

MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations

Sneha Rath^{1,2,3,†}, Rohit Sharma^{1,2,3,†}, Rahul Gupta^{1,2,3,†}, Tslil Ast^{1,2,3}, Connie Chan^{1,2,3}, Timothy J. Durham^{1,2,3}, Russell P. Goodman^{1,2,3}, Zenon Grabarek^{1,2,3}, Mary E. Haas^{1,2,3}, Wendy H. W. Hung^{1,2,3}, Pallavi R. Joshi^{1,2,3}, Alexis A. Jourdain^{1,2,3}, Sharon H. Kim^{1,2,3}, Anna V. Kotrys^{1,2,3}, Stephanie S. Lam^{1,2,3}, Jason G. McCoy^{1,2,3}, Joshua D. Meisel^{1,2,3}, Maria Miranda^{1,2,3}, Apekshya Panda^{1,2,3}, Anupam Patgiri^{1,2,3}, Robert Rogers^{1,2,3}, Shayan Sadre^{1,2,3}, Hardik Shah^{1,2,3}, Owen S. Skinner^{1,2,3}, Tsz-Leung To^{1,2,3}, Melissa A. Walker^{1,2,3}, Hong Wang^{1,2,3}, Patrick S. Ward^{1,2,3}, Jordan Wengrod^{1,2,3}, Chen-Ching Yuan^{1,2,3}, Sarah E. Calvo^{1,2,3,*} and Vamsi K. Mootha^{1,2,3,*}

¹Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA, ²Howard Hughes Medical Institute and Departments of Molecular Biology and Medicine, Massachusetts General Hospital, Boston, MA 02114, USA and ³Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

Received September 18, 2020; Revised October 12, 2020; Editorial Decision October 13, 2020; Accepted October 14, 2020

ABSTRACT

The mammalian mitochondrial proteome is under dual genomic control, with 99% of proteins encoded by the nuclear genome and 13 originating from the mitochondrial DNA (mtDNA). We previously developed MitoCarta, a catalogue of over 1000 genes encoding the mammalian mitochondrial proteome. This catalogue was compiled using a Bayesian integration of multiple sequence features and experimental datasets, notably protein mass spectrometry of mitochondria isolated from fourteen murine tissues. Here, we introduce MitoCarta3.0. Beginning with the MitoCarta2.0 inventory, we performed manual review to remove 100 genes and introduce 78 additional genes, arriving at an updated inventory of 1136 human genes. We now include manually curated annotations of sub-mitochondrial localization (matrix, inner membrane, intermembrane space, outer membrane) as well as assignment to 149 hierarchical ‘MitoPathways’ spanning seven broad functional categories relevant to mitochondria. MitoCarta3.0, including sub-mitochondrial localization and MitoPathway annotations, is freely available at <http://www.broadinstitute.org/mitocarta> and should serve as a

continued community resource for mitochondrial biology and medicine.

INTRODUCTION

Mitochondria are a central hub of energy metabolism and play critical roles in redox balance, cofactor and vitamin processing, signaling and cell death. This organelle derives from an ancestral bacterial endosymbiont (1) and contains its own genome (mtDNA) that encodes 13 proteins, all involved in the respiratory chain. However, all other mitochondrial proteins are encoded by nuclear genes, translated in the cytoplasm, and imported into the organelle. Mutations in any of 400 mtDNA- or nuclear-encoded mitochondrial genes can give rise to debilitating monogenic, inborn errors of mitochondrial function (2).

As a foundation for developing a systematic understanding of mitochondrial biology, we previously aimed to identify all protein components resident in the mammalian mitochondrion (Figure 1). First, in 2008, we constructed the MitoCarta inventory (3) based on a Naïve Bayes integration of the following seven experimental and sequence features: (i) in-depth tandem mass spectrometry (MS/MS) of mitochondria isolated from 14 mouse tissues, (ii) prediction of mitochondrial targeting signals using TargetP (4), (iii) homology to yeast proteins known to localize to mitochondria, (iv) presence of protein domains enriched in

*To whom correspondence should be addressed. Tel: +1 617 643 3059; Email: vamsi@hms.harvard.edu
Correspondence may also be addressed to Sarah E. Calvo. Email: scalvo@broadinstitute.org.

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

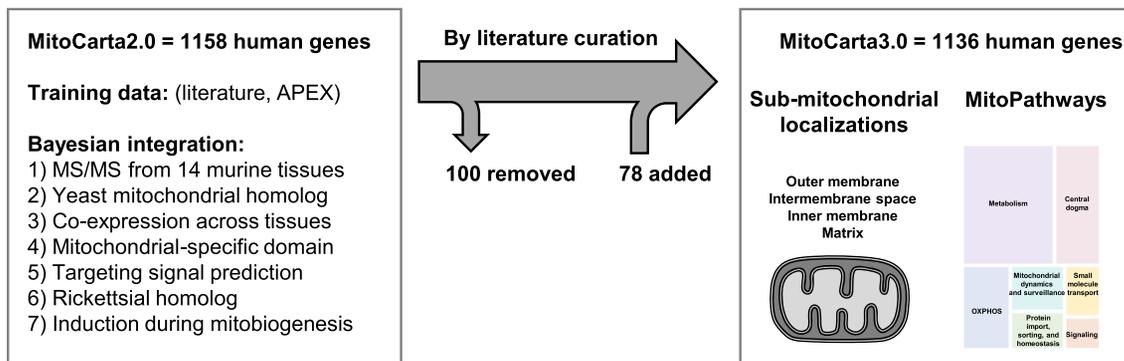


Figure 1. Overview of MitoCarta3.0. MitoCarta2.0 was previously built by Bayesian integration of the enumerated features. MitoCarta3.0 presents a revised inventory which is further annotated for sub-mitochondrial localization based on integrated analysis and MitoPathway membership based on literature curation.

mitochondrial proteins, (v) homology to proteins encoded by the genome of *Rickettsia prowazekii*, an extant bacterium related to the ancestor of modern mitochondria (vi) mRNA co-expression with known mitochondrial proteins across tissues, and (vii) transcriptional induction during mitochondrial biogenesis. The accuracy of each feature was determined using large training sets of known mitochondrial and non-mitochondrial human and mouse proteins. We then combined the features weighted by their accuracy to calculate a MitoCarta Bayes score for each human or mouse gene. In 2015, we created MitoCarta2.0 (5) using similar methods on updated protein models and incorporating evidence from proximity proteomics (6,7). The MitoCarta2.0 inventory of 1158 human or mouse genes was created using a 5% false discovery rate threshold (FDR). For use in interpreting high-throughput screens or large-scale patient sequence data, we suggested using a less stringent MitoCarta2.0 threshold such as 15% FDR. MitoCarta and MitoCarta2.0 have spurred basic science discoveries, such as the mitochondrial calcium uniporter, and aided in the prioritization of disease genes in whole exome and genome sequencing studies (8–17).

MitoCarta and MitoCarta2.0 have been widely used over the past decade, but the inventory is incomplete for certain classes of proteins and is also known to contain co-purifying proteins. The use of 14 murine tissues for mitochondrial proteomics is expected to miss genes that are expressed with high spatial or temporal specificity. As an illustrative example, *IRG1* is expressed almost exclusively in macrophages following activation, and detailed studies demonstrated that it encodes a mitochondrial aconitate decarboxylase that produces itaconate, an immunometabolite (18). The inventory was also underpowered to detect mitochondrial outer membrane proteins because such proteins typically lack the mitochondrial targeting sequence and co-purify with contaminating organelles in MS/MS experiments, though new proximity proteomics help to better spotlight such proteins (19). Certain protein components of organelles such as the endoplasmic reticulum (ER) or peroxisome can tether and thus indirectly co-purify with mitochondria. Therefore, detailed manual review of the literature can potentially help to prune and enrich this inventory.

Over the recent years, we have received many requests for annotations of sub-mitochondrial localization and mi-

tochondrial pathways. The mitochondrion has a double membrane organization and understanding whether proteins are localized to the mitochondrial outer membrane (MOM), the intermembrane space (IMS), the inner membrane (MIM), or the matrix can be valuable to end-user biologists. Additionally, biological pathway annotations can aid in discerning mitochondrial signatures within large-scale ‘omics’ datasets and screens. Such annotations are generally available through existing broad databases, such as NCBI Gene Ontology (GO) (20), Kyoto Encyclopedia of Genes and Genomes (KEGG) (21) and REACTOME (22). However, these pathway annotations were not generated specifically for mitochondria and can lack specificity and accuracy. For example, the KEGG Oxidative Phosphorylation (OXPHOS) pathway annotation lacks eight canonical human subunits and all assembly factors, while it erroneously includes 23 subunits of the lysosomal ATPase. Therefore, a dedicated set of mitochondria-centric pathway annotations that are manually curated to capture domain-specific knowledge are needed for in-depth studies of the organelle.

To address these limitations, we now present MitoCarta3.0 which contains an updated inventory of 1136 human mitochondrial proteins along with annotations of their sub-mitochondrial localization and pathway membership, all based on literature curation. MitoCarta3.0 is freely available at www.broadinstitute.org/mitocarta.

ANNOTATION PROCESS

We performed a manual review of the MitoCarta2.0 inventory, simultaneously flagging proteins for removal or inclusion, while annotating them to pathways and sub-compartments. Pairs of annotators tackled specific pathways, and multiple rounds of iteration were performed for cross-checking. MitoCarta membership, sub-mitochondrial localization, and pathway assignment were performed in parallel. A core team of ‘librarians’ ensured annotation consistency, facilitated data integration and performed a final review of all annotations.

To update the inventory, we systematically nominated candidates for addition to or removal from MitoCarta2.0 and then conducted independent literature-guided review of each candidate. To add new mitochondrial proteins to

the inventory, we short-listed hundreds of gene products with evidence of mitochondrial localization from the following data sources: (i) pathway review and primary literature, (ii) APEX-based proximity proteomics of the cytosol-facing mitochondrial outer membrane proteome (19) and (iii) genes with primary literature showing mitochondrial localization based on GO (20) and UniProt (23). Separately, we flagged genes for removal from the inventory if the proteins they encode had weak MitoCarta2.0 evidence, lacked literature support of mitochondrial localization, or were convincingly shown to localize elsewhere in multiple studies. Genes with strong MitoCarta2.0 Bayes scores were retained even if they had evidence of extra-mitochondrial localizations, allowing for the possibility of dual localization. In our review of each nominated candidate, we required multiple lines of high-quality experimental support for inclusion or removal, such as fractionation, protease sensitivity profiles, high-resolution microscopy with validated antibodies, experimental validation of a mitochondrial targeting sequence, and functional and genetic evidence. At least three independent reviews were considered and reconciled to arrive at MitoCarta3.0.

To assign sub-mitochondrial localizations to all MitoCarta3.0 proteins, we leveraged available experimental datasets and intrinsic protein features that jointly prioritize a specific compartment (Figure 2A). Annotators considered in aggregate all of these pieces of evidence to arrive at a best judgment for sub-mitochondrial localization. Specifically, we considered the following lines of evidence: (i) recovery of the protein in APEX-based proximity labelling from the matrix, IMS, or MOM, (ii) presence of a mitochondrial targeting sequence predicted by TargetP (24) (matrix, MIM or sometimes IMS), (iii) presence of helical transmembrane domains predicted by TMHMM (4) (MIM or MOM), (iv) presence of twin-Cx9C motifs suggesting IMS localization, (v) review of traditional experiments from primary literature acquired from the GO database (20) or manual literature search, and (vi) consideration of the pathway within which the gene product operates. In specific cases, we also used prediction of targeting signals from TPpred2 (25) and MitoFates (26). These features were logically and manually combined with prior knowledge of protein function during the annotation process, in order to assign the vast majority of MitoCarta3.0 proteins to the most likely sub-mitochondrial compartment. We note that proteins that form tight membrane-bound complexes were all typically assigned the same membrane compartment, for example all subunits of complex V were assigned MIM including members of the matrix-facing F1-ATPase, and all members of the calcium uniporter, including soluble regulatory subunits MICU1/2/3, were assigned to the inner membrane.

In parallel, we compiled a list of 149 pathways that capture the most relevant mitochondrial biology and then annotated MitoCarta3.0 genes to these 'MitoPathways'. Most of these MitoPathways are highly specific to mitochondrial biology. We adhered to two guiding principles to ensure utility from the perspective of an end-user biologist: (i) Each pathway is defined to clearly reflect a distinct mitochondrial process and (ii) pathway membership tends to be restrictive rather than inclusive (e.g. the TCA cycle does not include the numerous anaplerotic inputs or cataplerotic outputs).

At the same time, we did not impose an arbitrary limit on the number of pathways to which a gene can be annotated.

MITOCARTA3.0

An updated inventory of mitochondrial proteins

MitoCarta3.0 contains 1136 human genes, including 1058 genes previously in MitoCarta2.0 (Figure 1) and a highly similar list of 1140 mouse genes. Most of the differences between the two inventories can be attributed to species-specific paralogs. For example, while humans have one gene (*CYCS*) that encodes cytochrome *c*, mice have two genes, *Cyes* and *Cyct*, the latter representing a paralog expressed in the testis. Below we describe the human MitoCarta3.0 inventory.

We added 78 genes to the inventory. New additions were based on literature review and included genes with tissue-specific expression (e.g. *MCCD1* in kidney, *MTFR2* in testis), genes induced following stimuli (e.g., *IRG1* in macrophages following activation), and duplicated paralogs (e.g., *GLUD2*, *ACSM2B*), as well as three recently identified protein models not previously recorded in the NCBI Entrez Gene database: (i) *HTD2*, a bicistronic gene within *RPP14* which encodes a 168aa protein involved in fatty acid synthesis (27), (ii) *PIGBOS*, a 54aa microprotein in the MOM which interacts with the endoplasmic reticulum (ER) to regulate unfolded protein response (28) and (iii) *RP11.469A15.2*, a 62aa protein within a previously defined lincRNA (29). Separately, our review of mitochondrial pathways inspired specific additions to the inventory. For example in reviewing apoptosis, we decided to add all BH-domain containing proteins with established roles in the vicinity of the mitochondrial outer membrane, even though some proteins are embedded in the outer membrane while others are peripheral to it.

We removed 100 genes from the inventory. We excluded genes that received strong MitoCarta2.0 scores primarily because of homology to a mitochondrial paralog but themselves had overwhelming literature or experimental evidence of an alternate location (e.g. peroxisomal LONP2 and cytosolic ACO1, IDH1 and SHMT1). We also excluded genes if they had strong literature evidence of exclusive extra-mitochondrial localization and weak MitoCarta2.0 scores. These included canonical ER proteins such as EMC2, components exclusive to cytosolic translation (RARS, TARS, RPS14, RPS15A, RPS18, RPL10A, RPL34, RPL35A), the glycolytic complex that likely associates in the vicinity of the mitochondrial outer membrane (30) (GK, GPI, HK1, HK2), and proteins from other compartments whose low-abundance detection in isolated mitochondria could easily be explained by contamination. We introduced a separate tag termed 'mito-interacting' to annotate proteins that associate with the mitochondrion but do not reside within the double-membraned organelle. These include cytoplasmic signaling partners of MAVS, components of the ER mitochondrial-associated membrane (MAM), the glycolytic complex, and proteins required for mitochondrial trafficking, among others. Although we do not exhaustively curate all 'mito-interacting' proteins, the tag now offers a way to capture many proteins

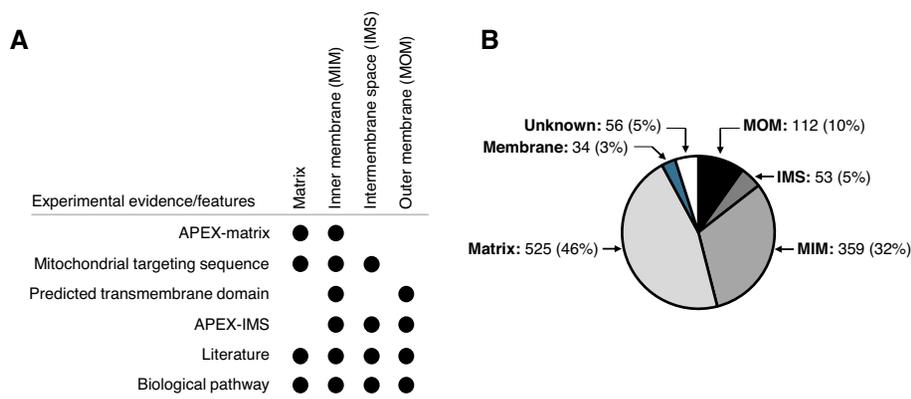


Figure 2. Sub-mitochondrial localization of MitoCarta3.0 proteins. (A) Schematic of possible sub-mitochondrial localizations that can be inferred from each type of published experimental evidence, computationally predicted feature, or knowledge of biological pathway. (B) Pie chart of MitoCarta3.0 proteins assigned to one or more sub-mitochondrial compartments.

removed from the inventory and others that were considered for addition but only associate peripherally with the genuine mitochondrial proteome.

Sub-mitochondrial localization of MitoCarta3.0 proteins

The mitochondrion is a double membrane organelle with proteins residing in the matrix, inner membrane, intermembrane space, and/or the outer membrane. During our annotation process we combined multiple tiers of evidence from the literature, protein domains, and pathway membership to arrive at the most likely sub-organelle localization. This process yielded assignment of 525 (46%) MitoCarta3.0 proteins to the matrix, 359 (32%) to the inner membrane, 53 (5%) to the IMS, and 112 (10%) to the outer membrane (Figure 2B). Proteins that could not be assigned to one of these four specific compartments were binned into the more general categories: 34 (3%) mitochondrial membrane or 56 (5%) unknown sub-mitochondrial localization. Notably, of the newly added genes to MitoCarta3.0, a disproportionately high percentage encoded outer membrane proteins (32% versus 10% expected). Thus, leveraging the recent outer-membrane proximity ligation dataset during our curation led to a targeted expansion of the outer membrane proteome in this inventory update.

To assess the accuracy of our sub-mitochondrial localizations, we compared our annotations to a recently published mitochondrial proximity interaction network (31). Considering all 11,569 bait-prey interactions between MitoCarta3.0 proteins, which covers nearly half of MitoCarta3.0 (527/1136 genes), we find that 98% of interactions are consistent with our annotated localizations. Specifically, 69% of interactors had identical sub-compartments, and 28% were in neighboring compartments compatible with an interaction.

MitoPathways: MitoCarta3.0 genes organized into functional pathways

In parallel, we sought to capture and functionally organize the breadth of mitochondrial processes using a custom ontology that spans seven broad categories of path-

ways. We first defined the seven categories to identify major classes of mitochondrial processes: (i) mitochondrial central dogma, (ii) protein import, sorting and homeostasis, (iii) oxidative phosphorylation (OXPHOS), (iv) metabolism, (v) small molecule transport, (vi) mitochondrial dynamics and surveillance, and (vii) signaling. Going one level deeper, we broke down each category into its constituent parts guided by widely accepted chemical, structural and functional classifications. Consequently, metabolism was broken down by type of biomolecule, OXPHOS by individual complexes of the electron transport chain, while mitochondrial central dogma by the three defining functional sub-categories—mtDNA maintenance, mtRNA metabolism and translation. Finally, within these subcategories we often further delineated more specific processes as necessary, for example under translation are five distinct pathways namely, mitochondrial ribosome, ribosome assembly, translation factors, mt-tRNA synthetases, and fMet processing. We also compiled genesets of intrinsic protein features that are relevant to mitochondrial pathways, such as EF-hand proteins, Fe-S-containing proteins, heme-containing proteins, and selenoproteins. All together, we manually compiled a mitochondrial ontology consisting of 149 hierarchical pathways spanning seven broad categories, which we collectively call ‘MitoPathways’ (Figure 3).

We annotated MitoCarta3.0 proteins to any MitoPathways in which they have been shown to function directly. This yielded a total of 1696 specific gene-pathway pairs, with only 101 proteins not annotated by a MitoPathway. Importantly, 87% of genes were annotated to one or two pathways alone, which reflects the intended selectivity of our annotations. About 41% of MitoCarta3.0 genes are involved in metabolism, followed by 20% in central dogma, 15% in OXPHOS, and less than 10% each in protein import, sorting and homeostasis, mitochondrial dynamics, small molecule transport, and signaling (Figure 3).

WEBSITE INTERFACE

The MitoCarta3.0 website (<http://www.broadinstitute.org/mitocarta>) provides the inventory of mitochondrial pro-

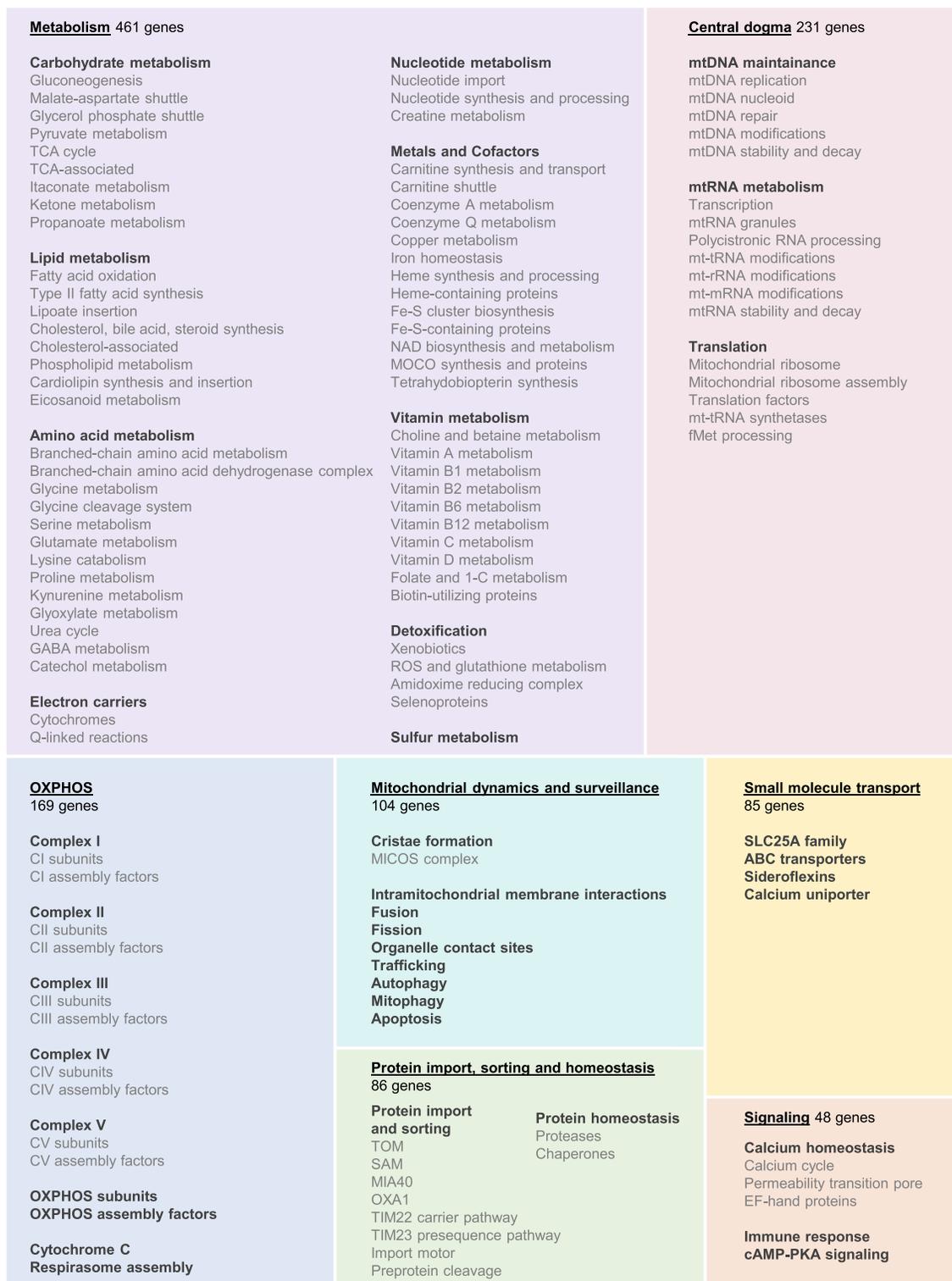


Figure 3. Annotation of MitoCarta3.0 genes to MitoPathways. Listed are a total of 149 MitoPathways organized in a three-tier hierarchy. Genes can be assigned to one or more of these MitoPathways. At the top level are seven overarching categories (color-coded boxes with bold, underlined labels), followed by the second (bold only) and third level of MitoPathways (lighter grey). The underlying treemap was generated in Excel using the number of MitoCarta3.0 genes annotated to each top-level category; this number is also listed in each box.

teins, evidence of mitochondrial localization, protein expression across 14 mouse tissues, sub-mitochondrial localization and pathway assignments. In addition, FASTA files of protein sequences, BED files of genomic positions, and detailed Excel data are available for download.

COMPARISON TO OTHER MITOCHONDRIAL DATABASES

MitoCarta3.0 includes a reference inventory of mitochondrial proteins that is more specific than large-scale databases of cellular localization and existing mitochondria-specific reference databases. Broad databases such as GO version 2020–08-31 reported 1625 human genes associated with the mitochondria cellular compartment (998 of which are in MitoCarta3.0), while the COMPARTMENTS database version 2020-08-31 (32) reported 1296 genes with strong mitochondrial evidence by an integrated score exceeding 4 (909 of which are in MitoCarta3.0). While excellent for their breadth, these more general databases often lack the specificity of mitochondria-centric databases. In contrast, the Integrated Mitochondrial Protein Index (IMPI) is a resource that, similar to MitoCarta, uses machine learning to score evidence of mitochondrial localization based on proteomics, antibody staining, and mitochondrial targeting sequence prediction. The IMPI version Q2 2018, released as part of MitoMiner 4.0 (33), contains 1626 Ensembl genes with 1064 in MitoCarta3.0. Compared to the Naïve Bayes machine learning methods in MitoCarta, IMPI's machine learning methodology can allow redundant datasets that are not conditionally independent, however these can lead to overfitting of training data and yield less interpretable scores. Thus, MitoCarta3.0 provides a highly specific reference of the mitochondrial proteome with interpretable scores and manual review.

MitoCarta3.0 additionally includes annotation of sub-mitochondrial localization. Such annotations are also available on GO cellular component but these are limited to literature reports and do not manually combine sequence features, experimental data, and pathway knowledge as in MitoCarta3.0. The COMPARTMENTS database provides annotation of sub-organelle localization for 60% of their mitochondrial proteins (integrated score exceeding four) and is updated weekly.

Finally, MitoCarta3.0 provides mitochondria-centric pathway annotations that are complementary to broader pathway databases such as KEGG pathways, REACTOME, and GO biological processes. Our 149 MitoPathways are customized for mitochondrial biology, and the manual review of all mitochondrial proteins enables a more complete annotation. In contrast, KEGG, REACTOME and GO include all cellular compartments into their biological pathways which is helpful to put mitochondrial pathways in a larger context but can also be inaccurate. For example, KEGG metabolic pathways do not annotate cellular localization and thus do not distinguish parallel pathways that exist in the cytoplasm versus mitochondria. Further, because they are inherently based on enzyme commission identifiers, KEGG annotations do not distinguish between orthologous enzymes functioning in different pathways, for example since the lysosomal and mitochondrial ATP syn-

thases have the same enzyme commission identifier, both are assigned to the Oxidative Phosphorylation pathway. GO biological processes are comprehensive and continuously updated to reflect the latest literature, however they suffer from highly redundant pathways and incomplete annotations. Thus, the manual annotation of MitoPathways provides a complementary and mitochondria-centric annotation of biological functions.

CONCLUSION

MitoCarta3.0 provides an updated inventory of mitochondrial proteins and their sub-organelle compartments and pathways.

While we anticipate that the updated MitoCarta3.0 inventory, localization and pathway annotations will be widely useful to the community, we acknowledge its limitations. First, it is possible that some annotations are incorrect or will prove to be so as new data emerge, especially for less well studied proteins. Second, our inventory is static and is not continually updated. Third, while many proteins are dual-localized, our mitochondria-centric inventory does not provide annotations for other extra-mitochondrial localizations and does not indicate cases in which only specific isoforms are mitochondrial. Finally, given the dense inter-connectedness of biology, we used our best judgment regarding where to draw pathway boundaries, however we acknowledge this can be very subjective.

Despite these limitations, MitoCarta3.0 provides a valuable research tool to study mitochondrial proteins and pathways. It provides an accessible gateway to learn and interpret the functions of previously studied mitochondrial proteins. At the same time, it serves as a foundation for the use of high-throughput datasets in molecular biology and human genetics to illuminate the function of understudied or undiscovered mitochondrial proteins.

ACKNOWLEDGEMENTS

We thank numerous members of the mitochondrial research community that have provided specific and broad constructive feedback on MitoCarta.

FUNDING

National Institutes of Health [K00CA212468 to S.R., T32AG000222 to R.G., K08DK115881 to R.P.G., R01AR071942 to Z.G., T32AG000222 to S.H.K., F32GM128259 to S.S.L., F32GM133047 to O.S.S., T32DK007028 to P.S.W. and R35GM122455 to V.K.M.]; NSF GRFP (to P.R.J.); Jane Coffin Childs fellowship (to J.D.M.); Deutsche Forschungsgemeinschaft [431313887 to M.M.]; Massachusetts General Hospital Department of Neurology Transformative Scholar Award (to M.A.W.); Dollis Huntington Endowment Fund for Cancer Research (to J.C.W.); V.K.M. is an Investigator of the Howard Hughes Medical Institute. Funding for open access charge: National Institutes of Health [R35GM122455 to V.K.M.]. *Conflict of interest statement.* O.S.S. is a paid consultant for Proteinaceous Inc. V.K.M. is a paid advisor to Janssen Pharmaceuticals and 5am Ventures and owns equity in Raze Therapeutics.

REFERENCES

- Andersson,S.G.E., Zomorodipour,A., Andersson,J.O., Sicheritz-Ponten,T., Alsmark,C.M., Podowski,R.M., Naslund,K.A., Eriksson,A., Winkler,H.H. and Kurland,C.G. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, **396**, 136–143.
- Rahman,S. (2020) Mitochondrial disease in children. *J Int Med*, **287**, 609–633.
- Pagliarini,D.J., Calvo,S.E., Chang,B., Sheth,S.A., Vafai,S.B., Ong,S.-E., Walford,G.A., Sugiana,C., Boneh,A., Chen,W.K. *et al.* (2008) A mitochondrial protein compendium elucidates complex I disease biology. *Cell*, **134**, 112–123.
- Krogh,A., Larsson,B., Von Heijne,G. and Sonnhammer,E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.
- Calvo,S.E., Clauser,K.R. and Mootha,V.K. (2016) MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res.*, **44**, D1251–D1257.
- Rhee,H.W., Zou,P., Udeshi,N.D., Martell,J.D., Mootha,V.K., Carr,S.A. and Ting,A.Y. (2013) Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*, **339**, 1328–1331.
- Hung,V., Zou,P., Rhee,H.W., Udeshi,N.D., Cracan,V., Svinkina,T., Carr,S.A., Mootha,V.K. and Ting,A.Y. (2014) Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol. Cell*, **55**, 332–341.
- Baughman,J.M., Perocchi,F., Girgis,H.S., Plovanich,M., Belcher-Timme,C.A., Sancak,Y., Bao,X.R., Strittmatter,L., Goldberger,O., Bogorad,R.L. *et al.* (2011) Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter. *Nature*, **476**, 341–345.
- Nilsson,R., Schultz,I.J., Pierce,E.L., Soltis,K.A., Naranuntarat,A., Ward,D.M., Baughman,J.M., Paradkar,P.N., Kingsley,P.D., Culotta,V.C. *et al.* (2009) Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. *Cell Metab.*, **10**, 119–130.
- Perocchi,F., Gohil,V.M., Girgis,H.S., Bao,X.R., McCombs,J.E., Palmer,A.E. and Mootha,V.K. (2010) MICU1 encodes a mitochondrial EF hand protein required for Ca(2+) uptake. *Nature*, **467**, 291–296.
- Strittmatter,L., Li,Y., Nakatsuka,N.J., Calvo,S.E., Grabarek,Z. and Mootha,V.K. (2014) CLYBL is a polymorphic human enzyme with malate synthase and beta-methylmalate synthase activity. *Hum. Mol. Genet.*, **23**, 2313–2323.
- Lanning,N.J., Looyenga,B.D., Kauffman,A.L., Niemi,N.M., Sudderth,J., DeBerardinis,R.J. and MacKeigan,J.P. (2014) A mitochondrial RNAi screen defines cellular bioenergetic determinants and identifies an adenylate kinase as a key regulator of ATP levels. *Cell Rep.*, **7**, 907–917.
- To,T., Cuadros,A., Shah,H., Hung,W., Li,Y., Kim,S., Rubin,D., Boe,R., Rath,S., Eaton,J. *et al.* (2019) A compendium of genetic modifiers of mitochondrial dysfunction reveals intra-organelle buffering. *Cell*, **179**, 1222–1238.
- Calvo,S.E., Compton,A.G., Hershman,S.G., Lim,S.C., Lieber,D.S., Tucker,E.J., Laskowski,A., Garone,C., Liu,S., Jaffe,D.B. *et al.* (2012) Molecular diagnosis of infantile mitochondrial disease with targeted next-generation sequencing. *Sci. Transl. Med.*, **4**, 118ra110.
- Calvo,S.E., Tucker,E.J., Compton,A.G., Kirby,D.M., Crawford,G., Burtt,N.P., Rivas,M., Guiducci,C., Bruno,D.L., Goldberger,O.A. *et al.* (2010) High-throughput, pooled sequencing identifies mutations in NUBPL and FOXRED1 in human complex I deficiency. *Nat Genet.*, **42**, 851–858.
- Lieber,D.S., Calvo,S.E., Shanahan,K., Slate,N.G., Liu,S., Hershman,S.G., Gold,N.B., Chapman,B.A., Thorburn,D.R., Berry,G.T. *et al.* (2013) Targeted exome sequencing of suspected mitochondrial disorders. *Neurology*, **80**, 1762–1770.
- Falk,M.J., Pierce,E.A., Consugar,M., Xie,M.H., Guadalupe,M., Hardy,O., Rappaport,E.F., Wallace,D.C., LeProust,E. and Gai,X. (2012) Mitochondrial disease genetic diagnostics: optimized whole-exome analysis for all MitoCarta nuclear genes and the mitochondrial genome. *Discov. Med.*, **14**, 389–399.
- Michelucci,A., Cordes,T., Ghelfi,J., Pailot,J., Reiling,N., Goldmann,O., Binz,T., Wegner,A., Tallam,A., Rausell,A. *et al.* (2013) Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 7820–7825.
- Hung,V., Lam,S., Udeshi,N., Svinkina,T., Guzman,G., Mootha,V.K., Carr,S.A. and Ting,A.Y. (2017) Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *eLife*, **6**, e24463.
- The Gene Ontology Consortium. (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
- Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.
- Jassal,B., Matthews,L., Viteri,G., Gong,C., Lorente,P., Fabregat,A., Sidiropoulos,K., Cook,J., Gillespie,M., Haw,R. *et al.* (2020) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **48**, D498–D503.
- The UniProt Consortium. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
- Emanuelsson,O., Nielsen,H., Brunak,S. and Von Heijne,G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.*, **300**, 1005–1016.
- Savojardo,C., Martelli,P.L., Fariselli,P. and Casadio,R. (2014) TPpred2: improving the prediction of mitochondrial targeting peptide cleavage sites by exploiting sequence motifs. *Bioinformatics*, **30**, 2973–2974.
- Fukasawa,Y., Tsuji,J., Fu,S., Tomii,K., Horton,P. and Imai,K. (2015) MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol. Cell. Proteomics*, **14**, 1113–1126.
- Autio,K.J., Kastaniotis,A.J., Pospiech,H., Miinalainen,I., Schonauer,M., Dieckmann,C. and Hiltunen,K. (2008) An ancient genetic link between vertebrate mitochondrial fatty acid synthesis and RNA processing. *FASEB J.*, **22**, 569–578.
- Martinez,T.F., Novak,S.W., Donaldson,C.J., Tan,D., Vaughan,J.M., Chang,T., Diedrich,J.K., Andrade,L., Kim,A., Zhang,T. *et al.* (2019) Regulation of the ER stress response by a mitochondrial microprotein. *Nat. Commun.*, **10**, 4883.
- Chen,J., Brunner,A., Cogan,J.Z., Nunez,J.K., Fields,A.P., Adamson,B., Itzhak,D., Li,J.Y., Mann,M., Leonetti,M.D. *et al.* (2020) Pervasive functional translation of noncanonical human open reading frames. *Science*, **367**, 1140–1146.
- Brandina,I., Graham,J., Lemaitre-Guillier,C., Entelis,N., Krashennnikov,I., Sweetlove,L., Tarassov,I. and Martin,R.P. (2006) Enolase takes part in a macromolecular complex associated to mitochondria in yeast. *Biochim. Biophys. Acta - Bioenerg.*, **1757**, 1217–1228.
- Antonicka,H., Lin,Z., Janer,A., Aaltonen,M.J., Weraarpachai,W., Gingras,A. and Shoubridge,E.A. (2020) A high-density human mitochondrial proximity interaction network. *Cell Metab.*, **32**, P479–P497.
- Binder,J.X., Pletscher-Frankild,S., Tsaou,K., Stolte,C., O'Donoghue,S.I., Schneider,R. and Jensen,L.J. (2014) COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database (Oxford)*, **2014**, bau012.
- Smith,A.C. and Robinson,A.J. (2019) MitoMiner v4.0: an updated database of mitochondrial localization evidence, phenotypes and diseases. *Nucleic Acids Res.*, **47**, D1225–D1228.