Fibroblast phenylalanine concentration as a surrogate biomarker of cellular number

Clara Oliva [a], Angela Arias [a] Montserrat Ruiz[b,c], Aurora Pujol[b,c,d], Gloria Garrabou[d,e,f,g], Judith Canto-Santos[d,e], Roser Urreizti[a,d,], Laura Castilla-Vallmanya[d,h], Helena Rodriguez-Gonzalez[a], Cristina Jou[d,i], Mercedes Casado[a,d], Aida Ormazabal[a,d], *Rafael Artuch[a,d]

[a]Clinical Biochemistry Department. Sant Joan de Déu Research Institute, Passeig Sant Joan de Déu, 2, 08950, Esplugues de Llobregat, Barcelona, Spain

[b]Neurometabolic Diseases Laboratory, Bellvitge Biomedical Research Institute (IDIBELL), Avinguda de la Gran via de l'Hospitalet, 199, 08908 L'Hospitalet de Llobregat, Barcelona, Spain

[c]Catalan Institution of Research and Advanced Studies (ICREA), Universitat Autònoma de Barcelona, Av. de Serragalliners, s/n, 08193 Bellaterra, Barcelona, Spain.

[d] Centro de Investigación Biomédica en Red de Enfermedades Raras, ISCIII, Avda. de Monforte de Lemos, 5, 28029 Madrid, Spain.

[e]Muscle Research and Mitochondrial Function Laboratory, CELLEX-August Pi i Sunyer Biomedical Research Institute, C/ del Roselló, 149, 08036, Barcelona, Spain.

[f]Faculty of Medicine and Health Sciences-University of Barcelona, C/ de Casanova, 143, 08036, barcelona Spain.

[g]Internal Medicine Department-Hospital Clinic of Barcelona, C/ del Roselló, 149, 08036, Barcelona, Spain.

[h]Department of Genetics, Microbiology and Statistics, Institute of Biomedicine (IBUB), Faculty of Biology, Universitat de Barcelona, Avinguda Diagonal, 643, 08028 Barcelona, Spain.

[i] Biobank and Pathology department, Hospital Sant Joan de Deu, Passeig Sant Joan de Déu, 2, 08950, Esplugues de Llobregat, Barcelona, Spain.

*****Correspondence: Dr. Rafael Artuch. e-mail; Rafael.artuch@sjd.es.

ORCID number: 0000-0002-3422-9685

Abbreviations: HDF; human dermal fibroblasts. %RSD; the Relative Standard Deviation. Phe; phenylalanine. PCA; principal component analysis.

1. Introduction

Cellular metabolomic studies in human dermal fibroblasts (HDF) are used to elucidate biological mechanisms associated with diseases. Data from cell lines are easier to control and interpret than those derived from animal models or human biological fluid studies due to decreased heterogeneity caused by environmental-genetic interactions [1]. However, several critical methodological issues have been identified in *in vitro* metabolomic studies. Differences in culture media composition, cell harvesting, and halting cell metabolism procedures may change metabolomic profiles [2,3]. Additional factors, such as cell characteristics (age, sex, race, skin biopsy location), fibroblast passage number, or the degradation of labile metabolites during metabolite extraction, can also influence metabolomic results [4–6]. Thus, it is well recognized that the metabolic status inside cells will change depending on all of the aforementioned conditions. For a better interpretation of biological variations, one of the main challenges of intracellular metabolomics studies is to reduce variability in the results due to technical factors [7–10]. Furthermore, for targeted metabolomic studies, standardization of specific preanalytical and analytical protocols has been suggested [1] as the only way to obtain robust and reliable results.

Postanalytical procedures and standardization of data are also key points in metabolomics studies [11]. Targeted metabolomic studies are intended to report concentrations in absolute values for selected metabolites. In this sense, data normalization to correct metabolite concentrations according to the available cell number is one of the cornerstones in the interpretation of in vitro metabolomic studies. Two general normalization approaches are used: sample-based and data-based [12,13]. Regarding sample-based normalization, different parameters have been used as surrogate biomarkers of cell number. These include direct cell number count, tissue weight and total protein concentration [12,13]. Cell counting would be the ideal normalization method, but performance problems such as inhomogeneity of the cell suspension, random variations introduced during aliquoting and transfer process, or the fact that this method delays the quenching procedure resulting in alterations in the metabolic profile, strongly advises to look for alternative surrogates [13]. Total protein quantification by different methods is the one most usually used [2,12,14]. These procedures have the advantages that are simple and available in most laboratories. As disadvantages, lack of adequate protein standards (albumin is usually used, although the cellular protein content is radically different), the use of low-sensitivity colorimetric methods and the use of proportionally large sample

volumes are limitations that are inherent to these procedures. There are several reports that have proposed new methods for sample-based normalization by using biomarkers such as DNA concentration, inositol, pantothenate, phospholipid signals or the use of 17 dansyl-labeled amino acid standards in different cell models, among others [12,13,15–18].

With these antecedents, the aim of the present study was to quantify amino acids by a classical analytical procedure in cultured HDFs obtained from a control population. We estimated the number of cells by analysing total protein concentration and compared it with different intracellular amino acid concentrations as new potential sample-based normalization factors.

2. Material and methods

2.1 Fibroblast samples

An overview of the fibroblast preparation and analysis workflow is depicted in supplementary figure 1. To obtain derived skin fibroblasts, 44 skin biopsy samples from paediatric and adult healthy controls were collected. For most paediatric controls, skin biopsies were taken during orthopaedic surgical interventions. Exclusion criteria included the presence of any chronic disease. Four Samples from laboratory A were purchased from Coriell Institute (Camden, USA).

Skin biopsies from different anatomical areas were done upon 6 mm of diameter punch (Table 1). Dermal fibroblasts were obtained from skin explants by physical disaggregation. Fibroblasts were harvested between passages 4 and 13, and cell pellets were kept dry at -80 C° for further analyses. Fibroblasts were grown by 4 different laboratories (A-D) using their own protocols. All of them used DMEM growth medium with 10% foetal bovine serum and 1% penicillin/streptomycin and shared harvesting (trypsinization) and confluence protocols (80-90%). The main differences of HDF culture conditions are summarized in **Table 1.**

2.2 Sample preparation and analysis

Frozen pellets from cultured HDFs were resuspended in 200 µL of phosphate-buffered saline. Fibroblasts were lysed by sonication, using cycles of 10 seconds (40% amplitude, on ice). Once complete homogenization was achieved, a 10 µL aliquot was taken for total protein analysis as an estimate of the cellular number using the Lowry method [19]. The range of total protein values was 0.90-7.55 g/L, indicating remarkable differences in

fibroblast number (supplementary table 1). The remaining sample was frozen at -80 °C until analysis, as previously reported [2,20,21].

2.3 Amino acid analysis

Samples were thawed and centrifuged for 10 min at 10 °C. Amino acid levels were measured in the clear supernatants by UPLC-MS/MS, as previously described [22]. Briefly, 10 µL of the fibroblast supernatants were mixed with 25 µL of the internal standard solution (a mix of 17 amino acids labelled by isotopes ^{13}C and ^{15}N; Cambridge Isotope Laboratories, Inc. REF MSK-A2-S) and 40 µL of methanol/0.1% formic acid to precipitate the proteins. Then, samples were centrifuged at 600 x g for 10 min at 10 °C. For the derivatization reaction, 5 µL of the supernatant was mixed with 35 µL of borate buffer and 10 µL of AQC solution (3 mg/mL in acetonitrile; AccQ·Tag Ultra Derivatization Kit, REF 186003836). The samples were analysed using a Waters ACQUITY UPLC H-class. Chromatographic separation was performed at 55 °C with a CORTECS C18 2.1 × 150, 1.6 µm column (Waters). A gradient of water (mobile phase A) and acetonitrile (mobile phase B), both containing 0.1% formic acid, was used at a flow rate of 0.5 mL/min. Once the amino acids were separated, identification and quantification were performed in a Waters Xevo TQD triple-quadrupole mass spectrometer using positive electrospray ionization in the multiple reaction monitoring mode [22]. In supplementary figure S2 there is an example of a sample's chromatogram. Amino acid quantification in biological specimens is accredited by the ISO15189 (ENAC agency) and subjected to external quality control schemes (ERNDIM program [23]), which means that data obtained from the analysis are consistent with those of other participant laboratories in the control scheme. Data are available on request.

2.4 Statistical analysis

All data were analysed using RStudio (Version 4.1.2) [24]. To search for correlations among the different individual amino acid concentrations and the total protein content, Spearman's correlation coefficient analysis with the stats R package applied was used, as the data showed a non-Gaussian distribution. Individual amino acid values were normalized by both total protein concentration (µmol amino acid/g of protein) and by all the amino acid values (µmol/µmol) to compare the variability among the different measurements. Raw and normalized data are stated in Tables S1 and S2. To find the normalization factor that reduced the variability of sample metabolite determination in HDFs, the coefficient of variation (%RSD = standard deviation/average*100) of all

samples was calculated as average/standard deviation*100. Paired sample Wilcoxon test was performed to compare differences between %RSD of the different normalization methods.

The differences among data from the 4 laboratories where HDF cultures were performed and among the different data normalization methods were assessed by exploratory data analysis using unsupervised methods (principal component analysis, PCA, and clustering method with heatmap) with the R packages FactoMineR, factoextra, vegan and cluster. We selected PCA since it is one of the easiest and most common way of evaluating the variations [13]. Distance calculation was performed in Euclidean distance, and the optimal number of clusters was determined using the silhouette method. The *Wilcoxon test* was applied as a supervised method to confirm the unsupervised results.

2.5 Ethics statement

All participants gave their informed consent for inclusion prior to participation. The study was approved by the Ethics Committee of Hospital Sant Joan de Déu, and it was conducted in accordance with the guidelines established in the WMA Declaration of Helsinki. The samples from controls were donated by the HSJD and CIBERER Biobanks, who provided written informed consent from patients and/or their parents. Projects were revised and approved by the Institutional Review Boards from Hospital Sant Joan de Déu and Universidad de Barcelona (References FIS20/00340 and IRB00003099, respectively).

3. Results

Total intracellular amino acid levels and corrected by total protein and phenylalanine (Phe) concentrations, as surrogate biomarkers of cell number, and the individual %RSD for all amino acids are shown in supplementary Tables S1 and S2. The Spearman correlation test showed that Phe displayed the second highest correlation with all other measured amino acids (only isoleucine displayed a slightly higher r value of 0.81), with a mean r value of 0.8 (range 0.41–0.99; $p<0.01$). Total protein concentration showed a mean r value of 0.67 (range 0.48–0.78; $p<0.01$). Spearman's correlation coefficient between total protein and Phe concentration was 0.72 ($p <0.001$). The detailed data of correlation coefficients among all the amino acids and total protein concentration are shown in supplementary Table S3.

Following measurement of amino acid levels and total protein concentration, the %RSD was calculated for all data using various normalization methods. The lowest %RSD was

obtained when amino acid levels were normalized by Phe concentration (mean=42%), with isoleucine was (mean=), whereas the mean %RSD using total protein normalization was 57% (supplementary table S2). The difference in %RSD observed between data normalized by protein and by Phe was statistically significant (p-value < 0.05). Phe is an essential-proteogenic amino acid that is only metabolized by phenylalanine hydroxylase enzyme (EC: 1.14.16.1), an enzyme mainly expressed in the liver and to a lesser extent in the kidneys, with very residual activity in other tissues including HDF [25]. Hence, Phe is not metabolized in HDF. Consequently, Phe may act as a biomarker of cellular number in HDF. In fact, PCA analysis of amino acid levels normalized by total protein content showed a trend towards separation among the different laboratories (Figure 1A). Although clustering analysis (using RStudio's vegan, cluster and factoextra packages) classified two main groups of samples (agglomerative coefficient 0.93), these groups did not correspond to any specific laboratory (see heatmaps in Figure 1C). The Wilcoxon test with Bonferroni correction did not show significant differences between data from different laboratories. However, when amino acid values were normalized by Phe, laboratory A showed data aggregation by both PCA and clustering analysis (Figure 1B and 1D). More specifically, clustering analysis showed 2 main groups (agglomerative coefficient 0.96), one of which contained the samples from laboratory A (except samples 5 and 6), and the other of which contained the samples from all other laboratories (see heatmaps). When amino acid levels were normalized by Phe, a Wilcoxon test with Bonferroni correction revealed that several amino acid levels were significantly different among laboratory A and all other laboratories (arginine, lysine, valine, and leucine showed lower values while serine, glutamine, glycine, aspartate, glutamate, threonine, proline, ornithine, tyrosine, and tryptophan showed higher values in laboratory A when compared with the others). Based on these data, the %RSD was recalculated to exclude data from laboratory A. The resulting median %RSD of all amino acid levels normalized by Phe was subsequently reduced from 42% to 31%, whereas the %RSD of amino acid levels normalized by total protein concentration was only reduced from 57% to 54%. When data from the other laboratories were excluded and %RSD was recalculated, similar reductions in %RSD were not observed (data not shown).

4. Discussion

Here, we propose a method of intracellular amino acid concentration normalization that uses Phe as a surrogate biomarker of HDF cellular number that can be complementary to

the total protein concentration of homogenates. To this end, we used a previously reported targeted metabolomic procedure that was strictly controlled by external quality control schemes and accreditation procedures (ISO-15189 norm and external quality assessment scheme ERNDIM; www.ERNDIM.org) and which possessed the advantages of high reproducibility.

It has been recognized that preanalytical variations in cellular metabolomics contributes to variability in measurements across laboratories that use different skin biopsy collection procedures and varying HDF culture protocols [2,26,27]. Other variables also contribute to the high variability observed in intracellular metabolomic studies, such as the method utilized for metabolomic data normalization by cellular number. Different parameters have been used as surrogate biomarkers of cellular number. These include total protein concentration, cell number count, and tissue weight [12,13]. In order to study another method of data normalization that might minimize inter-laboratory variation, HDF samples were cultured by different laboratories, using different cell culture conditions (Table 1), as occurs in the daily practice (no consensus protocols are regularly applied in laboratories dealing with fibroblasts culturing). We followed previously published preanalytical sample management protocols [2].

The different methods employed for total protein quantification (BCA, Bradford, and Lowry) generally show good agreement, and Lowry was considered adequate for the purposes of the present study [28]. Other authors have proposed alternatives to total protein analysis for data normalization [15,16]. The advantage of Phe analysis is that it is a robust procedure that is usually available in laboratories working in targeted metabolomic analyses. Disadvantages would rely mainly on the need of having more sophisticated technology compared with that needed for total protein quantification.

The initial correlation study displayed the best correlation values between various essential amino acids. Remarkably, Phe showed the second highest correlation with total protein concentration and with other amino acid levels (only isoleucine displayed slightly higher correlation values when compared with those of Phe). More importantly, %RSD was significantly reduced when amino acid levels were normalized by Phe compared to those normalized by total protein concentration or other amino acids (Table S2). Considering that we are studying fibroblasts from a healthy population, the %RSD reduction may allow detecting subtle metabolic disturbances when real patients are compared with controls. As mentioned above, Phe may be a useful surrogate biomarker of total cellular number in HDF, since it is an essential amino acid used in intracellular

protein biosynthesis that is not catabolized by fibroblasts, as a key difference with isoleucine. Although isoleucine displayed the highest mean correlation with the other amino acids, the lowest variability was obtained when data were normalized by Phe, reinforcing its choice as surrogate of cell number. According to Table 1, the passage number, age of controls, and other variables (such as the cell number obtained for each culture) were different among the 4 laboratories that cultured HDFs, suggesting that these differences may contribute to the sample variability observed in the present study, especially for some individual amino acids. This was further confirmed by the observed range of total protein concentrations of the pellets, indicating noticeable variations in the cellular count among laboratories and individual samples. Although a batch effect was not seen when data were normalized by total protein concentration, data from laboratory A did display a batch effect when normalized by Phe. This finding was further confirmed by supervised statistical analysis. Subtle methodological variations in fibroblasts culturing conditions might explain these differences, since analytical and post-analytical variables were exactly the same. Furthermore, some findings regarding amino acid variations make sense since the branched chain amino acids valine and leucine (both of them share common transport systems and catabolic pathways) were both lower in Lab A, while the aromatic amino acids tryptophan and tyrosine showed higher values for Lab A (they also share common transport systems). In the same way, glycine and serine, that are closely related amino acids (glycine is synthesized from serine), also showed higher values in Lab A. Moreover, after removing this dataset, the %RSD further decreased by 11% in the Phe-normalized data whereas it only decreased by 3% in total protein concentration-normalized data. Therefore, the substantial reduction in the variability of the data when normalized by Phe was not observed when data was normalized by protein concentration.

A limitation is we did not study effects of chemical treatment, genetic modification or drug interventions that could potentially alter intracellular metabolism, and this would require further validation.

5. Conclusions

Phenylalanine can be an additional and suitable biomarker of the estimated cellular number in cultured HDFs. This conclusion is supported by the strong correlation between Phe and other amino acid levels and the reduction in %RSD observed in the entire cohort of HDFs from controls. Taking together, normalization by Phe may contribute to a

reduction in targeted metabolomic amino acid results variation due to methodological factors.

Author Contributions: Conceptualization, C.O. and R.A.; methodology, A.A. M.C., J.M., and A.O..; validation, M.C. and A.O.; formal analysis, C.O. and R.A.; investigation, A.A., M.R., J.M., A.P., G.G., J.C-S, R.U., L.C., H.R-G, C.J; resources, M.R., A.P., G.G., R.U., C.J., R.A.; data curation, C.O.; writing—original draft preparation, C.O., R.A. writing—review and editing, All, authors.; visualization, C.O., R.A.,; supervision, R.A..; funding acquisition, A.P. G.G. and R.A.. All authors have read and agreed to the published version of the manuscript."Data Availability Statement: The metabolomics data reported in this paper are available in supplementary tables 1 and 2.

Declaration of Competing Interest: The authors declare no conflict of interest.

Data availability: Data are fully available in supplementary tables 1 and 2.

Funding: This research was funded by: Instituto de Salud Carlos III (PI20-00340, PI18-00498, PI2100935, PI20/00759) (co-funded by European Regional Development Fund "A way to make Europe"), the Centre for Biomedical Research on Rare Diseases (CIBERER) (ACCI19-759), the Secretariat for Universities and Research of the Ministry of Business and Knowledge of the Government of Catalonia (2017SGR1206) to AP and Carmen de Torres grants (Institut de Recerca Sant Joan de Déu). CIBERER is an initiative from the Health Institute Carlos III.

Ethics approval

Institutional Review Board Statement: This study was approved by the Ethical Committee of the Hospital San Joan de Déu (HSJD) in accordance with the Helsinki Declaration of 1964, as revised in 2001. The samples from patients and controls were donated by the HSJD and CIBERER Biobanks and they provided written informed consent from patients or their parents. The study was approved by the ethics committee with the reference PIC-97-16.s.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

References

[1] M. Čuperlović-Culf, D.A. Barnett, A.S. Culf, I. Chute, Cell culture metabolomics: applications and future directions, Drug Discov Today. 15 (2010) 610–621. https://doi.org/10.1016/J.DRUDIS.2010.06.012.

[2] J. Wilkins, D. Sakrikar, X.M. Petterson, I.R. Lanza, E. Trushina, A comprehensive protocol for multiplatform metabolomics analysis in patient-derived skin fibroblasts, Metabolomics. 15 (2019). https://doi.org/10.1007/S11306-019-1544-Z.

[3] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics., Mass Spectrom Rev. 26 (n.d.) 51–78. https://doi.org/10.1002/mas.20108.

[4] X. Luo, L. Li, Metabolomics of Small Numbers of Cells: Metabolomic Profiling of 100, 1000, and 10000 Human Breast Cancer Cells, Anal Chem. 89 (2017) 11664–11671. https://doi.org/10.1021/ACS.ANALCHEM.7B03100.

[5] A.P.H. Danielsson, T. Moritz, H. Mulder, P. Spégel, Development and optimization of a metabolomic method for analysis of adherent cell cultures, Anal Biochem. 404 (2010) 30–39. https://doi.org/10.1016/J.AB.2010.04.013.

[6] A.G. Foote, Z. Wang, C. Kendziorski, S.L. Thibeault, Tissue specific human fibroblast differential expression based on RNAsequencing analysis, BMC Genomics. 20 (2019). https://doi.org/10.1186/s12864-019-5682-5.

[7] C.A. Sellick, R. Hansen, A.R. Maqsood, W.B. Dunn, G.M. Stephens, R. Goodacre, A.J. Dickson, Effective quenching processes for physiologically valid metabolite profiling of suspension cultured Mammalian cells, Anal Chem. 81 (2009) 174–183. https://doi.org/10.1021/AC8016899.

[8] I.F. Duarte, J. Marques, A.F. Ladeirinha, C. Rocha, I. Lamego, R. Calheiros, T.M. Silva, M.P.M. Marques, J.B. Melo, I.M. Carreira, A.M. Gil, Analytical approaches toward successful human cell metabolome studies by NMR spectroscopy, Anal Chem. 81 (2009) 5023–5032. https://doi.org/10.1021/AC900545Q.

[9] J. Yuan, B.D. Bennett, J.D. Rabinowitz, Kinetic flux profiling for quantitation of cellular metabolic fluxes, Nat Protoc. 3 (2008) 1328–1340. https://doi.org/10.1038/NPROT.2008.131.

[10] J.B. Ritter, Y. Genzel, U. Reichl, Simultaneous extraction of several metabolites of energy metabolism and related substances in mammalian cells: optimization using experimental design, Anal Biochem. 373 (2008) 349–369. https://doi.org/10.1016/J.AB.2007.10.037.

[11] J.E. Wulff, M.W. Mitchell, J.E. Wulff, M.W. Mitchell, A Comparison of Various Normalization Methods for LC/MS Metabolomics Data, Advances in Bioscience and Biotechnology. 9 (2018) 339–351. https://doi.org/10.4236/ABB.2018.98022.

[12] B.B. Misra, Data normalization strategies in metabolomics: Current challenges, approaches, and tools., Eur J Mass Spectrom (Chichester). 26 (2020) 165–174. https://doi.org/10.1177/1469066720918446.

[13] Y. Wu, L. Li, Sample normalization methods in quantitative metabolomics., J Chromatogr A. 1430 (2016) 80–95. https://doi.org/10.1016/j.chroma.2015.12.007.

[14] M. Ruiz, M. Jové, A. Schlüter, C. Casasnovas, F. Villarroya, C. Guilera, F.J. Ortega, A. Naudí, R. Pamplona, R. Gimeno, S. Fourcade, M. Portero-Otín, A. Pujol, Altered glycolipid and glycerophospholipid signaling drive inflammatory cascades in adrenomyeloneuropathy., Hum Mol Genet. 24 (2015) 6861–76. https://doi.org/10.1093/hmg/ddv375.

[15] A. Ruiz-Aracama, A. Peijnenburg, J. Kleinjans, D. Jennen, J. van Delft, C. Hellfrisch, A. Lommen, An untargeted multi-technique metabolomics approach to studying intracellular metabolites of HepG2 cells exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin, BMC Genomics. 12 (2011). https://doi.org/10.1186/1471-2164-12-251.

[16] B. Cao, J. Aa, G. Wang, X. Wu, L. Liu, M. Li, J. Shi, X. Wang, C. Zhao, T. Zheng, S. Guo, J. Duan, GC-TOFMS analysis of metabolites in adherent MDCK cells and a novel strategy for identifying intracellular metabolic markers for use as cell amount indicators in data normalization, Anal Bioanal Chem. 400 (2011) 2983–2993. https://doi.org/10.1007/S00216-011-4981-8.

[17] L.P. Silva, P.L. Lorenzi, P. Purwaha, V. Yong, D.H. Hawke, J.N. Weinstein, Measurement of DNA concentration as a normalization strategy for metabolomic data from adherent cell lines., Anal Chem. 85 (2013) 9536–42. https://doi.org/10.1021/ac401559v.

[18] Y. Wu, L. Li, Dansylation metabolite assay: a simple and rapid method for sample amount normalization in metabolomics., Anal Chem. 86 (2014) 9428–33. https://doi.org/10.1021/ac503359v.

[19] O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, Protein measurement with the folin phenol reagent, Journal of Biological Chemistry. 193 (1951) 265–275. https://doi.org/10.1016/S0021-9258(19)52451-6.

[20] S. Zukunft, C. Prehn, C. Röhring, G. Möller, M. Hrabě de Angelis, J. Adamski, J. Tokarz, High-throughput extraction and quantification method for targeted metabolomics in murine tissues, Metabolomics. 14 (2018). https://doi.org/10.1007/S11306-017-1312-X.

[21] Z. León, J.C. García-Cañaveras, M.T. Donato, A. Lahoz, Mammalian cell metabolomics: experimental design and sample preparation, Electrophoresis. 34 (2013) 2762–2775. https://doi.org/10.1002/ELPS.201200605.

[22] M. Casado, C. Sierra, M. Batllori, R. Artuch, A. Ormazabal, A targeted metabolomic procedure for amino acid analysis in different biological specimens by ultra-high-performance liquid chromatography-tandem mass spectrometry, Metabolomics. 14 (2018). https://doi.org/10.1007/S11306-018-1374-4.

[23] European Research Network for Evaluation and Improvement of Screening, Diagnosis and Treatment of Inherited Disorders of Metabolism. https://www.erndimqa.nl/ (accessed May 2, 2022).

[24] RStudio Team, RStudio: Integrated Development Environment for R, (2020). http://www.rstudio.com/.

[25] GTEx Portal. https://www.gtexportal.org/home/gene/PAH (accessed May 2, 2022).

[26] K. Dettmer, N. Nürnberger, H. Kaspar, M.A. Gruber, M.F. Almstetter, P.J. Oefner, Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols., Anal Bioanal Chem. 399 (2011) 1127–39. https://doi.org/10.1007/s00216-010-4425-x.

[27] H. Bi, K.W. Krausz, S.K. Manna, F. Li, C.H. Johnson, F.J. Gonzalez, Optimization of harvesting, extraction, and analytical protocols for UPLC-ESI-MS-based metabolomic analysis of adherent mammalian cancer cells., Anal Bioanal Chem. 405 (2013) 5279–89. https://doi.org/10.1007/s00216-013-6927-9.

[28] M. Betti, C. Ciacci, S. Abramovich, F. Frontalini, Protein Extractions from Amphistegina lobifera: Protocol Development and Optimization, Life (Basel). 11 (2021). https://doi.org/10.3390/LIFE11050418.

Table 1 Main HDF culture conditions from the 4 laboratories involved in the study.

Lab	Sample Number	Growth medium composition	Passage number	Ages	Fibroblast precedence
A	10	1 g/L glucose, L-glutamine and pyruvate	5-13	3 - 25 years	alar surface of the arm
B	16	4.5 g/L glucose, L-glutamine	6	2 – 44 years	alar surface of the arm or abdominal
C	6	1 g/L glucose, L-glutamine and pyruvate	7-10	4 – 69 years	alar surface of the arm
D	12	4.5 g/L glucose, L-glutamine and amphotericin	4-6	4 months - 17 years	Shoulders (9), hand (2) or leg (1)

Figure 1 (in colour). PCA and heatmap analysis for data normalized by protein (Figure 1A and 1C respectively) and by Phe (Figure 1B and 1D respectively).

PCA's colours/symbols (1A and 1B) correspond to the laboratory where each sample belongs and the plot confidence ellipses around group mean points. X and Y axis represents 1st and 2nd dimension respectively with its percentage of explanation between brackets. Heatmap figures (1C and 1D) show the aggrupation between samples from different laboratories (right and left). Each letter at the right means the group where belongs each sample analysed, and left graphics shows the proximity between them. The colour represents the power of influence of each amino acid in samples classification in a scale from 0 (lowest) to 50 (highest). Although a trend towards Lab A clustering is visually observed in panel A (data normalized by protein values), this was not confirmed after clustering analysis, while it did clustered when normalized by Phe values (panel B). Regarding heatmap, we can see that there are two main groups that cluster with either the normalization used. In the case of da-ta normalized by proteins (panel C), both clusters include samples from different laboratories, while when data were normalized by Phe (panel D), one of the clusters only have samples form laboratory A.

Supplementary materials:

Supplementary Table S1: Fibroblast total protein (g/L) and amino acid concentrations (umol/L) in the 44 samples cultured in different laboratories.

Supplementary Table S2: Intracellular amino acid levels corrected by Phe (upper data set) and total protein concentrations (expressed as umol/g protein; lower data set), as surrogate biomarkers of cell number. Calculation of each amino acid %RSD (and total mean value) are also stated.

Supplementary Table S3: Data of correlation coefficients among all the amino acids and total protein concentrations.

Supplementary Figure S1: Fibroblast workflow study figure. The sample number indicated in this figure are the same numbers used in figure 1A and 1B.

Supplementary Figure S2: A real example of mass spectrometry analysis is depicted (sample 34). On the top, we present a table with each amino acid, precursor and fragment transitions (Trace), retention time (RT), peak height (Height), peak area (Area), internal standard area (IS area), response and calculated concentration (Conc.). Down (left and right panels), an example of a Phe chromatogram and the calibration curve are presented.