Cell, Vol. 115, 629–640, November 26, 2003, Copyright ©2003 by Cell Press

Integrated Analysis of Protein Composition, Tissue Diversity, and Gene Regulation in Mouse Mitochondria

Vamsi K. Mootha,[1,2,3] Jakob Bunkenborg,[1] Jesper V. Olsen,[1,5] Majbrit Hjerrild,[1] Jacek R. Wisniewski,[1] Erich Stahl,[2] Marjan S. Bolouri,[2] Heta N. Ray,[2] Smita Sihag,[2] Michael Kamal,[2] Nick Patterson,[2] Eric S. Lander,[2,4,6,*] and Matthias Mann[1,5,6,*]

[1]MDS Proteomics
Odense 5230
Denmark

[2]Whitehead Institute/MIT Center for Genome Research
Cambridge, Massachusetts 02139

[3]Department of Medicine
Brigham and Women's Hospital
Harvard Medical School
Boston, Massachusetts 02115

[4]Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02138

[5]Center for Experimental BioInformatics (CEBI)
University of Southern Denmark
Odense 5230
Denmark

Summary

Mitochondria are tailored to meet the metabolic and signaling needs of each cell. To explore its molecular composition, we performed a proteomic survey of mitochondria from mouse brain, heart, kidney, and liver and combined the results with existing gene annotations to produce a list of 591 mitochondrial proteins, including 163 proteins not previously associated with this organelle. The protein expression data were largely concordant with large-scale surveys of RNA abundance and both measures indicate tissue-specific differences in organelle composition. RNA expression profiles across tissues revealed networks of mitochondrial genes that share functional and regulatory mechanisms. We also determined a larger "neighborhood" of genes whose expression is closely correlated to the mitochondrial genes. The combined analysis identifies specific genes of biological interest, such as candidates for mtDNA repair enzymes, offers new insights into the biogenesis and ancestry of mammalian mitochondria, and provides a framework for understanding the organelle's contribution to human disease.

Introduction

Mammalian mitochondria are ubiquitous organelles responsible for 90% of ATP production in respiring cells. They are best known for housing the oxidative phosphorylation (OXPHOS) machinery as well as enzymes needed for free fatty acid metabolism and the Kreb's cycle. Key steps of heme biosynthesis, ketone body generation, and hormone synthesis also reside within this organelle (Stryer, 1988). The mitochondrion generates the majority of cellular reactive oxygen species (ROS) and has specialized scavenging systems to protect itself and the cell from these toxic by-products. Furthermore, the organelle is crucial for cellular calcium signaling and hosts key machinery for programmed cell death, serving as a gatekeeper for apoptosis (Hockenbery et al., 1993; Kluck et al., 1997). Given its contribution to cellular physiology, it is not surprising that this organelle can play an important role in human disease, such as diabetes, obesity, cancer, aging, neurodegeneration, and cardiomyopathy (Wallace, 1999).

Mitochondria contain their own DNA (mtDNA) which is a compact genome encoding only 13 polypeptides. Through reductive evolution, the complement of genes constituting the original eubacterial predecessors of modern-day mitochondria have been either lost or transferred from mtDNA to the nuclear genome (Andersson et al., 1998). Through an expansive process, the mitochondrion has also acquired new proteins and functionality. The exact number of mammalian mitochondrial proteins is not currently known, but estimates based on comparisons to the closest eubacterial relative of mammalian mitochondria, *Rickettsia prowazakeii* (Andersson et al., 1998), comparisons to *Saccharomyces cerevisiae* (Kumar et al., 2002), and two-dimensional gel electrophoresis studies of isolated mammalian mitochondria (Lopez et al., 2000; Rabilloud et al., 1998) suggest that the organelle contains approximately 1200 proteins. Although several recent studies have utilized proteomic and genetic approaches to expand the inventory of mammalian mitochondrial proteins, only 600–700 mitochondrial proteins are currently known (Da Cruz et al., 2003; Lopez et al., 2000; Ozawa et al., 2003; Taylor et al., 2003; Westermann and Neupert, 2003).

Classic electron microscopy studies have demonstrated morphologic differences in mitochondria from different cell types (Ghadially, 1997). Moreover, the mitochondria capacity, mtDNA copy number, enzymatic stoichiometry, carbon substrate utilization patterns, and biosynthetic pathways can be specialized (Stryer, 1988; Veltri et al., 1990; Vijayasarathy et al., 1998). Despite this apparent physiologic diversity, little is known about the molecular basis for these differences and the extent to which mitochondrial composition varies across different tissues. Emerging proteomics and genomics technologies afford the opportunity to survey these properties. Here, we use mass-spectrometry-based proteomics (Aebersold and Mann, 2003) to profile mitochondrial composition across four different tissues, and we then correlate and extend the proteomic results with mRNA expression data across a much larger set of tissues.

Our proteomic and RNA expression profiling study identified hundreds of gene products that are either localized to this organelle or tightly coregulated with mitochondrial genes, providing new hypotheses about the

*Correspondence: lander@genome.wi.mit.edu (E.S.L.), mann@bmb.sdu.dk (M.M.)

[6]These authors contributed equally to this work.

molecular composition of this organelle and how networks of genes may confer specialized function to mitochondria. The survey serves as an initial step toward elucidating the transcriptional, translational, and protein targeting mechanisms likely operative in achieving tissue specific differences in mitochondrial form and function.

Results and Discussion

Proteomic Survey of Mouse Mitochondria

We carried out a systematic survey of mitochondrial proteins from brain, heart, kidney, and liver of C57BL6/J mice (see Experimental Procedures). Each of these tissues provides a rich source of mitochondria. The isolation consisted of density centrifugation followed by Percoll purification. To assess the purity of our preparations, we performed immunoblot analysis of the organelles using markers for mitochondria as well as for contaminating organelles (Supplemental Figure S1 available online at http://www.cell.com/cgi/content/full/115/5/629/DC1). As can be seen in Supplemental Figure S1, the liver, heart, and kidney preparations were highly enriched in mitochondria, while brain preparations tended to show persistent contamination by synaptosomes, which themselves are a rich source of neuronal mitochondria (see Fernandez-Vizarra et al., 2002).

Mitochondrial proteins from each tissue were solubilized and size separated by gel filtration chromatography into a batch of approximately 15–20 fractions (see Experimental Procedures). These proteins were then digested and analyzed by liquid chromatography mass spectrometry/mass spectrometry (LC-MS/MS). More than 100 LC-MS/MS experiments were performed (see Experimental Procedures).

The acquired tandem mass spectra were then searched against the NCBI nonredundant database (October 2002) consisting of mammalian proteins using a probability-based method (Perkins et al., 1999). We used stringent criteria for accepting a database hit. Specifically, only peptides corresponding to complete tryptic cleavage specificity with scores greater than 25 were considered (see Experimental Procedures). Furthermore, we only accepted fragmentation spectra which also exhibited a correct, corresponding peptide sequence tag (Mann and Wilm, 1994) consisting of at least three amino acids.

Using these criteria, 4766 proteins were identified. This list contains a high degree of redundancy, because a protein may have been found in adjacent gel-filtration fractions and in different tissues, and may correspond to different database entries corresponding to nearly identical proteins which have not been distinguished by mass spectrometry. To produce a nonredundant list of identified proteins, we used a permissive clustering routine (see Experimental Procedures) based on BLAST (Altschul et al., 1990) to collapse the 4766 hits to a distinct set of 399 mouse protein clusters (see Figure 1A; Supplemental Table S1 available at above URL).

Previously Annotated Mitochondrial Proteins

We created a list of previously annotated mouse and human mitochondrial proteins by pooling all the mouse and human proteins from MITOchondria Project (MITOP, http://mips.gsf.de/proj/medgen/mitop/), a public database of curated mitochondrial proteins, as well as all proteins annotated as mitochondrial in NCBI's LocusLink database (http://www.ncbi.nlm.nih.gov/LocusLink/) (January 2003, see Experimental Procedures). After elimination of redundancy, the list contains 428 distinct mouse proteins that are either directly annotated as mitochondrial or whose human homolog is annotated as mitochondrial (Figure 1A). We did not include the human proteins recently reported to be mitochondrial by Taylor et al. (2003) in a study published after the construction of our list of previously annotated proteins. These proteins instead serve as a control against which to compare the proteins identified in our proteomic analysis. The list of 428 previously annotated mitochondrial proteins is by no means comprehensive—many mitochondrial proteins are simply not cataloged in these public databases. However, it does provide a reasonable, high confidence list of previously annotated proteins against which to benchmark our proteomic survey.

Newly Identified Mitochondrial Proteins

The set of 399 proteins identified in our proteomic survey include 236 of the 428 proteins previously annotated to be mitochondrial (55%) and 163 proteins not previously annotated as associated with the this organelle (Figure 1A). Combining the previous and new sets, we obtain a list of 591 mitochondria-associated proteins (mito-A) that are physically associated with this organelle (Supplemental Table S1).

The 399 proteins identified in the proteomic survey span a wide range of molecular weight and isoelectric points (Figures 1B and 1C), although proteins from the inner mitochondrial membrane are underrepresented (Figure 1D). The incomplete coverage (55%) is most likely due to the finite sensitivity of the mass spectrometric methodology, which acts as a bias against proteins of low abundance. This explanation is supported by analysis of RNA expression of the genes encoding the detected and undetected proteins. Considering the subset of the 428 previously annotated proteins for which RNA expression was reported in an atlas of mRNA expression in mouse (Su et al., 2002), the distribution of RNA expression level was about 5-fold higher for the genes whose products were detected in our proteomic survey as compared to those that were not ($p = 1 \times 10^{-21}$) (Figure 1E). This suggests that the proteomics strategy preferentially detected the higher abundance proteins.

The 163 proteins not previously annotated as mitochondrial potentially represent new mitochondrial proteins, either in the conventional sense of being present within the organelle or in a broader sense of being tethered to the mitochondrial outer membrane. To test this notion, we sought independent evidence that these 163 proteins are actually mitochondrial (Supplemental Table S2). First, we compared the list to proteins identified in a recent survey of human heart mitochondria (Taylor et al., 2003). Human homologs of 88 of the 163 proteins were identified in this recently published study. Of the remaining 75 proteins, 19 (25%) have strong mitochondrial targeting sequences based on bioinformatic analysis of protein targeting sequences (see Supplemental

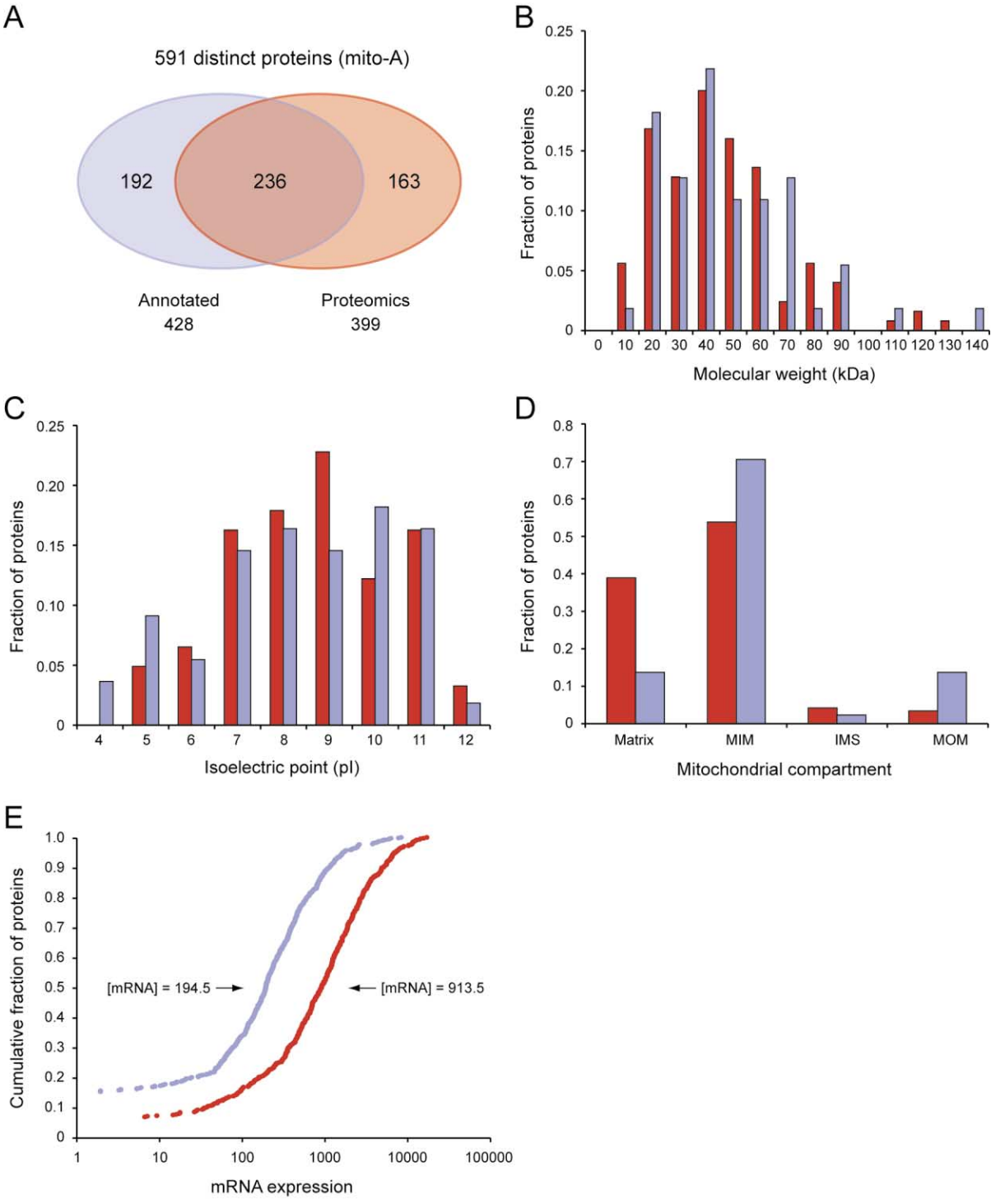


Figure 1. Previously Known and Newly Identified Mitochondrial Proteins (mito-A)

(A) Proteomic survey of mitochondria from mouse brain, heart, kidney, and liver resulted in the identification of 399 protein clusters, 236 of which were previously annotated as being mitochondrial. The distributions for (B) molecular weight, (C) isoelectric point, and (D) mitochondrial compartments are plotted for proteins detected (red) or not detected (blue) by our proteomic survey. Isoelectric point, molecular weight, and subcellular distribution data came from the MITOchondria Project (MITOP [Scharfe et al., 2000]). (E) Cumulative distribution of mRNA abundance for those genes whose protein product was detected (red) or not detected (blue) by proteomics. The median expression levels for both groups are indicated. MIM, mitochondrial inner membrane; IMS, intermembrane space; and MOM, mitochondrial outer membrane.

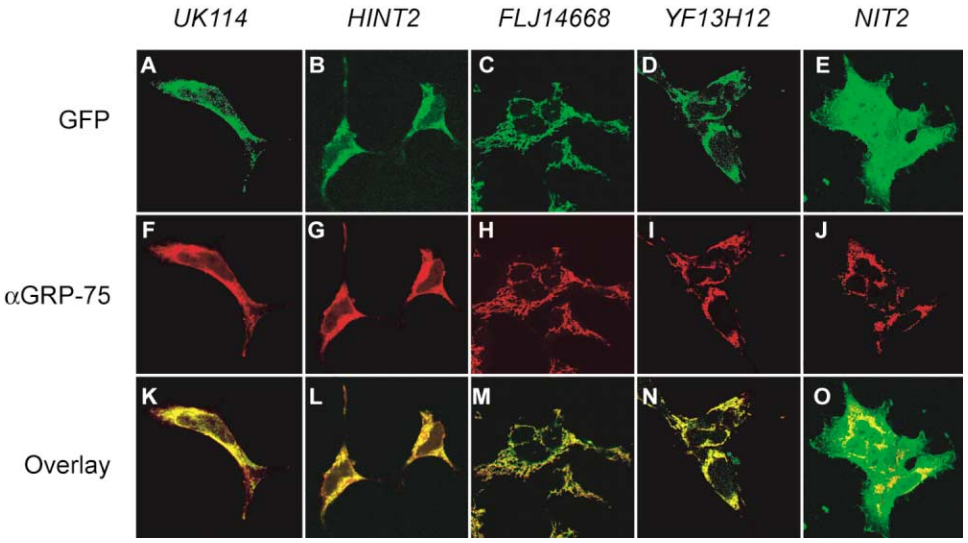


Figure 2. Subcellular Localization of Proteins Not Previously Associated with the Mitochondrion

GFP fusion proteins for human homologs of five of the newly identified proteins (A–E) were expressed in human 293 cells, counterstained with an antibody (α-GRP-75) directed against a known mitochondrial marker (F–J), and imaged by confocal microscopy. Panels (K)–(O) show the overlay of the two images. (A) UK114 (translational inhibitor protein p14.5), homolog of Hrsp12 (GenPept accession 6680277). (B) HINT2 (histidine triad nucleotide binding protein 2), homolog of 1190005L05Rik (GenPept accession 12835711). (C) FLJ14668 (hypothetical protein), homolog of 2010309E21Rik (GenPept accession 13385042). (D) YF13H12 (protein expressed in thyroid), homolog of 0610025L15Rik (GenPept accession 12963539). (E) NIT2 (Nit protein 2), homolog of Nit2 (GenPept accession 12963555).

Table S2 and Experimental Procedures) (Nakai and Horton, 1999), a proportion slightly lower than the known mitochondrial proteins (38%). Of those that do not have strong mitochondrial targeting sequences, seven show RNA expression patterns tightly correlated with known mitochondrial genes. For example, polymerase delta interacting protein 38 (encoded by *Pdip38*), which was detected only in liver mitochondria, and the gene product of *Rnaseh1*, which was found only in the kidney, have strong mitochondrial targeting scores. The protein 2010100O12Rik, which was detected in mouse liver and in kidney, appears to be an integral membrane protein whose gene expression is extremely tightly correlated with the known mitochondrial genes. Hence, the majority of the 163 newly identified mito-A members have multiple tiers of evidence supporting that they are mitochondrial.

To provide direct experimental evidence, we chose human homologs of five of the 163 newly identified mouse mito-A proteins and created GFP-tagged fusions to determine their subcellular localization by confocal microscopy (Figure 2). Four of these five showed exclusive mitochondrial staining, while one showed diffuse mitochondrial and cytosolic staining. Taken together, our analyses show that of the 163 mito-A proteins, 113 have at least one additional tier of support (Supplemental Table S2), suggesting that the list of newly identified proteins is indeed highly enriched in mitochondrial proteins.

The list of 163 proteins above includes many proteins of unknown function (Supplemental Table S2). For example, very little is known about the five proteins whose localization we confirmed. NIT2 (Figure 2E) and HINT2 (Figure 2B), human homologs of proteins we identified, are both evolutionarily conserved enzymes with putative roles in nucleotide metabolism and possibly in tumor suppression (Brenner et al., 1999). UK114 (Figure 2A) is the human homolog of mouse protein Hrsp12, previously described as a liver protein that occurs as a dimer and is differentially expressed following heat shock (Samuel et al., 1997). YF13H12 (Figure 2D) and FLJ14668 (Figure 2C) are human homologs of mouse proteins we identified that also appear to be exclusively mitochondrial based on microscopy studies. Other proteins identified in our study are poorly characterized, but based on their protein domains, could play very interesting roles in the mitochondrion. For example, the AAA-ATPase domain containing protein Tob3 may play a role in the assembly or degradation of mitochondrial protein complexes (Lupas and Martin, 2002). This list also includes a number of well-characterized proteins not traditionally associated with the organelle, including the glycolytic enzymes hexokinase, aldolase, and glyceraldehyde 3 phosphate dehydrogenase. Previous studies have suggested that these enzymes may be tethered to outer mitochondrial proteins, and several other recent proteomics studies have detected these proteins in their mitochondrial preparations (Taylor et al., 2003). Close proximity of this glycolytic machinery to the outer membrane of the mitochondrion would serve an obvious biological function, since it produces pyruvate, which feeds into the Kreb's cycle in the mitochondrion. Our list also includes several proteins traditionally associated with the lysosome (e.g., cathepsin and saposin), which may play a role in mitochondrial protein degradation. However, it is possible that these latter proteins merely represent contamination by other organelles.

Human homologs of two proteins identified by the proteomic survey are clearly involved in human disease.

The first is *LET1*, which is deleted in nearly all patients with Wolf-Hirschhorn syndrome (WHS) (Endele et al., 1999). We identified this protein in mouse brain, heart, kidney, and in liver, and a recent study confirmed its mitochondrial localization (Taylor et al., 2003). The second is *LRPPRC*, encoding an mRNA binding protein, whose human homolog we recently identified as being mutated in a human cytochrome *c* oxidase deficiency, Leigh Syndrome, French Canadian variant (Mootha et al., 2003a).

Clearly, additional studies are needed to fully validate the subcellular localization of all these 163 proteins (Supplemental Table S2) and to determine their function. While several bioinformatic tools are currently available for detecting mitochondrial targeting sequences (Nakai and Horton, 1999; Neupert, 1997), such predictions still suffer from poor sensitivity and specificity. With the growing inventory of mito-A proteins, it may be possible to discover new protein targeting motifs and mechanisms.

Concordance of mRNA Abundance and Protein Detection

Next, we sought to determine whether protein detection in our proteomics experiments is broadly concordant with mRNA abundance of the corresponding gene measured by oligonucleotide microarrays. The traditional approach to relating mRNA abundance to protein abundance is to calculate a simple correlation coefficient. However, protein detection by mass spectrometry and RNA expression analysis with microarrays can result in noisy data. For example, the protein product of a given gene may give rise to few or unfavorable tryptic peptides for mass spectrometric identification. Similarly, the oligonucleotide probes on the microarray may be imperfect detectors for certain genes. Previous efforts to analyze such noisy data with simple correlation analyses have resulted in positive but weak associations (Griffin et al., 2002; Lian et al., 2001) between mRNA and protein, while analyses with more robust statistics have yielded stronger correlations (Gygi et al., 1999).

To decrease the effect of noisy data, we developed an RNA/protein concordance test that takes advantage of the availability of mRNA and protein measures across four tissues (see Experimental Procedures). If a given protein is detected in liver but not in heart, for example, we say that the mRNA abundance is concordant if the mRNA expression level in liver exceeds that in heart. The mRNA/protein concordance test overcomes those technical artifacts that are uniform for a given gene across different tissues. For a given gene, we can count the total number of concordant measures for all pairs of tissues and compare to the expected distribution of concordance in the null case in which there is no association between mRNA and protein detection (see Experimental Procedures).

We applied this analysis to proteins identified in well-matched brain, heart, kidney, and liver batches for which we also had mRNA expression measures. We found that 426 of the 569 pairwise comparisons were concordant, allowing us to strongly reject the null hypothesis that there is no association between protein detection and mRNA abundance ($p = 3.0 \times 10^{-14}$). Hence, on a bulk level, mRNA expression levels are indeed correlated to detection by proteomics. The fully discordant cases may represent genes whose mRNA and protein products are regulated via posttranscriptional mechanisms (Klausner and Harford, 1989), although some may reflect noise in the measurements.

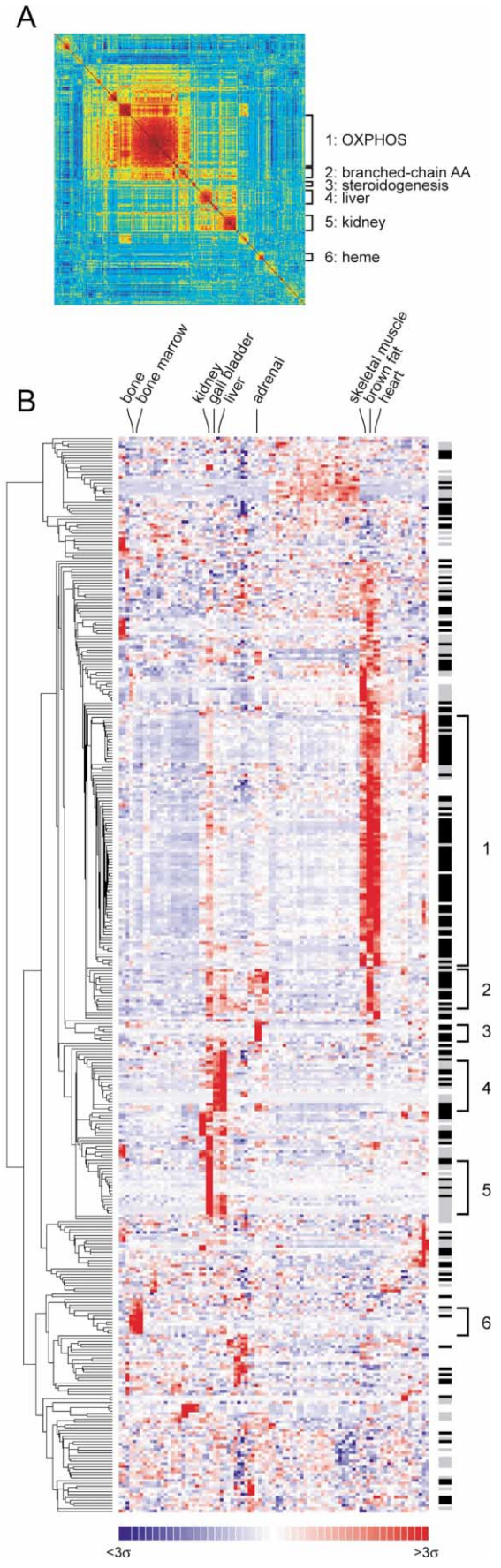
Abundance Differences across Tissues

We next sought to investigate the degree to which mito-A transcripts and proteins exhibit compositional differences across tissues. Of course, apparent absence of a gene product in these experiments cannot be distinguished from a very low level of expression, so these surveys should be interpreted as revealing differences in the abundance of mitochondrial components across tissues.

Of the 236 previously known mitochondrial proteins that were detected in our proteomic survey (Figure 1A), about 40% were detected in all four tissues. mRNA expression measures were available for 168 of these genes (Su et al., 2002). Using a previously established criterion that a gene is "expressed" (see Experimental Procedures), we found that 57% of these genes were expressed in all four tissues.

The fact that only about one-half of gene products are detected in all four tissues could reflect true differences in the abundance of these components or an artifact from random under-sampling of the tissues by our methodologies. To distinguish these possibilities, we considered five well-matched experimental tissue batches: two independent liver samples and one sample from each of brain, heart, and kidney. We then computed the conditional probability that a protein detected in the first liver sample is also detected in a specified one of the other samples. The conditional probability of detecting the protein in another sample is 92% for the second liver sample (indicating good, although not perfect reproducibility) but averages only 79% for brain, heart and kidney. The probability of detection in a distinct tissue is therefore $\sim$85% as large as the probability of redetection in the same tissue. The diversity of mitochondrial protein composition across different tissues is thus substantially greater than can be accounted for by experimental noise alone, indicating that there are differences in protein composition between the tissues.

We therefore sought to model the degree to which mito-A transcripts and proteins are shared across different tissues. We can define P_i as the probability that a given protein is found in a set of $i + 1$ tissues, conditional on being found in a specific set of i tissues, averaged over all distinct subsets of tissues (see Experimental Procedures). Focusing on protein expression in four, well-matched tissue batches, we find that $P_1 = 0.79$, $P_2 = 0.89$, and $P_3 = 0.93$. And likewise, using the RNA expression data, we find $P_1 = 0.89$, $P_2 = 0.93$, and $P_3 = 0.94$. These results are broadly consistent with a simple theoretical model in which half of the mitochondrial components are present in all tissues and the other half being tissue specific such that they occur in a given tissue with 50% probability. Out of a hundred mitochondrial proteins, two tissues would then each contain the 50 ubiquitous mitochondrial proteins as well as 25 tissue-specific proteins, of which half would be shared (i.e., 62.5/75 or 83% proteins shared). In this way, this



simple model would result in $P_1 = 0.83$, $P_2 = 0.90$, and $P_3 = 0.94$ (see Experimental Procedures), very close to the degree of protein and transcript sharing across tissues.

The notion that only a subset of mitochondrial proteins are shared (that is, present at detectable expression levels) among mitochondria from two different tissues is consistent with previous studies demonstrating morphological and functional specialization of this organelle. The consistency of RNA and protein expression analysis is important, since proteomics, but not RNA expression analysis, allows us to control for organelle copy number, which can vary across cell types.

Subnetworks of Mitochondrial Genes

Numerous studies have shown that functionally related sets of genes can often exhibit patterns of correlated gene expression (DeRisi et al., 1997). We were interested in determining whether subsets of the 591 mito-A genes might exhibit distinct patterns of expression across different tissues. For 386 of the 591 mito-A genes, mRNA expression measures were available in a mouse gene expression compendium containing data across 45 tissues (Su et al., 2002).

We calculated pairwise correlations and performed hierarchical clustering of these 386 gene expression profiles (Figure 3). There are several striking mitochondrial gene modules (Figure 3A), which we define as clusters of genes showing strong expression correlation across the 45 tissues (see Supplemental Table S3 for annotations of these genes). These modules include previously known as well as newly identified members of mito-A (see bar labeling in Figure 3B). As shown in Figure 3B, mitochondrial gene expression profiles vary tremendously from tissue to tissue, suggesting a regulatory diversity that is consistent with the compositional diversity noted above.

Each of these gene modules is characterized by tightly correlated gene expression across the tissue compendium, but some are heavily enriched by members of well-known biochemical pathways. Members of these modules likely share transcriptional regulatory mechanisms as well as cellular functions. And because many of the newly identified mitochondrial genes (Figure 3B) lie within these modules, they provide an initial step

Figure 3. Modules of Mitochondrial Genes

(A) Pairwise correlation matrix for the 386 mitochondrial genes represented on the GNF mouse tissue compendium (Su et al., 2002). Red represents strong positive correlation, blue represents strong negative correlation. Dominant gene modules are labeled 1–6 with annotations. (B) mRNA expression profile for 386 mitochondrial genes (rows) across 45 different mouse tissues performed in duplicate (columns) in the GNF mouse compendium. Genes and tissues were hierarchically clustered and visualized using DCHIP (Schadt et al., 2001). Selected tissues are labeled at the top of the panel. Evidence that a gene encodes a mitochondrial protein is indicated by the bars placed to the right of the correlogram: white, previously annotated but not found in proteomics; gray, not previously annotated but identified by proteomics; and black, previously annotated and found in proteomics. Annotations of these 386 genes are available in Supplemental Table S3 (available online at http://www.cell.com/cgi/content/full/115/5/629/DC1).

toward an understanding of their function. Of the 104 probe-sets corresponding to newly identified mitochondrial proteins, 38 fall within one of these modules, providing them with a preliminary functional context (Supplemental Table S3).

Modules Enriched in Genes of Oxidative Phosphorylation

Perhaps the most striking subnetwork of mitochondrial genes is module 1, consisting of 90 genes related to oxidative phosphorylation (OXPHOS), β-oxidation, and the TCA cycle, and are highly expressed in brown fat, skeletal muscle, and heart (Figure 3B). This module includes 13 probe-sets corresponding to 12 newly identified mito-A genes. Previous work has identified the bovine homolog of one of these proteins, Grim19, as a component of complex I of the electron transport chain (Fearnley et al., 2001). The other proteins, which, to our knowledge, have not been associated with oxidative metabolism, include Usmg5, Np15, D10Ertd214e, 2010100O12Rik, 2610207I16, Rik1110018B13Rik, 2610205H19Rik, 0610041L09Rik, 0610006O17Rik, 2310005O14Rik, and Gbas.

We recently showed that tightly correlated members of the OXPHOS biochemical pathway exhibit reduced gene expression in human diabetes (Mootha et al., 2003b). It will be interesting to determine whether this property extends to this module, as well as what regulatory mechanisms account for this striking pattern of correlated gene expression.

Other Gene Modules

Several of the other gene modules have clear functional associations. For example, module 2 contains 15 genes, a large fraction of which are involved in branched chain amino acid metabolism. This module also contains two of the four known biotin-dependent carboxylases. These pathways are highly expressed in brown fat—but not skeletal muscle and heart—as well as in liver, kidney, adrenal, and testis, raising hypotheses about tissue capacities for amino acid metabolism.

It has long been known that adrenal mitochondria play a central role in steroidogenesis. Several of the enzymes involved in this pathway, including steroidogenic acute regulatory protein (Star), ferredoxin reductase, and ferredoxin, are all found in module 3. Ferredoxin reductase is the sole mammalian P450 NADPH reductase, transferring electrons from NADPH, via ferredoxin, to cholesterol. Under substrate limiting conditions, it is known that electrons from this system can generate a large load of reactive oxygen species (ROS) that can be quenched by scavenging enzymes (Hwang et al., 2001). Interestingly, module 3 also includes the ROS scavenger peroxiredoxin 3, which may serve this function. Two known heat shock proteins, Hspe1 and Hspd1, are also coordinately expressed in this module, though their role in steroid metabolism is not known.

Module 6 includes genes involved in heme biosynthesis that form a tight cluster highly expressed in bone and in bone marrow. Of the four mitochondrial enzymes involved in heme biosynthesis (Stryer, 1988), aminolevulinic acid synthetase, ferrochelatase, and coproporphyrinogen oxidase are found within this module. Several genes encoding heme-containing proteins or involved with heme metabolism are also expressed in this cluster, as well as a newly identified mitochondrial protein, 1110021D01Rik.

The mitochondrial modules represent a first step toward a systematic, functional characterization of mitochondrial genes. The modules can be used for functional discovery as well as for discovering *cis*-elements involved in organelle remodeling.

Mitochondrial Gene Expression Neighborhood

The above studies focused on those genes whose products are physically localized or associated with the mitochondrion and attempted to characterize subnetworks within this group. We next sought to systematically identify those genes that are coregulated with this set. We refer to this "mitochondrial neighborhood" as mito-CR, for mitochondria-co-regulated. The mito-CR set may contain genes not in the mito-A set and may encode proteins that are not physically associated with mitochondria but which function coordinately with mitochondrial processes.

To define the mitochondrial neighborhood, we used the neighborhood index (N_{100}), a previously described statistic that measures a given gene's expression similarity to a target gene set (Mootha et al., 2003a). For a given gene, the mitochondria neighborhood index is defined as the number of mito-A genes among its nearest 100 expression neighbors. We applied neighborhood analysis to all genes in the mouse expression atlas (Figure 4), which includes a total of 10,043 genes, including 386 of the mito-A genes. We sought a threshold for N_{100} that would define the boundary of the neighborhood. We found that an N_{100} value of at least 15 (see Experimental Procedures) would be expected to occur by chance approximately 1 in 20 times, after correcting for multiple hypothesis testing (corresponding to a global p value of $\sim$0.05).

A total of 643 genes have $N_{100} \geq 15$. We define this as the expression neighborhood of the mito-A set, and we interpret these genes as being coregulated with mitochondrial genes (see the entire rank ordered list in Supplemental Table S3). This group corresponds to only 6.4% of all the genes studied, but it contains 45% of the mito-A genes (7-fold enrichment). The list includes 48 that are newly mitochondrial based on our proteomic survey and 18 that were previously known to be mitochondrial but not detected by our proteomic survey.

Importantly, the expression neighborhood mito-CR includes 470 genes that are not present in the mito-A set itself. Some of these genes may encode proteins that are physically present in mitochondria but were missed in our proteomic survey, while others may encode proteins that are functionally related to mitochondria but not physically associated. The neighborhood mito-CR thus provides a catalog of genes that are likely functionally relevant to mitochondrial biology and is complementary to the proteomic approach that identified proteins resident in this organelle.

Transcriptional Regulators within the Mitochondrial Neighborhood

Because tissue-specific transcription factors are often involved in specifying tissue differentiation, we rea-

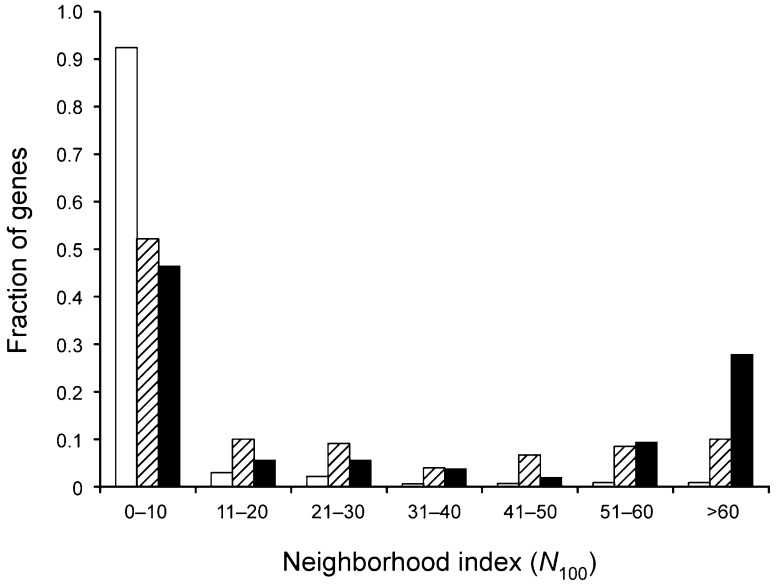


Figure 4. Mitochondria Neighborhood Analysis

The mitochondria neighborhood index (N_{100}) is defined as the number of mito-A genes that occur within the nearest 100 expression neighbors of a given gene (Mootha et al., 2003a). The distribution of N_{100} is plotted for all genes (white), mito-A genes that are not identified as ancestral (hashed), and for the ancestral mito-A genes (black).

soned that the expression neighborhood might contain genes encoding transcriptional regulators of organelle biogenesis (Table 1). While none of these factors have previously been shown to exhibit expression patterns correlated with mitochondrial genes, several have previously been implicated in mitochondrial biology. For example, Pparγ, Pparα, and Esrrα are nuclear receptors that are involved in adipogenesis and fatty acid metabolism and coactivated by PGC-1α, a regulator of mitochondrial biogenesis (Puigserver and Spiegelman, 2003). A number of other transcription factors, including Nfix, Tbx6, and Klf9 exhibit patterns of correlated expression

Table 1. Genes in the Mitochondria Expression Neighborhood with Putative Roles in DNA Maintenance and Repair

Gene name	Gene symbol	N_{100}
Transcriptional regulators		
MyoD family inhibitor	*Mdfi*	63
nuclear factor I/X	*Nfix*	60
zinc finger protein 288	*Zfp288*	56
T-box 6	*Tbx6*	49
Cofactor required for Sp1 transcriptional activation subunit 2	*Crsp2*	47
RIKEN cDNA 9130025P16 gene	*9130025P16Rik*	46
Kruppel-like factor 9	*Klf9*	43
EGL nine homolog 1	*Egln1*	39
Estrogen related receptor, alpha	*Esrra*	36
nuclease sensitive element binding protein 1	*Nsep1*	34
sirtuin 1 (silent mating type information regulation 2, homolog) 1	*Sirt1*	31
peroxisome proliferator activated receptor alpha	*Ppara*	29
metastasis associated 1-like 1	*Mta1l1*	28
NK2 transcription factor related, locus 5 (Drosophila)	*Nkx2-5*	27
cardiac responsive adriamycin protein	*Crap*	24
homeo box D8	*Hoxd8*	21
nuclear receptor subfamily 1, group I, member 2	*Nr1i2*	21
nuclear receptor subfamily 1, group H, member 3	*Nr1h3*	20
cellular nucleic acid binding protein	*Cnbp*	19
transcription factor 2	*Tcf2*	19
Est2 repressor factor	*Erf*	19
nuclear receptor subfamily 5, group A, member 1	*Nr5a1*	18
nuclear factor, erythroid derived 2,-like 1	*Nfe2l1*	18
zinc finger protein 30	*Zfp30*	17
peroxisome proliferator activated receptor gamma	*Pparg*	17
cAMP responsive element binding protein 1	*Creb1*	15
SRY-box containing gene 6	*Sox6*	15
CCAAT/enhancer binding protein (C/EBP), alpha	*Cebpa*	15
DNA repair		
mutL homolog 1	*Mlh1*	29
mutS homolog 5	*Msh5*	24
excision repair cross-complementing rodent repair deficiency, complementation group 1	*Ercc1*	15

that make them candidates for involvement in organelle remodeling. Surprisingly, the nutrient sensor Sir2 is also found within the mitochondrial expression neighborhood. *Sir2* encodes an NAD(+)-dependent histone deacetylase involved in gene silencing, chromosomal stability, and aging. Chromatin remodeling enzymes rely on coenzymes derived from metabolic pathways, including those generated by the mitochondrion. Our observations suggest that *Sir2* and mitochondrial gene expression are coordinately regulated, providing a potential regulatory link between the mitochondrion and the nutrient sensing activities of Sir2.

DNA Repair Enzymes within the Mitochondrial Neighborhood

Identifying proteins involved in mtDNA repair has been extremely challenging. These proteins are believed to occur in low abundance, and when found in mitochondrial preparations, it is difficult to preclude the possibility of nuclear contamination. Although mtDNA mismatch repair activity has been reported in human cells (Mason et al., 2003), a mammalian mtDNA mismatch repair enzyme has not yet been identified. This has been puzzling, since yeast mitochondria have a mutS homolog (Chi and Kolodner, 1994).

The mitochondria expression neighborhood contains genes encoding two mammalian mismatch repair enzymes, *Msh5* and *Mlh1*. Msh5, a mammalian MutS homolog, has previously been described to be required for meiotic progression (Edelmann et al., 1999), but no association with mitochondria has been noted. Interestingly, Mlh1, a mammalian MutL homolog, has previously been reported to be involved in repair of DNA following oxidative stress (Hardman et al., 2001). Supporting the notion that Msh5 and Mlh1 function in mitochondria, we find evidence by bioinformatic analysis (Claros, 1995; Nakai and Horton, 1999) that these two proteins contain reasonable mitochondrial targeting sequences.

While our findings by no means prove that these enzymes are involved in mtDNA repair, their strong correlated expression with mitochondrial genes and their mitochondrial targeting sequences make them very attractive candidates for mediating these repair activities.

Mitochondrial Gene History

The dual origin hypothesis suggests that the modern mitochondrial proteome can be divided into two groups, consisting of proteins derived from their eubacterial ancestry, while the remaining proteins have been acquired over the last 2 billion years (Andersson et al., 1998; Karlberg et al., 2000). We consider a gene to be an ancestral mitochondrial gene if it has a detectable ortholog in *Rickettsia prowazekii*, the nearest eubacterial relative to mammalian mitochondria (Andersson et al., 1998). Of the mito-A genes for which we had gene expression measures, 54 can be identified as being ancestral (see Experimental Procedures). We find that the ancestral mitochondrial genes tend to have higher local enrichment by mitochondrial genes, as assayed by the neighborhood index (Figure 4). Interestingly, previous studies have suggested that mRNA populations encoding ancestral mitochondrial proteins tend to be translated at polysomes associated with the mitochondrial outer membrane (Marc et al., 2002). The current result (Figure 4) hints that ancestral mitochondrial genes may exhibit a pattern of gene expression distinct from the other mitochondrial proteins, hence providing an additional signature of their history.

Conclusion

We have performed a large-scale proteomic survey of mitochondria purified from four different mouse tissues and have analyzed the results in the context of existing annotations and publicly available gene expression profiles. Integration of these datasets provides a first step toward a functional annotation of these newly identified proteins as well as an understanding of the regulatory organization of all mitochondrial genes.

Our proteomic analysis is best thought of as a survey of the abundant mitochondrial proteins. Clearly, the mito-A list is incomplete. Based on comparisons to previously known mitochondrial genes, our proteomic survey appears to have a sensitivity of 55% and thus would be predicted to have missed 133 novel mitochondrial proteins; this would suggest that the true number of mito-A genes is at least 725. Because the well annotated proteins likely represent the more abundant proteins, amenable to analysis by traditional biochemical approaches, this estimate likely represents a lower bound on the mitochondrial proteome. Future proteomic surveys of the mitochondrion aimed at expanding the inventory of mitochondrial proteins may benefit from higher dimensional chromatography and improved sample preparation, more sensitive and quantitative mass spectrometry technologies (Aebersold and Mann, 2003), and perhaps use of genetic strategies (Ozawa et al., 2003). When combined with genome-wide expression microarrays, it should be possible to more comprehensively reconstruct pathways within the mitochondrion and to determine the extent to which mitochondrial diversity extends to other cell types and to lower abundance gene products.

Proteomics and RNA expression profiling provide complementary insights. The mito-A list consists of 591 genes whose products reside in or in close association with the mitochondrion, while the mitochondrial expression neighborhood includes a large group of 643 genes whose transcription profiles are tightly correlated to those of mito-A. The expression neighborhood mito-CR contains a large fraction of the mito-A genes assayed in the expression survey (including some that had not been detected in the proteomic survey), as well as many additional genes. Some of these additional genes may encode products that actually reside in the mitochondria, while others may encode products that reside elsewhere but are related to mitochondrial biogenesis and function. In the future, it will be valuable to combine insights from complementary approaches, as sensitivity and specificity measures can be improved by combining different sources of experimental evidence.

At present, the mechanisms that achieve cell-type-specific differences in mitochondrial form and function are not known. How a mitochondrion remodels in response to changes in nutrient status and energy demands or in disease states, such as cancer and diabe-

tes, is poorly understood. It is likely that transcriptional mechanisms work in concert with mRNA processing and protein-targeting mechanisms to carefully achieve appropriate enzymatic stoichiometries required for each mitochondrion. Deciphering these mechanisms is an important challenge. Mitochondrial modules serve as an excellent starting point for identifying important *cis*-regulatory elements, and the genes whose protein and RNA expression levels are discordant may guide the identification of new posttranscriptional regulatory mechanisms. Finally, an expanded list of mitochondrial proteins may assist in identifying new organelle targeting sequences.

Given the central role of the mitochondrion in the life and death of the cell, it is likely that the mitochondria-associated genes and those in the expression neighborhood represent a rich source of candidate genes for human disease as well as targets for future drug development. Such therapies may exploit the apparent compositional and regulatory diversity within this organelle to provide treatment specificity for pathways operative in human disease.

Experimental Procedures

Organelle Purification and Sample Preparation

Six- to eight-week-old male mice were subjected to an 8 hr fast and then euthanized. Brain, heart, kidney, and livers were harvested immediately and placed in ice-cold saline. Mitochondria were isolated using differential centrifugation as previously described and purified with a Percoll gradient (Mootha et al., 2003a). To test the purity of these preparations, we performed Western blot analysis as previously described, using antibodies directed against known mitochondrial proteins (cytochrome *c*, COXIV, and VDAC) as well as antibodies directed against calreticulin (a marker for the endoplasmic reticulum) and for SNAP25 (a marker for synaptosomes). The proteins were then solubilized, size separated, and digested as previously described (Mootha et al., 2003a).

Tandem Mass Spectrometry

Liquid chromatography tandem mass spectrometry (LC-MS/MS) was performed on QSTAR pulsar quadrupole time of flight mass spectrometers (AB/MDS Sciex, Toronto) as described previously (Mootha et al., 2003a). Tandem mass spectra were searched against the NCBInr database (October 2002) with tryptic constraints and initial mass tolerances $<$0.13 Da in the search software Mascot (Matrix Sciences, London). Only peptides achieving a Mascot score above 25 and containing a sequence tag of at least three consecutive amino acids were accepted.

Curation of Previously Annotated Mitochondrial Proteins

We used two key sources to identify previously annotated proteins. First, we downloaded the human and mouse protein sequences at MITOchondria Project (Scharfe et al., 2000). We also downloaded the 199 human and 290 mouse protein sequences annotated at LocusLink (http://www.ncbi.nlm.nih.gov/LocusLink) as having a mitochondrial subcellular localization based on gene ontology terminology (GO:0005739) (January 2003). We also included in our master list the 13 mtDNA encoded proteins, based on LocusLink annotation.

A Nonredundant List of Mitochondrial Proteins

FASTA sequences corresponding to the previously annotated mitochondrial proteins, newly identified mitochondrial proteins, and the mouse Reference Sequences (August 2003) (Maglott et al., 2000) were merged. These were then collapsed into distinct protein clusters using a command-line version of blastclust (http://www.ncbi.nlm.nih.gov/BLAST/). We required that members of a cluster demonstrate 70% sequence identity over 50% of the total length, not requiring a reciprocal relationship to exist. Clusters containing multiple reference sequences were then broken using a higher stringency blastclust, in which we required 90% identity over 50% of the length and then manually reviewed. Some clusters were eliminated if they consisted of sequences that were annotated as fragments. Each protein cluster consists of accessions corresponding to previously annotated mitochondrial proteins, as well as accessions of proteins identified directly in the proteomics experiments. Hence, some clusters are supported by proteomics alone or by annotations alone, while others have support from both (Figure 1A). Most of these clusters also contain a reference sequence, which serves as an exemplar representative sequence for that cluster. Some clusters did not have reference sequences, so a mouse protein sequence was manually identified through iterative NCBI BLAST routines.

This procedure resulted in a total of 601 protein clusters (Supplemental Table S1 available online at http://www.cell.com/cgi/content/full/115/5/629/DC1). Ten clusters consisted of actin, hemoglobin, keratin, lysozyme, trypsin, or tubulin. These were flagged as expected contaminants. While they are included in the list in Supplemental Table S1, they were eliminated from all subsequent analyses. Hence, there are a total of 591 mito-A protein clusters (Figure 1A). Of these 591 clusters, 399 were previously annotated as being mitochondrial in LocusLink, in MITOP, or based on the name of the gene.

Note that the data presented in Supplemental Table S1 comes from a total of 12 experimental batches, where each batch corresponds to a single tissue from a single mouse. We performed more proteomics experiments on mouse liver, and all this data is included in our Supplemental Table S1. However, in analyses of mRNA:protein concordance and in analysis of the compositional diversity, we limited our analyses to four well matched batches, corresponding to mouse brain, heart, kidney, and liver.

Cell Culture and Transfection

GFP-tagged proteins were generated for five human homologs of the identified proteins using the Gateway cloning system (Life Technology) as described by the manufacturer. Approximately 6×10^5 HEK 293 cells were seeded on coverslips in a 6 well-plate and incubated overnight in DMEM supplemented with 10% FBS, 100 U/ml penicillin and 100 μg/ml streptomycin at 37°C in a humidified 5% carbon dioxide atmosphere. Six microliters Genejammer (Stratagene) in 100 μl DMEM was incubated 10 min at room temperature and 1 μg DNA was added. The mixture was then incubated for a further 10 min. Nine hundred microliters of DMEM with 10% FBS and the transfection mixture were combined and added to the cells. After 3 hr, 1 ml of DMEM with 10% FBS and antibiotics were added. These transfected cells were then incubated for 48 hr.

Immunofluorescence Microscopy

Transfected cells were washed with PBS and fixed with 4% paraformaldehyde in phosphate buffered saline (PBS) for 15 min at room temperature. Cells were washed three times with 100 mM glycine in PBS and permeabilized by a three minute incubation in PBS with 0.2% Triton X-100. Then the cells were incubated in 1% BSA to prevent nonspecific staining. Mitochondria were stained with α-GRP-75 antibody (Stressgen) diluted 1:200 in 1% BSA in PBS for one hour. Cells were washed three times with PBS and incubated with 10 μg/ml of the secondary antibody Alexa Fluor 568 goat anti-mouse IgM A21043 (Molecular Probes) for 30 min. After three washes with PBS the coverslips were mounted in anti-fade mounting media and the subcellular distribution of these proteins analyzed by confocal microscopy.

RNA/Protein Concordance Test

We developed the RNA/protein concordance test to determine whether there is significant association between protein detection in a proteomics experiment and mRNA abundance in a microarray experiment.

Consider the pair of tissues, i,j, where $i,j \in$ {brain, heart, kidney, liver}. For a given gene, G, we let $M(G,k)$ represent the gene expression level of gene G in tissue k. Let $P(G,k)$ be an indicator variable that is 0 if the protein product of gene G is not found in tissue k, and 1 if the protein product is found in tissue k. We set

$$x(i,j) = \begin{cases} 1, \text{ if } M(G,i) > M(G,j) \text{ and } P(G,j) > P(G,j) \\ -1, \text{if } M(G,i) > M(G,j) \text{ and } P(G,i) < P(G,j) \\ 0, \text{ otherwise} \end{cases}$$

and define the concordance for gene G, C_G, by

$$C_G = \sum_{i,j} x(i,j).$$

In the null case in which there is no association between protein detection and mRNA abundance, the expected concordance for a gene is 0. The variance in concordance for G, denoted by v_G, depends on the number of tissues, k, in which that gene's product was detected. If the protein product was detected in $k = 0$ or 4 tissues, then C_G and v_G are both 0. If the protein product was detected in exactly one or three tissues, then the possible concordance measures are $+3$, $+1$, 0, -1, or -3. Again, because the expected concordance is 0, the variance under the null model is simply $[(+3)^2 + (+1)^2 + (0)^2 + (-1)^2 + (-3)^2]/4 = 5$. Finally, if the protein product was detected in exactly two tissues, then the possible concordance measures for the gene are $+4$, $+2$, 0, 0, -2, or -4, and hence, the null variance is $[(+4)^2 + (+2)^2 + (0)^2 + (0)^2 + (-2)^2 + (-4)^2]/6 = 20/3$. More generally, if the protein was detected in k tissues out of n that were surveyed, it can be shown that $v_G = k(n-k)(n+1)/3$.

We compute the observed concordance and null variance for every gene and sum over all genes. Our test statistic then becomes

$$C = \sum_G c_G \Big/ \sqrt{\sum_G v_g},$$

which is approximately normally distributed with mean 0 and variance 1 in the null case where there is no association between RNA abundance and protein detection.

Compositional Diversity Across Tissues

Mitochondrial gene products show distinct patterns of expression based on protein and RNA expression. These patterns of distribution motivate a simple model that describes core mitochondrial proteins versus those that are specialized to any set of cell types. Consider a set of $i + 1$ tissues, S_{i+1}, as well as a distinct subset S_i, i.e., $S_i \subset S_{i+1}$, where $i > 0$. We are interested in the probability that a given gene product is found in S_{i+1} conditional that it is found in S_i, or simply $T(S_{i+1}, S_i) = P$ (gene product is found in S_{i+1}|gene product is found in S_i). We define P_i as the average $T(S_{i+1}, S_i)$ over all selections of $S_i \subset S_{i+1}$. When we assessed compositional diversity using RNA expression levels, we interpreted an RNA expression level greater than 200 as "expressed" (Su et al., 2002).

These average conditional probabilities P_i can also be modeled. Imagine that a fraction f of all mitochondrial proteins are ubiquitous (i.e., expressed in all cell types with probability 1) and that a fraction $1 - f$ are not ubiquitous, but rather, appear in a given tissue with probability p. Then $P_i = (f + (1 - f)p^{i+1})/(f + (1 - f)p^i)$.

DNA Microarray Analysis

To identify Affymetrix probe-sets corresponding to each protein cluster, we mapped the exemplar protein accession to its LocusLink ID, then to its Unigene cluster, and then identified the corresponding Affymetrix MG-U74Av2 probe set using the NetAffx website (http://www.affymetrix.com) and its annotation tables (August 2003). Note that this automated mapping does not guarantee every protein is mapped to a probe-set ID; the majority of mito-A exemplars could be mapped to Affymetrix probe sets, but we know that the automated procedure has failed to provide a corresponding probe-set. Note that the mapping is largely 1:1, but there are some many:many mappings.

The GNF mouse expression atlas (Su et al., 2002) was downloaded from its website (http://expression.gnf.org). In comparisons of protein detection and mRNA abundance, we used the mRNA expression level for a given tissue averaged over the replicates, since the GNF mouse expression atlas includes duplicates for each tissue. Because we performed the proteomic survey on whole brain, we simply compared to the average expression of all brain samples in the GNF mouse atlas. Hierarchical clustering was performed using DCHIP (Schadt et al., 2001), using $1 - r$ as the distance metric, where r is the Pearson correlation coefficient, and the relative expression levels are displayed.

Neighborhood analysis was performed using a stand-alone Perl script that was previously described (Mootha et al., 2003a). We used the GNF mouse expression atlas for these analyses. Of the 10,043 genes represented in this atlas, 386 correspond to the mito-A genes. These 386 genes form the target gene set in neighborhood analysis. For each query gene in the atlas, we rank order all other genes in the atlas on the basis of Euclidean distance of gene expression. The neighborhood index, N_{100}, is defined as the number of mito-A genes within the top 100 ranking genes. If the 386 mito-A genes were a random subset of the 10,043 genes, then the probability of detecting at least 15 mito-A genes in a random sample of 100 genes is 6.7×10^{-6}, corresponding to a Bonferroni corrected p-value (for the 10,043 measures made) of 0.07.

Identification of Ancestral Mitochondrial Genes

We downloaded the consensus FASTA sequences for the genes represented on the Affymetrix MG-U74Av2 oligonucleotide array from the NetAFFX (Liu et al., 2003) website (http://www.affymetrix.com). We performed a blastx comparison of these sequences against the *Rickettsia prowazekii* protein sequences, downloaded from the NCBI, and then performed a tblastn comparison of the bacterial protein sequences against the consensus FASTA sequences. In both analyses, default blast parameters were used in conjunction with the BLOSUM62 scoring matrix. We defined an ancestral gene as one achieving a BLASTX $E < 0.01$ and having a reciprocal best match in the above BLAST analysis.

Acknowledgments

We are grateful to B. Gewurz, A. Paulovich, K. Lindblad-Toh, M. Zody, P. Tamayo, M. Reich, J. Hirschhorn, M. Daly, D. Bogenhagen, and colleagues at MDS Proteomics for valuable assistance, fruitful discussions, and thoughtful comments on the manuscript. We thank L. Gaffney for preparing illustrations. V.K.M. was supported by a physician postdoctoral fellowship from Howard Hughes Medical Institute. M.M.'s laboratory at the University of Southern Denmark is supported by a grant by the Danish National Research Foundation.

Received: June 25, 2003
Revised: October 16, 2003
Accepted: November 6, 2003
Published: November 25, 2003

References

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. Nature *422*, 198–207.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H., and Kurland, C.G. (1998). The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature *396*, 133–140.

Brenner, C., Bieganowski, P., Pace, H.C., and Huebner, K. (1999). The histidine triad superfamily of nucleotide-binding proteins. J. Cell. Physiol. *181*, 179–187.

Chi, N.W., and Kolodner, R.D. (1994). Purification and characterization of MSH1, a yeast mitochondrial protein that binds to DNA mismatches. J. Biol. Chem. *269*, 29984–29992.

Claros, M.G. (1995). MitoProt, a Macintosh application for studying mitochondrial proteins. Comput. Appl. Biosci. *11*, 441–447.

Da Cruz, S., Xenarios, I., Langridge, J., Vilbois, F., Parone, P.A., and Martinou, J.C. (2003). Proteomic analysis of the mouse liver mitochondrial inner membrane. J. Biol. Chem. *278*, 41566–41571.

DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science *278*, 680–686.

Edelmann, W., Cohen, P.E., Kneitz, B., Winand, N., Lia, M., Heyer, J., Kolodner, R., Pollard, J.W., and Kucherlapati, R. (1999). Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. Nat. Genet. *21*, 123–127.

Endele, S., Fuhry, M., Pak, S.J., Zabel, B.U., and Winterpacht, A. (1999). LETM1, a novel gene encoding a putative EF-hand Ca(2+)-binding protein, flanks the Wolf-Hirschhorn syndrome (WHS) critical region and is deleted in most WHS patients. Genomics *60*, 218–225.

Fearnley, I.M., Carroll, J., Shannon, R.J., Runswick, M.J., Walker, J.E., and Hirst J. (2001). GRIM-19, a cell death regulatory gene product, is a subunit of bovine mitochondrial NADH:ubiquinone oxidoreductase (complex I). J. Biol. Chem. *276*, 38345–38348.

Fernandez-Vizarra, E., Lopez-Perez, M.J., and Enriquez, J.A. (2002). Isolation of biogenetically competent mitochondria from mammalian tissues and cultured cells. Methods *26*, 292–297.

Ghadially, F.N. (1997). Ultrastructural Pathology of the Cell and Matrix, Volume 1 (Boston: Butterworth-Heinemann).

Griffin, T.J., Gygi, S.P., Ideker, T., Rist, B., Eng, J., Hood, L., and Aebersold, R. (2002). Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. Mol. Cell. Proteomics *1*, 323–333.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. *19*, 1720–1730.

Hardman, R.A., Afshari, C.A., and Barrett, J.C. (2001). Involvement of mammalian MLH1 in the apoptotic response to peroxide- induced oxidative stress. Cancer Res. *61*, 1392–1397.

Hockenbery, D.M., Oltvai, Z.N., Yin, X.M., Milliman, C.L., and Korsmeyer, S.J. (1993). Bcl-2 functions in an antioxidant pathway to prevent apoptosis. Cell *75*, 241–251.

Hwang, P.M., Bunz, F., Yu, J., Rago, C., Chan, T.A., Murphy, M.P., Kelso, G.F., Smith, R.A., Kinzler, K.W., and Vogelstein, B. (2001). Ferredoxin reductase affects p53-dependent, 5-fluorouracil-induced apoptosis in colorectal cancer cells. Nat. Med. *7*, 1111–1117.

Karlberg, O., Canback, B., Kurland, C.G., and Andersson, S.G. (2000). The dual origin of the yeast mitochondrial proteome. Yeast *17*, 170–187.

Klausner, R.D., and Harford, J.B. (1989). cis-trans models for post-transcriptional gene regulation. Science *246*, 870–872.

Kluck, R.M., Bossy-Wetzel, E., Green, D.R., and Newmeyer, D.D. (1997). The release of cytochrome c from mitochondria: a primary site for Bcl-2 regulation of apoptosis. Science *275*, 1132–1136.

Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., et al. (2002). Subcellular localization of the yeast proteome. Genes Dev. *16*, 707–719.

Lian, Z., Wang, L., Yamaga, S., Bonds, W., Beazer-Barclay, Y., Kluger, Y., Gerstein, M., Newburger, P.E., Berliner, N., and Weissman, S.M. (2001). Genomic and proteomic analysis of the myeloid differentiation program. Blood *98*, 513–524.

Liu, G., Loraine, A.E., Shigeta, R., Cline, M., Cheng, J., Valmeekam, V., Sun, S., Kulp, D., and Siani-Rose, M.A. (2003). NetAffx: Affymetrix probesets and annotations. Nucleic Acids Res. *31*, 82–86.

Lopez, M.F., Kristal, B.S., Chernokalskaya, E., Lazarev, A., Shestopalov, A.I., Bogdanova, A., and Robinson, M. (2000). High-throughput profiling of the mitochondrial proteome using affinity fractionation and automation. Electrophoresis *21*, 3427–3440.

Lupas, A.N., and Martin, J. (2002). AAA proteins. Curr. Opin. Struct. Biol. *12*, 746–753.

Maglott, D.R., Katz, K.S., Sicotte, H., and Pruitt, K.D. (2000). NCBI's LocusLink and RefSeq. Nucleic Acids Res. *28*, 126–128.

Mann, M., and Wilm, M. (1994). Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal. Chem. *66*, 4390–4399.

Marc, P., Margeot, A., Devaux, F., Blugeon, C., Corral-Debrinski, M., and Jacq, C. (2002). Genome-wide analysis of mRNAs targeted to yeast mitochondria. EMBO Rep. *3*, 159–164.

Mason, P.A., Matheson, E.C., Hall, A.G., and Lightowlers, R.N. (2003). Mismatch repair activity in mammalian mitochondria. Nucleic Acids Res. *31*, 1052–1058.

Mootha, V.K., Wei, M.C., Buttle, K.F., Scorrano, L., Panoutsakopoulou, V., Mannella, C.A., and Korsmeyer, S.J. (2001). A reversible component of mitochondrial respiratory dysfunction in apoptosis can be rescued by exogenous cytochrome c. EMBO J. *20*, 661–671.

Mootha, V.K., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Delmonte, T., Villeneuve, A., Sladek, R., Xu, F., et al. (2003a). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. Proc. Natl. Acad. Sci. USA *100*, 605–610.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., et al. (2003b). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.

Nakai, K., and Horton, P. (1999). PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. *24*, 34–36.

Neupert, W. (1997). Protein import into mitochondria. Annu. Rev. Biochem. *66*, 863–917.

Ozawa, T., Sako, Y., Sato, M., Kitamura, T., and Umezawa, Y. (2003). A genetic approach to identifying mitochondrial proteins. Nat. Biotechnol. *21*, 287–293.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551–3567.

Puigserver, P., and Spiegelman, B.M. (2003). Peroxisome proliferator-activated receptor-γ coactivator 1α (PGC-1α): transcriptional coactivator and metabolic regulator. Endocr. Rev. *24*, 78–90.

Rabilloud, T., Kieffer, S., Procaccio, V., Louwagie, M., Courchesne, P.L., Patterson, S.D., Martinez, P., Garin, J., and Lunardi, J. (1998). Two-dimensional electrophoresis of human placental mitochondria and protein identification by mass spectrometry: toward a human mitochondrial proteome. Electrophoresis *19*, 1006–1014.

Samuel, S.J., Tzung, S.P., and Cohen, S.A. (1997). Hrp12, a novel heat-responsive, tissue-specific, phosphorylated protein isolated from mouse liver. Hepatology *25*, 1213–1222.

Schadt, E.E., Li, C., Ellis, B., and Wong, W.H. (2001). Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. J. Cell. Biochem. Suppl. (*Suppl*), 120–125.

Scharfe, C., Zaccaria, P., Hoertnagel, K., Jaksch, M., Klopstock, T., Dembowski, M., Lill, R., Prokisch, H., Gerbitz, K.D., Neupert, W., et al. (2000). MITOP, the mitochondrial proteome database: 2000 update. Nucleic Acids Res. *28*, 155–158.

Stryer, L. (1988). Biochemistry, 3rd Edition (New York: W.H. Freeman and Company).

Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A., et al. (2002). Large-scale analysis of the human and mouse transcriptomes. Proc. Natl. Acad. Sci. USA *99*, 4465–4470.

Taylor, S.W., Fahy, E., Zhang, B., Glenn, G.M., Warnock, D.E., Wiley, S., Murphy, A.N., Gaucher, S.P., Capaldi, R.A., Gibson, B.W., and Ghosh, S.S. (2003). Characterization of the human heart mitochondrial proteome. Nat. Biotechnol. *21*, 281–286.

Veltri, K.L., Espiritu, M., and Singh, G. (1990). Distinct genomic copy number in mitochondria of different mammalian organs. J. Cell. Physiol. *143*, 160–164.

Vijayasarathy, C., Biunno, I., Lenka, N., Yang, M., Basu, A., Hall, I.P., and Avadhani, N.G. (1998). Variations in the subunit content and catalytic activity of the cytochrome c oxidase complex from different tissues and different cardiac compartments. Biochim. Biophys. Acta *1371*, 71–82.

Wallace, D.C. (1999). Mitochondrial diseases in man and mouse. Science *283*, 1482–1488.

Westermann, B., and Neupert, W. (2003). 'Omics' of the mitochondrion. Nat. Biotechnol. *21*, 239–240.