Sequential Fractionation Strategy Identifies Three Missing Proteins in The Mitochondrial Proteome of Commonly Used Cell Lines

Journal:	*Journal of Proteome Research*
Manuscript ID	pr-2018-00422d.R2
Manuscript Type:	Letter
Date Submitted by the Author:	n/a
Complete List of Authors:	Ronci, Maurizio; University "G. d'Annunzio" of Chieti-Pescara, Department of Medical, Oral and Biotechnological Sciences ; IRCSS Fondazione Santa Lucia, Proteomics and Metabonomics Unit Pieroni, Luisa; IRCSS Fondazione Santa Lucia, Proteomics and Metabonomics Unit Greco, Viviana; IRCSS Fondazione Santa Lucia, Proteomics and Metabonomics Unit Scotti, Luca; University of Chieti-Pescara, DSMOB Marini, Federica; Universita Cattolica del Sacro Cuore Carregari, Victor; IRCSS Fondazione Santa Lucia, Proteomics and Metabonomics Unit Cunsolo, Vincenzo; University of Catania, Chemical Science Foti, Salvatore; Universita degli Studi di Catania, department of Chemical Science Aceto, Antonio; University "G. d'Annunzio" of Chieti-Pescara, Department of Medical, Oral and Biotechnological Sciences Urbani, Andrea; Universita Cattolica del Sacro Cuore

Sequential Fractionation Strategy Identifies Three Missing Proteins in The Mitochondrial Proteome of Commonly Used Cell Lines

Maurizio Ronci[1,2], Luisa Pieroni[2], Viviana Greco[3,4], Luca Scotti[1], Federica Marini[3,4], Victor C Carregari[3,4], Vincenzo Cunsolo[5], Salvatore Foti[5], Antonio Aceto[1], Andrea Urbani[3,4]*

1 Department of Medical, Oral and Biotechnological Sciences, University G. D'Annunzio Chieti;

2 IRCCS Fondazione Santa Lucia, Proteomics and Metabonomics Unit, Rome;

3 Institute of Biochemistry and Clinical Biochemistry, Università Cattolica del Sacro Cuore, Rome;

4 Department of Laboratory Diagnostic and Infectious Diseases, Fondazione Policlinico Universitario Agostino Gemelli-IRCCS, Rome;

5 Department of Chemical Sciences, University of Catania;

Corresponding Author

* Dr. Maurizio Ronci

Department of Medical, Oral and Biotechnological Sciences, University G. D'Annunzio Chieti; Analytical Biochemistry and Proteomics Unit CeSI-Met.

Via Luigi Polacchi, 11/13

66013 Chieti ITALY

Phone # +3908713554517;

E-mail: maurizio.ronci@unich.it

Abstract

Mitochondria are undeniably the cell powerhouse, directly affecting cell survival and fate. Growing evidences suggest that mitochondrial protein repertoire affects both metabolic activity and plays an important role in determining cell proliferation/differentiation or quiescence shift. Consequently, the bioenergetic status of a cell is associated with the quality and abundance of the mitochondrial populations and proteomes. Mitochondrial morphology changes in the development of different cellular functions associated with metabolic switches. It is therefore reasonable to speculate that different cell lines do contain different mitochondrial associated proteins and the investigation of these pools may well represent a source for mining missing proteins (MPs). A very effective approach to increase the number of IDs through mass spectrometry consists in reducing the complexity of the biological samples by fractionation. The present study aims at investigating the mitochondrial proteome of five phenotypically different cell lines, possibly expressing some of the MPs, through an enrichment-fractionation approach, at organelle and protein level. We demonstrate a substantial increase in the proteome coverage, which in turns increases the likelihood of detecting low abundant proteins, often falling in the category of MPs, and resulting in the identification of three MPs. All MS data have been deposited to the MAssIVE data repository (https://massive.ucsd.edu) with the data set identifier MSV000082409 and PXD010446.

KEYWORDS: High-density datasets; missing proteins; mitochondrial proteome; mt-HPP; protein fractionation; METTL12; FAM163A; RGS13

Introduction

Current knowledge grown experimentally in the field of proteomics originates greatly from conventional studies on conventional samples using conventional techniques such as commercial cell lines and standard mass spectrometry approaches. The concept of 'missing proteins' (MPs) emerged in recent years after recognizing that many of the proteins encoded in the genome were never identified in direct experiments, also when using the definitive technologies [1, 2]. Definitive methods are based of the direct detection of molecules thus without the involvements of secondary reaction which would provide an indirect detection of a given chemical entities. Immunochemistry based methods represent a well-established example of indirect detection while spectroscopic and spectrometric technologies are considered key example of definitive methods to discover new generation biomarkers for clinical practice[3].

Nevertheless, the capability of spectrometric methods to detect a given protein goes along with the accessibility of proteotypic transitions, a crucial aspect regarding the nature of proteins[4, 5]. Proteins cannot be found by any MS methods if no detectable peptide is produced, if the protein is highly hydrophobic or if the protein is subsumed or indistinguishable in sequence from other known proteins [6, 7]. In addition to this, the MPs without transcript evidence (PE3/4) may be encoded by genes only expressing them at extremely low concentrations or in not common organs, tissues, or cell types or after induction (e.g. infection or inflammation for beta-defensins) [8-10]. A beautiful example of this concept is testis, a tissue that has recently emerged as the leading example of tissue-specific protein expression, with 999 proteins enriched and 89 missing proteins detected [11-13]. However, MPs with transcript evidence (PE2) are not supposed to be "missing" at protein level. It is worth noting that the criteria for claiming MPs identification are becoming more and more stringent and especially for novel proteins, researchers have agreed on using more rigorous guidelines. The latest set of HPP guidelines, the HPP Mass Spectrometry Data Interpretation Guidelines Version 2.1,

has been introduced to address the stringency of data required to identify MPs [14]. As of January 2018, the total MPs computed by neXtProt based on experimental information from different types of studies is 2,186 PE2,3,4 predicted proteins and 574 uncertain/dubious entries PE5 (Credit to Lydie Lane and Gilbert S. Omenn). It should be emphasized that numerous genomic variants have not yet been incorporated into the reference databases, exposing the community to the risk of misidentifying the unrecognized variant proteins and peptides and suggesting the use of more sophisticated bioinformatics tools such as error tolerant type of database search (SPIDER).

Here we report the application of a standard analytical approach to a novel challenge, the identification of MPs, employing direct mass spectrometry on protein extracted from mitochondria isolated from five different cell lines. Since its establishment in 2014, The Mitochondrial Human Proteome Project (mt-HPP) [15] led by the Italian Proteomics Association (ItPA) has been studying human mitochondrial proteins [16]. The interest in mitochondrial proteome is mainly connected to the regulation of life and death of cells operated by this key organelle [17]. Mitochondria are very versatile, and they are involved in a plethora of cellular functions such as energy production, regulation of death pathways, and calcium buffering. Their dysfunction has been related to many pathological conditions ranging from neurodegenerative disorders to cancer and metabolic diseases [18-20]. The vast majority of human mitochondrial disorders are caused by defects of nuclear genes, which is not surprising since the mtDNA encodes only 13 proteins and 2 functional peptides. Advances in MS based proteomics have only recently made it possible to systematically identify the complement of over 1,000 proteins that comprise the mammalian mitochondrial proteome. The entire mitochondrial proteome is composed by proteins that are encoded by mitochondrial and nuclear genomes, together with all the physical and functional interactors, involved in mitochondrial dynamics and metabolism[21]. Hence, their dynamic proteome is crucial for the definition of the healthy state, the dynamics of the mitochondrial network, and, in turn, of the cellular conditions.

According to the overall goals of both HPP initiatives, C-HPP aimed at eliminating proteins that show no evidence at protein levels and B/D HPP aimed at answering important biological questions, the present study aims to investigate the mitochondrial proteome of different cell types that may be different in composition, and possibly enriched in some of the MPs. Herein, by employing a double fractionation approach and state-of-the-art mass spectrometry and following the Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1, we identify three proteins annotated as 'missing' in the latest version of the neXtProt database.

Materials and Methods

Materials and Reagents

40% acrylamide/Bis solution, N,N,N',N'-tetramethylethylenediamine (TEMED), molecular mass standards and electrophoresis apparatus for one-dimensional electrophoresis were from Bio-Rad Laboratories, Inc., Hercules CA.

β-mercaptoethanol, dithiothreitol (DTT), ammonium persulfate, 3-[3-cholamidopropyl dimethylammonium]-1-propanosulfonate (CHAPS), acetonitrile(ACN), trifluoroacetic acid (TFA), sodium dodecyl sulphate (SDS), iodoacetamide (IAA), formic acid (FA) and all other chemicals used all along the experimental work were current pure analytical grade products and purchased from Sigma-Aldrich Corporation, St Louis, MO. Water and acetonitrile (OPTIMA®LC/MS grade) for LC/MS analyses were purchased from Fisher Scientific, UK. Modified porcine trypsin was purchased from Promega (Madison, WI, USA). Cell culture media and other cell culture reagents were from Euroclone, Sigma, Corning, and Gibco Life Technologies.

Cell culture

Neuroblastoma SH-SY5Y (ECACC 94030304; Lot No. 11C016), osteosarcoma U-2 OS (ATCC HTB-96), embryo kidney Hek293 (ATCC CRL-1573), cervix adenocarcinoma Hela (ATCC CCL-2) and hepatocellular carcinoma HepG2 (ATCC HB-8065) cells were maintained at 37 °C under humidified conditions and 5% CO2 and grown in High Glucose DMEM, 10% FBS except for HepG2 for which Low Glucose was used. Absence of mycoplasma was periodically checked by EZ-PCR Mycoplasma Test Kit (Biological Industries) according to manufacturer's instructions.

Mitochondria enrichment and mitochondrial protein sample preparation

Cells were starved and washed in PBS solution to remove trypsin residual and centrifuged. Mechanical lysis of cells was performed in ice bath, resuspending the pellets in the mitochondrial buffer containing 210 mM D-mannitol, 70 mM sucrose, 1 mM EDTA, 10 mM Hepes, pH 7.5 plus 1× Protease Inhibitor Cocktail, P8340 Sigma-Aldrich) and using a hand driven Potter-Elvejehm homogenizer. 30 strokes were enough to break > 90% of cells. Homogenate were centrifuged at low speed (1,000g × 10 min @4°C) to remove intact cells, nuclei and cell debris. Supernatant were subsequently centrifuged at higher speed (7,000g × 10min @4°C) to enrich crude mitochondria. The purified fraction was separated by sucrose gradient and a high-speed ultracentrifugation. Finally, mitochondria were collected and recovered at gradient interface. The final mitochondrial pellets were aliquoted and stored at -80°C until use.

Mitochondria were resuspended in lysis buffer (Sucrose 250 mM, EDTA 1 mM pH 8, Hepes 20 mM pH 7.5, 0.5% n-Dodecyl β-D-maltoside) on ice for 30 min. Lysates were centrifuged (13,000 RPM × 20 min@ 4°C) and supernatant recovered and stored on ice. Protein concentration were estimated by Bradford assay. 50 µg of proteins of each cell line were solubilized in Laemlli Buffer (DTT 50 mM, Tris pH 6.8, 0.1 M, SDS 8%, Glycerol 20%) and loaded onto an SDS-PAGE gel, composed by a 4% polyacrylamide stacking gel (500 mM Tris-HCl, pH 6.8, 0.1%, m/v, SDS) over a 12% resolving polyacrylamide gel (in 1.5M Tris-HCl, pH 8.8, 0.1%, m/v, SDS buffer).

Electrophoresis was performed at 50 V for 30 min and 150 V until the bromophenol blue front reaches the bottom of the gel. Staining and de-staining were performed with Colloidal Coomassie Blue and a solution water, methanol, and acetic acid in a ratio of 50/40/10 (v/v/v), respectively. Gel lanes were equally sliced into 20 bands and were washed in water, washing solution (25 mM Ammonium bicarbonate, 50% ACN) and finally with 100% ACN. Dried slices were treated with 10 mM DDT and alkylated with 55 mM iodoacetamide. Digestion was performed overnight with trypsin for all cell lines. Peptides were extracted and dried before LC-MS/MS analysis.

Mass spectrometry and bioinformatics

Fractionated mitochondrial peptides were reconstituted in 0.1% FA and analyzed using Orbitrap Fusion Tribrid (Q-OT-qIT) mass spectrometer (Thermo Scientific, San Jose, CA). Samples were loaded onto an Acclaim Nano Trap C18 Column (100 µm x 2 cm, 5 µm, 100Å) and after 3 min trapping separated by a PepMap RSLC C18 EASY-Spray column, 50cm x 75µm 2µm ID, with a Thermo Scientific Dionex UltiMate 3000 RSLC nano system (Sunnyvale, CA). Peptides eluted with linear gradient from 1 to 40 % B in 90 min at flow rate 0.300 µl/min. Survey scans of peptide precursors from 380 to 2000 m/z were performed at 240K resolution with a 2×105 ion count target and a maximum injection time of 50 ms. HCD fragmentation with normalized collision energy of 35, and rapid scan MS analysis in the ion trap. The MS2 ion count target was set to 1·104 and the maximum injection time was 150 ms. Only precursors with charge state 2–7 and an intensity above the threshold of 5·103 were sampled for MS2. The dynamic exclusion duration was set to 60 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor selection was on. The instrument was run in top speed mode with 3 seconds cycles, meaning the instrument would continuously perform MS2 events until the list of non-excluded precursors diminishes to zero or 3 s. Raw data were processed using PEAKS Studio 7.5 (Bioinformatics Solutions Inc., Waterloo, ON Canada) and searched using both the PEAKS search engine and the SPIDER peptide mutation

and homology search tool against neXtProt human database (July 2018; 42,228 total entries). Parent Mass Error Tolerance was set to 10.0 ppm and Fragment Mass Error Tolerance to 0.6 Da. Other search parameters were trypsin enzyme specificity, two missed cleavages per peptide, fixed Carbamidomethylation of Cys and variable Oxidation of Met, Deamidation of Gln and Asn (NQ), with two variable PTM per peptide. Non-specific cleavage was allowed to only one end of the peptide. FDR estimation was enabled, and precursor options corrected. To follow the HPP Mass Spectrometry Data Interpretation Guidelines Version 2.1 and in view of our goal of finding MPs, we set the FDR threshold on PSMs to 0.5%, typically resulting in FDR on peptides around 1% and we filtered out all the proteins identified with only one peptide, resulting in FDR at the protein level typically less than 0.5%. The resulting proteins were exported and compared to the latest protein existence information (2,760 entries - January 2018 update) in nextProt and the corresponding matching peptides were further investigated by inspecting for proteotypicity using the online unicity-checker tool (https://www.nextprot.org/tools/peptide-uniqueness-checker). Nested Venn diagrams were obtained using the free software VennPainter v1.2.0, a tool for the Comparison and Identification of Candidate Genes Based on Venn Diagrams[22] .

Data Availability

All MS proteomics data have been deposited to the MAssIVE data repository (https://massive.ucsd.edu) and are available for download through the ProteomeXchange Consortium with the data set identifiers MSV000082409 and PXD010446.

Results and Discussion

In the last build of PeptideAtlas (Human 2018-01b) the number of PSMs and the number of Distinct Peptides reached 217,845,775 and 1,418,040 respectively, accounting for more than 23% and 14% increase compared to the precedent build, only twelve months older. Even if this high

increase translates in only 0.17% increase in the number of proteins, (35 more), it is clear that the 'quest for missing proteins' has become a crucial aspect of the proteomics research worldwide. The existence of a high number of proteins with evidence at transcript level is experimentally based on antibodies, but antibodies are often not fully reliable. There is therefore the need of confirming the existence of many classes of MPs by cross-validation through mass spectrometry-based experiments.

Enrichment and fractionation strategy

The most common reductionist strategy employed to characterize subcellular organelles consists in homogenizing the cells in absence of detergents and fractionating the lysates by sequential differential centrifugation. In our studies, the efficacy of the protocol for mitochondria enrichment by differential centrifugation and subsequent purification on sucrose gradient is routinely assessed by measuring the amounts of Histone H3, β-tubulin, VDAC 1, Cytrate Synthase and Cox5B as markers of nuclear, cytosolic, mitochondrial outer membrane (OMM), mitochondrial matrix (MM) and mitochondrial inner membrane (IMM) respectively, as reported in our previous work[16]. Differential centrifugation followed by sucrose gradient purification resulted a suitable trade-off between purity and yield for mitochondrial enrichment.

The proteins extracted from the purified mitochondria were separated by mono-dimensional (1D) SDS-PAGE, which provides a separation based on the molecular mass. Figure 1 shows a typical acrylamide/bis-acrylamide gel at 14%, stained with Coomassie blue, in which a total mitochondrial protein extract has been separated according to their apparent molecular weight.

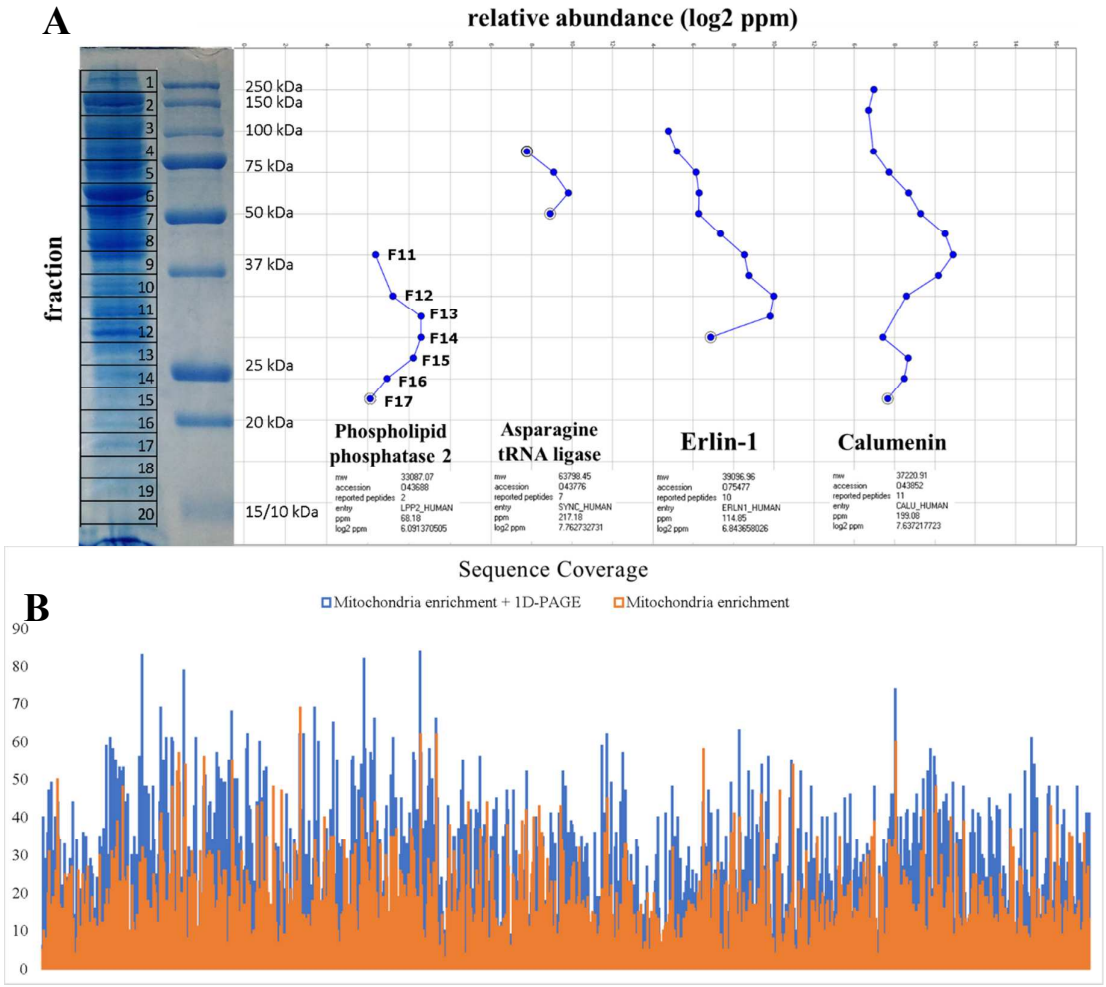


Figure 1. Panel A, Example of 14% PAGE gel, stained with Coomassie blue, in which the proteins extracted from the purified mitochondria were separated (left) and the corresponding protein marker lane. An even distribution of the proteins with mass ranging from 300 to 25 kDa and a lower abundance of the proteins below 25 kDa can be noted. The right side of the figure shows the relative amount of four selected protein as a function of the fraction - band - in which that protein was identified. Panel B, Comparison of the sequence coverage of the common proteins identified after mitochondrial enrichment only (orange) and after mitochondrial enrichment and 20x 1D-PAGE fractionation.

The results indicate an even distribution of the proteins with mass ranging from 300 to 25 kDa, and lower abundance of proteins below 25 kDa. From the same figure, on the right side, the matching between the theoretical mass of some of the identified proteins (reported as an example) and the mass of the band in which that same protein has been identified can be noted. Interestingly, from our

previous work, the total number of proteins identified from sucrose gradient enriched mitochondria in Hek293 cells resulted around one thousand. The same number more than double adding the 1D-PAGE fractionation. Obviously, together with the number of IDs also the protein sequence coverage increases. This is evident from the increased sequence coverage of the proteins identified in both analyses, as shown in figure 1B.

This rather deep approach was employed to analyze the mitochondrial proteome extracted from HeLa, U2OS, SH-SY5Y, HepG2 and Hek293 cells. The main features of the selected cell lines, e.g. histological source, number of chromosomes etc., are shown in Table 1.

Protein Identification

The enrichment/fractionation approach coupled to high resolution orbitrap MS returned a high number of protein IDs. 250 to 700 protein groups are expected to be identified in each gel band. According to the latest HPP guidelines, at least two non-nested uniquely mapping peptides of length ≥ 9 aa must be considered for MPs identification. Therefore, the bands were merged and searched in a single run with the PEAKS DB search engine and then with the SPIDER algorithm [23] to address for possible single amino acids variants (SAAVs). Biological samples commonly contain proteins with slightly different sequences compared to those in the databases. This is frequently caused by polymorphism, database errors, cross species database searching. The SPIDER algorithm tries to match the de novo sequence tags with the database proteins. When a significant similarity is found, the algorithm tries to use both de novo sequencing errors and homology peptide mutations to explain the differences.

The results were filtered to retain only proteins identified by two or more unique peptides, the FDR at the PSM level was set to 0.5% resulting in an FDR at the peptide level of around 1% and calculated FDR at the protein level from 0.1 to 1.1% (Table 2a and 2b). In the same table are also reported the number of total PSMs match, the number of peptides and proteins together with the

number of false positive hits. Several proteins ranging from 1,920 to 3,242 were identified using the standard DB search algorithm and from 1,962 to 3,292 using the SPIDER search algorithm. Supplementary Table S1 reports the statistics of DB search and SPIDER search for the five analysed cell lines, together with the list of the identified proteins as exported by the software. In total, 18 MPs were identified by at least two unique peptides (≥9 aa) through the DB search algorithm and 20 through the SPIDER search (Figure 2 A, B).

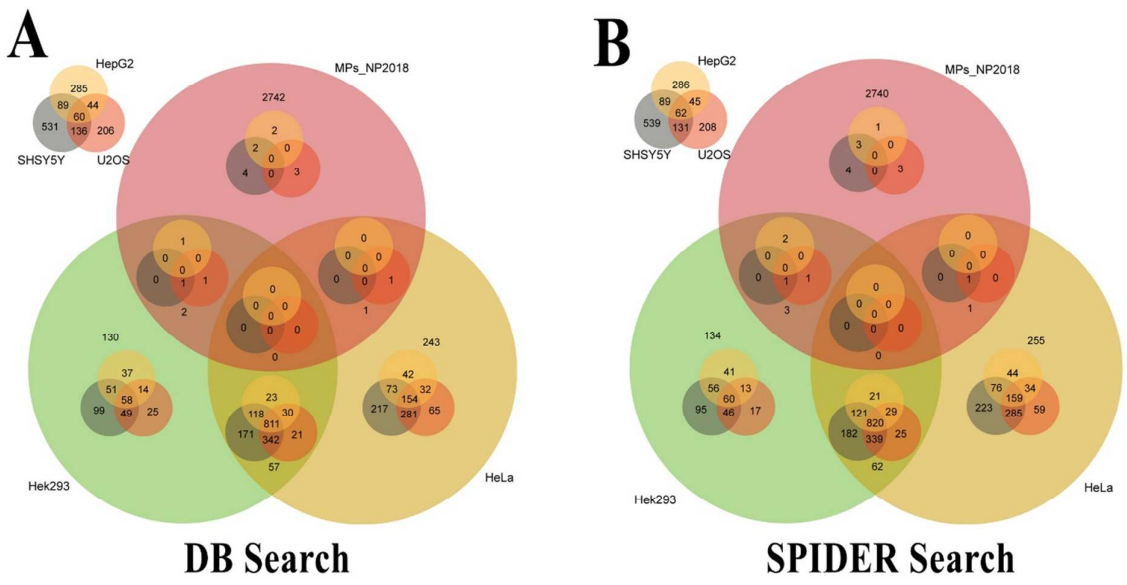


Figure 2. Nested Venn diagrams of the identified proteins for the five cell lines and the set of MPs using the DB Search (panel A) and the SPIDER search (panel B). The DB search resulted in 6 identified MPs in U2OS cells, 5 in Hek293, 5 in HepG2, 7 in SH-SY5Y and 2 in HeLa corresponding to 18 unique shared elements. The SPIDER search resulted in 6 identified MPs in U2OS cells, 7 in Hek293, 6 in HepG2, 9 in SH-SY5Y and 2 in HeLa corresponding to 20 unique shared elements.

To confirm that they fulfilled either partially or in total the requirements suggested by HUPO in the latest guidelines 2.1 the MPs candidates were further investigated. False PSMs are very hard to detect and remove by any algorithm to achieve zero false discovery rate, and therefore the target-decoy method has been widely used to estimate the FDR in proteomics. In PEAKS DB, the target and decoy sequences are not treated as separate entries in the database. PEAKS construct a decoy

database with the same size of the target database by concatenating every entry to a random sequence of the same length. The target and decoy databases are then searched either separately or together by the software. After the search, the target and decoy identifications are separated by checking whether they are from the first or the second half of each concatenated sequence. The FDR in PEAKS Studio is calculated as follows: the PSM FDR is equal to the decoy PSMs divided by the target PSMs multiplied by 100, the Peptide FDR is equal to the decoy peptide hits divided by the target peptide hits multiplied by 100 and the Protein FDR is equal to the decoy protein hits divided by the target protein hits multiplied by 100. It should be kept in mind that for the computation of the peptide FDR the redundant peptides (PTMs modified or differing only for one isoleucine/leucine) are merged and therefore considered to be a single identification [24]. In any case, not all peptides and proteins surviving the threshold are "confidently identified".

83 unique peptides, matching to the 18 candidate MPs, were identified through the standard DB search algorithm and 157, matching to the 20 candidate MPs, with the SPIDER Search. They were checked for unicity using the online tool available on the neXtProt website. The results are provided in Supplementary Table S2. 19 peptides survived the uniqueness filter for the DB search peptides and 19 for the SPIDER search. At the protein level, these peptides correspond to six candidate MPs in the current study, namely METTL12, FAM163A, RGS13, RASA4/4B, KRT87P and POTEKP. METTL12, FAM163A and RGS13 (Table 3a) were identified with a number of unique and non-nested peptides ≥ 2. The spectra of the corresponding peptides showed also a convincing matching when compared to the spectra of the same peptides in SRMAtlas (Supporting Figure S1, S4 and S8) and therefore they could be probably upgraded to PE1 at next neXtProt release.

RASA4 and RASA4B (Table 3b) can only be distinguished and validated by peptides covering residue 533 and since they were identified through the same set of peptides already present in PeptideAtlas/neXtProt, they cannot be considered as identified MPs. Moreover, two of the

supporting peptide spectra do not have a high S/N ratio as requested by the Guidelines (Supporting Figure S6 and S7). The same is true for KRT87P and POTEKP (Table 3b) which are annotated as dubious proteins (PE5). The former was identified through several nested peptides and through two mutated peptides by the SPIDER Search. The latter has the peptide EITALAPSTMK matching at 100% with 6 abundant human proteins including actin (ACTG2, ACTG1, ACTB, ACTA1, ACTA2, ACTC1) and the two other peptides already reported in PeptideAtlas/neXtProt, but not considered as sufficiently credible to reclassify this gene as protein-coding by curators. The spectra of their corresponding peptides (Supporting Figure S2, S3 and S5) should be evaluated by the experts, but since many of them are not present in SRMAtlas they will have to be replicated and perhaps compared with spectra from synthetic homologous peptides.

Assumption on the role of the reported proteins

The data provided in this study arise from the enrichment of a subcellular fraction which may well include rafts from mitochondria neighbouring sub-cellular structures. As such, further investigation on the annotations of the three newly identified MPs showed some interesting points of discussion. In fact, the match between the data on RNA expression levels in Human Protein Atlas and the proteomics data in Peptide Atlas and GPM for some of them showed excellent agreement. All identified MPs displayed RNA expression levels consistent with the cell line in which they were detected (see Table 3a and 3b). Other brief considerations can be done for all of them: for example, regulator of G-protein signaling 13, which has detectable RNA levels only in lymphoid myeloma cells (Daudi, Karpas-707, HEL) and in SHSY5Y (RNA expression level 3.1 TPM), was only detected in SHSY5Y cells in the present study. RNA expression analysis of Methyltransferase like 12 (detected in U-2 OS cells) shows that this protein is ubiquitously expressed at low levels in many cell lines but not in SHSY5Y, HeLa and Hek293, and it has been recently described as a novel human mitochondrial lysine-specific methyltransferases (KMT) which acts on citrate synthase[25, 26]

having probably a role in the modulation of this important mitochondrial metabolic enzyme. RNA expression analysis of protein FAM163A also known as neuroblastoma-derived secretory protein, shows that this protein is ubiquitously expressed at very low levels in most human tissues while it is higher in neuroblastoma cells, which is consistent with our detection. RAS p21 protein activator 4 and 4B were detected in HepG2 and SHSY5Y through the same set of peptides, therefore it is not possible to confidently state if one or the other or both were present. Interestingly, by sequence similarity, Uniprot associates these proteins to a function related to the RAS-MAPK pathway, in agreement with the recent finding of Serasinghe MN and colleagues[27], regarding the link between functional mitochondria division and the RAS induced oncogenic cellular transformation mediated by MAPK signalling. The RNA expression level for RAS p21 protein activator 4 and 4B is also reported to be enhanced in U-138 MG, U-2 OS and SH-SY5Y cells, in agreement with our results, and even with uncertain sub cellular localization, they appear to be close to cell junctions and vesicles, which again makes it consistent with the enrichment of mitochondrial fraction.

Putative beta-actin-like protein 3 (identified in HeLa cells) and Putative keratin-87 protein (identified in Hek293) are annotated as 'uncertain' (PE5) and based on current knowledge no function or localization information is available. A special consideration should be made for the peptides containing single amino acids variants, found through the SPIDER Search only for the two previously mentioned PE5 proteins (figure S3 and S5). SPIDER performs homology search based on de novo sequencing tags and it represent a powerful tool for cross-species searches and for finding point mutations. The conventional protein database search approach suffers from the very well-known issue that prevents a protein from being detected if not included in the database and therefore a SPIDER type of search should always be performed. However, the real advantage of the SPIDER Search do not appear from the present study, since the spectra of the matching peptides are not present in SRMAtlas and hence cannot be verified.

Conclusions

With the neXt-50 challenge of C-HPP a new MPs hunting era has started toward the completion of a high-resolution draft of the human proteome. With this study we demonstrated that a deep proteomics analysis can be achieved even through conventional procedures and instrumentation. We report the identification of three MPs within the nomenclature of the HPP. The evidences provided are consistent with the available bibliographic data and provide mass spectrometric features and proteotypic transitions to further explore these proteins in functional investigations. Our approach might be useful to detect MPs on several other cell lines after the enrichment of mitochondria but also of other cellular compartments. An in-depth assessment of the physical and functional mitochondrial proteome could fill our gaps in the understanding of complex mitochondrial processes.

Tables

Table 1. Main features of the selected cell lines

Cell line	*Reference*	*Tissue*	*Cell type/properties*	*Disease*	*Clinical data*	*Kariotype Modal number*
HeLa	ATCC-CCL2	Cervix	epithelial/ adherent	adenocarcinoma	Female, black 31years	82
Hek293	ATCC CRL-1573	Embryonic kidney	epithelial/ adherent	n.d.	Fetus	64
HepG2	ATCC HB-8065	Liver	epithelial/ adherent	hepatocellular carcinoma	Male, Caucasian 15 years	55
SH-SY5Y	ECACC94030304	Neuronal (Bone marrow)	epithelial/ adherent	neuroblastoma	Female, 4 years	47
U-2 OS	ATCC HTB-96	Bone (Tibia)	epithelial/ adherent	osteosarcoma	Female, Caucasian, 15 years	Hypertriploid range

Table 2. Statistics of filtered results for the two types of database search

a) DB search

Cell line	Peptide-Spectrum Matches (PSM)	Peptide Sequences	Proteins/Protein Groups	Decoy hits (peptides)	Decoy hits (protein)	FDR (PSM)	FDR (peptides)	FDR (protein)
HeLa	92,396	31,415	4,134/2,702	99	43	0.5 %	1.4 %	1.1 %
Hek293	42,846	17,322	3,334/2,013	8	4	0.5 %	1.2 %	0.1 %
HepG2	38,415	16,856	3,278/1,920	20	10	0.5 %	1.1 %	0.3 %
SH-SY5Y	77,556	36,694	4,992/3,242	50	21	0.5 %	1.0 %	0.4 %
U-2 OS	53,912	24,856	3,548/2,310	40	16	0.5 %	1.0 %	0.5 %

b) SPIDER search

Cell line	Peptide-Spectrum Matches (PSM)	Peptide Sequences	Proteins/Protein Groups	Decoy hits (peptides)	Decoy hits (protein)	FDR (PSM)	FDR (peptides)	FDR (protein)
HeLa	94,417	32,636	4,375/2,747	110	47	0.5 %	1.5 %	1.1 %
Hek293	44,030	18,224	3,429/2,034	15	7	0.5 %	1.2 %	0.2 %
HepG2	39,438	17,650	3,418/1,952	20	10	0.5 %	1.1 %	0.3 %
SH-SY5Y	80,553	38,657	5,151/3,287	56	24	0.5 %	1.0 %	0.5 %
U-2 OS	54,788	25,428	3,598/2,325	44	18	0.5 %	1.0 %	0.5 %

Table 3a. Details on the identified missing proteins with the corresponding accession number and gene name, cell type, matching peptides, chromosome number and RNA and main localization information from Human Protein Atlas (if available).

Protein name	Accession (Gene name)	Identified in	Chr #	Matching Peptides (bold = proteotypic) (italic = nested)	RNA cell category (HPA)	Main location (HPA)
Methyltransferase-like protein 12, mitochondrial	NX_A8MUP2, METTL12	U-2 OS	11	**VLNPQGTLIQFSDEDPDVR, GITYFAYLIQGSH, AYQLLSECLR**	Mixed (TPM U-2 OS = 0.9)	Not available
Protein FAM163A	NX_Q96GL9, FAM163A	SH-SY5Y	1	**LSFAPTYYK, LAAPQSYPVTWPGSGR**	Cell line enriched (TPM SH-SY5Y = 118.8)	Not available
Regulator of G-protein signaling 13	NX_O14921, RGS13	SH-SY5Y	1	**YGPVVYAAYLK, NIQEPTETCFEEAQK**	Cell line enriched (TPM Daudi = 37.3, TPM SH-SY5Y = 3.1)	Not available

Table 3b. Details on the identified proteins which are PE5 (KRT87P and POTEKP) or PE2 but not distinguishable from each other (RASA4/4B).

Protein name	Accession (Gene name)	Identified in	Chr #	Matching Peptides (bold = proteotypic) (italic = nested)	RNA cell category (HPA)	Main location (HPA)
RAS p21 protein activator 4B	NX_C9J798, RASA4B	HepG2, SH-SY5Y	7	ELSGGAEAGTVPTSPGK, DITGSSDPYCIVK, VSINNTGLLGSYHPGVFR, AHLGALLSALSR	Cell line enhanced (TPM U-2 OS = 16.6, TPM SH-SY5Y = 14.3)	Localized to the Cell Junctions (uncertain) In addition localized to the Vesicles (uncertain)
RAS p21 protein activator 4	NX_O43374, RASA4	HepG2, SH-SY5Y	7	ELSGGAEAGTVPTSPGK, DITGSSDPYCIVK, VSINNTGLLGSYHPGVFR, AHLGALLSALSR	Cell line enhanced (TPM U-2 OS = 23.3, TPM SH-SY5Y = 34.2)	Localized to the Vesicles (uncertain)
Putative keratin-87 protein	NX_A6NCN2, KRT87P	HepG2, SH-SY5Y	12	***KSDLEANVEALTQEIDFLR***, ***SDLEANVEALTQEIDFLR,*** ***KSDLEANVEALTQEIDFLRR,*** ***KSDLEANA(sub V)EALTQET(sub I)DFLR,*** ***SDLEANA(sub V)EALTQET(sub I)DFLR,*** ***SDLEANVEALTQEIDFLRR,*** ***LEANVEALTQEIDFLR,***	N/A	N/A
Putative beta-actin-like protein 3	NX_Q9BYX7, POTEKP	HeLa	2	**CPEALFQPCFLGMESCGIHKTTFNSIVK** **AGFAGDDAPQAVFPSIVGRPR** **EITALAPST(sub I)MK**	N/A	Not available

Supporting Information

The following supporting information is available free of charge at ACS website http://pubs.acs.org

Supporting Figure S1. Annotated spectra of the peptides matching to FAM163A identified in ShSy5y cells and corresponding spectra downloaded from the SRMAtlas database.

Supporting Figure S2. Annotated spectra of the peptides matching to KRT87P identified in HepG2 cells. Corresponding spectra not available in the SRMAtlas database.

Supporting Figure S3. Annotated spectra of the peptides matching to KRT87P identified in ShSy5y cells via SPIDER Search.

Supporting Figure S4. Annotated spectra of the peptides matching to METTL12 identified in U2-OS cells and corresponding spectra downloaded from the SRMAtlas database.

Supporting Figure S5. Annotated spectra of the peptides matching to POTEKP identified in HeLa cells. Corresponding spectra only available for peptide n. 2 in the SRMAtlas database.

Supporting Figure S6. Annotated spectra of the peptides matching to RASA4/4B identified in HepG2 cells and corresponding spectra downloaded from the SRMAtlas database (not available for peptide n. 2).

Supporting Figure S7. Annotated spectra of the peptides matching to RASA4/4B identified in ShSy5y cells and corresponding spectra downloaded from the SRMAtlas database (not available for peptide n. 2).

Supporting Figure S8. Annotated spectra of the peptides matching to RGS13 identified in ShSy5y cells and corresponding spectra downloaded from SRMAtlas database.

Supporting Table S1. The first sheet of the excel file reports the statistics of database search (top section) and the list of the identified proteins per cell line (bottom section). In the middle section the corresponding missing proteins are shown. Sheet 2 to sheet 9 contain the list of proteins identified per cell line as exported by the software PEAKS Studio 7.5.

Supporting Table S2. List of all peptides matching to proteins annotated as missing and uncertain (version 2018). The peptides resulted to be unique for that specific protein are highlighted in green. The column 'Accession' reports the accession number of the protein(s) containing the selected peptide with the MP red colored and bold. The first sheet shows the results of the DB Search algorithm and the second sheets the results of the SPIDER Search.

Author Contributions

All authors have given approval to the final version of the manuscript.

Acknowledgements

The Mitochondrial Human Proteome Project (mt-HPP) is supported by the Italian Proteomics Association (ItPA). Mass spectrometry was performed in the laboratory of VC. MR is supported by the Italian Ministry of University and Research (grant N. FIRB RBFR13PDFX) and the Italian Ministry of Health (Grant N. GR-2011-02351086); AU is supported by Catholic University of the Sacred heart intramural research D1.1 and by MIUR PRIN 20158EB2CM.

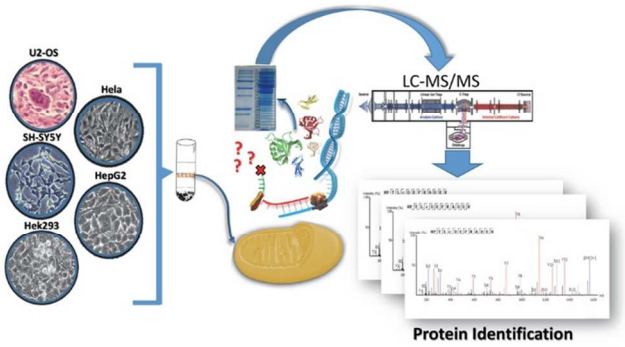
References

1. Paik, Y. K.; Jeong, S. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Cho, S. Y.; Lee, H. J.; Na, K.; Choi, E. Y.; Yan, F.; Zhang, F.; Zhang, Y.; Snyder, M.; Cheng, Y.; Chen, R.; Marko-Varga, G.; Deutsch, E. W.; Kim, H.; Kwon, J. Y.; Aebersold, R.; Bairoch, A.; Taylor, A. D.; Kim, K. Y.; Lee, E. Y.; Hochstrasser, D.; Legrain, P.; Hancock, W. S., The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat Biotechnol* **2012,** 30, (3), 221-3.
2. Paik, Y. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Marko-Varga, G.; Aebersold, R.; Bairoch, A.; Yamamoto, T.; Legrain, P.; Lee, H. J.; Na, K.; Jeong, S. K.; He, F.; Binz, P. A.; Nishimura, T.; Keown, P.; Baker, M. S.; Yoo, J. S.; Garin, J.; Archakov, A.; Bergeron, J.; Salekdeh, G. H.; Hancock, W. S., Standard guidelines for the chromosome-centric human proteome project. *J Proteome Res* **2012,** 11, (4), 2005-13.
3. Donald H. Chace, J. R. B., Mark W. Duncan, Dietrich Matern, Michael R. Morris, Darryl Erik Palmer-Toy, Alan L. Rockwood, Gary Siuzdak, Andrea Urbani, Alfred L. Yergey, Y. Michael Chan, Mass Spectrometry in the Clinical Laboratory: General principles and guidance; approved guideline (C50-A). Clinical and Laboratory Standards Institute:Wayne, PA, USA, **2006**.
4. Steen, H.; Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **2004,** 5, (9), 699-711.
5. Nesvizhskii, A. I.; Aebersold, R., Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* **2005,** 4, (10), 1419-40.
6. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** 422, (6928), 198-207.
7. Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K.; Muthusamy, B.; Leal-Rojas, P.; Kumar, P.; Sahasrabuddhe, N. A.; Balakrishnan, L.; Advani, J.; George, B.; Renuse, S.; Selvan, L. D.; Patil, A. H.; Nanjappa, V.; Radhakrishnan, A.; Prasad, S.; Subbannayya, T.; Raju, R.; Kumar, M.; Sreenivasamurthy, S. K.; Marimuthu, A.; Sathe, G. J.; Chavan, S.; Datta, K. K.; Subbannayya, Y.; Sahu, A.; Yelamanchi, S. D.; Jayaram, S.; Rajagopalan, P.; Sharma, J.; Murthy, K. R.; Syed, N.; Goel, R.; Khan, A. A.; Ahmad, S.; Dey, G.; Mudgal, K.; Chatterjee, A.; Huang, T. C.; Zhong, J.; Wu, X.; Shaw, P. G.; Freed, D.; Zahari, M. S.; Mukherjee, K. K.; Shankar, S.; Mahadevan, A.; Lam, H.; Mitchell, C. J.; Shankar, S. K.; Satishchandra, P.; Schroeder, J. T.; Sirdeshmukh, R.; Maitra, A.; Leach, S. D.; Drake, C. G.; Halushka, M. K.; Prasad, T. S.; Hruban, R. H.; Kerr, C. L.; Bader, G. D.; Iacobuzio-Donahue, C. A.; Gowda, H.; Pandey, A., A draft map of the human proteome. *Nature* **2014,** 509, (7502), 575-81.
8. Zhang, Y.; Li, Q.; Wu, F.; Zhou, R.; Qi, Y.; Su, N.; Chen, L.; Xu, S.; Jiang, T.; Zhang, C.; Cheng, G.; Chen, X.; Kong, D.; Wang, Y.; Zhang, T.; Zi, J.; Wei, W.; Gao, Y.; Zhen, B.; Xiong, Z.; Wu, S.; Yang, P.; Wang, Q.; Wen, B.; He, F.; Xu, P.; Liu, S., Tissue-Based Proteogenomics Reveals that Human Testis Endows Plentiful Missing Proteins. *J Proteome Res* **2015,** 14, (9), 3583-94.
9. Vogel, C.; Marcotte, E. M., Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **2012,** 13, (4), 227-32.
10. Zhu, J.; Chen, G.; Zhu, S.; Li, S.; Wen, Z.; Bin, L.; Zheng, Y.; Shi, L., Identification of Tissue-Specific Protein-Coding and Noncoding Transcripts across 14 Human Tissues Using RNA-seq. *Sci Rep* **2016,** 6, 28400.

11. Carapito, C.; Duek, P.; Macron, C.; Seffals, M.; Rondel, K.; Delalande, F.; Lindskog, C.; Freour, T.; Vandenbrouck, Y.; Lane, L.; Pineau, C., Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods. *J Proteome Res* **2017,** 16, (12), 4340-4351.
12. Vandenbrouck, Y.; Lane, L.; Carapito, C.; Duek, P.; Rondel, K.; Bruley, C.; Macron, C.; Gonzalez de Peredo, A.; Coute, Y.; Chaoui, K.; Com, E.; Gateau, A.; Hesse, A. M.; Marcellin, M.; Mear, L.; Mouton-Barbosa, E.; Robin, T.; Burlet-Schiltz, O.; Cianferani, S.; Ferro, M.; Freour, T.; Lindskog, C.; Garin, J.; Pineau, C., Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. *J Proteome Res* **2016,** 15, (11), 3998-4019.
13. Uhlen, M.; Fagerberg, L.; Hallstrom, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigyarto, C. A.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P. H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Ponten, F., Proteomics. Tissue-based map of the human proteome. *Science* **2015,** 347, (6220), 1260419.
14. Deutsch, E. W.; Overall, C. M.; Van Eyk, J. E.; Baker, M. S.; Paik, Y. K.; Weintraub, S. T.; Lane, L.; Martens, L.; Vandenbrouck, Y.; Kusebauch, U.; Hancock, W. S.; Hermjakob, H.; Aebersold, R.; Moritz, R. L.; Omenn, G. S., Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. *J Proteome Res* **2016,** 15, (11), 3961-3970.
15. Urbani, A.; De Canio, M.; Palmieri, F.; Sechi, S.; Bini, L.; Castagnola, M.; Fasano, M.; Modesti, A.; Roncada, P.; Timperio, A. M.; Bonizzi, L.; Brunori, M.; Cutruzzola, F.; De Pinto, V.; Di Ilio, C.; Federici, G.; Folli, F.; Foti, S.; Gelfi, C.; Lauro, D.; Lucacchini, A.; Magni, F.; Messana, I.; Pandolfi, P. P.; Papa, S.; Pucci, P.; Sacchetta, P.; Italian Mt-Hpp Study Group-Italian Proteomics, A., The mitochondrial Italian Human Proteome Project initiative (mt-HPP). *Mol Biosyst* **2013,** 9, (8), 1984-92.
16. Alberio, T.; Pieroni, L.; Ronci, M.; Banfi, C.; Bongarzone, I.; Bottoni, P.; Brioschi, M.; Caterino, M.; Chinello, C.; Cormio, A.; Cozzolino, F.; Cunsolo, V.; Fontana, S.; Garavaglia, B.; Giusti, L.; Greco, V.; Lucacchini, A.; Maffioli, E.; Magni, F.; Monteleone, F.; Monti, M.; Monti, V.; Musicco, C.; Petrosillo, G.; Porcelli, V.; Saletti, R.; Scatena, R.; Soggiu, A.; Tedeschi, G.; Zilocchi, M.; Roncada, P.; Urbani, A.; Fasano, M., Toward the Standardization of Mitochondrial Proteomics: The Italian Mitochondrial Human Proteome Project Initiative. *J Proteome Res* **2017,** 16, (12), 4319-4329.
17. Picard, M.; Wallace, D. C.; Burelle, Y., The rise of mitochondria in medicine. *Mitochondrion* **2016,** 30, 105-16.
18. Modica-Napolitano, J. S.; Singh, K. K., Mitochondrial dysfunction in cancer. *Mitochondrion* **2004,** 4, (5-6), 755-62.
19. Kasahara, A.; Scorrano, L., Mitochondria: from cell death executioners to regulators of cell differentiation. *Trends Cell Biol* **2014,** 24, (12), 761-70.
20. Lowell, B. B.; Shulman, G. I., Mitochondrial dysfunction and type 2 diabetes. *Science* **2005,** 307, (5708), 384-7.
21. Fasano, M.; Alberio, T.; Babu, M.; Lundberg, E.; Urbani, A., Towards a functional definition of the mitochondrial human proteome. *EuPA Open Proteom* **2016,** 10, 24-27.
22. Lin, G.; Chai, J.; Yuan, S.; Mai, C.; Cai, L.; Murphy, R. W.; Zhou, W.; Luo, J., VennPainter: A Tool for the Comparison and Identification of Candidate Genes Based on Venn Diagrams. *PLoS One* **2016,** 11, (4), e0154315.

23. Han, Y.; Ma, B.; Zhang, K., SPIDER: software for protein identification from sequence tags with de novo sequencing error. *J Bioinform Comput Biol* **2005,** 3, (3), 697-716.
24. Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B., PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics* **2012,** 11, (4), M111 010587.
25. Malecki, J.; Jakobsson, M. E.; Ho, A. Y. Y.; Moen, A.; Rustan, A. C.; Falnes, P. O., Uncovering human METTL12 as a mitochondrial methyltransferase that modulates citrate synthase activity through metabolite-sensitive lysine methylation. *J Biol Chem* **2017,** 292, (43), 17950-17962.
26. Finkemeier, I., Identification of the missing mitochondrial methyltransferase of citrate synthase. *FEBS Lett* **2017,** 591, (12), 1653-1656.
27. Serasinghe, M. N.; Wieder, S. Y.; Renault, T. T.; Elkholi, R.; Asciolla, J. J.; Yao, J. L.; Jabado, O.; Hoehn, K.; Kageyama, Y.; Sesaki, H.; Chipuk, J. E., Mitochondrial division is requisite to RAS-induced transformation and targeted by oncogenic MAPK pathway inhibitors. *Mol Cell* **2015,** 57, (3), 521-36.

For TOC Only


U2-OS
Hela
SH-SY5Y
HepG2
Hek293
LC-MS/MS
Protein Identification