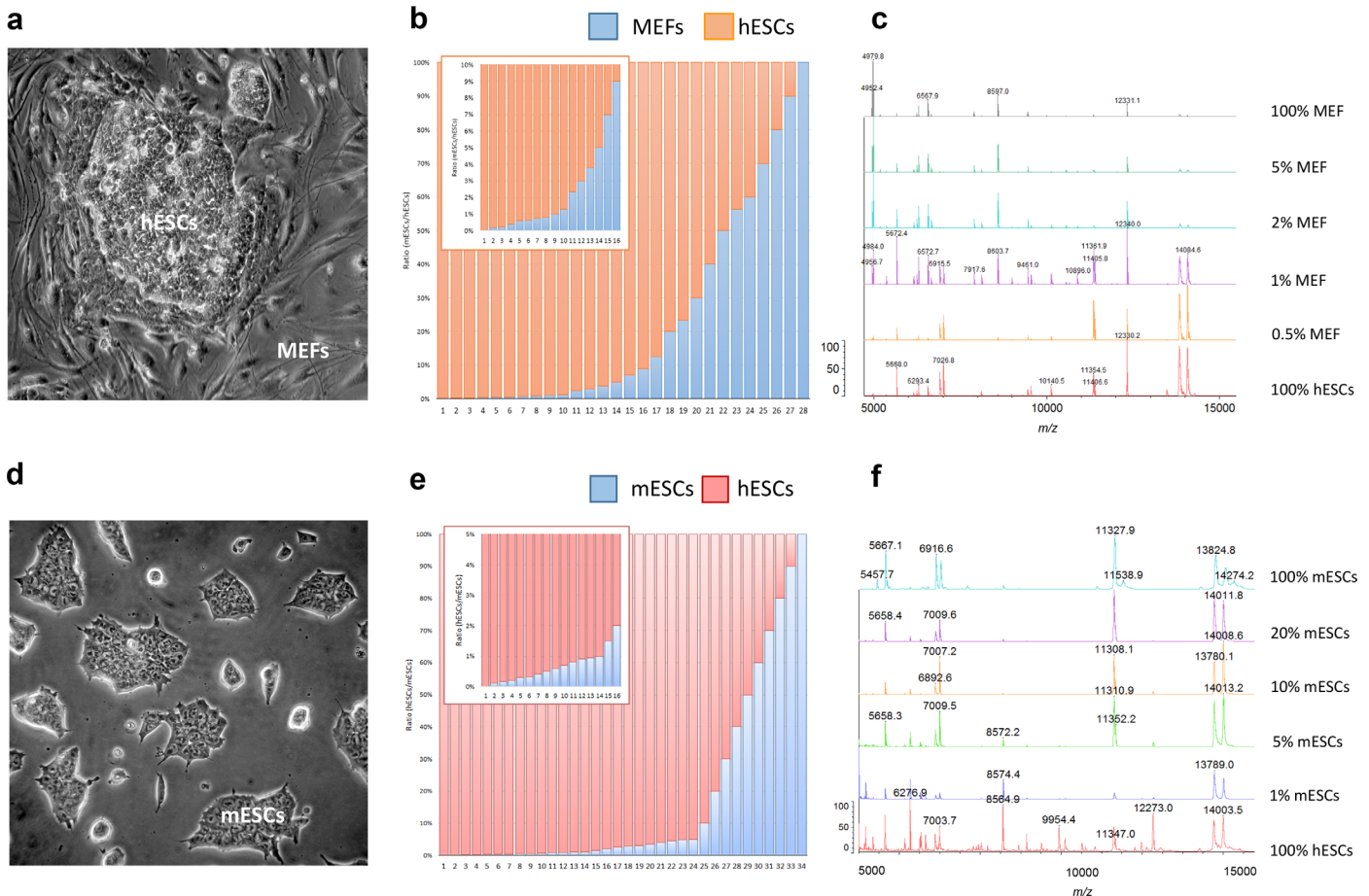
where  $N$  is the number of mass spectra,  $M$  is the number of outputs,  $o^*_{pk}$  is the ANN calculated, and  $o_{pk}$  is the experimental output value. The optimal ANN architecture was confirmed by plotting the RMS value against the number of neurons in the hidden layer(s) and number of training cycles (epochs).

## Results And Discussion

### Preparation of two-component mixtures of different cell types and intact cell MALDI-TOF mass spectrometry

We prepared calibration datasets consisting of twenty-eight defined two-component mixtures of hESCs + MEFs, thirty-four mixtures of hESCs + mESCs (Fig 1A and 1D), and pure cell populations, with total cell numbers of  $1 \times 10^6$  (Fig 1B and 1E). We then recorded the mass spectra of two-component cell mixtures and pure cell populations in the 2000–20000  $m/z$  range without previous fractionation or extraction. We pre-processed the mass spectra by (i) resampling to 30000  $m/z$  values (homogenizing in a chosen range and reducing the number of  $m/z$  values), (ii) aligning (removing the systematic shifts in mass spectra of repeated experiments), (iii) baseline subtraction, (iv) smoothing, and (v) normalization to a vector of unit length ( $\sum X_i = 1$ ), where  $X_i$  are the intensities of the peaks of the mass spectrum [42] (Fig 1C and 1F, S1 Fig). Next, we organized the spectral data into a matrix with dimensions  $m \times n$ , where  $m$  represents the number of the mass spectrum of the particular cell mixture and  $n$  are the  $m/z$  values. The  $i$ -th row of the matrix represents the mass spectrometric fingerprint of the  $i$ -th mixture. In order to decrease the latent noise in pre-processed mass spectra, we selected only the peaks with intensity higher than an arbitrarily set threshold ( $1 \times 10^{-3}$ ) for further analysis. Thus, the data matrices of hESCs + MEFs and hESCs + mESCs were reduced from the original  $n \times 30000$  to  $n \times 84$  and  $n \times 122$ , respectively. Next, we identified peaks with the highest intrinsic variability in the datasets, as described elsewhere [43–46]. In brief, we selected informative peaks by comparative determination of standard deviations of individual peaks normalized to the total variance of the dataset, Lasso regression, and sparse partial least squares regression. The informative peaks are visualized in S3 Fig. That allowed us to finally reduce the data matrices of hESCs + MEFs and hESCs + mESCs to  $n \times 10$  and  $n \times 30$ , respectively (S1 Table).

Then, we used these datasets to perform principal component analysis aiming to discriminate the pure hESC and MEF cell populations and the cell mixtures containing 50% of each cell

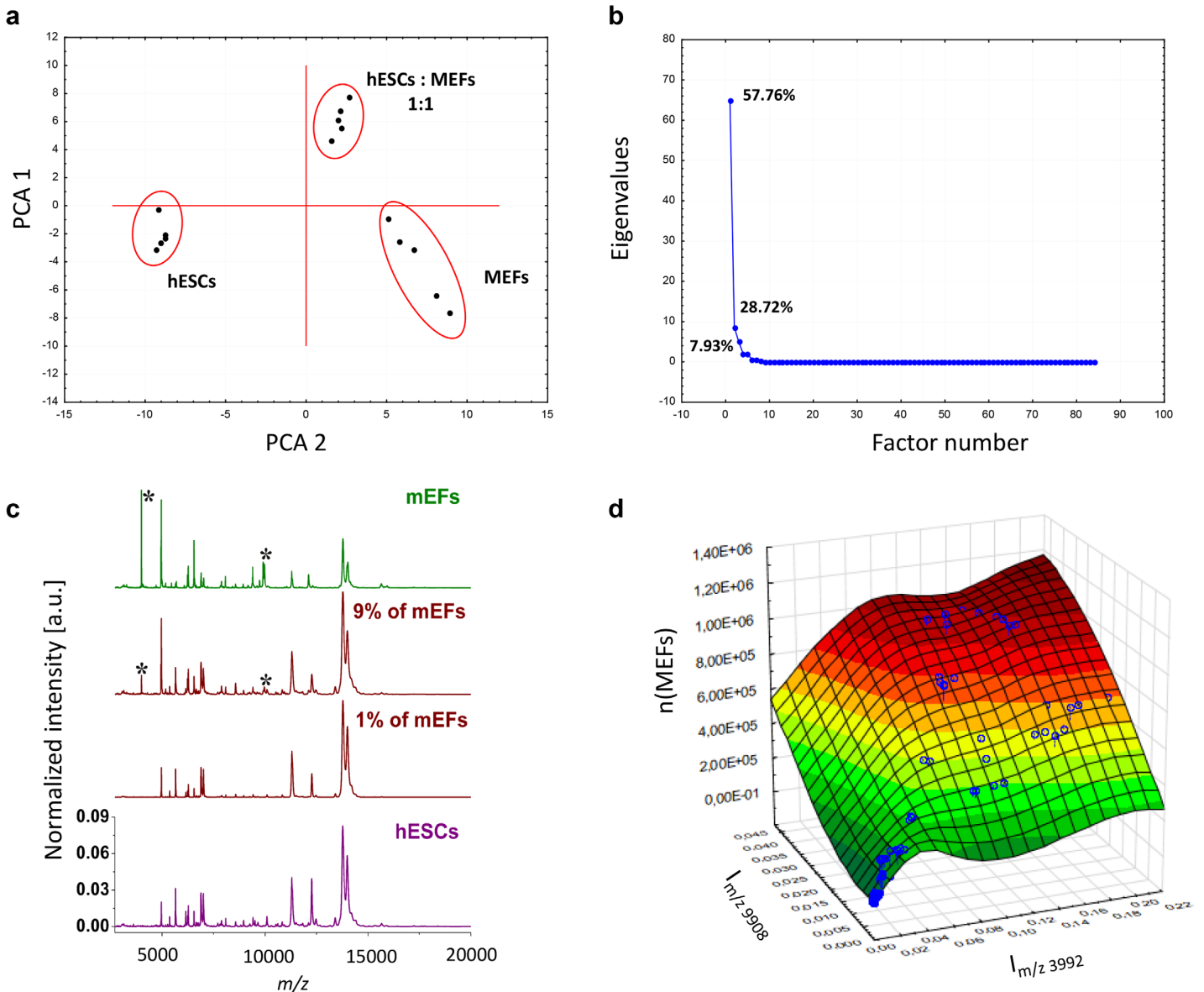


**Fig 1.** (A) Colony of human embryonic stem cells (hESCs) cultured on a feeder layer of mouse embryonic fibroblasts. (B) Experimental ratios of two-component mixtures of hESCs + MEFs. (C) Representative mass spectra of selected two-component mixtures of hESCs + MEFs. (D) Colony of mouse embryonic stem cells (mESCs) in a feeder-free culture. (E) Experimental ratios of two-component mixtures of hESCs + mESCs. (F) Representative mass spectra of selected two-component mixtures of hESCs + mESCs.

doi:10.1371/journal.pone.0147414.g001

type. Eigenvalue analysis showed the presence of three factors contributing up to 94% of the overall variability. Plotting the principal components revealed three clearly separated clusters and provided proof of principle for the discriminative information of the MEF and hESC mass spectra (Fig 2A and 2B). Similar discrimination was achieved for mESCs and hESCs (data not shown).

Next, we visually compared mass spectra obtained from the pure hESCs and MEFs, and their two-component mixtures. Despite the high similarity of the mass spectra, we identified peaks unique to hESCs and MEFs ( $m/z$  3992 and 9908) (Fig 2C) appearing reproducibly over various mixtures. We presumed that if these two marker peaks are informative for MEFs, their intensities should be proportional to the content of MEFs in the two-component mixtures. However, we did not identify any linear trend between the normalized intensities of these two marker peaks and the percentage of MEFs, especially in two-component mixtures with low concentrations of MEFs (Fig 2D). In the case of highly similar pluripotent cell types, the mass spectra of hESCs + mESCs mixtures lacked any spectral patterns specific for individual pure cell lines (Fig 1F). Therefore, assessment of the individual biomarker peaks was not suitable for precise and unambiguous quantification.



**Fig 2.** (A) Principal component analysis of mass spectra dataset containing intensities of 84  $m/z$  for pure MEF and hESC populations and their 1:1 mixture. (B) Scree plot documenting the presence of three factors contributing predominantly to the overall variability in the analyzed dataset. (C) Pre-processed MALDI-TOF mass spectra for pure hESCs and MEFs and a hESCs + MEFs two-component mixture containing 99% hESCs and 1% MEFs. The spectra were normalized to vector of unit length (a.u.). Asterisks indicate peaks at  $m/z$  3992 and 9908. (D) Surface plot of intensities of peaks at  $m/z$  3992 and 9908 versus the number of MEFs in hESCs + MEFs two-component mixtures.

doi:10.1371/journal.pone.0147414.g002

## Quantitative determination of contamination levels

Because of the data complexity, it was difficult to handle the mass spectral datasets by simple linear analyses. We first examined the data by a method of partial least squares with projection to latent structures regression (PLS) on the full data matrix of the complete mass spectra. PLS has been developed and extended by Herman and Svante Wold, respectively, [47] for quantitative analysis of highly complex multivariate data and is used preferentially in chemometrics. Despite a correlation between predicted and actual cell percentages in the two-component mixtures, the prediction precision by PLS was rather low, with substantial root mean square error

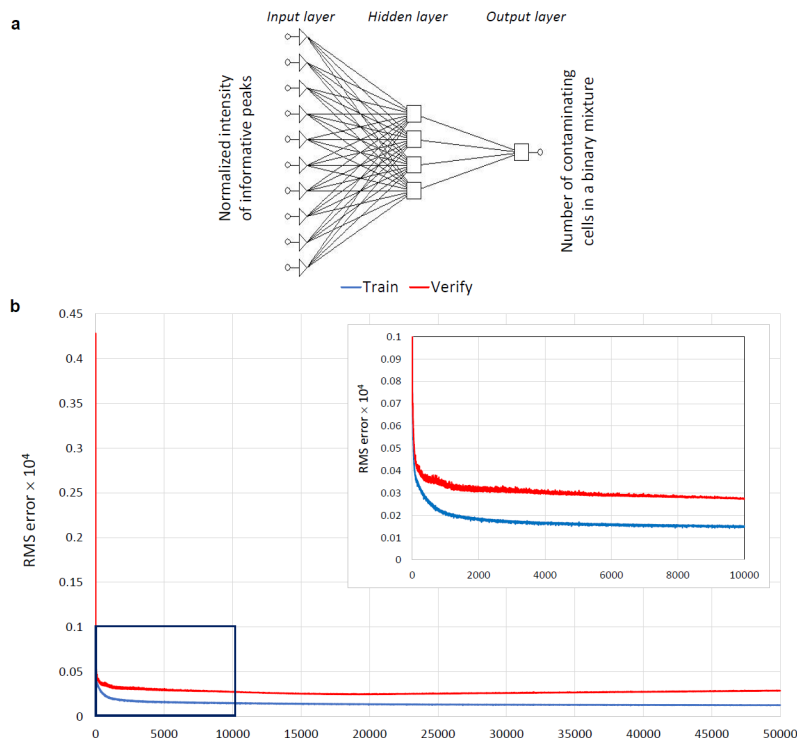
**Table 1.** Values of RMS calculated as differences between predicted and observed values,  $k$  (regression coefficients), and  $R^2$  (determination coefficient).

	hESCs: MEFs		hESCs: mESCs	
	PLS	ANN	PLS	ANN
<b>RMS</b>	$51.7 \times 10^3$	$3.16 \times 10^3$	$96.6 \times 10^3$	$7.1 \times 10^3$
<b>k</b>	0.976	0.996	0.601	0.975
<b>R<sup>2</sup></b>	0.9759	0.9992	0.6010	0.9822

doi:10.1371/journal.pone.0147414.t001

(RMS) showing signs of systematic trends (Table 1, S3 Fig). Therefore, we asked whether non-linear approaches and artificial intelligence methods, such as ANNs, could make predictions with more precision. ANNs were previously reported to provide effective analysis and classification of biological, clinical or bioanalytical, and chemometric non-linear data (for review see [30]), and were found particularly suitable for analysis of MS data [28].

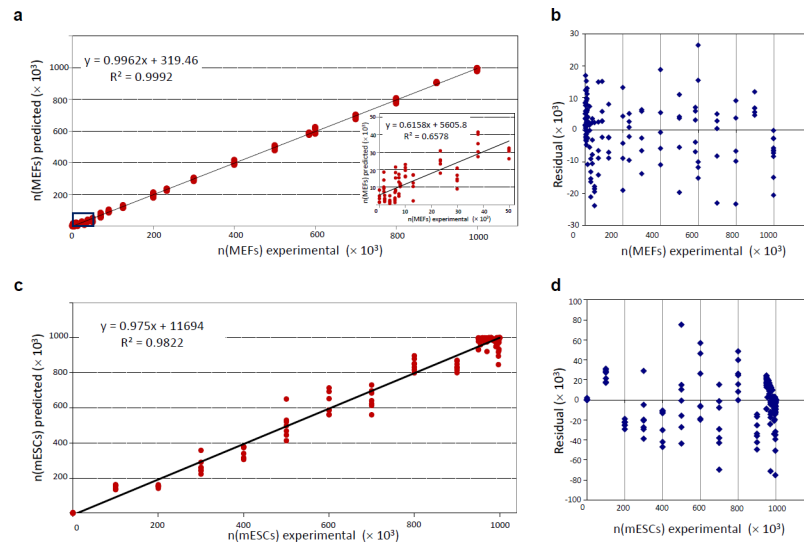
In our analysis the intensities of the selected peaks comprised the ANN-input data and the number of MEFs and/or mESCs cells in the hESC calibration mixtures was the ANN-output data. For the training step of the ANN, we tested several algorithms and found the back-propagation algorithm to be the most suitable (data not shown). We determined the optimal architecture containing four neurons in one hidden layer (Fig 3A) by plotting the RMS against the number of nodes (data not shown) and we validated over 100000 training cycles (epochs), without overfitting the model (Fig 3B). We used the leave-one-out cross-validation method to test the “generalization” ability of the designed network to predict the single cases excluded



**Fig 3.** (A) Optimal ANN architecture (one Input layer, one Hidden layer with four neurons, and one Output layer). (B) Training and leave-one-out verification plot of the RMS versus the number of training cycles (epochs). First 50 000 iterations are shown. The inset shows a detailed plot for the first 10 000 training cycles.

doi:10.1371/journal.pone.0147414.g003





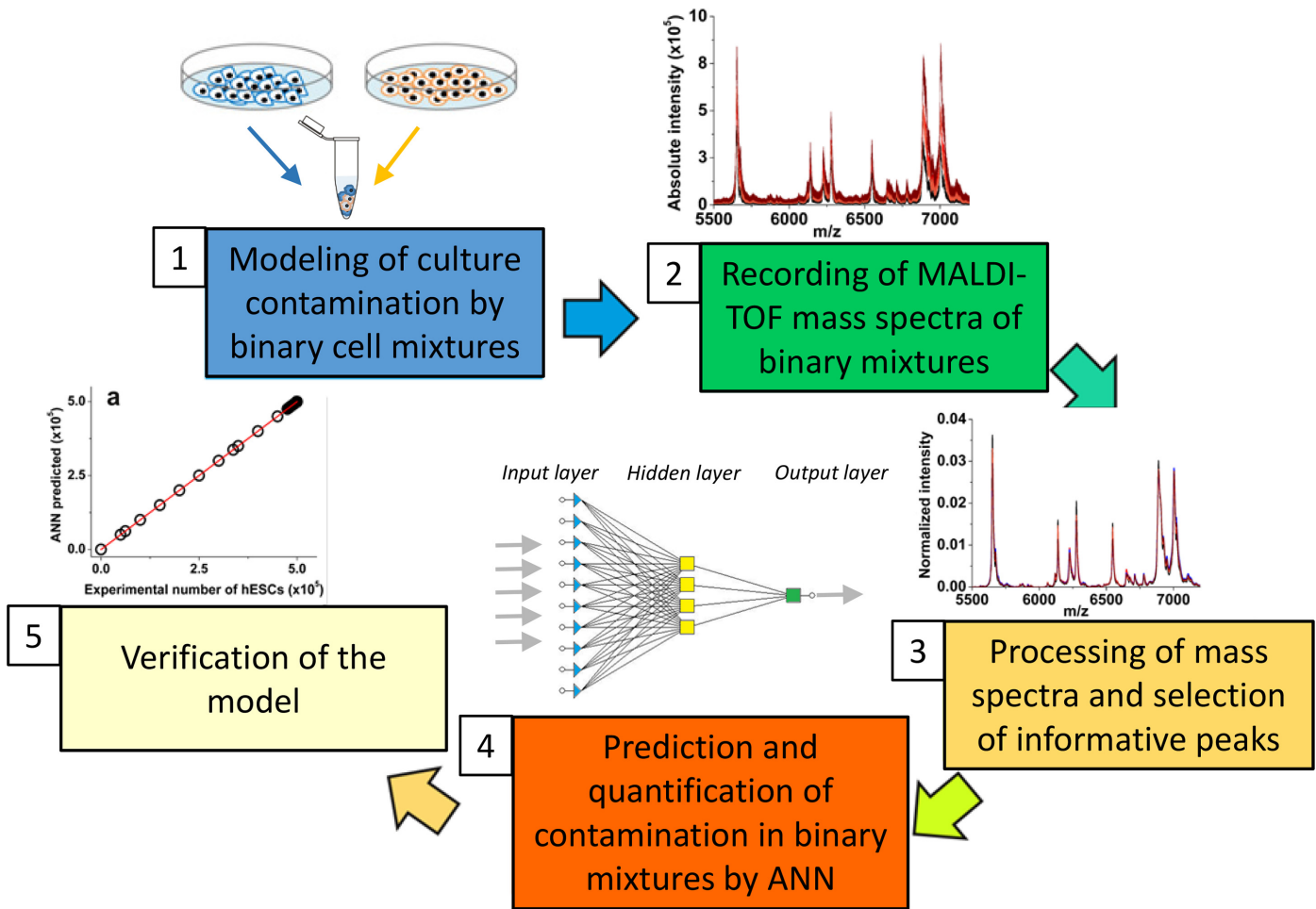
**Fig 4.** (A) Correlation between ANN-predicted number of cells and the experimental number of MEFs in two-component mixtures of hESCs + MEFs. The inset shows the correlation between experimental and predicted values in low concentration ranges of MEFs up to the  $50 \times 10^3$  cells in the two-component mixtures. (B) Overview of Residuals (difference between ANN-predicted number of cells and the experimental values) versus the experimental number of MEF cells in two-component mixtures of hESCs + MEFs. (C) Correlation between ANN-predicted number of cells and the experimental number of mESCs in two-component mixtures of hESCs + mESCs. (D) Overview of Residuals (difference between ANN-predicted number of cells and the experimental values) versus the experimental number of mESC cells in the two-component mixtures of hESCs + mESCs.

doi:10.1371/journal.pone.0147414.g004

from the training data set. We used the RMS as a measure of the prediction accuracy. The network was able to evaluate the input data and correctly predict the number of MEFs (Fig 4A) and mESCs (Fig 4C) in the hESC suspensions over the whole range of evaluated ratios. Prediction by ANN was correct even at low percentages of contaminating cells in suspension. To validate the model, the RMS was calculated as the differences between the predicted and experimental values. The residuals reached significantly lower values than in PLS predictions (Table 1). Moreover, the residual values were randomly distributed, and the absence of any systematic error or trend in residuals demonstrated correctness of the model (Fig 4B and 4D). To perform further validation of the method, we analyzed an independent dataset of fifty hESCs + MEFs mixtures. Using a training set described above, the ANN correctly determined the numbers of MEFs in hESCs suspension with high correlation between predicted and experimental values (S4 Fig). In summary, multivariate calibration coupled with a correctly trained ANN was able to determine the ratio of cell numbers in two-component mixtures.

We have identified the conditions and developed a step-by-step protocol for successful quantitation of two distinct cell types in a single two-component mixture by a multivariate calibration approach based on an ANN-coupled IC MALDI-TOF MS analysis. The major steps of the method include:

1. *a priori* knowledge or identification of the contaminating cell line
2. construction of two-component calibration mixtures of the given cell lines
3. mass spectra pre-processing and *m/z* selection
4. parallel recording of mass spectra of pure cell populations and calibration mixtures and building a library of spectral datasets for multivariate calibration



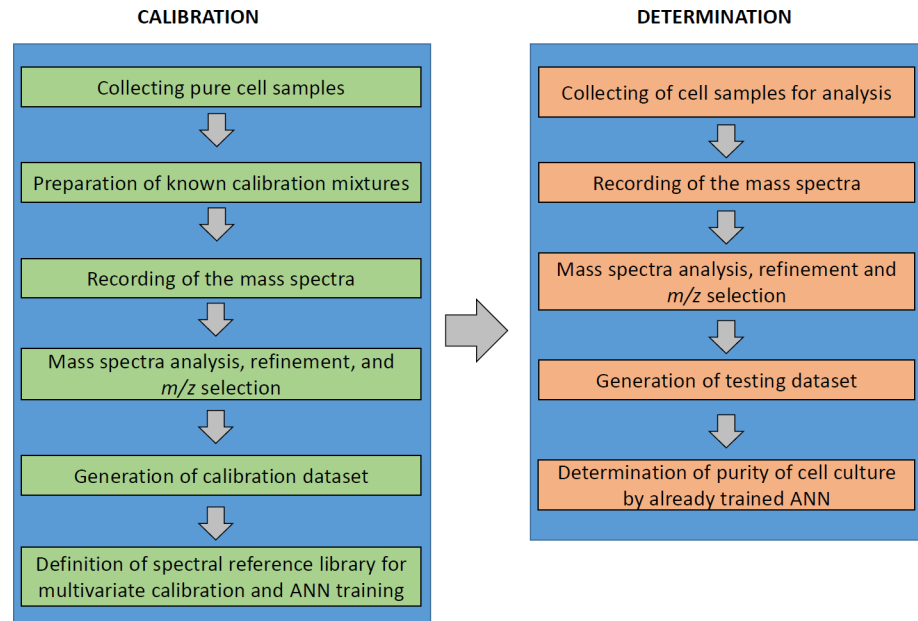
**Fig 5. Overview for quantitative ANN-coupled MS-based analysis of cross-contamination.**

doi:10.1371/journal.pone.0147414.g005

- estimation of the contamination levels in unknown samples using an ANN model trained on the calibration datasets

The application of ANNs allowed us to overcome the unwanted inconsistency and non-linearity of IC MALDI-TOF MS spectra and reveal hidden patterns in mass spectra to unambiguously identify and quantify MEFs or mESCs in the hESC culture. However, *a priori* knowledge of the contaminating cell line is a prerequisite for correct prediction and selection of the training dataset. The multivariate calibration-based ANN approach can be easily adapted to routine protocols for quantitative determination of cell culture homogeneity and consistency and for thorough MS analyses of cell parameters in various culture platforms, with all steps adaptable for any experimental, routine, or high-throughput culture setup [48,49] (Figs 5 and 6). Currently, methods involving assessment of cell quality in clinical grade cultures, biomedical research or bio-industry involve either genetic authentication confirming the cell identity or functional assays documenting the phenotype. The intact cell mass spectrometry coupled with ANN can reveal inconsistencies occurring in high-throughput or long-term cultures or co-cultures, by monitoring spectral patterns and their alterations.

In conclusion, we demonstrated for the first time that the multivariate calibration approach based on ANN-coupled IC MALDI-TOF MS analysis can provide quantitative information on



**Fig 6. Experimental schematic of the multivariate calibration-based ANN spectral analysis.**

doi:10.1371/journal.pone.0147414.g006

cell culture heterogeneity and authenticity and thus complement the portfolio of techniques that are available for characterization of mammalian cell cultures.

### Supporting Information

**S1 Fig. Comparison of raw (a) and pre-processed (b) mass spectra (five replicates) characterizing human embryonic stem cells (hESCs).** The pre-processed mass spectra were normalized to a vector of unit length ( $\sum X_i = 1$ , where  $X_i$  are the intensities of the peaks of the mass spectrum).

(TIF)

**S2 Fig. Standard deviations of the mean intensity of peaks of the particular m/z in the dataset normalized to the standard deviation of the mean intensity of the hESCs + MEFs (a) and hESCs + mESCs dataset (b).** Peaks used in the further analyses are indicated by respective m/z values.

(TIF)

**S3 Fig. Correlation between PLS-predicted number of cells and the experimental number of MEFs in two-component mixtures of (a) hESCs + MEFs and (b) hESCs + mESCs. Overview of Residuals (difference between PLS-predicted number of cells and the experimental values) versus the experimental number of cells in the two-component mixtures of (c) hESCs + MEFs and (d) hESCs + mESCs.**

(TIF)

**S4 Fig. Training and verification plot for an independent dataset of 50 cases of hESCs + MEFs mixtures (a).** Correlation between experimental and ANN-predicted numbers of MEFs in hESCs suspension.

(TIF)

**S1 Table. Overview of normalized mass-to-charge ratios (m/z) of informative peaks used for analysis.**

(TIF)

**Acknowledgments**

We thank Dr. Jiří Pacherník from the Department of Experimental Biology, Faculty of Science, Masaryk University for the kind provision of mouse embryonic stem cell line R1.

**Author Contributions**

Conceived and designed the experiments: TP JH AH PV. Performed the experiments: EV LK FA. Analyzed the data: LK LP JH PV. Contributed reagents/materials/analysis tools: TP JH AH PV. Wrote the paper: AH JH PV.

**References**

1. Capes-Davis A, Theodosopoulos G, Atkin I, Drexler HG, Kohara A, MacLeod RAF, et al. (2010) Check your cultures! A list of cross-contaminated or misidentified cell lines. *International Journal of Cancer* 127: 1–8.
2. Marx V (2014) Cell-line authentication demystified. *Nature Methods* 11: 483–+.
3. Masters JRW (2010) Cell line misidentification: the beginning of the end. *Nature Reviews Cancer* 10: 441–448. doi: [10.1038/nrc2852](https://doi.org/10.1038/nrc2852) PMID: [20448633](https://pubmed.ncbi.nlm.nih.gov/20448633/)
4. Nardone RM (2007) Eradication of cross-contaminated cell lines: A call for action. *Cell Biology and Toxicology* 23: 367–372. PMID: [17522957](https://pubmed.ncbi.nlm.nih.gov/17522957/)
5. Hynds RE, Giangreco A (2013) Concise Review: The Relevance of Human Stem Cell-Derived Organoid Models for Epithelial Translational Medicine. *Stem Cells* 31: 417–422. doi: [10.1002/stem.1290](https://doi.org/10.1002/stem.1290) PMID: [23203919](https://pubmed.ncbi.nlm.nih.gov/23203919/)
6. Mehling M, Tay S (2014) Microfluidic cell culture. *Current Opinion in Biotechnology* 25: 95–102. doi: [10.1016/j.copbio.2013.10.005](https://doi.org/10.1016/j.copbio.2013.10.005) PMID: [24484886](https://pubmed.ncbi.nlm.nih.gov/24484886/)
7. Nienow AW (2006) Reactor engineering in large scale animal cell culture. *Cytotechnology* 50: 9–33. doi: [10.1007/s10616-006-9005-8](https://doi.org/10.1007/s10616-006-9005-8) PMID: [19003068](https://pubmed.ncbi.nlm.nih.gov/19003068/)
8. Baradez MO, Lekishvili T, Marshall D (2015) Rapid phenotypic fingerprinting of cell products by robust measurement of ubiquitous surface markers. *Cytometry Part A* 87A: 624–635.
9. Didion JP, Buus RJ, Naghashfar Z, Threadgill DW, Morse HC 3rd, de Villena FP (2014) SNP array profiling of mouse cell lines identifies their strains of origin and reveals cross-contamination and widespread aneuploidy. *BMC Genomics* 15: 847. doi: [10.1186/1471-2164-15-847](https://doi.org/10.1186/1471-2164-15-847) PMID: [25277546](https://pubmed.ncbi.nlm.nih.gov/25277546/)
10. Brougham DF, Ivanova G, Gottschalk M, Collins DM, Eustace AJ, O'Connor R, et al. (2011) Artificial neural networks for classification in metabolomic studies of whole cells using 1H nuclear magnetic resonance. *J Biomed Biotechnol* 2011: 158094. doi: [10.1155/2011/158094](https://doi.org/10.1155/2011/158094) PMID: [20886062](https://pubmed.ncbi.nlm.nih.gov/20886062/)
11. Houska J, Pena-Mendez EM, Hernandez-Fernaund JR, Salido E, Hampl A, Havel J, et al. (2014) Tissue profiling by nanogold-mediated mass spectrometry and artificial neural networks in the mouse model of human primary hyperoxaluria 1. *Journal of Applied Biomedicine* 12: 119–125.
12. Lui H, Zhao JH, McLean D, Zeng HS (2012) Real-time Raman Spectroscopy for In Vivo Skin Cancer Diagnosis. *Cancer Research* 72: 2491–2500. doi: [10.1158/0008-5472.CAN-11-4061](https://doi.org/10.1158/0008-5472.CAN-11-4061) PMID: [22434431](https://pubmed.ncbi.nlm.nih.gov/22434431/)
13. Munteanu B, von Reitzenstein C, Hansch GM, Meyer B, Hopf C (2012) Sensitive, robust and automated protein analysis of cell differentiation and of primary human blood cells by intact cell MALDI mass spectrometry biotyping. *Analytical and Bioanalytical Chemistry* 404: 2277–2286. doi: [10.1007/s00216-012-6357-0](https://doi.org/10.1007/s00216-012-6357-0) PMID: [22955673](https://pubmed.ncbi.nlm.nih.gov/22955673/)
14. Buchanan CM, Malik AS, Cooper GJS (2007) Direct visualisation of peptide hormones in cultured pancreatic islet alpha- and beta-cells by intact-cell mass spectrometry. *Rapid Communications in Mass Spectrometry* 21: 3452–3458. PMID: [17918213](https://pubmed.ncbi.nlm.nih.gov/17918213/)
15. Lokhov P, Balashova E, Dashtiev M (2009) Cell proteomic footprint. *Rapid Communications in Mass Spectrometry* 23: 680–682. doi: [10.1002/rcm.3928](https://doi.org/10.1002/rcm.3928) PMID: [19184978](https://pubmed.ncbi.nlm.nih.gov/19184978/)
16. Maurer K, Eschrich K, Schellenberger W, Bertolini J, Rupf S, Remmerbach TW (2013) Oral brush biopsy analysis by MALDI-ToF Mass Spectrometry for early cancer diagnosis. *Oral Oncology* 49: 152–156. doi: [10.1016/j.oraloncology.2012.08.012](https://doi.org/10.1016/j.oraloncology.2012.08.012) PMID: [23000400](https://pubmed.ncbi.nlm.nih.gov/23000400/)

17. Munteanu B, Hopf C (2013) Emergence of whole-cell MALDI-MS biotyping for high-throughput bioanalysis of mammalian cells? *Bioanalysis* 5: 885–893. doi: [10.4155/bio.13.47](https://doi.org/10.4155/bio.13.47) PMID: [23590472](https://pubmed.ncbi.nlm.nih.gov/23590472/)
18. Zhang X, Scalf M, Berggren TW, Westphall MS, Smith LM (2006) Identification of mammalian cell lines using MALDI-TOF and LC-ESI-MS/MS mass spectrometry. *Journal of the American Society for Mass Spectrometry* 17: 490–499. PMID: [16488154](https://pubmed.ncbi.nlm.nih.gov/16488154/)
19. Dong HJ, Shen W, Cheung MTW, Liang YM, Cheung HY, Allmaier G, et al. (2011) Rapid detection of apoptosis in mammalian cells by using intact cell MALDI mass spectrometry. *Analyst* 136: 5181–5189. doi: [10.1039/c1an15750g](https://doi.org/10.1039/c1an15750g) PMID: [22059228](https://pubmed.ncbi.nlm.nih.gov/22059228/)
20. Hanrieder J, Wicher G, Bergquist J, Andersson M, Fex-Svenningsen A (2011) MALDI mass spectrometry based molecular phenotyping of CNS glial cells for prediction in mammalian brain tissue. *Analytical and Bioanalytical Chemistry* 401: 135–147. doi: [10.1007/s00216-011-5043-y](https://doi.org/10.1007/s00216-011-5043-y) PMID: [21553124](https://pubmed.ncbi.nlm.nih.gov/21553124/)
21. Munteanu B, Meyer B, von Reitzenstein C, Burgermeister E, Bog S, Pahl A, et al. (2014) Label-Free in Situ Monitoring of Histone Deacetylase Drug Target Engagement by Matrix-Assisted Laser Desorption Ionization-Mass Spectrometry Biotyping and Imaging. *Analytical Chemistry* 86: 4642–4647. doi: [10.1021/ac500038j](https://doi.org/10.1021/ac500038j) PMID: [24559101](https://pubmed.ncbi.nlm.nih.gov/24559101/)
22. Povey JF, O'Malley CJ, Root T, Martin EB, Montague GA, Feary M, et al. (2014) Rapid high-throughput characterisation, classification and selection of recombinant mammalian cell line phenotypes using intact cell MALDI-ToF mass spectrometry fingerprinting and PLS-DA modelling. *Journal of Biotechnology* 184: 84–93. doi: [10.1016/j.jbiotec.2014.04.028](https://doi.org/10.1016/j.jbiotec.2014.04.028) PMID: [24858576](https://pubmed.ncbi.nlm.nih.gov/24858576/)
23. Volta P, Riccardi N, Lauceri R, Tonolla M (2012) Discrimination of freshwater fish species by Matrix-Assisted Laser Desorption/Ionization-Time Of Flight Mass Spectrometry (MALDI-TOF MS): a pilot study. *Journal of Limnology* 71: 164–169.
24. Chiu NH, Jia Z, Diaz R, Wright P (2015) Rapid differentiation of in vitro cellular responses to toxic chemicals by using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Environ Toxicol Chem* 34: 161–166. doi: [10.1002/etc.2774](https://doi.org/10.1002/etc.2774) PMID: [25319019](https://pubmed.ncbi.nlm.nih.gov/25319019/)
25. Kober SL, Meyer-Alert H, Grienitz D, Hollert H, Frohme M (2015) Intact cell mass spectrometry as a rapid and specific tool for the differentiation of toxic effects in cell-based ecotoxicological test systems. *Anal Bioanal Chem*.
26. Asakawa D, Sakakura M, Takayama M (2012) Matrix effect on in-source decay products of peptides in matrix-assisted laser desorption/ionization. *Mass Spectrom (Tokyo)* 1: A0002.
27. Bas D, Boyaci IH (2007) Modeling and optimization I: Usability of response surface methodology. *Journal of Food Engineering* 78: 836–845.
28. Goodacre R, Neal MJ, Kell DB (1996) Quantitative analysis of multivariate data using artificial neural networks: A tutorial review and applications to the deconvolution of pyrolysis mass spectra. *Zentralblatt Fur Bakteriologie-International Journal of Medical Microbiology Virology Parasitology and Infectious Diseases* 284: 516–539.
29. Li H, Zhang YX, Polaskova P, Havel J (2002) Enhancement of precision in the analysis of medicines by capillary electrophoresis using artificial neural networks. *Acta Chimica Sinica* 60: 1264–1268.
30. Amato F, Lopez A, Pena-Mendez EM, Vanhara P, Hampl A, Havel J (2013) Artificial neural networks in medical diagnosis. *Journal of Applied Biomedicine* 11: 47–58.
31. Basheer IA, Hajmeer M (2000) Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods* 43: 3–31. PMID: [11084225](https://pubmed.ncbi.nlm.nih.gov/11084225/)
32. International Stem Cell Initiative, Adewumi O, Aflatoonian B, Ahrlund-Richter L, Amit M, Andrews PW, et al. (2007) Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat Biotechnol* 25: 803–816. PMID: [17572666](https://pubmed.ncbi.nlm.nih.gov/17572666/)
33. Kotasova H, Vesela I, Kucera J, Houdek Z, Prochazkova J, Kralickova M, et al. (2012) Phosphoinositide 3-kinase inhibition enables retinoic acid-induced neurogenesis in monolayer culture of embryonic stem cells. *Journal of Cellular Biochemistry* 113: 563–570. doi: [10.1002/jcb.23380](https://doi.org/10.1002/jcb.23380) PMID: [21948563](https://pubmed.ncbi.nlm.nih.gov/21948563/)
34. Barta T, Vinarsky V, Holubcova Z, Dolezalova D, Verner J, Pospisilova S, et al. (2010) Human embryonic stem cells are capable of executing G1/S checkpoint activation. *Stem Cells* 28: 1143–1152. doi: [10.1002/stem.451](https://doi.org/10.1002/stem.451) PMID: [20518019](https://pubmed.ncbi.nlm.nih.gov/20518019/)
35. Holubcova Z, Matula P, Sedlackova M, Vinarsky V, Dolezalova D, Barta T, et al. (2011) Human embryonic stem cells suffer from centrosomal amplification. *Stem Cells* 29: 46–56. doi: [10.1002/stem.549](https://doi.org/10.1002/stem.549) PMID: [20960514](https://pubmed.ncbi.nlm.nih.gov/20960514/)
36. Adewumi O, Aflatoonian B, Ahrlund-Richter L, Amit M, Andrews PW, Beighton G, et al. (2007) Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nature Biotechnology* 25: 803–816. PMID: [17572666](https://pubmed.ncbi.nlm.nih.gov/17572666/)

37. Amps K, Andrews PW, Anyfantis G, Armstrong L, Avery S, Baharvand H, et al. (2011) Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. *Nature Biotechnology* 29: 1132–U1113. doi: [10.1038/nbt.2051](https://doi.org/10.1038/nbt.2051) PMID: [22119741](https://pubmed.ncbi.nlm.nih.gov/22119741/)
38. Hilario M, Kalousis A, Pellegrini C, Muller M (2006) Processing and classification of protein mass spectra. *Mass Spectrometry Reviews* 25: 409–449. PMID: [16463283](https://pubmed.ncbi.nlm.nih.gov/16463283/)
39. Mevik B-H, Wehrens R (2007) The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software*, 18: 1–24.
40. Amato F, Gonzalez-Hernandez JL, Havel J (2012) Artificial neural networks combined with experimental design: a "soft" approach for chemical kinetics. *Talanta* 93: 72–78. doi: [10.1016/j.talanta.2012.01.044](https://doi.org/10.1016/j.talanta.2012.01.044) PMID: [22483879](https://pubmed.ncbi.nlm.nih.gov/22483879/)
41. Pivetta T, Isaia F, Trudu F, Pani A, Manca M, Perra D, et al. (2013) Development and validation of a general approach to predict and quantify the synergism of anti-cancer drugs using experimental design and artificial neural networks. *Talanta* 115: 84–93. doi: [10.1016/j.talanta.2013.04.031](https://doi.org/10.1016/j.talanta.2013.04.031) PMID: [24054565](https://pubmed.ncbi.nlm.nih.gov/24054565/)
42. Alfassi ZB (2004) On the normalization of a mass spectrum for comparison of two spectra. *Journal of the American Society for Mass Spectrometry* 15: 385–387. PMID: [14998540](https://pubmed.ncbi.nlm.nih.gov/14998540/)
43. Filzmoser P, Gschwandtner M, Todorov V (2012) Review of sparse methods in regression and classification with application to chemometrics. *Journal of Chemometrics* 26: 42–51.
44. Friedman J, Hastie T, Tibshirani R (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33: 1–22. PMID: [20808728](https://pubmed.ncbi.nlm.nih.gov/20808728/)
45. Chun H, Keles S (2010) Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 72: 3–25.
46. Rasmussen MA, Bro R (2012) A tutorial on the Lasso approach to sparse modeling. *Chemometrics and Intelligent Laboratory Systems* 119: 21–31.
47. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58: 109–130.
48. Dittrich P, Ibanez AJ (2015) Analysis of metabolites in single cells-what is the best micro-platform? *Electrophoresis*.
49. Xie W, Gao D, Jin F, Jiang Y, Liu H (2015) Study of Phospholipids in Single Cells Using an Integrated Microfluidic Device Combined with Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal Chem* 87: 7052–7059. doi: [10.1021/acs.analchem.5b00010](https://doi.org/10.1021/acs.analchem.5b00010) PMID: [26110742](https://pubmed.ncbi.nlm.nih.gov/26110742/)