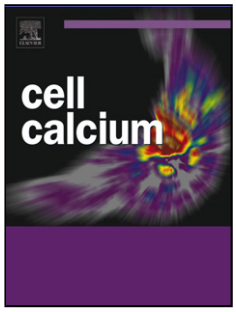


Towards a systems-level understanding of mitochondrial biology

Hilda Carolina Delgado de la Herran (Conceptualization) (Writing - original draft)<ce:contributor-role>Writing -review and editing) (Visualization), Yiming Cheng (Formal analysis), Fabiana Perocchi (Conceptualization) (Writing - review and editing) (Supervision) (Project administration) (Funding acquisition)

PII: S0143-4160(21)00018-X

DOI: https://doi.org/10.1016/j.ceca.2021.102364

Reference: YCECA 102364

To appear in: *Cell Calcium*

Received Date: 22 November 2020

Revised Date: 22 January 2021

Accepted Date: 23 January 2021

Please cite this article as: Delgado de la Herran HC, Cheng Y, Perocchi F, Towards a systems-level understanding of mitochondrial biology, *Cell Calcium* (2021), doi: https://doi.org/10.1016/j.ceca.2021.102364

Title

Towards a systems-level understanding of mitochondrial biology

Authors

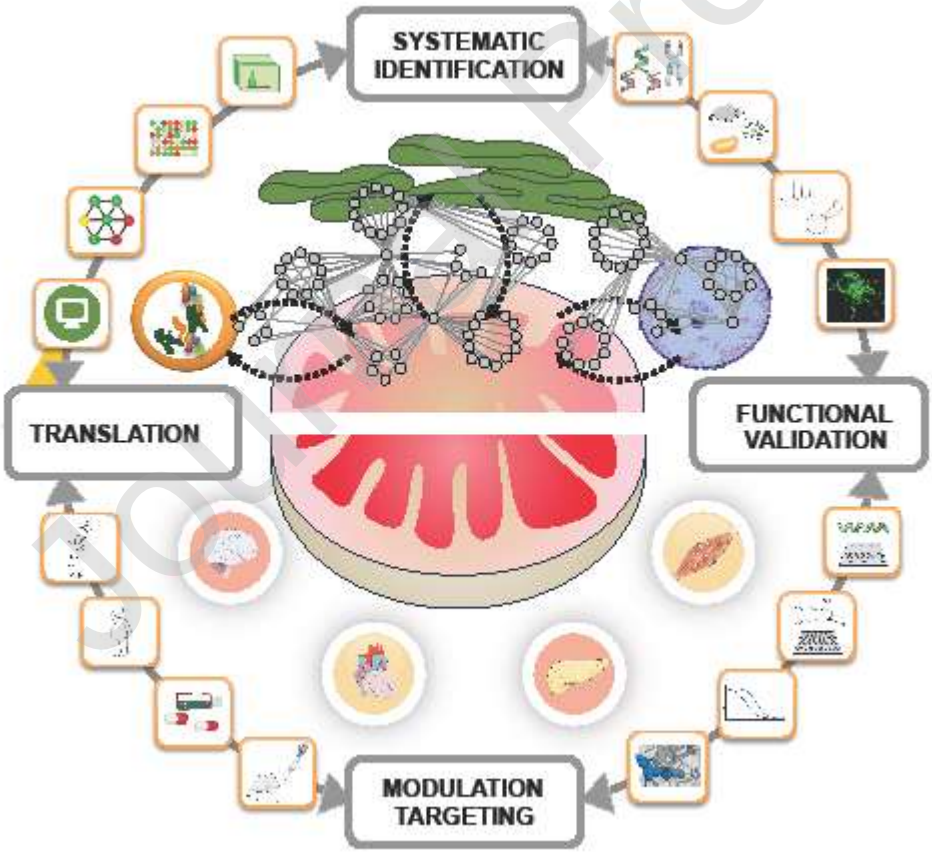
Hilda Carolina Delgado de la Herran[a] , Yiming Cheng[a], Fabiana Perocchi[a,b,*]

[a]Institute for Diabetes and Obesity, Helmholtz Diabetes Center (HDC), Helmholtz Zentrum München, 85764 Neuherberg, Germany

[b]Munich Cluster for Systems Neurology, 81377 Munich, Germany

*Corresponding author: fabiana.perocchi@helmholtz-muenchen.de

Graphical abstract



Highlights

- Roughly 20-40% of human mitochondrial proteins remain functionally orphan to date.
- Over 15% of the mitochondrial proteome shows cell-type-specific differences.
- Systems-wide research approaches catalyze the identification and functional characterization of human mitochondrial proteins.
- Systematic data integration can overcome the limited specificity and coverage of available experimental and computational tools.

Abstract

Human mitochondria are complex and highly dynamic biological systems, comprised of over a thousand parts and evolved to fully integrate into the specialized intracellular signaling networks and metabolic requirements of each cell and organ. Over the last two decades, several complementary, top-down computational and experimental approaches have been developed to identify, characterize and modulate the human mitochondrial system, demonstrating the power of integrating classical reductionist and discovery-driven analyses in order to de-orphanize hitherto unknown molecular components of mitochondrial machineries and pathways. To this goal, systematic, multiomics-based surveys of proteome composition, protein networks, and phenotype-to-pathway associations at the tissue, cell and organellar level have been largely exploited to predict the full complement of mitochondrial proteins and their functional interactions, therefore catalyzing data-driven hypotheses. Collectively, these multidisciplinary and integrative research approaches hold the potential to propel our understanding of mitochondrial biology and provide a systems-level framework to unraveling mitochondria-mediated and disease-spanning pathomechanisms.

Abbreviations

mt-DNA, mitochondrial DNA; MTS_s, mitochondrial targeting sequences; IMM, mitochondrial inner membrane; OMM, mitochondrial outer membrane; IMS, mitochondrial intermembrane

space; PPI, protein-protein interactions; MCUC, mitochondrial calcium uniporter channel; MAMs, mitochondria-associated membranes

1. Introduction

Mitochondria are essential organelles for cellular and organismal life in virtually all eukaryotes (**Figure 1**). Present-day human mitochondria originated from the integration of an endosymbiotic α-proteobacterium into a host cell, therefore exchanging their independence for a semi-autonomous life [1,2]. By the late 1990s, comparative genomics analyses of α-proteobacteria genomes and quantitative two-dimensional gels of highly purified mitochondria suggested that the mammalian mitochondrial proteome consists of ~1,000–1,500 distinct proteins [3,4]. The majority of those proteins derive from the eukaryotic genome, whereas the prokaryotic genome was significantly reduced during the transition from endosymbiotic bacterium to organelle [2] (**Figure 2A**). To date, only a handful of protein-coding genes – thirteen in mammals – are still retained in the mitochondrial DNA (mt-DNA) of almost all eukaryotes. Therefore, most of the mitochondrial proteome is encoded from the nuclear genome, translated in the cytosol, and then targeted and imported into the organelle.

Strikingly, only 1% of mammalian mitochondrial proteins are allocated to ATP synthesis, highlighting that the organelle's functions reach far beyond energy production (**Figure 1**). Indeed, mitochondria are at the core of multiple cellular pathways, including the biosynthesis of precursors for cholesterol, estrogen, testosterone and hemoglobin; the regulation of redox and ion homeostasis; the activation of antiviral responses and cell death. Adding an additional layer of complexity, mitochondrial functions are tied to the specialized tasks and physiology of different cell types, tissues, and organisms [5]. For instance, only between 40-70% of the human mitochondrial proteome is conserved in commonly used model systems such as unicellular eukaryotes (e.g. *S. cerevisiae*) and invertebrates (e.g. *C. elegans*, *D. melanogaster*) (**Figure 2B**). Furthermore, over 15% of the mitochondrial system shows

tissue-specificity [6] and profound differences even among cell types of the same tissue [6,7].

As a consequence of their dual genetic origin and pleiotropic roles, mitochondria need to tune their biogenesis and activity to the metabolic requirements of each cell. To fulfil this task, the organelle engages in bi-directional signaling with other subcellular compartments by remodeling shape, size, motility, metabolism, protein composition – and more – throughout cell development, differentiation, and proliferation [8,9]. In neurons, for example, mitochondria travel along the axon from the cell body to the synapses and backwards. Transport, positioning, and docking of the organelle at specific locations represent important mechanisms to provide local ATP supply and to buffer cytosolic calcium (Ca^{2+}) for proper axonal branching, local protein translation, neuronal polarization, and synaptic transmission [10]. Furthermore, mitochondria can directly and promptly engage in physical and functional inter-organelle cross-talk, mediating the biosynthesis and exchange of metabolites and the homeostasis of several ions [11]. In light of all evidence, a crucial question arises: How can we tackle such a complex system? Understanding mitochondrial function and dysfunction becomes extremely challenging when taking a one-component-at-a-time approach. Advances in 'omics' technologies such as whole-genome editing, functional genomics, proteomics, and bioinformatics have recently allowed assessing mitochondrial function by holistic and systems-wide research strategies that integrate more than one technique, model system and discipline (**Figure 3**). Several reviews have delved into specific applications of mitochondria systems biology for compiling mitochondrial parts lists, characterizing their tissue heterogeneity and evolutionary origins, and for identifying disease genes and pathomechanisms [5,12–19]. Here, we attempt to provide a comprehensive overview of the current state-of-the-art experimental and computational tools for studying human mitochondria.

2. *De novo* identification of mitochondrial proteins

An essential first step towards understanding mitochondrial functions is to achieve a complete knowledge of the mitochondrial proteome parts list, which is referred to as the subset of the whole-cell proteome localizing specifically to the organelle. To that end, top-down systems-level approaches have been instrumental for large-scale and unbiased prediction of mitochondrial proteins from different organs and organisms. Below, we present several key computational and experimental strategies that have catalyzed the identification of mitochondria-localized proteins.

2.1 *In silico strategies*

Several freely available and user-friendly databases can be queried for supporting evidence of a protein's mitochondrial localization (**Table 1**). The Mitochondria Protein Atlas [20], for example, provides manually curated and updated inventories of experimentally validated human mitochondrial proteins, including information on their sub-mitochondrial localization, function, structure, interactions, and involvement in human diseases. Besides repositories of known mitochondria-localized proteins, numerous *in silico* approaches have been developed to identify novel mitochondrial proteins based on the prediction of specific targeting signals for sorting proteins into the organelle, as well as a variety of complementary clues on protein primary and secondary structures, physicochemical properties, sequence motifs, and homology to proteins with a known mitochondrial localization in other species. Indeed, all nuclear-encoded mitochondrial proteins are directed to and imported into sub-mitochondrial compartments based on the recognition of mitochondrial targeting signals (MTS_s) by specific translocator complexes [21,22]. As an example, soluble matrix and inner membrane (IMM) proteins often contain within the first one hundred residues a presequence that is cleaved by mitochondrial peptidases for retention into the organelle or for membrane insertion. Many attempts have been made to predict mitochondrial protein localization by analyzing the biochemical properties of N-terminal MTS_s that usually exhibit biased amino acid composition, internal protease recognition sites, and show positively charged amphiphilicity. In a study by Vaca Jacome et *al.* [23], 356 proteins were found with a cleavable N-terminal

presequence by systematic trimethoxyphenyl phosphonium (TMPP)-based labelling of U937 human monocytic mitochondria coupled to liquid chromatography with tandem mass spectrometry (LC-MS/MS). In a subsequent survey, Calvo et *al.* [24] identified cleaved N-terminal presequences out of 327 mouse kidney and liver mitochondrial proteins through a subtiligase-based protein biotinylation approach. Although those N-proteome analyses of isolated mitochondrial fractions also included false positives due to, for example, co-purification of other organelles, they highlighted that at least 30% of the mammalian proteome might be targeted to mitochondria via cleavable N-terminal presequences. Besides, other mitochondrial import mechanisms also exist that involve the recognition of C-terminal or, rarely, internally located MTS_s, especially for outer membrane (OMM) or intermembrane space (IMS) proteins [21]. Thus, several computational approaches aiming at systematically identifying MTS_s over whole organism proteomes have been developed for the automated discovery of mitochondrial proteins (**Table 1**). The vast majority of available *in silico* tools employ supervised machine learning algorithms. The latter are trained to discriminate between mitochondrial and non-mitochondrial proteins using reference sets of true positive (proteins unambiguously localized to mitochondria) and true negative (proteins convincingly annotated to other subcellular compartments) for benchmarking. Most common models predict mitochondrial protein localization using as input either known biochemical features of MTS_s or directly the overall protein amino acid sequence. Among those, PSORTII [25], TargetP 2.0 [26], Predotar [27], MitoFates [28], and TPpred3 [29] focus on the identification of cleavable N-terminal presequences. Given that the biochemical properties of mitochondrial N-terminal presequences are well known, those models can reach a high sensitivity (true positive rate) but their predictive power remains limited, considering that not all mitochondrial proteins contain a MTS. Among overall sequence-based approaches (e.g., ngLOC [30], DeepLoc [31], LocTree3 [32], and CELLO II [33]), SubMitoPred [34] and DeepMito [35] also allow predicting the specific localization of a protein into sub-mitochondrial compartments (OMM, IMS, IMM, and matrix). Altogether, prediction accuracy greatly varies across different *in silico* tools, depending on machine learning algorithms,

training and testing datasets, and biological features input for learning, making it advisable to compare results from different queries.

2.2 Experimental strategies

Major advances in the sensitivity and throughput of mass spectrometry (MS) and imaging technologies coupled to genome editing have made possible to survey to systematically identifying novel mitochondrial proteins. One such endeavor is the Cell Atlas database, which is part of the Human Protein Atlas (HPA, www.proteinatlas.org) [36], an open-access resource including the annotation of expression and subcellular localization for 12,390 proteins across a panel of 26 human cell lines. By using fluorescence microscopy of native proteins with an immunologically detectable epitope, 1,098 proteins could be annotated as mitochondria-localized at high-resolution, of which 46% were independently validated by additional experimental strategies (e.g., gene silencing, fluorescent protein-tagging, different antibodies) or by evidence from external databases. Overall, immunofluorescence (IF)-based approaches offer the advantage to analyze subcellular and spatio-temporal protein distribution *in situ* and in single cells, thus also enabling the identification of cell-to-cell protein variability and multi-organelle localization. However, due to the lack of available antibodies for all human proteins, the number of truly localized mitochondrial proteins might be currently underestimated. As a complementary approach, organellar proteomics, using mitochondria-enriched fractions as input material for state-of-the-art MS-based analyses, have proved instrumental to nearly double the number of known yeast [12,37], mouse [6,38], and human mitochondrial proteins [39–41]. Several protocols are available to obtain highly pure mitochondrial preparations [42], which mainly differ in the methods used for selective disruption of the plasma membrane (e.g., sonication, mechanical homogenization, nitrogen cavitation) and organelle enrichment (e.g., differential centrifugation, high-affinity magnetic immunocapture followed by ultracentrifugation on a density gradient). In 2008, using Percoll density gradient purified mitochondria, Pagliarini et *al.* [6] performed a systematic and comprehensive survey of mitochondrial protein expression across 14 mouse tissues by

reversed-phase LC-MS/MS. Altogether, over 3,800 proteins were identified, with an average of 1,500 expressed in mitochondrial fractions from each tissue. Unexpectedly, at least 15% of the organelle parts list showed tissue specific expression, suggesting that while a core set of mitochondrial proteins perform ubiquitous tasks, the rest must fulfil the specific functional and metabolic requirements of each organ. Afterwards, studies from Fecher et *al.* [7] and Bayraktar et *al.* [43] demonstrated that such diversity across tissues does not simply derive from global changes in bulk mitochondrial proteomes but reflects cell type-specific heterogeneity in mitochondrial protein sets from the same organ. By engineering a reporter mouse line, named MitoTag, that expresses an OMM-targeted green fluorescent protein (GFP) in a Cre recombinase-dependent manner, Fecher et *al.* systematically dissected mitochondrial proteome variability among three different cell types of the cerebellum. Here, the GFP-OMM epitope was used as a handle for immunocapturing tagged mitochondria directly from major inhibitory (PC, Purkinje cells) and excitatory (GC, granule cells) neurons, and astrocytes within their tissue context. Strikingly, comparative LC-MS/MS analysis of cell type-specific mitochondrial fractions showed that only about 85% of the identified proteins were shared among PC, GC and astrocytes. The rest reflected differentially regulated mitochondrial pathways and functions, providing a set of markers for monitoring cell type-specific mitochondrial changes in healthy and diseased mouse and human brains. Using a similar experimental strategy, Bayraktar et *al.* profiled both proteome and metabolome of hepatocyte mitochondria within liver tissue. A reporter mouse model (also called MITO-Tag) that expressed an OMM-targeted HA epitope tag under the control of the Albumin promoter was exploited to specifically and rapidly immunocapture mitochondria (10 min after tissue homogenization). A total of 511 proteins and a variety of hepatocyte metabolites were found to be highly enriched in mitochondria compared to whole-liver proteome and metabolome. Altogether, results from both studies highlighted the utility of Mito-Tag mouse models as tools for characterizing the mitochondrial system *in vivo* and upon physiological and pathological perturbations. Furthermore, Mito-Tag mice can be employed to isolate mitochondria from virtually any cell type, without the need for cell sorting and lengthy

purification protocols. However, they require the lysis of tissue and mitochondrial samples, which inevitably results in substantial distortion of the *in vivo* mitochondrial physiological state, compared for example to microscopy-based analyses. Nevertheless, both approaches are unable to survey dynamic changes in protein composition and distribution *in situ* and simultaneously for all mitochondrial proteins, at either tissue or cell type-specific levels. To this goal, synthetic biology strategies have recently opened the way for spatially and temporally resolved snapshots of mitochondrial proteomes within living cells by combining the strengths of microscopy and MS technologies [44]. A series of studies from the Ting group profiled the composition of individual mitochondrial sub-compartments by targeting the ascorbate peroxidase APEX to either the matrix [45], IMS [46] or OMM [47] of human embryonic kidney (HEK) cells. In presence of biotin-phenol and H_2O_2, APEX catalizes within 1 min the generation of phenoxyl radicals that can covalently react to electron-rich amino acids while the cell is still intact. Those radicals are short-lived, highly reactive, membrane-impermeant, and have a small labeling radius, leading to high spatial resolution. Biotin-labelled proteins are then recovered by streptavidin-based enrichment on cell extracts and identified by tandem MS-based proteomics. This approach was especially instrumental for mapping the IMS proteome, which cannot be otherwise characterized by traditional biochemical approaches based on density centrifugation. As a result, the Ting group identified a total of 495, 127, and 137 matrix, IMS, and OMM proteins, respectively, of which roughly half had previously unknown sub-mitochondrial localization, providing a rich resource of orphan proteins and proteins without a previous functional link to mitochondria. However, the approach has been only validated in cells and organs *ex vivo* [48,49], questioning its utility for *in vivo* tagging of mitochondrial proteomes.

2.3 Integrative biology strategies

The aforementioned genome and proteome-scale approaches in biochemistry, genetics, imaging and bioinformatics have undoubtedly led to the identification of novel human mitochondrial proteins. However, their predictive power was hampered by limited specificity

and coverage, as each method suffered from intrinsic methodological limitations and was biased towards different subsets of mitochondrial proteins. As an example, nowadays MS-based proteomics can quantify over 6000 proteins from nanograms of a whole-cell extract [50,51]. If on the one side the ever-increasing resolution and detection limits of mass spectrometers make deep organellar proteome analysis an extremely powerful discovery tool, on the other discriminating between true mitochondria-resident and contaminant proteins that either co-sediment or are physically interacting with the organelle remains challenging. Unfortunately, minimizing contamination by maximizing the purity of mitochondrial preparations can only offer a partial solution at the risk of compromising organelle's integrity. On the contrary, computational searches for cleavable, N-terminal targeting signals yield highly specific lists of mitochondrial proteins but show limited sensitivity, given that not all mitochondria-resident proteins contain a pre-sequence. Furthermore, experimental and *in silico*-derived catalogs of mitochondrial proteins have shown modest overlap [13,17]. Therefore, it is plausible that the systematic integration of all data types could compensate for the shortcomings of each individual approach and increase true positive rate. The latter hypothesis was tested in the years 2004 [37] and 2006 [52] by the Steinmetz group to predict mitochondrial proteins in *S. cerevisiae*. In two consecutive studies, the authors performed machine-learning based integration of over twenty computational and experimental genome-wide datasets, interrogating different biological properties of the mitochondrial system, from evolutionary conservation and gene regulation, to protein abundance and physical protein-protein interactions (PPI). The model was trained to discriminate between a positive reference set of known yeast mitochondrial proteins and the remaining yeast proteome in order to rank all datasets according to their power in identifying true mitochondria-localized proteins. Among all, IF-based analyses of sub-cellular localization by protein-tagging, MS on isolated mitochondria, and orthology mapping to known mitochondrial proteins in other species showed the highest sensitivity and specificity, in contrast to transcriptome analysis. As expected, data integration outperformed the predictive power of each individual dataset and yielded a comprehensive and accurate

inventory of the yeast mitochondrial proteome, including 91% of the reference set and an additional 346 candidates, of which nearly half were still uncharacterized [52]. Later on, the same approaches have been applied to predict new mitochondrial proteins in mammals. For example, Pagliarini et *al.* [6] used a naïve Bayes algorithm to combine six genome-scale datasets of mitochondrial localization with MS-based analyses of mitochondria isolated from 14 mouse tissues and large-scale IF experiments of GFP-tagged proteins tagging to compute a likelihood score of mitochondrial localization for each mouse protein. At a cut-off of 10% false discovery rate (FDR), the resulting inventory of 1098 mouse proteins, termed MitoCarta v1.0, was estimated to be 85% complete and containing nearly 300 genes without previous mitochondrial annotation in the Gene Ontology (GO) database. A human MitoCarta of 1013 proteins was then generated based on sequence homology and updated twice (MitoCarta v2.0 [53] and v3.0 [54]) to a final list of to 1136 human mitochondrial proteins by manual literature curation, including sub-mitochondrial compartment and pathways annotations. Using a different machine-learning method (support vector machines), Smith et *al.* [55] developed an integrated mitochondrial protein index (IMPI) from the integration of 56 datasets from a variety of resources, including MitoCarta v2.0, MS of purified cell fractions, GFP-tagging and microscopy, and computational prediction of MTS_s. The resulting database, named MitoMiner (IMPI version Q2 2018) [55], contains 1626 human genes that encode for mitochondrial proteins, 442 are novel candidates, and roughly two-third overlap with MitoCarta v2.0. Overall, both MitoCarta and MitoMiner provide the most specific and comprehensive catalogs of mammalian mitochondrial proteins to date and represent valuable platforms to investigate tissue-specific expression of mitochondrial proteins and their links to disease.

Integrative biology approaches have also proved successful in unraveling the genetic identity of whole mitochondrial complexes, as exemplified by the discovery of the mitochondrial calcium uniporter components [56] (**Figure 4A**). Ca^{2+} handling by mitochondria was first described more than 50 years ago, when DeLuca & Engstrom [57] and Vasington & Murphy [58] independently reported that mitochondria isolated from rat

kidney could rapidly uptake millimolar amounts of Ca^{2+} from the extra-mitochondrial space by passive Ca^{2+} transport down its electrochemical gradient. Mitochondrial Ca^{2+} uptake was attributed to a Ca^{2+} uniporter located in the IMM, which was later shown to be a highly selective Ca^{2+} channel [59]. The mitochondrial calcium uniporter channel, here referred to as MCUC, was dependent on the pH gradient and the negative potential established by the respiratory chain across the inner mitochondrial membrane and inhibited by nanomolar concentrations of ruthenium red and its derivative Ru360 [60]. However, despite the biophysical properties of mitochondrial Ca^{2+} uptake have been characterized for decades, the genetic identity of MCUC has evaded traditional biochemical strategies that aimed at purifying it from animal tissues, as well as genome-wide RNA interference (RNAi)-based loss of function screens. Interestingly, besides rat kidney, MCUC-mediated Ca^{2+} uptake was measured in mitochondria from virtually all mammalian tissues and in several species from other kingdoms (e.g., protozoa and plants), yet, despite rigorous, repeated attempts, it was never observed in mitochondria from *S. cerevisiae* [61]. Those evidence were used by Perocchi et *al.* [62] to define a “physiological signature” of MCUC, being high capacity, located to the IMM, powered by the mitochondrial membrane potential, and inhibited by ruthenium red. The authors hypothesized that human genes encoding for MCUC should exhibit a “phylogenetic signature” matching the physiological profile across taxa, namely present in vertebrates and kinetoplastids, but absent in yeast. By integrating clues from comparative physiology, evolutionary genomics and MitoCarta, they predicted 18 human proteins fulfilling the above criteria. RNAi against these top MCUC candidates identified a previously uncharacterized gene, CBARA1, as the first uniporter component, which was renamed Mitochondrial Calcium Uptake 1 (MICU1). Silencing *MICU1* abolished mitochondrial Ca^{2+} uptake in human cells and mouse liver and conserved calcium-binding, EF-hands domains were found to be essential for its activity, suggesting that the protein could act as a Ca^{2+} sensing regulatory subunit of MCUC and opening the way to the full molecular characterization of the uniporter channel.

3. From proteins to functions: deorphanizing the unknown

Approximately 20-40% of predicted human mitochondrial proteins still remain functionally uncharacterized to date, namely, no information can be found in PubMed or in other manually curated data sources and no associations are available from experimental evidence in the gene ontology (GO) database [14,19,63]. Moreover, a number of true mitochondrial proteins may have been overlooked, due to the complex and dynamic nature of the mitochondrial proteome, with more than 200 dually-localized components and roughly one-third of the proteome distributed in a tissue-specific manner. To fill this gap, several computational and experimental approaches have been developed to predict protein-to-protein and gene-to-phenotype associations based on the reconstruction of biological networks, whereby functional inference can be obtained for known and uncharacterized proteins that are closely linked.

3.1 *In silico strategies*

Clues about the function of an orphan protein are often inferred based on the “guilt-by-association” principle: the underlying assumption is that a target protein is likely to play a functional role in the same biological process of its neighbors. Experimental, computational, or combined experimental/computational analyses have been employed to map functional associations, including direct physical links between proteins in a complex, interactions between enzymes in a pathway, or regulatory networks between transcription factors and their target genes. One of the most comprehensive and unbiased source of functional associations is the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database [64], which currently includes over 2000 million interactions between 24.6 million proteins across more than 5000 organisms. STRING enables predicting PPI through machine learning-based integration of complementary evidence types, such as correlated gene expression profiles (co-expression) across a large set of samples and conditions, shared patterns of presence and absence of homologs in different species (co-evolution), co-occurrence in automated text mining analyses, physical interactions in large-scale high

throughput affinity purification screens, and previous knowledge in manually curated databases. Functional links are further inferred through homology transfer across thousands of organisms. To this goal, *S. cerevisiae* has served as the model organism of choice for global mapping of protein networks, largely because of the abundance and quality of diverse high-throughput PPI datasets. Yeast two-hybrid screens, tandem affinity purifications, and protein-fragment complementation assays resulted in an *en masse* detection of binary PPI and stable physical protein complexes for over 70% of the yeast proteome and were used to generate the first global model of a mitochondrial protein network [52]. Here, the Steinmetz group employed STRING to place roughly 90% of yeast mitochondrial proteins in a functional context through the systematic retrieval of over 9000 linkages. Hierarchical average-linkage clustering based on confidence scoring of network associations yielded 164 functionally distinct modules, containing known and orphans mitochondrial proteins, as well as extra-mitochondrial proteins not physically localized to the organelle but still critical to its function. This module map was shown to be more accurate and comprehensive than publically available annotation of protein complexes and metabolic pathways based on a single species and data type. Importantly, it provided the first clues to the function of over 150 uncharacterized yeast mitochondrial components and allowed surveying properties of the mitochondrial system that would not be easily deduced from its parts list. As an example, by overlaying onto the module map genome-wide expression profiles and mutant growth phenotypes under fermentable and non-fermentable conditions, the authors could spotlight several module-to-phenotype correlations leading to the identification of novel functional groups involved in oxidative metabolism and cell viability. Importantly, most yeast mitochondrial modules were highly conserved to humans and five were enriched in disease genes, allowing investigating disease susceptibility of mitochondrial functional processes and prioritizing candidate genes for putative mitochondrial disorders [17]. A decade later, Yim et *al.* [65] extended similar integrative strategies to generate global mitochondrial interactomes for different species based on manual curation of annotated functional associations and PPI. As a result, the authors developed MitoXplorer, a gene expression

and mutant data mining platform that users can query to analyze how mitochondrial networks remodel in response to various experimental and disease conditions. Likely owing to the mosaic evolutionary origin of the mitochondrial proteome, prediction of functional associations based on comparative genomics has proved instrumental for the *de novo* genetic underpinning of several mitochondrial processes, the deorphanization of mitochondrial proteins, as well as the identification of novel components of mitochondrial protein complexes and pathways. The study from Gabaldon et *al.* [66] on the evolutionary history of the NADH:ubiquinone oxidoreductase (Complex I) represented one of the first examples of how evolutionary genomics and phylogenetic analyses could be exploited to predict previously unidentified subunits of mitochondrial protein complexes. By charting the distribution of mammalian Complex I subunit orthologs across 17 eukaryotic species from various phylogenetic groups, Gabaldón et *al.* identified NDUFAF2 as a previously unknown component of Complex I and disease gene in human progressive encephalopathy. Three years later, the Mootha group applied a similar approach to profile the phylogenetic distribution of all MitoCarta protein orthologs across 42 sequenced eukaryotes, leading to the prediction of novel candidates for Complex I assembly and inherited Complex I deficiency in humans [6]. In 2011, two studies provided a prime example of the power of combining comparative genomics with *in silico* predictions of functional associations to deorphanize mitochondrial proteins, leading to the discovery of the pore-forming, and Ca^{2+}-conducting subunit of MCUC [67,68]. Building up on the discovery of MICU1, Baughman et *al.* [67] searched for MICU1 co-evolving and co-expressed genes across 500 organisms and 81 mouse cell types and tissues, respectively, whose proteins would be equally abundant in mitochondria of 14 mouse tissues. Independently, De Stefani et *al.* [68] looked for MitoCarta proteins with at least two predicted transmembrane domains, ubiquitously expressed in mammalia tissues, conserved in kinetoplastids and lacking orthologs in *S. cerevisiae*. Both studies spotlighted a poorly characterized protein, CCDC109A, which also co-immunoprecipitated with MICU1 in human cells. Silencing of CCDC109A abrogated mitochondrial Ca^{2+} uptake, whereas its overexpression enhanced ruthenium red-sensitive

Ca^{2+} influx, reduced cytosolic Ca^{2+} transients, and sensitized cells to apoptotic challenges. In addition, highly conserved residues within the putative pore-forming domain linking the two transmembrane regions were found necessary for Ca^{2+} permeation and for conferring sensitivity to ruthenium red. Finally, its expression in planar lipid bilayers was sufficient to reconstitute ion channel activity in solutions containing only Ca^{2+} as the permeant ion, confirming the role of CCDC109A as the MCUC pore-forming component and thereof renamed MCU (Mitochondrial Calcium Uniporter).

Currently, several computational tools are available for the systematic prediction of protein function based on phylogenetic profiling, likely owing to substantial progress in sequencing and annotation of whole genomes. One such a discovery tool, named ProtPhylo (www.protphylo.org), was generated by Cheng et *al.* [69] to identify protein-to-protein and phenotype-to-protein functional associations. Here, phylogenetic profiles were computed for over 9.7 million non-redundant protein sequences across 2048 organisms in all three domains of life (1678 bacteria, 115 archaea and 255 eukaryotes) by using multiple orthology inference algorithms. ProtPhylo then rank all proteins within the organism of interest based on the similarity of their phylogenetic profiles to the query protein and allows user to prioritize the number of testable hypotheses based on complementary evidence of PPI, subcellular localization, protein domains and membrane spanning regions. Considering the ever-increasing amount of large-scale, unbiased datasets surveying the mitochondrial system, databases such as STRING and novel user-friendly computational platforms hold the promise to increase both accuracy and coverage of protein networks for the automated prediction of mitochondrial protein function.

3.2 Experimental strategies

Being able to confidently assign proteins to annotated complexes and biological processes can provide clues to their function. However, only 27% and 56% of the human mitochondrial proteome can be functionally linked in either macromolecular assemblies or molecular pathways based on automated data retrieval and manual curation of literature by CORUM

3.0 (Comprehensive Resource of Mammalian Protein Complexes) [70] and KEGG (Kyoto Encyclopedia of Genes and Genomes) [71] databases, respectively. Therefore, it is of outmost importance to develop experimental strategies to generate high-coverage functional networks of predicted mitochondrial parts lists, e.g. by screening for physical PPI and genetic interactions. Most of the experimental analyses performed so far have defined stable and transient PPI between mitochondrial proteins of interest using purified mitochondrial samples as starting material for either blue native polyacrylamide gel electrophoresis (BN-PAGE) or affinity purification in combination with western blotting and MS. Only a few studies, though, have applied those techniques for a systematic and unbiased reconstruction of mitochondrial interaction networks (**Figure 4B**). One such example is the complexome profiling performed by Heide et *al.* [72] on intact rat heart mitochondria. BN-PAGE was employed as a mild, nondestructive method to separate native, soluble and membrane-bound complexes up to 30 MDa from enriched heart mitochondrial fractions based on charge, mass, and structure. A total of 464 proteins were identified by LC-MS/MS analysis of 60 even gel slices, of which 85% were previously known to be mitochondria-localized. Based on the relative abundance and specific migration profile in different gel slices, all identified proteins were hierarchically clustered to define the molecular composition of several protein assemblies, including very large and abundant multiprotein complexes and assembly intermediates, such as those of the oxidative phosphorylation (OXPHOS) system. As a result, an orphan protein, TMEM126B, was identified as a novel subunit of the mitochondrial Complex I assembly factor complex, providing a proof-of-principle to the direct application of complexome profiling for protein function prediction. Interestingly, several proteins were detected in a number of different slices, likely representing transient intermediates of macromolecular complexes and reflecting the ability of such an approach to shed light not only on the protein composition of the final stable complex but also onto the dynamics of its assembly.

A further step towards a comprehensive and accurate characterization of mitochondrial protein complexes was made possible by combining cross-linking mass spectrometry (XL-

MS) with organelle-wide analyses. In XL-MS, small organic molecules containing two reactive groups at either end of a spacer arm (cross-linkers) are employed to chemically join functional groups of specific amino acids by a covalent bond. After cleavage by tryptic digestion and MS analysis, residue-to-residue cross-links are identified by peptide sequencing. Therefore, besides protein complex composition, XL-MS also enables locating native inter and intra-molecular contacts between proteins. Moreover, owing to the constraints introduced by the length of the spacer arm, architectures and conformations of protein complexes can also be probed, a task that is otherwise experimentally challenging given that roughly a third of the mitochondrial proteome is assembled into macromolecular membrane complexes. Recently, Liu et *al.* [73] applied XL-MS to intact mouse heart mitochondria, mapping over 3,000 unique connections between 359 mitochondrial proteins in MitoCarta v2.0. Notably, 60% of the detected cross-links were formed between distinct proteins, reflecting the high protein density environment of intact mitochondria. Overall, the interactome showed high sensitivity, covering 75% of annotated mitochondrial protein complexes in CORUM, and provided *in situ* evidence for the assembly of OXPHOS components in super complexes. Moreover, contrary to the analysis of mitochondrial complexosomes by BN-PAGE, XL-MS of intact mitochondria allowed to probe the spatial distribution of protein interactions in all sub-mitochondrial compartments, thanks to the use of membrane permeable cross-links. However, organelle-wide XL-MS approaches still suffers from limited depth, as the captured cross-links mostly involve highly abundant proteins. Other conceptually similar techniques based on the distance proximity of proteins in situ have recently provided deep coverage of mitochondrial interactomes within the cellular context. As an example, Antonicka et *al.* [74] used the proximity-dependent biotinylation assay, BioID [75], to map potential interactors of 100 mitochondrial protein baits in HEK-293 cells. Baits, including both known and poorly characterized mitochondrial components and spanning all mitochondrial sub-compartments, were fused to a mutant *Escherichia coli* biotin ligase $BirA^{R118G}$ (known as BioID), which biotinylates neighboring proteins (preys) within an estimated radius of 10 nm. MS-based detection of biotinylated

proteins from each bait and of background labeling by matrix and IMS-targeted BioIDs was then employed to identify high-confidence, specific proximity interactions. As a result, the authors defined a mitochondrial network of over 15,000 unique associations between 100 baits and 1465 enriched preys. Notably, of those, only roughly 50% were annotated as mitochondria-localized by Mitocarta v2.0. Most of the non-mitochondrial preys were proximal to OMM-located baits facing the cytosolic environment, consisting of proteins annotated as having multiple localizations in the HPA database and likely involved in interoganelle contact sites. Next, clustering based on correlated patterns of connectivity across all baits predicted several distinct modules for the matrix, IMS/IMM and OMM sub-compartments, with clear functional annotations based on GO terms. Reassuringly, those included known mitochondrial protein complexes such as the small and large mitochondrial ribosomes and assembly intermediates of the OXPHOS system, as well as proteins involved in mitochondrial fusion and fission. Importantly, several proteins of unknown function were also included in the modules map, validating the potential of this resource for the functional characterization of orphan mitochondrial proteins. Furthermore, clustering of non-mitochondrial preys across OMM baits showed that distinct subsets of mitochondrial proteins might mediate the specific cross-talk of mitochondria with other organelles, such as the ER and the peroxisome. Two studies have further explored proximity labeling to specifically map ER-mitochondria contact sites in HEK-293T cells using modified, “split” versions of either BioID or its derivative TurboID, which requires only 1 to 10 min of labeling time (compared to 16 hours for BioID-based labeling). In both systems, named Contact-ID [76] and Split-TurboID [77], the enzyme is split into two inactive fragments, each containing half of the FRB-FKBP dimerization system and facing the cytosol from either the ER or the OMM. When brought to proximity by a PPI or, as in the following examples, by organelle membrane-membrane apposition, and in presence of rapamycin, a functional biotin ligase is reconstituted that upon biotin addition catalyzes spatially restricted biotinylation. Altogether, Contact-ID and Split-TurboID-based proteomic profiling of ER-mitochondria contacts identified roughly one-hundred proteins, including annotated components of mitochondrial-

associated membranes (MAMs), as well as proteins with known OMM or ER membrane localization. The candidate lists were enriched in biological processes that are consistent with previously reported regulatory roles of ER-mitochondria contacts in Ca^{2+} signaling, sterol metabolism, and mitochondrial fission. Among many novel ER and OMM-localized proteins and ER-mitochondria contact site candidates, several were also experimentally validated. Unexpectedly, although both Contact-ID and Split-TurboID-derived datasets showed high specificity when compared to lists of known MAMs and based on GO term enrichment analysis, they only shared 15% of the proteins. This modest overlap is likely arising from differences in construct design and labeling time, resulting in the biotinylation of proteins at closer or wider distance between the two organelles. Nevertheless, both approaches validated proximity labelling as a valuable tool for profiling not only functional relationships within an organelle but also with other cellular structures.

Besides PPI profiling, genome-wide loss-of-function screens upon genetic as well as environmental perturbations have also proved effective in characterizing the biological role of mitochondrial proteins (**Figure 4C**). As an example, given the dual genetic origin of the mitochondrial proteome, understanding the molecular basis of mitonuclear signaling pathways regulating mitochondrial processes (e.g., OXPHOS assembly) and how those remodel to buffer environmental changes remains an outstanding question of great translation value. To define an integrated mitochondrial stress footprint, Quirós et *al.* [78] applied a multiomics approach in HeLa cells challenged by four different stressors, doxycycline, actinonin, FCCP, and MitoBlock-6, which impair mitochondrial translation, OXPHOS proteins stability, mitochondrial membrane potential, and protein import, respectively. Notably, they found that the mitochondrial unfolded protein response, one of the best-characterized retrograde stress responses in invertebrates, was not activated under the conditions used. Instead, all stressors induced a pronounced decrease of mitochondrial ribosomal proteins, OXPHOS components, and epigenetic regulators, that was substantially dependent from changes at proteome rather than the transcriptome level. At the same time, the expression of genes involved in the biosynthesis of amino acid, in particular of serine,

and carbon metabolism was up-regulated, suggesting the activation of alternative cataplerotic pathways to convert and replenish tricarboxylic acid cycle intermediates for the synthesis of glycolytic intermediates. These findings were consistent with stress-induced changes at the metabolome level, and altogether highlighted a compensatory response aimed at rewiring cellular metabolism and preventing oxidative damage through the synthesis of key metabolites and lipids. *De novo* motif analysis in the common upregulated gene sets showed that half of the co-regulated genes were targets of the transcription factor ATF4, a key component of the cellular integrated stress response (ISR), whose transcript and protein levels were also found to be upregulated upon stress induction. Based on these results, ATF4 was proposed to act as the main molecular effector of the mammalian mitonuclear response and mitochondrial stress signature, by inducing the expression of cytoprotective genes while attenuating global translation. However, the precise mitochondrial signaling pathway relaying mitochondrial stress and malfunction to the nucleus remained unaddressed. Further insights were gained by performing a genome-wide random mutagenesis screen on haploid HAP1 cells that expressed the C/EBP homologous protein CHOP as a fusion protein with mNeon and were challenged with three distinct mitochondrial stressors, tunicamycin, 2-cyano-3,12-dioxo-oleana-1,9(11)-dien-28-oic acid (CDDO), or Carbonyl cyanide m-chlorophenyl hydrazone (CCCP) [79]. CHOP is a transcription factor of the ISR, whose activation by ATF4 acts as a cellular checkpoint through the initiation of apoptotic and non-apoptotic cell death programs. By sequencing mutations in cell populations with either an enhanced or diminished CHOP expression in all conditions, Fessler et *al.* [79] identified a poorly characterized mitochondrial protein, DELE1, as a global positive regulator of CHOP induction, and thus of ISR. Follow-up experiments clearly validated DELE1 as a hit and demonstrated the existence of a proteolytic signaling axis, whereby mitochondrial stress would induce proteolysis of DELE1 by OMA1 with the consequent release of the processed form of DELE1 to the cytosol and interaction with components of the ISR pathway.

4. Perspectives

The identification and functional characterization of the mitochondrial system represents an extraordinarily important milestone for mitochondrial biology and human diseases. Although major progress has been made in compiling accurate and exhaustive lists of mitochondria-localized proteins, still after a century from the discovery of mitochondria, 20-40% of the system remains functionally orphan, even in well-studied model organisms such as budding yeast. Furthermore, mitochondrial protein networks mediating signal transduction in and out of the organelle are almost completely uncharted. This gap greatly hampers the understanding, diagnosis and treatment of mitochondria-related pathologies, given that mitochondrial dysfunction has been linked to an extremely wide spectrum of disease phenotypes, including the decreased activity of metabolic enzymes, impaired respiratory capacity, and increased oxidative damage. Corroborating this notion over 300 genes encoding for mitochondrial proteins have already been implicated in a wide range of human diseases (as in the Online Mendelian Inheritance in Man, OMIM, database), from metabolic syndrome-related disorders (e.g., diabetes and obesity) to neurodegenerative diseases (e.g., Parkinson‘s and Alzheimer‘s diseases). How can impairment on ubiquitous mitochondria-mediated processes result in such diverse disease outcomes? The answer lies in the tissue and cell type-specificity of mitochondrial proteomes, networks, and intracellular cross-talks. To this goal, all recently developed systems-wide approaches to investigate mitochondrial biology - and more yet to come – hold the promise to aid in the identification of potential candidates for mitochondrial-related diseases, the understanding of disease pathomechanisms and possible new pharmacological targets.

Author Statement

Hilda Carolina Delgado de la Herran: Conceptualization, Writing – Original Draft, Writing – Review & Editing, Visualization. **Yiming Cheng:** Formal Analysis. **Fabiana Perocchi:** Conceptualization, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition.

Competing interests The authors declare no competing interests.

Acknowledgments

FP, HCD, and YC acknowledge support from the Munich Center for Systems Neurology (SyNergy EXC 2145 /ID 390857198) and the ExNet-0041-Phase2-3 („SyNergy-HMGU“) through the Initiative and Network Fund of the Helmholtz Association. The authors would like to thank Dr. Arduino for help on figure design.

References

[1] A.A. Pittis, T. Gabaldón, Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry, Nature. 531 (2016) 101–104. https://doi.org/10.1038/nature16941.

[2] A.J. Roger, S.A. Muñoz-Gómez, R. Kamikawa, The Origin and Diversification of Mitochondria, Curr Biol. 27 (2017) R1177–R1192. https://doi.org/10.1016/j.cub.2017.09.015.

[3] M.F. Lopez, B.S. Kristal, E. Chernokalskaya, A. Lazarev, A.I. Shestopalov, A. Bogdanova, M. Robinson, High-throughput profiling of the mitochondrial proteome using affinity fractionation and automation, Electrophoresis. 21 (2000) 3427–3440. https://doi.org/10.1002/1522-2683(20001001)21:16<3427::AID-ELPS3427>3.0.CO;2-L.

[4] S.G.E. Andersson, A. Zomorodipour, J.O. Andersson, T. Sicheritz-Pontén, U.C.M. Alsmark, R.M. Podowski, A.K. Näslund, A.-S. Eriksson, H.H. Winkler, C.G. Kurland, The genome sequence of Rickettsia prowazekii and the origin of mitochondria, Nature. 396 (1998) 133–140. https://doi.org/10.1038/24094.

[5] J. Nunnari, A. Suomalainen, Mitochondria: In Sickness and in Health, Cell. 148 (2012) 1145–1159. https://doi.org/10.1016/j.cell.2012.02.035.

[6] D.J. Pagliarini, S.E. Calvo, B. Chang, S.A. Sheth, S.B. Vafai, S.-E. Ong, G.A. Walford, C. Sugiana, A. Boneh, W.K. Chen, D.E. Hill, M. Vidal, J.G. Evans, D.R. Thorburn, S.A. Carr, V.K. Mootha, A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology, Cell. 134 (2008) 112–123. https://doi.org/10.1016/j.cell.2008.06.016.

[7] C. Fecher, L. Trovò, S.A. Müller, N. Snaidero, J. Wettmarshausen, S. Heink, O. Ortiz, I. Wagner, R. Kühn, J. Hartmann, R.M. Karl, A. Konnerth, T. Korn, W. Wurst, D. Merkler, S.F. Lichtenthaler, F. Perocchi, T. Misgeld, Cell-type-specific profiling of brain mitochondria reveals functional and molecular diversity, Nature Neuroscience. 22 (2019) 1731–1742. https://doi.org/10.1038/s41593-019-0479-z.

[8] L. Ellenrieder, H. Rampelt, T. Becker, Connection of Protein Transport and Organelle Contact Sites in Mitochondria, Journal of Molecular Biology. 429 (2017) 2148–2160. https://doi.org/10.1016/j.jmb.2017.05.023.

[9] B. Glancy, Visualizing Mitochondrial Form and Function within the Cell, Trends Mol Med. 26 (2020) 58–70. https://doi.org/10.1016/j.molmed.2019.09.009.

[10] M.J. Devine, J.T. Kittler, Mitochondria at the neuronal presynapse in health and disease, Nature Reviews Neuroscience. 19 (2018) 63–80. https://doi.org/10.1038/nrn.2017.170.

[11] M. Rossini, P. Pizzo, R. Filadi, Better to keep in touch: investigating inter-organelle cross-talk, FEBS J. (2020). https://doi.org/10.1111/febs.15451.

[12] H. Gonczarowska-Jorge, R.P. Zahedi, A. Sickmann, The proteome of baker's yeast mitochondria, Mitochondrion. 33 (2017) 15–21. https://doi.org/10.1016/j.mito.2016.08.007.

[13] S.E. Calvo, V.K. Mootha, The Mitochondrial Proteome and Human Disease, Annual Review of Genomics and Human Genetics. 11 (2010) 25–44. https://doi.org/10.1146/annurev-genom-082509-141720.

[14] A.Y. Sung, B.J. Floyd, D.J. Pagliarini, Systems Biochemistry Approaches to Defining Mitochondrial Protein Function, Cell Metabolism. 31 (2020) 669–678. https://doi.org/10.1016/j.cmet.2020.03.011.

[15] J. Rahman, S. Rahman, Mitochondrial medicine in the omics era, The Lancet. 391 (2018) 2560–2574. https://doi.org/10.1016/S0140-6736(18)30727-X.

[16] X. Hu, Y.-M. Go, D.P. Jones, Omics Integration for Mitochondria Systems Biology, Antioxidants & Redox Signaling. 32 (2019) 853–872. https://doi.org/10.1089/ars.2019.8006.

[17] F. Perocchi, E. Mancera, L.M. Steinmetz, Systematic screens for human disease genes, from yeast to human and back, Mol Biosyst. 4 (2008) 18–29. https://doi.org/10.1039/b709494a.

[18] R.S. Aiyar, J. Gagneur, L.M. Steinmetz, Identification of mitochondrial disease genes through integrative analysis of multiple datasets, Methods. 46 (2008) 248–255. https://doi.org/10.1016/j.ymeth.2008.10.002.

[19] R.H. Malty, M. Jessulat, K. Jin, G. Musso, J. Vlasblom, S. Phanse, Z. Zhang, M. Babu, Mitochondrial Targets for Pharmacological Intervention in Human Disease, J. Proteome Res. 14 (2015) 5–21. https://doi.org/10.1021/pr500813f.

[20] M. Uhlén, E. Björling, C. Agaton, C.A.-K. Szigyarto, B. Amini, E. Andersen, A.-C. Andersson, P. Angelidou, A. Asplund, C. Asplund, L. Berglund, K. Bergström, H. Brumer, D. Cerjan, M. Ekström, A. Elobeid, C. Eriksson, L. Fagerberg, R. Falk, J. Fall, M. Forsberg, M.G. Björklund, K. Gumbel, A. Halimi, I. Hallin, C. Hamsten, M. Hansson, M. Hedhammar, G. Hercules, C. Kampf, K. Larsson, M. Lindskog, W. Lodewyckx, J. Lund, J. Lundeberg, K. Magnusson, E. Malm, P. Nilsson, J. Odling, P. Oksvold, I. Olsson, E. Oster, J. Ottosson, L. Paavilainen, A. Persson, R. Rimini, J. Rockberg, M. Runeson, A. Sivertsson, A. Sköllermo, J. Steen, M. Stenvall, F. Sterky, S. Strömberg, M. Sundberg, H. Tegel, S. Tourle, E. Wahlund, A. Waldén, J. Wan, H. Wernérus, J. Westberg, K. Wester, U. Wrethagen, L.L. Xu, S. Hober, F. Pontén, A human protein atlas for normal and cancer tissues based on antibody proteomics, Mol Cell Proteomics. 4 (2005) 1920–1932. https://doi.org/10.1074/mcp.M500279-MCP200.

[21] The versatility of the mitochondrial presequence processing machinery: cleavage, quality control and turnover | SpringerLink, (n.d.). https://link.springer.com/article/10.1007%2Fs00441-016-2492-9 (accessed November 18, 2020).

[22] A. Chacinska, C.M. Koehler, D. Milenkovic, T. Lithgow, N. Pfanner, Importing mitochondrial proteins: machineries and mechanisms, Cell. 138 (2009) 628–644. https://doi.org/10.1016/j.cell.2009.08.005.

[23] A.S. Vaca Jacome, T. Rabilloud, C. Schaeffer-Reiss, M. Rompais, D. Ayoub, L. Lane, A. Bairoch, A. Van Dorsselaer, C. Carapito, N-terminome analysis of the human mitochondrial proteome, PROTEOMICS. 15 (2015) 2519–2524. https://doi.org/10.1002/pmic.201400617.

[24] S.E. Calvo, O. Julien, K.R. Clauser, H. Shen, K.J. Kamer, J.A. Wells, V.K. Mootha, Comparative Analysis of Mitochondrial N-Termini from Mouse, Human, and Yeast, Molecular & Cellular Proteomics. 16 (2017) 512–523. https://doi.org/10.1074/mcp.M116.063818.

[25] P. Horton, K. Nakai, Better prediction of protein cellular localization sites with the k nearest neighbors classifier, Proc Int Conf Intell Syst Mol Biol. 5 (1997) 147–152.

[26] J.J.A. Armenteros, M. Salvatore, O. Emanuelsson, O. Winther, G. von Heijne, A. Elofsson, H. Nielsen, Detecting sequence signals in targeting peptides using deep learning, Life Science Alliance. 2 (2019). https://doi.org/10.26508/lsa.201900429.

[27] I. Small, N. Peeters, F. Legeai, C. Lurin, Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences, PROTEOMICS. 4 (2004) 1581–1590. https://doi.org/10.1002/pmic.200300776.

[28] Y. Fukasawa, J. Tsuji, S.-C. Fu, K. Tomii, P. Horton, K. Imai, MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites, Molecular & Cellular Proteomics. 14 (2015) 1113–1126. https://doi.org/10.1074/mcp.M114.043083.

[29] C. Savojardo, P.L. Martelli, P. Fariselli, R. Casadio, TPpred3 detects and discriminates mitochondrial and chloroplastic targeting peptides in eukaryotic proteins, Bioinformatics. 31 (2015) 3269–3275. https://doi.org/10.1093/bioinformatics/btv367.

[30] B.R. King, S. Vural, S. Pandey, A. Barteau, C. Guda, ngLOC: software and web server for predicting protein subcellular localization in prokaryotes and eukaryotes, BMC Research Notes. 5 (2012) 351. https://doi.org/10.1186/1756-0500-5-351.

[31] J.J. Almagro Armenteros, C.K. Sønderby, S.K. Sønderby, H. Nielsen, O. Winther, DeepLoc: prediction of protein subcellular localization using deep learning, Bioinformatics. 33 (2017) 3387–3395. https://doi.org/10.1093/bioinformatics/btx431.

[32] T. Goldberg, M. Hecht, T. Hamp, T. Karl, G. Yachdav, N. Ahmed, U. Altermann, P. Angerer, S. Ansorge, K. Balasz, M. Bernhofer, A. Betz, L. Cizmadija, K.T. Do, J. Gerke, R. Greil, V. Joerdens, M. Hastreiter, K. Hembach, M. Herzog, M. Kalemanov, M. Kluge, A. Meier, H. Nasir, U. Neumaier, V. Prade, J. Reeb, A. Sorokoumov, I. Troshani, S. Vorberg, S. Waldraff, J. Zierer, H. Nielsen, B. Rost, LocTree3 prediction of localization, Nucleic Acids Res. 42 (2014) W350-355. https://doi.org/10.1093/nar/gku396.

[33] C.-S. Yu, Y.-C. Chen, C.-H. Lu, J.-K. Hwang, Prediction of protein subcellular localization, Proteins: Structure, Function, and Bioinformatics. 64 (2006) 643–651. https://doi.org/10.1002/prot.21018.

[34] R. Kumar, B. Kumari, M. Kumar, Proteome-wide prediction and annotation of mitochondrial and sub-mitochondrial proteins by incorporating domain information, Mitochondrion. 42 (2018) 11–22. https://doi.org/10.1016/j.mito.2017.10.004.

[35] C. Savojardo, N. Bruciaferri, G. Tartari, P.L. Martelli, R. Casadio, DeepMito: accurate prediction of protein sub-mitochondrial localization using convolutional neural networks, Bioinformatics. 36 (2020) 56–64. https://doi.org/10.1093/bioinformatics/btz512.

[36] P.J. Thul, L. Åkesson, M. Wiking, D. Mahdessian, A. Geladaki, H.A. Blal, T. Alm, A. Asplund, L. Björk, L.M. Breckels, A. Bäckström, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hober, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C.M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J.M. Schwenk, Å. Sivertsson, E. Sjöstedt, M. Skogs, C. Stadler, D.P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Pontén, K. von Feilitzen, K.S. Lilley, M. Uhlén, E. Lundberg, A subcellular map of the human proteome, Science. 356 (2017). https://doi.org/10.1126/science.aal3321.

[37] H. Prokisch, C. Scharfe, D.G.C. Ii, W. Xiao, L. David, C. Andreoli, M.E. Monroe, R.J. Moore, M.A. Gritsenko, C. Kozany, K.K. Hixson, H.M. Mottaz, H. Zischka, M. Ueffing, Z.S. Herman, R.W. Davis, T. Meitinger, P.J. Oefner, R.D. Smith, L.M. Steinmetz, Integrative Analysis of the Mitochondrial Proteome in Yeast, PLOS Biology. 2 (2004) e160. https://doi.org/10.1371/journal.pbio.0020160.

[38] S.D. Cruz, I. Xenarios, J. Langridge, F. Vilbois, P.A. Parone, J.-C. Martinou, Proteomic Analysis of the Mouse Liver Mitochondrial Inner Membrane, J. Biol. Chem. 278 (2003) 41566–41571. https://doi.org/10.1074/jbc.M304940200.

[39] F. Marini, V.C. Carregari, V. Greco, M. Ronci, F. Iavarone, S. Persichilli, M. Castagnola, A. Urbani, L. Pieroni, Exploring the HeLa Dark Mitochondrial Proteome, Front. Cell Dev. Biol. 8 (2020). https://doi.org/10.3389/fcell.2020.00137.

[40] S.W. Taylor, E. Fahy, B. Zhang, G.M. Glenn, D.E. Warnock, S. Wiley, A.N. Murphy, S.P. Gaucher, R.A. Capaldi, B.W. Gibson, S.S. Ghosh, Characterization of the human heart mitochondrial proteome, Nat. Biotechnol. 21 (2003) 281–286. https://doi.org/10.1038/nbt793.

[41] S.P. Gaucher, S.W. Taylor, E. Fahy, B. Zhang, D.E. Warnock, S.S. Ghosh, B.W. Gibson, Expanded Coverage of the Human Heart Mitochondrial Proteome Using Multidimensional Liquid Chromatography Coupled with Tandem Mass Spectrometry, J. Proteome Res. 3 (2004) 495–505. https://doi.org/10.1021/pr034102a.

[42] J. Wettmarshausen, F. Perocchi, Isolation of Functional Mitochondria from Cultured Cells and Mouse Tissues, Methods Mol. Biol. 1567 (2017) 15–32. https://doi.org/10.1007/978-1-4939-6824-4_2.

[43] E.C. Bayraktar, L. Baudrier, C. Özerdem, C.A. Lewis, S.H. Chan, T. Kunchok, M. Abu-Remaileh, A.L. Cangelosi, D.M. Sabatini, K. Birsoy, W.W. Chen, MITO-Tag Mice enable rapid isolation and multimodal profiling of mitochondria from specific cell types in vivo, Proc Natl Acad Sci USA. 116 (2019) 303–312. https://doi.org/10.1073/pnas.1816656115.

[44] J.D. Martell, T.J. Deerinck, Y. Sancak, T.L. Poulos, V.K. Mootha, G.E. Sosinsky, M.H. Ellisman, A.Y. Ting, Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy, Nat. Biotechnol. 30 (2012) 1143–1148. https://doi.org/10.1038/nbt.2375.

[45] H.-W. Rhee, P. Zou, N.D. Udeshi, J.D. Martell, V.K. Mootha, S.A. Carr, A.Y. Ting, Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging, Science. 339 (2013) 1328–1331. https://doi.org/10.1126/science.1230593.

[46] V. Hung, P. Zou, H.-W. Rhee, N.D. Udeshi, V. Cracan, T. Svinkina, S.A. Carr, V.K. Mootha, A.Y. Ting, Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging, Molecular Cell. 55 (2014) 332–341. https://doi.org/10.1016/j.molcel.2014.06.003.

[47] V. Hung, S.S. Lam, N.D. Udeshi, T. Svinkina, G. Guzman, V.K. Mootha, S.A. Carr, A.Y. Ting, Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation, ELife. 6 (2017) e24463. https://doi.org/10.7554/eLife.24463.

[48] G. Liu, A. Papa, A.N. Katchman, S.I. Zakharov, D. Roybal, J.A. Hennessey, J. Kushner, L. Yang, B.-X. Chen, A. Kushnir, K. Dangas, S.P. Gygi, G.S. Pitt, H.M. Colecraft,

M. Ben-Johny, M. Kalocsay, S.O. Marx, Mechanism of adrenergic Ca V 1.2 stimulation revealed by proximity proteomics, Nature. 577 (2020) 695–700. https://doi.org/10.1038/s41586-020-1947-z.

[49] Y.G. Santin, Uncovering the In Vivo Proxisome Using Proximity-Tagging Methods, BioEssays. 41 (2019) 1900131. https://doi.org/10.1002/bies.201900131.

[50] R. Aebersold, M. Mann, Mass-spectrometric exploration of proteome structure and function, Nature. 537 (2016) 347–355. https://doi.org/10.1038/nature19949.

[51] F. Meier, A.-D. Brunner, S. Koch, H. Koch, M. Lubeck, M. Krause, N. Goedecke, J. Decker, T. Kosinski, M.A. Park, N. Bache, O. Hoerning, J. Cox, O. Räther, M. Mann, Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer, Mol Cell Proteomics. 17 (2018) 2534–2545. https://doi.org/10.1074/mcp.TIR118.000900.

[52] F. Perocchi, L.J. Jensen, J. Gagneur, U. Ahting, C. von Mering, P. Bork, H. Prokisch, L.M. Steinmetz, Assessing Systems Properties of Yeast Mitochondria through an Interaction Map of the Organelle, PLOS Genetics. 2 (2006) e170. https://doi.org/10.1371/journal.pgen.0020170.

[53] S.E. Calvo, K.R. Clauser, V.K. Mootha, MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins, Nucleic Acids Res. 44 (2016) D1251–D1257. https://doi.org/10.1093/nar/gkv1003.

[54] MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations | Nucleic Acids Research | Oxford Academic, (n.d.). https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkaa1011/5974091 (accessed November 18, 2020).

[55] A.C. Smith, A.J. Robinson, MitoMiner v4.0: an updated database of mitochondrial localization evidence, phenotypes and diseases, Nucleic Acids Res. 47 (2019) D1225–D1228. https://doi.org/10.1093/nar/gky1072.

[56] D.M. Arduino, F. Perocchi, Pharmacological modulation of mitochondrial calcium homeostasis, J. Physiol. (Lond.). 596 (2018) 2717–2733. https://doi.org/10.1113/JP274959.

[57] H.F. DeLuca, G.W. Engstrom, Calcium Uptake by Rat Kidney Mitochondria, PNAS. 47 (1961) 1744–1750. https://doi.org/10.1073/pnas.47.11.1744.

[58] F.D. Vasington, J.V. Murphy, Ca++ Uptake by Rat Kidney Mitochondria and Its Dependence on Respiration and Phosphorylation, J. Biol. Chem. 237 (1962) 2670–2677.

[59] Y. Kirichok, G. Krapivinsky, D.E. Clapham, The mitochondrial calcium uniporter is a highly selective ion channel, Nature. 427 (2004) 360–364. https://doi.org/10.1038/nature02246.

[60] F.D. Vasington, P. Gazzotti, R. Tiozzo, E. Carafoli, The effect of ruthenium red on Ca2+ transport and respiration in rat liver mitochondria, Biochimica et Biophysica Acta (BBA) - Bioenergetics. 256 (1972) 43–54. https://doi.org/10.1016/0005-2728(72)90161-2.

[61] E. Carafoli, A.L. Lehninger, A survey of the interaction of calcium ions with mitochondria from different tissues and species, Biochem J. 122 (1971) 681–690. https://doi.org/10.1042/bj1220681.

[62] F. Perocchi, V.M. Gohil, H.S. Girgis, X.R. Bao, J.E. McCombs, A.E. Palmer, V.K. Mootha, MICU1 encodes a mitochondrial EF hand protein required for Ca 2+ uptake, Nature. 467 (2010) 291–296. https://doi.org/10.1038/nature09358.

[63] T. Gene Ontology Consortium, The Gene Ontology Resource: 20 years and still GOing strong, Nucleic Acids Res. 47 (2019) D330–D338. https://doi.org/10.1093/nar/gky1055.

[64] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C. von Mering, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (2019) D607–D613. https://doi.org/10.1093/nar/gky1131.

[65] A. Yim, P. Koti, A. Bonnard, F. Marchiano, M. Dürrbaum, C. Garcia-Perez, J. Villaveces, S. Gamal, G. Cardone, F. Perocchi, Z. Storchova, B.H. Habermann, mitoXplorer, a visual data mining platform to systematically analyze and visualize mitochondrial

expression dynamics and mutations, Nucleic Acids Res. 48 (2020) 605–632. https://doi.org/10.1093/nar/gkz1128.

[66] T. Gabaldón, D. Rainey, M.A. Huynen, Tracing the Evolution of a Large Protein Complex in the Eukaryotes, NADH:Ubiquinone Oxidoreductase (Complex I), Journal of Molecular Biology. 348 (2005) 857–870. https://doi.org/10.1016/j.jmb.2005.02.067.

[67] J.M. Baughman, F. Perocchi, H.S. Girgis, M. Plovanich, C.A. Belcher-Timme, Y. Sancak, X.R. Bao, L. Strittmatter, O. Goldberger, R.L. Bogorad, V. Koteliansky, V.K. Mootha, Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter, Nature. 476 (2011) 341–345. https://doi.org/10.1038/nature10234.

[68] D. De Stefani, A. Raffaello, E. Teardo, I. Szabò, R. Rizzuto, A forty-kilodalton protein of the inner membrane is the mitochondrial calcium uniporter, Nature. 476 (2011) 336–340. https://doi.org/10.1038/nature10230.

[69] Y. Cheng, F. Perocchi, ProtPhylo: identification of protein-phenotype and protein-protein functional associations via phylogenetic profiling, Nucleic Acids Res. 43 (2015) W160-168. https://doi.org/10.1093/nar/gkv455.

[70] M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, A. Ruepp, CORUM: the comprehensive resource of mammalian protein complexes—2019, Nucleic Acids Res. 47 (2019) D559–D563. https://doi.org/10.1093/nar/gky973.

[71] M. Kanehisa, Toward Pathway Engineering: A New Database of Genetic and Molecular Pathways, (1997). /paper/Toward-Pathway-Engineering%3A-A-New-Database-of-and-Kanehisa/d9ba4b63b75478aac0007afebe9bf229ea77d16a (accessed November 20, 2020).

[72] H. Heide, L. Bleier, M. Steger, J. Ackermann, S. Dröse, B. Schwamb, M. Zörnig, A.S. Reichert, I. Koch, I. Wittig, U. Brandt, Complexome Profiling Identifies TMEM126B as a Component of the Mitochondrial Complex I Assembly Complex, Cell Metabolism. 16 (2012) 538–549. https://doi.org/10.1016/j.cmet.2012.08.009.

[73] F. Liu, P. Lössl, B.M. Rabbitts, R.S. Balaban, A.J.R. Heck, The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes, Molecular & Cellular Proteomics. 17 (2018) 216–232. https://doi.org/10.1074/mcp.RA117.000470.

[74] H. Antonicka, Z.-Y. Lin, A. Janer, W. Weraarpachai, A.-C. Gingras, E.A. Shoubridge, A high-density human mitochondrial proximity interaction network, BioRxiv. (2020) 2020.04.01.020479. https://doi.org/10.1101/2020.04.01.020479.

[75] K.J. Roux, D.I. Kim, M. Raida, B. Burke, A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells, J Cell Biol. 196 (2012) 801–810. https://doi.org/10.1083/jcb.201112098.

[76] C. Kwak, S. Shin, J.-S. Park, M. Jung, T.T.M. Nhung, M.-G. Kang, C. Lee, T.-H. Kwon, S.K. Park, J.Y. Mun, J.-S. Kim, H.-W. Rhee, Contact-ID, a tool for profiling organelle contact sites, reveals regulatory proteins of mitochondrial-associated membrane formation, Proc Natl Acad Sci USA. 117 (2020) 12109–12120. https://doi.org/10.1073/pnas.1916584117.

[77] K.F. Cho, T.C. Branon, S. Rajeev, T. Svinkina, N.D. Udeshi, T. Thoudam, C. Kwak, H.-W. Rhee, I.-K. Lee, S.A. Carr, A.Y. Ting, Split-TurboID enables contact-dependent proximity labeling in cells, Proc Natl Acad Sci USA. 117 (2020) 12143–12154. https://doi.org/10.1073/pnas.1919528117.

[78] P.M. Quirós, M.A. Prado, N. Zamboni, D. D'Amico, R.W. Williams, D. Finley, S.P. Gygi, J. Auwerx, Multi-omics analysis identifies ATF4 as a key regulator of the mitochondrial stress response in mammals, J Cell Biol. 216 (2017) 2027–2045. https://doi.org/10.1083/jcb.201702058.

[79] E. Fessler, E.-M. Eckl, S. Schmitt, I.A. Mancilla, M.F. Meyer-Bender, M. Hanf, J. Philippou-Massier, S. Krebs, H. Zischka, L.T. Jae, A pathway coordinated by DELE1 relays mitochondrial stress to the cytosol, Nature. 579 (2020) 433–437. https://doi.org/10.1038/s41586-020-2076-4.

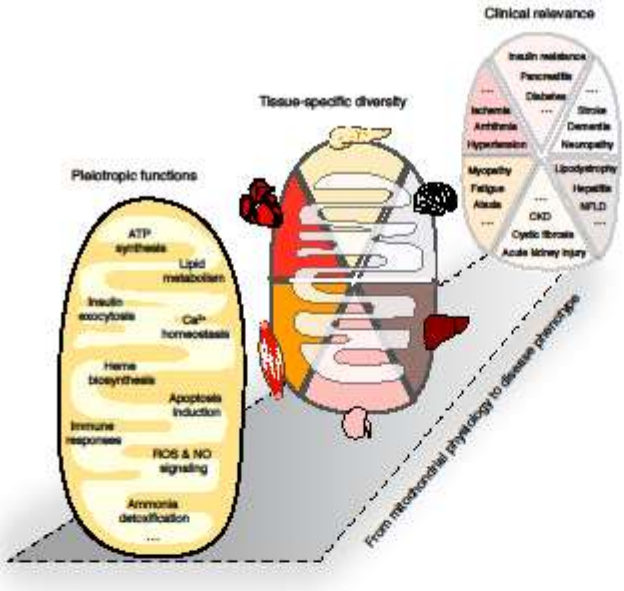
[80] N. Godin, J. Eichler, The Mitochondrial Protein Atlas: A Database of Experimentally Verified Information on the Human Mitochondrial Proteome, Journal of Computational Biology. 24 (2017) 906–916. https://doi.org/10.1089/cmb.2017.0011.

[81] D. Cotter, P. Guda, E. Fahy, S. Subramaniam, MitoProteome: mitochondrial protein sequence database and annotation system, Nucleic Acids Res. 32 (2004) D463-467. https://doi.org/10.1093/nar/gkh048.

[82] V. Ravichandran, A.S. Khan, P.E. Barker, G.B. Vasquez, S.J. Zullo, T.N. Bhat, G.L. Gilliland, Human Mitochondrial Protein Database: A Resource for Human Mitochondrial Proteomics, (2016). https://www.nist.gov/publications/human-mitochondrial-protein-database-resource-human-mitochondrial-proteomics (accessed November 19, 2020).

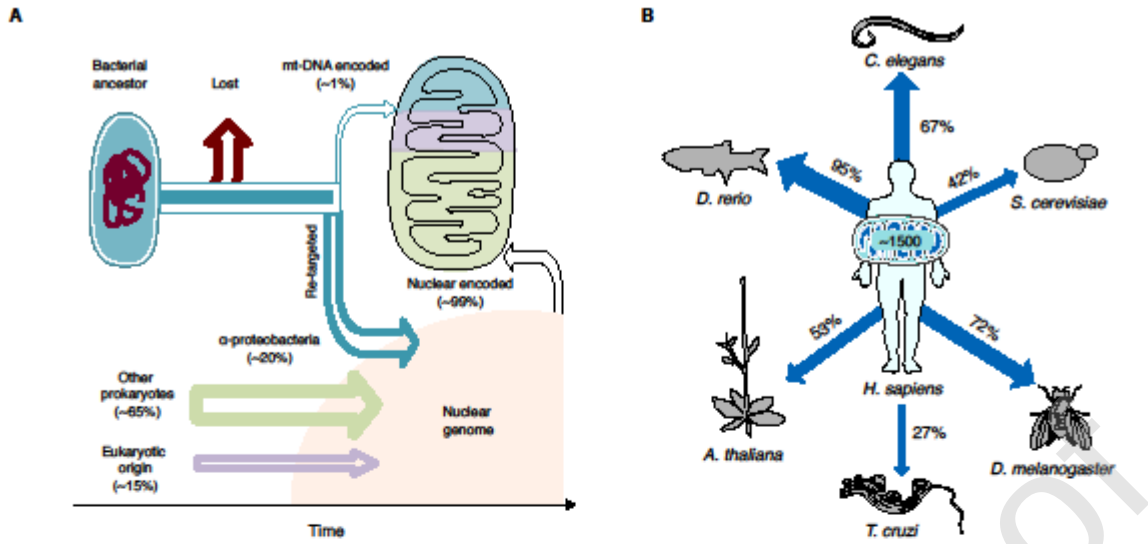
[83] M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome, Science. 347 (2015). https://doi.org/10.1126/science.1260419.

FIGURE LEGENDS



Delgado et al_Figure 1

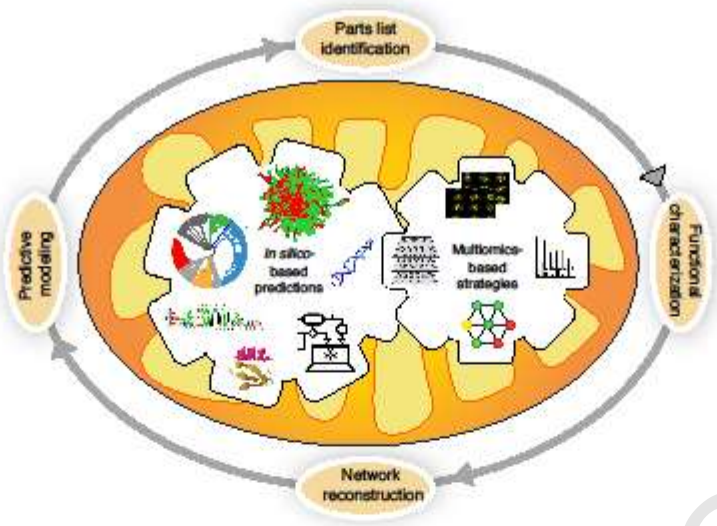
Figure 1. Complexity of the mitochondrial system. Mitochondria play a pleiotropic role in cell biology and physiology, which is reflected by the complexity and heterogeneity of their proteome, with over 1500 components that vary in both genetic and evolutionary origin and show tissue and cell type-specific expression. Mitochondrial diversity allows the system to meet the unique metabolic and physiological demands of each organ. Therefore, it is not surprising that more than 300 human proteins have been already linked to over 190 different disease phenotypes in the OMIM database, certainly an underestimation given that 20-40% of the mitochondrial proteome remain completely uncharacterized [14,19].
ROS, reactive oxygen species; *NO*, nitric oxide; *CKD*, chronic kidney disease; *NFLD*, Non-Alcoholic Fatty Liver Disease.



Delgado et al._Figure 2

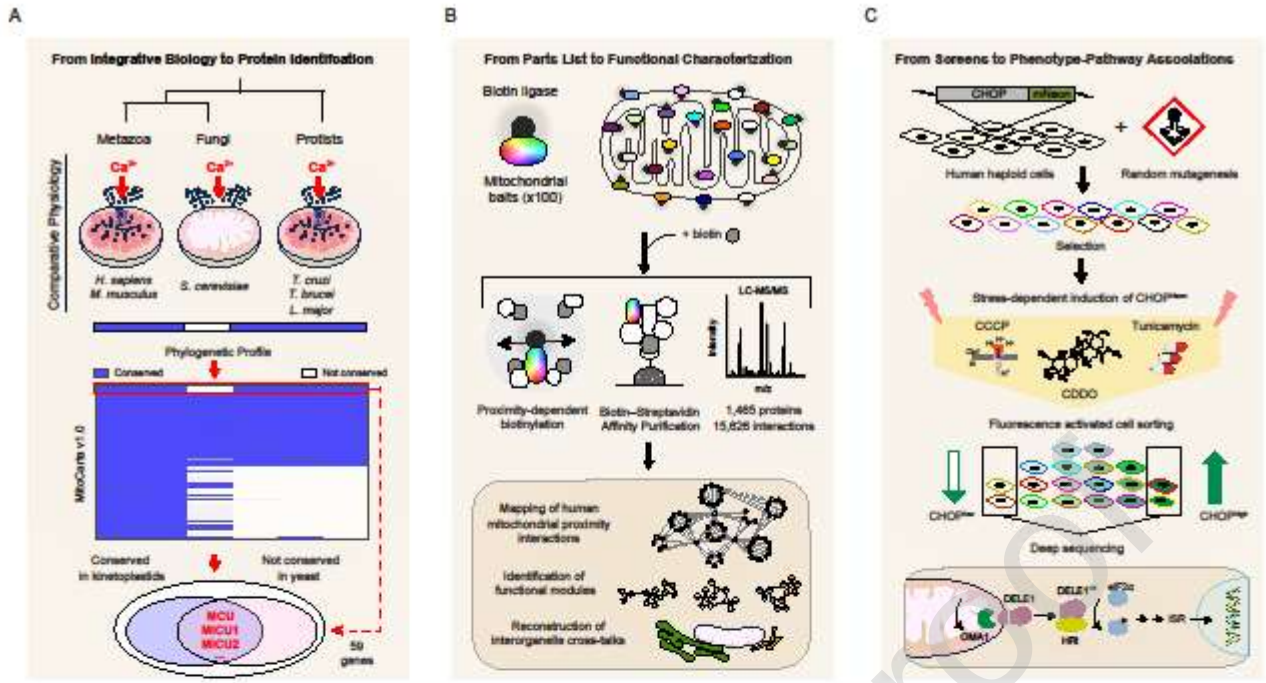
Figure 2. Evolutionary origin and conservation of the human mitochondrial proteome. (**A**) Percentage of present-day human mitochondrial proteome with alpha-proteobacteria, prokaryotic, and eukaryotic origin. (**B**) Percentage of human mitochondrial proteins with orthologs in each species. The list of human mitochondrial proteins was obtained from Malty

et *al.* [19] and Ensembl Compara v101 (Blastp e-value of 0.01) was used for homology inference.



Delgado et al_Figure 3

Figure 3. Systems-level approaches to study mitochondria. Defining and functionally characterizing the mitochondrial parts list involve systematic and integrative strategies that combine large-scale computational and experimental approaches with traditional biochemical, genetic and physiological analyses of mitochondrial function in different model organisms.



Delgado et al._Figure 4

Figure 4. Integrative, multiomics approaches to identify and deorphanize human mitochondrial proteins. Examples of *in silico* and experimental approaches developed to (**A**) identify de novo mitochondrial proteins involved in Ca^{2+} uptake [62], (**B**) chart intra and inter-organelle functional associations [74], (**C**) reconstruct signaling cascades regulating mitonuclear stress response pathways [79].

Table 1

Table 1. Systematic *in silico* and experimental approaches for identifying mitochondrial proteins

Tool	Method	Description	Predicted human mitochondrial proteins	Access location	Year	Ref.
COMPUTATIONAL						
PSORTII	Machine learning	Subcellular protein localization analysis. Prediction of mt-proteins is based on the biochemical features of N-terminal TS and presence of consensus CS	1712	https://psort.hgc.jp/form2.html	1999	[25]
ngLOC	Machine learning	Subcellular protein localization analysis. Prediction of mt-proteins is based on amino acid sequence	725	http://genome.unmc.edu/ngLOC/index.html	2012	[30]
LocTree3	Machine learning	Subcellular protein localization analysis. Prediction of mt-proteins is based on amino acid sequence and homology	1035	https://rostlab.org/services/loctree3/	2014	[32]

DeepLoc	Machine learning	Subcellular protein localization analysis. Prediction of mt-proteins is based on amino acid sequence	1553	http://www.cbs.dtu.dk/services/DeepLoc/	2017	[31]
SubMitoPred	Machine learning	Prediction of mitochondrial and sub-mitochondrial protein localization based on amino acid sequence and Pfam domains	Search based on the user input data	http://proteininformatics.org/mkumar/submitopred/	2018	[34]
DeepMito	Machine learning	Prediction mitochondrial and sub-mitochondrial protein localization based on amino acid sequence	254 (IMM), 124 (IMS), 499 (Matrix), 172 (OMM)	http://busca.biocomp.unibo.it/deepmito/	2020	[35]
MitoFates	Machine learning	Prediction of mt-protein localization based on the biochemical features of N-terminal TS and presence of consensus CS	1847	http://mitf.cbrc.jp/MitoFates/cgi-bin/top.cgi	2015	[28]
TargetP 2.0	Machine learning	Subcellular protein localization analysis. Prediction of mt-proteins is based on the biochemical features of N-terminal TS	648	http://www.cbs.dtu.dk/services/TargetP/	2019	[26]

TPpred3	Machine learning	Subcellular protein localization analysis. Prediction of mt-proteins is based on the biochemical features of N-terminal TS and presence of consensus CS	Search based on the user input data	https://tppred3.biocomp.unibo.it/tppred3	2015	[29]
Predotar	Machine learning	Subcellular protein localization analysis. Prediction of mt-proteins is based on the biochemical features of N-terminal TS	1426	https://urgi.versailles.inra.fr/predotar/	2004	[27]
CELLO II	Machine learning	Subcellular protein localization analysis; prediction of mitochondrial proteins is based on amino acid sequence and homology	Search based on the user input data	http://cello.life.nctu.edu.tw/	2006	[33]
Human MitoCarta 3.0	Machine learning and manual curation	Prediction of mt-proteins based on systematic data integration	1136	https://www.broadinstitute.org/mitocarta/mitocarta30-inventory-mammalian-mitochondrial-proteins-and-pathways	2020	[54]
Mitominer 4.0	Machine learning	Prediction of mt-protein based on systematic data integration	1626	http://mitominer.mrc-mbu.cam.ac.uk/release-4.0/begin.do	2018	[55]

Mitochondrial Protein Atlas	Manual curation	Database of human mt-proteins	911	http://lifeserv.bgu.ac.il/wb/jeichler/MPA/	2017	[80]
MitoProteome	Manual curation	Database of human mt-proteins	3625	http://www.mitoproteome.org/	2004	[81]
HMPDb	Automated data retrieval	Database of human mt-proteins	1465	https://bioinfo.nist.gov/hmpd/Search.html	2016	[82]
EXPERIMENTAL						
N-terminome	Biochemical	identification of human proteins with N-terminal, cleavable presequence	356	Supplementary Table 7	2015	[23]
Matrix and IMM	Biochemical	Proximity Labelling and MS	495	Supplementary Table 1	2013	[45]
IMS	Biochemical	Ratiometric Proximity Labelling and MS	127	Supplementary Table 2	2014	[46]
OMM	Biochemical	Ratiometric Proximity Labelling and MS	137	Supplementary File 1a	2017	[47]
Human Protein Atlas	Imaging	Database of subcellular protein localization	1098	https://www.proteinatlas.org/humanproteome/cell/organelle	2015	[83]

Abbreviations: mt, mitochondrial; TS, targeting sequence; CS, cleavage site; MS, mass-spectrometry; IMM, inner mitochondrial membrane; IMS, intermembrane space; OMM, outer mitochondrial membrane.