

A Mammalian Organelle Map by Protein Correlation Profiling

Leonard J. Foster,^{1,2} Carmen L. de Hoog,^{1,2} Yanling Zhang,^{3,4} Yong Zhang,^{3,4} Xiaohui Xie,⁵ Vamsi K. Mootha,^{5,6} and Matthias Mann^{1,3,*}

¹ Center for Experimental Bioinformatics (CEBI), Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

² Centre for Proteomics, Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

³ Department of Proteomics and Signal Transduction, Max-Planck Institute for Biochemistry, Martinsried, Germany D-82152

⁴ Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 101300, China

⁵ Broad Institute of Harvard and MIT, Cambridge, MA 02139, USA

⁶ Department of Systems Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA

*Contact: mmann@biochem.mpg.de

DOI 10.1016/j.cell.2006.03.022

SUMMARY

Protein localization to membrane-enclosed organelles is a central feature of cellular organization. Using protein correlation profiling, we have mapped 1,404 proteins to ten subcellular locations in mouse liver, and these correspond with enzymatic assays, marker protein profiles, and confocal microscopy. These localizations allowed assessment of the specificity in published organellar proteomic inventories and demonstrate multiple locations for 39% of all organellar proteins. Integration of proteomic and genomic data enabled us to identify networks of coexpressed genes, *cis*-regulatory motifs, and putative transcriptional regulators involved in organelle biogenesis. Our analysis ties biochemistry, cell biology, and genomics into a common framework for organelle analysis.

INTRODUCTION

For over a century, the eukaryotic cell has been analyzed biochemically and imaged in countless ways to arrive at our modern model of a protective plasma membrane surrounding several membrane-enclosed organelles. Many specialized functions are compartmentalized within these organelles, such as protein and lipid biosynthesis in the endoplasmic reticulum (ER) and Golgi apparatus or oxidative phosphorylation in the mitochondria. Cell biologists have taken a holistic approach to study organelles, using dyes and optical methods to define distinct morphological patterns. Conversely, biochemists have disrupted cells to separate compartments and complexes, usually by density gradient centrifugation, in order to characterize the chemical, physical, or enzymatic properties of each (Figure 1). To aid the interpretation of both biochemical and

microscopic examination of an organelle, certain proteins or enzyme activities that appear to localize exclusively to that organelle are considered markers, essentially defining that compartment.

Recently, proteomics (de Hoog and Mann, 2004) has been applied to study organelle composition. The genetic tractability of *Saccharomyces cerevisiae* has allowed a large fraction of yeast ORFs to be tagged for localization studies (Ross-Macdonald et al., 1999; Kumar et al., 2002; Huh et al., 2003), but such an approach is more challenging in mammalian systems due, in part, to artifacts from overexpression (Simpson et al., 2000). Mass spectrometry-based proteomics (Aebersold and Mann, 2003) is often employed to characterize the protein composition of organelle-enriched fractions. Indeed, protein catalogs are now available for virtually all cytoplasmic organelles as well as most of the major nuclear ones (reviewed in Yates et al., 2005). However, due to the high sensitivity of mass spectrometers and the difficulties inherent in purifying organelles to homogeneity, it has been challenging to distinguish bona fide organellar proteins from those that are contaminating. To address this problem, we previously introduced protein correlation profiling (PCP) (Andersen et al., 2003) to study the human centrosome. In that study, mass spectrometric intensity profiles from centrosomal marker proteins were used to define a consensus profile through a density centrifugation gradient, in direct analogy to Western blotting profiling of gradient fractions. Distribution curves generated from the intensities of tens of thousands of peptides from consecutive fractions established centrosomal proteins by their similarity to the consensus profile using mean squared deviation (χ^2 value). In the present study, we extend PCP to the entire cell (Figure 1A) to localize 1,400 proteins to ten cellular compartments (Figure 1B), and we confirm these data with biochemical and cell biological methods. Furthermore, we integrate the results with functional genomics datasets to gain insights into the biogenesis of cellular organelles.

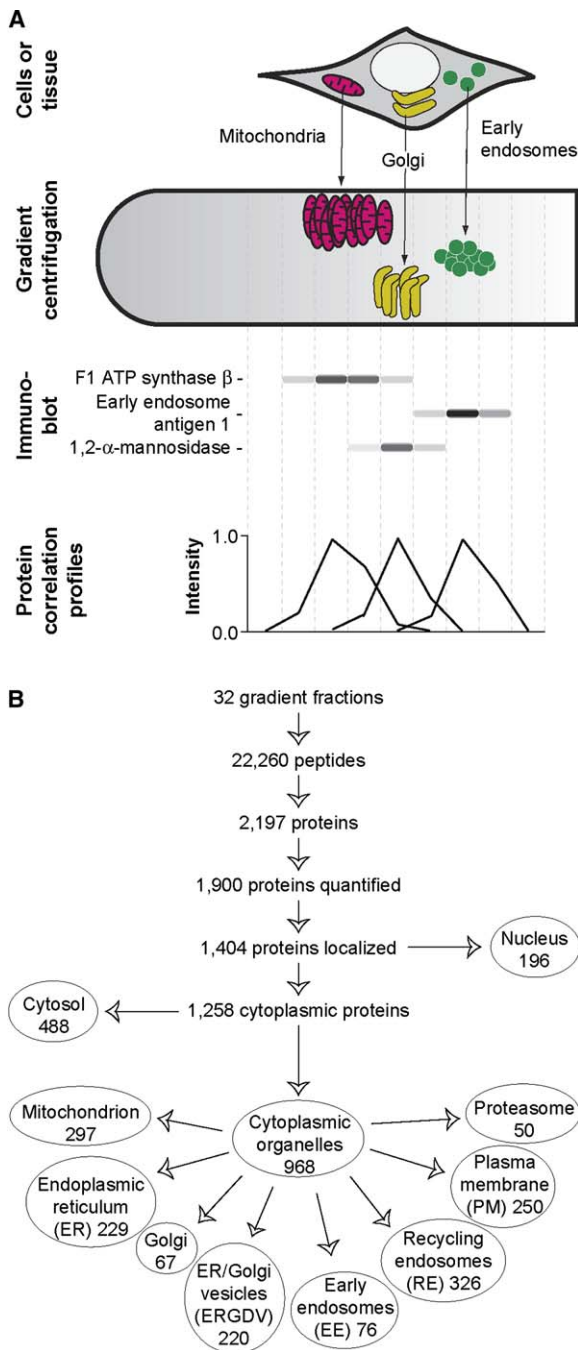


Figure 1. Organelle Profiling with Gradient Centrifugation
 (A) Cell lysates containing intact organelles can be resolved by continuous density gradient centrifugation. Fractions are often analyzed by Western blotting for specific marker proteins in order to define a profile for each organelle. In an analogous way, ion intensities for peptides from marker proteins, measured by mass spectrometric analysis of each gradient fraction and called Protein Correlation Profiles (PCPs), defined organelles.
 (B) Results flow from PCP analysis of mouse liver. The numbers of proteins localized to each organelle are shown along with the acronym used for each. Note that totals do not add up because of overlap between locations.

RESULTS

Organelle Profiling

Mouse liver homogenates were resolved by rate-zonal centrifugation (Figure 2A), and fractions were digested to peptides prior to analysis by liquid chromatography/tandem mass spectrometry (LC/MS/MS) using a linear ion-trap/Fourier transform hybrid mass spectrometer (Syka et al., 2004). Database searches with the fragmentation spectra identified 2,197 unique proteins (Table S1) from 22,256 distinct peptides (Table S2) that, by requiring two sequenced peptides per protein and less than 15 parts per million (p.p.m.) mass accuracy, contain a false positive rate of about one protein in 8,000 (Experimental Procedures). Common peptides identified in neighboring fractions were used to correlate elution times between each fraction (Andersen et al., 2003) so that ion current profiles across the gradient could be generated for the peptides and compiled into protein correlation profiles (Experimental Procedures). Due to the very large amount of analysis time involved, most data discussed here were generated from a single mouse liver, but a localization error rate of 6% was estimated from repeated PCP analysis of mitochondria (see Experimental Procedures and Table S8). Of the identified proteins, the quantitative data for 1900 were reliable enough to evaluate their sub-cellular locations (Figure 1B).

To determine the PCP of well-studied organelles, we examined the profiles of several well-characterized marker proteins, including 130 kDa Golgi phosphoprotein (GPP130, Golgi), AP-2 assembly subunit AP17 (plasma membrane [PM]), early endosome antigen 1 (EEA1, early endosomes [EE]), transferrin receptor 2 (TfR2, recycling endosome [RE]), calnexin (ER), p115 (ER/Golgi-derived vesicles [ERGDV]), and F₁-F₀ ATP synthase β subunit (mitochondria). Each of these markers peaked in different gradient fractions and had distinct profiles; thus at least these seven organelles could be distinguished with confidence (Figure 2B). Markers of other compartments were also observed, but their profiles matched closely to one of the seven mentioned above. In particular, ERGIC-53, a marker for the ER-Golgi intermediate compartment, overlapped very closely with the ER, as has been reported previously (Breuza et al., 2004). Likewise, the profiles of cation-independent mannose 6-phosphate receptor and adaptor-related protein 1 β , markers of the late endosome and trans-Golgi network, respectively, largely overlapped with TfR2 (Tables S3 and S4). This suggests that these compartments migrate similarly in rate-zonal centrifugation and is in agreement with the specialized conditions required for even partial segregation reported by others (Tulp et al., 1998; Hashimoto and James, 2000).

Interestingly, the profile for several proteasome subunits showed these proteins were enriched near 0.62 M sucrose, suggesting that the proteasome complex also migrated during centrifugation but not as quickly as any membrane bound organelle observed (Figure 2B). The profiles of proteasome subunits β 1, α 6, and α 7 were

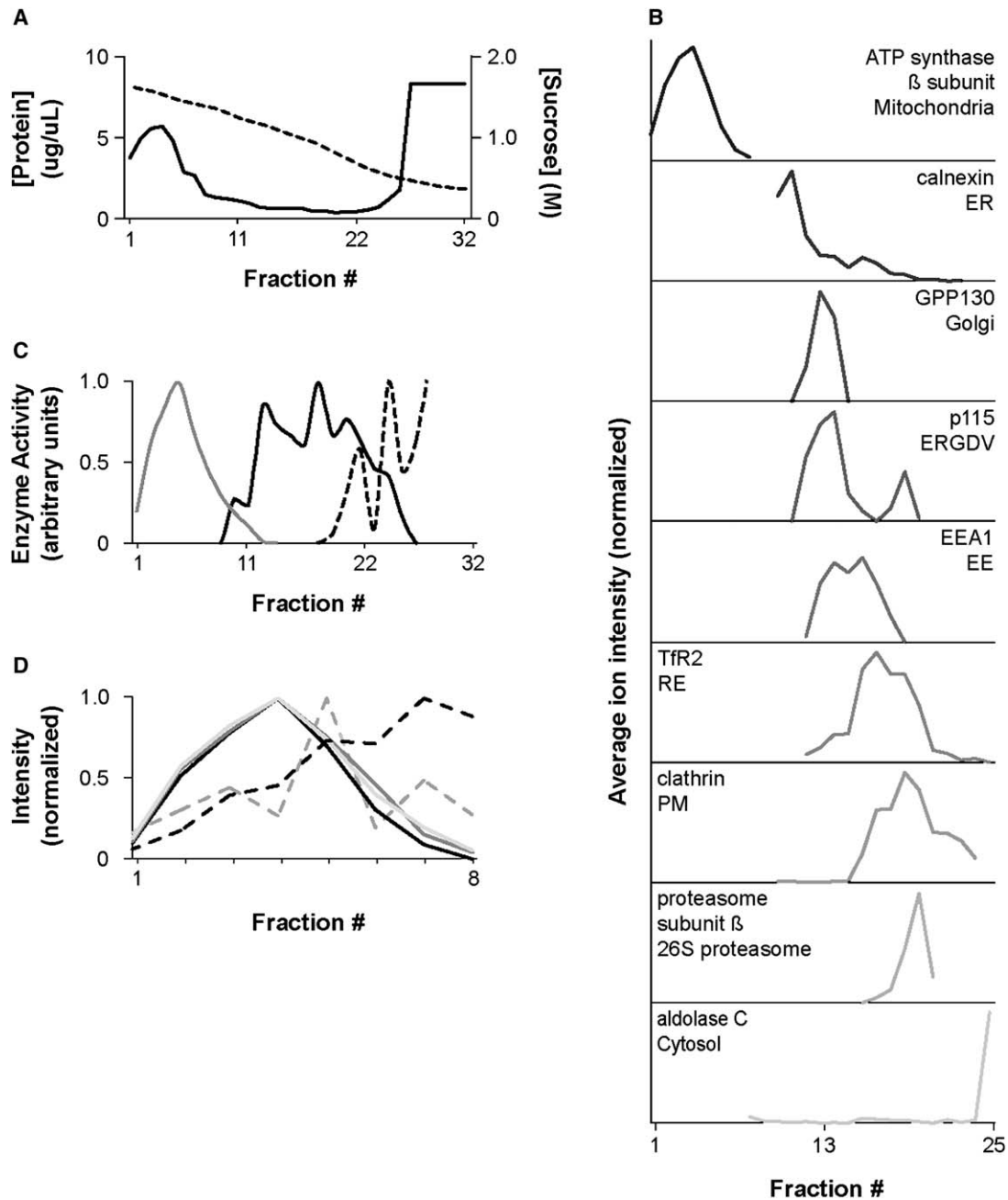


Figure 2. Sucrose Gradients and Protein Correlation Profiles

(A) Sucrose concentration (dashed line, right ordinate) and protein content (solid line, left ordinate) for all fractions from a representative gradient.

(B) PCP for indicated proteins and the organelles they represent.

(C) Enzyme activity for amine oxidase (gray curve), NADP-phosphatase (black curve), and γ -glutamyltransferase (dashed curve).

(D) Normalized PCP curves for glyceraldehyde-3-phosphate dehydrogenase (gray dashed), γ -actin (black dashed), sarcosine dehydrogenase (solid black), F_1 - F_0 ATP synthase α subunit (light gray), and malate dehydrogenase 2 (dark gray) in the region of the gradient containing mitochondria.

thus used as a standard curve for PCP comparisons. Ribosomes, much larger multimolecular complexes than proteasomes, migrated further in our gradients, apparently as both free complexes and attached to the ER. However, the PCP of free ribosomes overlapped those

of several organelles. Finally, since the initial homogenate was layered on top of the gradient and only large complexes or organelles migrate during centrifugation, cytosolic proteins remain highly enriched in the low-density fractions—typified by the glycolytic enzyme fructose

bisphosphate aldolase C (Figure 2B). Measurement of the activities for amine oxidase (mitochondrial marker), γ -glutamyltransferase (PM marker), and nicotinamide adenine dinucleotide phosphate (NADP)-phosphatase (Golgi marker) provided further confirmation of the separation of these organelles (Figure 2C).

Having established organellar profiles, we then evaluated the distribution of all 1,900 proteins using a more stringent cutoff criterion for the χ^2 test than we have used previously. Using the rules described more fully in Experimental Procedures, the PCPs of 1,258 proteins matched to at least one of the markers mentioned above and were distributed as follows: mitochondria—297, ER—229, Golgi—67, ERGDV—220, EE—76, RE—326, PM—250, proteasome—50, cytosol—488 (Figure 1B; for all 1,404 cytoplasmic and nuclear localizations, see Table S5). Where we sequenced peptides that allowed us to distinguish one isoform from another, we were in some cases able to discern differential localization. For instance, Rab1a was found in ERGDV, while Rab2b was found in the EE and RE/TGN (Table S5).

This dataset is available in the supplementary tables and also as a database at proteome.biochem.mpg.de/ormd.htm. It represents a major advance in large-scale protein localization, especially in mammals, and is likely to complement data obtained via other methods for several reasons: (1) the information is derived from native liver tissue, avoiding artifacts of cell culture; (2) the locations are derived for endogenous proteins rather than overexpression of exogenous, tagged proteins, avoiding pitfalls associated with fluorescence-based assays in mammalian systems (Wiemann et al., 2004); and (3) by only selecting proteins matching consensus profiles, background proteins are eliminated. Notably, this study achieved similar or deeper coverage of genuine organellar components than previous proteomics studies dedicated to specific compartments (see below). However, the PCP localizations presented here certainly do not yet cover the entire protein composition of any organelle, and further, more in-depth study of each will be required. While these data represent an unbiased set of localizations, we evaluated these assignments based on known information about each protein in order to estimate our own potential false positive rate (see Supplemental Experimental Procedures). Using the functional and spatial annotations in the UniProt and Gene Ontology databases, we asked whether the localizations we found matched previously reported data, excluding that collected in proteomic studies. Remarkably, in those cases where annotations contained unambiguous localization information, the PCP-based organelle assignments reported here were correct in 87% of all cases.

In a separate experiment, we quantified the relative abundances of 734 proteins in nuclear versus cytoplasmic preparations (Tables S3 and S4) based on integrated peptide ion intensities in LC/MS analyses of each. Major nuclear proteins, including histones, lamin, and fibrillarin, were enriched several hundred-fold in the nucleus, while the opposite was true for proteins such as glyceralde-

hyde-3-phosphate dehydrogenase (GAPDH) and translation initiation factors. The nuclear proteins observed in the gradient of cytoplasmic material likely represent newly synthesized products, given that they largely associated with biosynthetic organelles and that this level represents far less than 1% of the total within the cell. Thus this quantitative proteomic dataset can be mined for nuclear/cytoplasmic distributions of many protein families. For example, in the liver tissue we examined, proteasome subunits were, on average, 12 times more abundant in cytoplasm than in the nuclear fraction, with some regulatory subunits exclusive to the cytoplasm. This is in contrast to previous microscopic studies suggesting a more even distribution (Peters et al., 1994; Reits et al., 1997).

At least 90% of protein mass in liver comes from hepatocytes, with smaller contributions from Kupffer cells and various endothelial and epithelial cells. Thus, the protein localizations described here are largely representative of hepatocytes. While it is formally possible for an organelle from Kupffer cells to sediment at a different rate than the corresponding organelle in hepatocytes, we could find no systematic evidence of this.

The protein localizations reported here can now be used to assess outstanding issues in cell biology. For example, several groups have suggested that glycolytic enzymes associate with mitochondria, largely based on their presence in mitochondria-enriched fractions (Danial et al., 2003; Gaucher et al., 2004). The PCP evidence gathered here argues against this, especially for GAPDH (Figure 2D) and aldolase (Table S3), whose profiles clearly show that they do not physically associate with mitochondria, at least under the conditions used in our preparation. Thus, data tying these enzymes to mitochondria may instead reflect functional interactions.

At least 373 of the 968 proteins localized to cytoplasmic organelles or the proteasome were found in more than one compartment (not including cytosol or nucleus) (Table S5). In many cases, these multiple localizations might be predicted a priori given a basic knowledge of cell biology. Many proteins were found in more than one of the ER, Golgi, and/or ERGDV, likely reflecting the dynamic retro and anterograde traffic between these locations. In cases where one organelle is derived from another, such as the PM and RE, multiple localizations were also common. Sorting nexins (Snx) 1, 2, and 9 were in both the PM and RE as well as on the ER (Snx2) and ERGDV (Snx1 and 2). However, the unexpectedly large fraction of localizations to multiple organelles would likely go undiscovered with other large-scale methods as such assignments using microscopy require the use of multiple labels for colocalization and thus are essentially impossible in large-scale studies when no markers are used (Huh et al., 2003; Wiemann et al., 2004). For the subset of proteins whose subcellular locations have been well characterized in more focused studies of single proteins, our results agree 87% of the time, as discussed above, but in many cases PCP revealed multimodal distributions for proteins previously described as resident in single organelles.

Protein abundance is now recognized to be a key parameter describing a proteome, and a first proteome-wide study has recently been reported in yeast (Ghaemmaghami et al., 2003). We calculated a protein abundance index (PAI) that provides a rough measure of abundance for each protein measured in liver tissue. We used the extracted ion current PAI (xPAI) (Rappsilber et al., 2003), which represents the average ion current for the three most intense peptides from each protein. This resulted in a rough estimation of protein abundance, similar to protein staining methods, and demonstrates that we were able to detect proteins spanning a concentration range of about 10^5 between the most and least abundant proteins identified (Tables S3 and S4). These data should represent the vast majority of protein mass in the liver and may be useful in modeling major functions in this organ, such as detoxification or intermediary metabolism.

Correlation of PCP with Immunofluorescence

While the classical tool of the organelle biochemist is gradient centrifugation, for the organelle cell biologist it is fluorescence microscopy. Through the use of standard marker proteins, we have already classified the PCPs of several organelles but, as mentioned above, several proteins appeared to localize to more than one compartment. Therefore, we utilized indirect immunofluorescence and confocal microscopy to visualize how overlapping or non-overlapping PCPs translate into visual patterns. As an example of mostly overlapping PCPs, the profiles of TfR2 and secretory carrier-associated membrane protein-3 (SCAMP) indicated that both were highly enriched in the fractions enriched for the EE and RE compartments (Figure 3), consistent with their known functions. In accordance with this, the immunofluorescent patterns of these two proteins overlapped in an RE-like compartment as well as in large punctae reminiscent of early endosomes. However, their PCPs did not overlap completely, and this was reflected in the punctate green (SCAMP) pattern and plasmalemmal red (TfR2) staining seen in the merged image (Figure 3, first row). Neprilysin1, a protein localized to the ERGDV by PCP, showed less extensive overlap with clathrin HC, which PCP assigned mainly to the PM. However, the PCPs for both proteins had a shoulder extending into fraction 17, where TfR2 peaked, and this was reflected in their staining patterns (Azarani et al., 1998; Dell'Angelica et al., 1998), which were largely distinct from one another but displayed partial overlap in a perinuclear region (Figure 3, second row) similar to that seen for TfR2. On the other hand, the PCP for neprilysin1 showed little or no overlap with the PCPs from any proteasome subunits or cytochrome c, and double labeling of mouse liver cells showed no overlapping staining for these proteins (Figure 3, third and fourth rows). As demonstrated in these examples, immunofluorescence links the localizations determined biochemically by PCP with classical cell biological methods and indicate that much information can be derived from careful examination of individual PCPs.

Evaluation of Previously Published Large-Scale Datasets

Previous studies have utilized mass spectrometry-based proteomics to catalogue the components of organelles. In these studies, there are two possible sources for false-positive localizations: incorrect protein identifications or incorrect assignment of contaminating proteins. We have essentially eliminated the first source of error by using high-performance mass spectrometry, and we exclude contaminating proteins via PCP. Figure 2D demonstrates the difficulty in distinguishing copurifying proteins from organelle components in the absence of gradient profile information. To extend the mitochondria example, we identified 645 proteins in mitochondrial fractions, yet after PCP analysis only 297 of these had profiles matching the mitochondrial standard curve. The set of proteins validated by PCP, but not the remaining proteins, agreed very well with mitochondrial proteins described in the literature, excluding other proteomic studies.

In order to estimate the level of agreement between our study and published organelle catalogues, we asked whether our PCP data agreed with previously reported (Da Cruz et al., 2003; Mootha et al., 2003a; Andreoli et al., 2004; Cotter et al., 2004; Gaucher et al., 2004; Wu et al., 2004; Zhao et al., 2004) localizations (Experimental Procedures). The agreement between our data and other reports varied widely, even within one organelle (Table 1). The lower rates seen for online compendia of mitochondrial experimental datasets appear to be due to a common core of proteins in each dataset, with which our data do agree, and sets of proteins unique to the individual compendia, with which our data do not agree.

We next compared our localization data with a recent large-scale effort employing green fluorescent protein (GFP)-tagging in yeast (Huh et al., 2003), where it is possible to use the endogenous gene promoter to express the GFP fusion protein. We found a high level of agreement—approximately 74%—for organelle localizations, validating both datasets. This suggests, intriguingly, that proteins that are conserved over the billion or more years of evolution separating yeast and mammals also retain their “cellular home.” There was slightly less agreement on cytosolic localizations between our two studies (63%), and this may be due to the addition of the 27 kDa GFP molecule onto endogenous proteins or a limitation of microscopic methods where no markers were used simultaneously for confirmation of colocalization. Furthermore, in PCP each organelle is effectively separated from the cytosol prior to analysis, reducing the effect of a limited detectable concentration range in microscopic studies. For instance, poly(A) binding protein 4 is seen in the cytosol of yeast and in our data, but we also detect a fraction of this protein associated specifically with ER and Golgi (Table S5). While we count this instance as concordant localization, it illustrates the ability of PCP to detect multiple localizations.

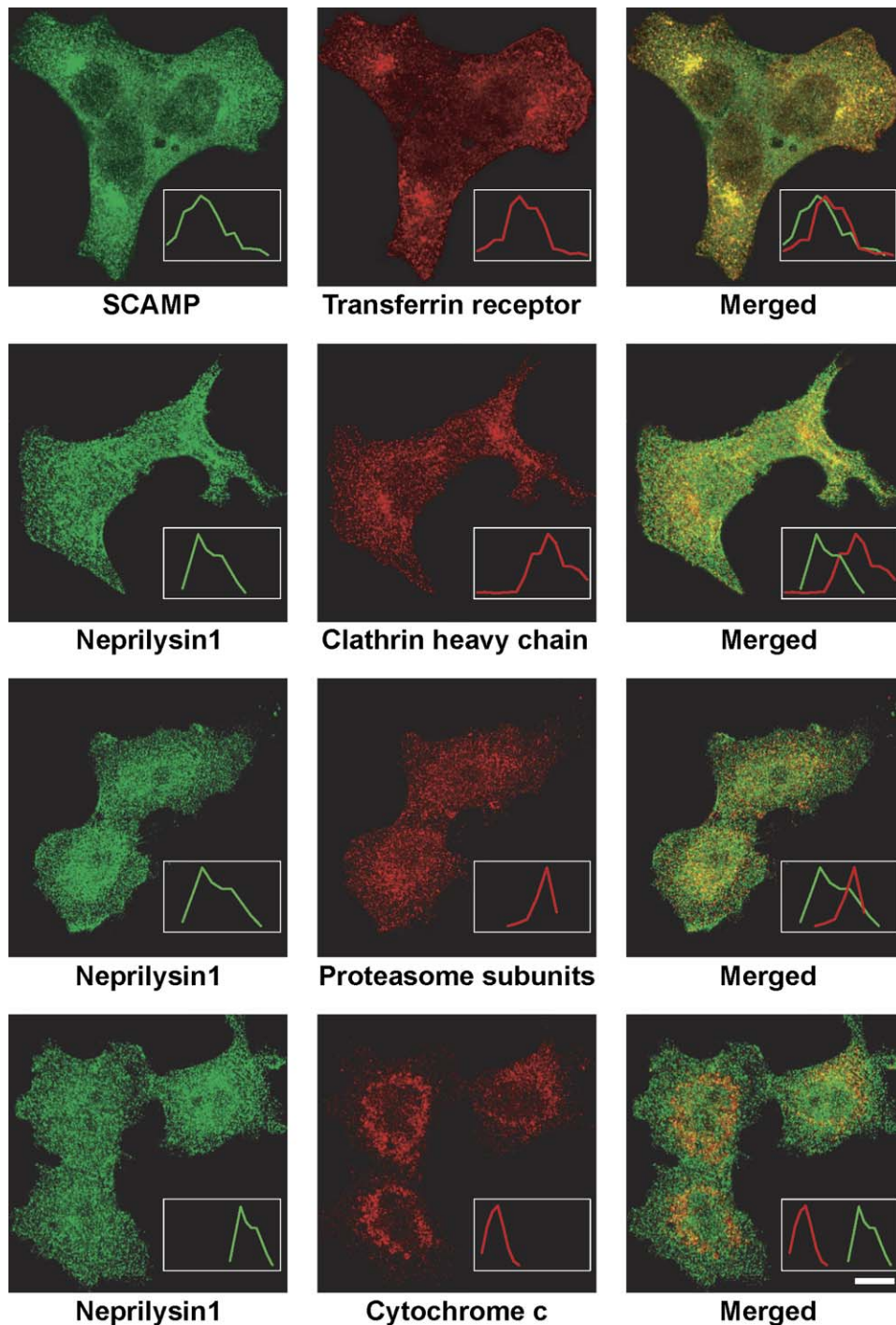


Figure 3. PCPs Translate Directly into Fluorescent Staining Patterns

Representative images of HepA mouse liver hepatoma cells stained with primary antibodies as indicated, followed by chicken anti-rabbit Alexa488 (green) and donkey anti-mouse Cy3 (red). Overlaid images of both channels are included in the right-hand column, with yellow representing overlapping signal. Scale bar represents 10 μm . PCPs for each protein shown in the image are included as inserts. Abscissae of PCPs are to scale.

Functional Genomics of Organelles

An organelle proteomic map provides a “cell biological scaffold” on which other functional genomics data can

be layered—the integration of many diverse data sets can thereby help achieve a more unified model of the cell (Glaser and Boone, 2004). We previously integrated RNA

Table 1. Assessment of Published Large-Scale Organelle Datasets

Organelle	Species	Source	# Reported ^a	# Matched ^b	Correct Localizations ^c	References
Golgi ^d	Rat	Liver tissue	421	176	36%	(Wu et al., 2004)
Plasma membrane ^e	Human	Lung cell line	897	211	49%	(Zhao et al., 2004)
Inner mitochondrial membrane ^f	Mouse	Liver tissue	182	89	93%	(Da Cruz et al., 2003)
Mitochondria	Mouse	Tissues	591	236	86%	(Mootha et al., 2003a)
Mitochondria	Human	Heart tissue	680	238	66%	(Gaucher et al., 2004)
Mitochondria	Human	Compilation ^g	847	310	63%	(Cotter et al., 2004)
Mitochondria	Mouse	Compilation ^g	601	114	61%	(Andreoli et al., 2004)

^a The total number of localized proteins reported in each study.

^b The number of proteins common to the published set and the PCP set (see [Experimental Procedures](#) for these criteria).

^c The fraction of proteins common to the published set and the PCP set that were assigned to the given organelle in the PCP set.

^d Reported Golgi proteins were considered concordant if they were found in the ERGDV or Golgi groups described here.

^e Reported plasma membrane proteins were considered correct if they were found in any of the PM, EE, or RE groups described here since the 15-min labelling reaction used to purify PM proteins in the referenced study may have allowed some labelling of the endosomal compartments.

^f Reported mitochondrial proteins were considered false positive if $\chi^2_{\text{mito}} > 0.2$ and $\chi^2_{\text{hemo}} - \chi^2_{\text{mito}} > -0.05$ (see [Experimental Procedures](#)).

^g Online compendia of several experimental and predicted datasets.

expression profiles with a proteomic survey of mouse mitochondria to gain insights into tissue diversity and gene regulation within this organelle (Mootha et al., 2003a). Here we apply some of these approaches as well as comparative genomics to better understand the transcriptional regulation of organelles.

First, for each organelle we sought to determine whether subsets of genes encoding its proteins show distinct patterns of gene expression across a battery of mouse tissues. Clustering of available RNA expression measures across 44 different tissues (Su et al., 2002) revealed (Figure 4) that subsets of organellar genes show striking coexpression patterns across different tissues. In many cases, these clusters were enriched in functionally related sets of genes, implying that poorly characterized genes falling within these clusters might share a similar function. For instance, in mitochondria we rediscovered many of the same functional groups we have observed previously, including a strong cluster corresponding to the genes of oxidative phosphorylation (Mootha et al., 2003a). For the ER, we found a tightly coexpressed cluster of 50 genes ($p = 9 \times 10^{-5}$, distance = 0.63), including 12 annotated cytochrome P450 enzymes and 21 other enzymes involved in catabolism of drugs and endogenous compounds. Also within the ER gene set, the ribosomal genes form a cluster of 28 genes ($p = 7 \times 10^{-5}$, distance = 0.40), of which 12 are components of the ribosome. In ERGDV a strong cluster of vesicle trafficking proteins emerged that contained 18 genes, and of these 10 ($p = 1 \times 10^{-6}$, distance = 0.47) were annotated as having a function in vesicle docking and/or fusion. These expression clusters are in agreement with the current understanding of their respective organelles and suggest a high level of transcriptional coregulation in the biogenesis and homeostasis of these organ-

elles. These data suggest several testable hypotheses regarding the location and function of unannotated proteins that display similar mRNA expression profiles to other proteins in their organelle datasets.

Next, we were interested in systematically identifying *cis*-regulatory motifs and transcriptional regulators that might be involved in mediated the biogenesis of each of the organelles. To do so, we used expression neighborhood analysis to systematically identify transcriptional regulators that are coexpressed with genes encoding each organelle, and we combined comparative analysis of mouse, human, dog, and rat with a motif discovery strategy to identify *cis*-motifs enriched upstream of genes encoding members of each organelle.

Neighborhood analysis is a computational strategy that scores each gene in the genome for the degree of its coexpression with a gene set of interest. Any gene whose neighborhood index exceeds a threshold achieved by a random gene set of the same size (see [Experimental Procedures](#) and [Supplemental Experimental Procedures](#)) is said to be within the “expression neighborhood” for that gene set. Members of the gene set’s neighborhood tend to be functionally related within the context of the expression dataset. We applied neighborhood analysis (Mootha et al., 2003a) to systematically identify other genes in the genome whose mRNA expression patterns match closely to those in each organelle, hypothesizing that such proteins could be additional residents of the given organelle or involved in the regulation/biogenesis of the compartment. We employed the organelle neighborhood index, N_{100} , to identify genes whose expression pattern neighborhoods are enriched in genes encoding components of each organelle (Mootha et al., 2003b). For example, if the 100 nearest expression neighbors of

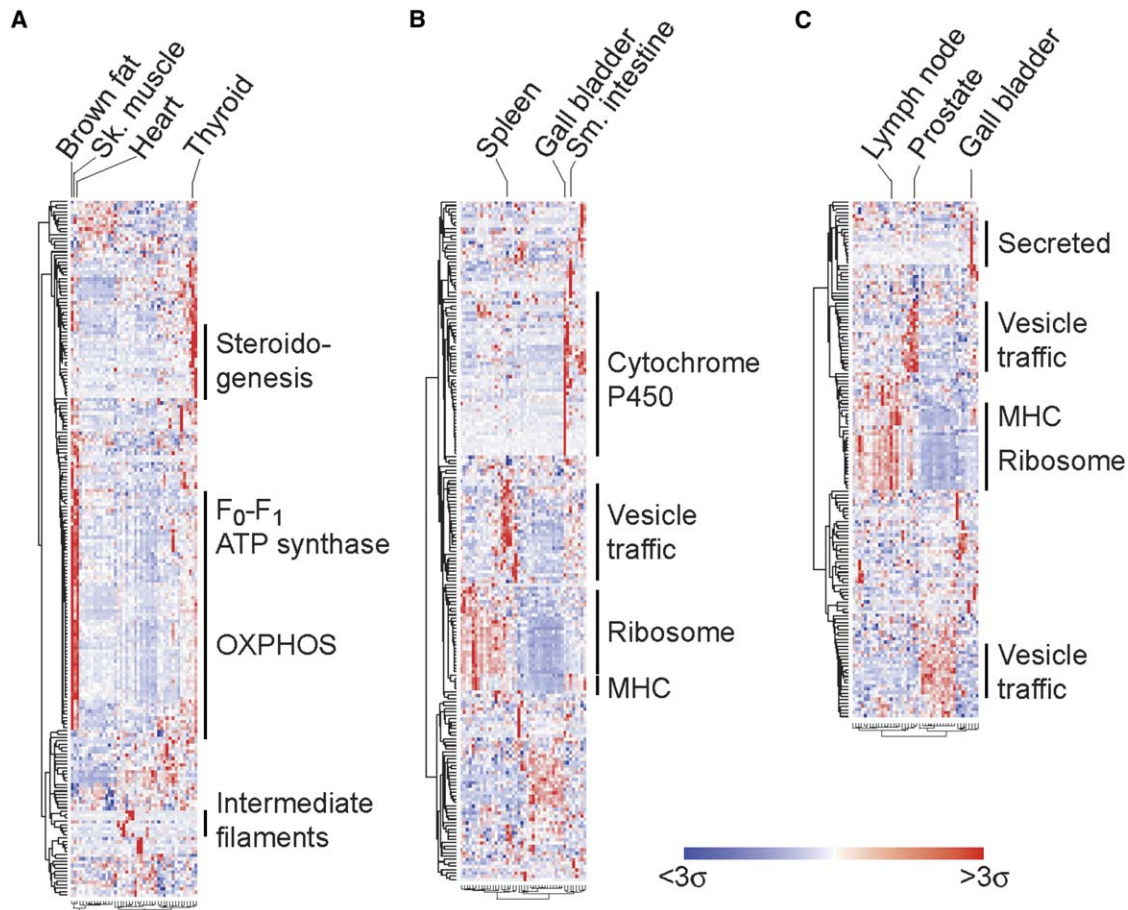


Figure 4. Hierarchical Clustering and Expression Neighborhood Analysis

Two-dimensional hierarchical clustering (distance metric = $1 - r$ where r is the Pearson correlation coefficient) of mRNA expression data from 44 mouse tissues (data for liver were removed) for genes corresponding to proteins localized to mitochondria (A) (208 genes), endoplasmic reticulum (B) (203 genes), or ER/Golgi-derived vesicles (C) (153 genes). Dominant gene modules are annotated to the right of each correlogram (MHC: major histocompatibility locus; OXPHOS: oxidative phosphorylation), and selected tissues are indicated along the top. Gradient bar in lower right reflects the color-coding scheme: Shades of blue represent tissue abundances down to three standard deviations (3σ) less than the average for that gene, and shades of red represent tissue abundances up to three standard deviations (3σ) more than the average for that gene.

a gene, G , include ten which encode ER proteins, then $N_{100}(G) = 10/100 = 0.10$. By calculating this value for each organelle, it became clear that genes encoding components of any given organelle tended to have higher N_{100} values than expected by chance ($p < 0.0001$). Mitochondrial genes tended to have the strongest coexpression (DeRisi et al., 1997; Mootha et al., 2003b), followed by proteasome components. This may reflect the more focused functions of those proteomes as opposed to, for instance, the more diverse roles of the plasma membrane proteome.

To identify potential transcriptional regulators for organelles, we searched for genes annotated as transcriptional regulators (GO:0003700 in the GO annotation scheme) within each organelle's expression neighborhood. As we have previously shown (Mootha et al., 2003a), this strategy can identify transcription factors that are coexpressed

with the organelle and hence are likely candidate regulators. Transcriptional regulators within each organelle's neighborhood are reported in Table 2. Within the mitochondrial neighborhood, we discovered a number of validated transcriptional regulators of organelle biogenesis, including myocyte enhancer factor-2 (Mef2a), peroxisome proliferator-activated receptor (PPAR), and estrogen-related receptor alpha (Esrra). The expression profile of the Sp1 transcription factor was found in the ERGDV expression neighborhood (Table 2), corresponding with the discovery of the binding motif for this protein in several ERGDV genes (Table 3). Additionally, hepatic nuclear factor 4 (Hnf4) was found to follow a similar expression pattern to ER genes in general and cytochrome p450 genes more specifically. This discovery lends additional experimental support to the role of Hnf4 as a general regulator of cytochrome P450 expression (Jover et al., 2001).

Table 2. *Trans*-Acting Transcription Factors within Organelle Neighborhoods

Organelle	Transcription Factor ^a
Mitochondrion	Cnot8
	Crsp2
	Nfix
	Esrr α
	Ppara
	Atf6
	Nr1h3
	Nr1i3
	Tbx6
	Mef2a
	Nfyc
	Mta2
	Endoplasmic reticulum
Atf5	
Nr1i3	
Hnf4a	
Usf1	
Pax2	
Endoplasmic reticulum and Golgi-derived vesicles	Btf3
	Atf5
	Hoxb5
	Myc
	Hhex
	Zfp207
	Sp1
	Rpl7
Recycling endosomes	Atf5
	Nr1h3
Proteasome	Pcbd1
	Hoxc5
Plasma membrane	Jund1
	Atf5
	Hnf4a

^aTranscription factors in the expression neighborhood ($p < 0.0001$; see [Experimental Procedures](#) and [Supplemental Experimental Procedures](#)) of each organelle based on GO annotations.

While neighborhood analysis enables us to identify putative *trans*-acting regulators of organelle biogenesis, we were also interested in the *cis*-regulatory elements that might be enriched upstream of the genes encoding

proteins resident in an organelle. To identify such *cis*-regulatory elements, we used a comparative genomics approach (Xie et al., 2005) that assessed each organelle separately. Specifically, we made use of mouse, human, rat, and dog genomes to ask whether any motifs are enriched in the promoter regions of genes corresponding to an organelle when compared to all other genes in the genome (see [Experimental Procedures](#) and [Supplemental Experimental Procedures](#)). Several known (Table 3) motifs among the 94 discovered (Table S7) include several that we and others had previously identified and experimentally validated. For the mitochondrion, our method is validated by recovering the promoter motifs for ERR α and nuclear respiratory factor 2 (NRF-2, a dimer of GABPA and GABPB) (Mootha et al., 2004; Scarpulla, 1997), both of which have been confirmed experimentally. Among the ER genes, two binding motifs for Myc transcription factors were also highly enriched (Table 2), raising the testable hypothesis that Myc and its binding site are involved in the biogenesis of these organelles. This is supported by previous work on Myc suggesting that it is involved in ribosomal biogenesis and assembly (Boon et al., 2001). A number of other high-scoring motifs were also identified, although at this time the factors binding to these sites are not known.

Of note, the neighborhood analysis and the motif analysis spotlight Esrra and Sp1 as well as their annotated binding motifs. This raises the hypothesis that these two factors might be involved in the regulatory network underlying the biogenesis of their respective organelles. It is important to note that *cis*- and *trans*-acting gene regulatory elements could not have been revealed through organelle proteomics alone since these components would only be present at the sites of transcription and translation.

DISCUSSION

Previous organellar proteomics studies have largely focused on one compartment, enriched in a single gradient fraction or centrifugation pellet. By retaining the separation information inherent in the gradient, we have used PCP to assign 1,404 proteins to ten subcellular locations with much higher confidence than has been achieved previously. Indeed, our results highlight the importance of incorporating an unbiased method to reduce false positives (Dunkley et al., 2004). These PCP results also indicate that even more organelles could be resolved with greater coverage of each, but higher resolution centrifugation gradients will likely be required in order to distinguish minor organelles. By their very nature, organelles such as endosomes and vesicles moving between the ER and Golgi are extremely dynamic, so a more focused PCP analysis could be combined with time-resolved organelle proteomic studies to gain insight into these systems. The data presented here also provide functional insights into a number of newly localized proteins. By integrating modern quantitative proteomics with classical biochemistry, we have developed a knowledge base of cellular organization

Table 3. Cis Promoter Motifs Enriched Upstream of Organelle Genes

Organelle	Motif ^a	Z-Score ^b	Annotation	Frequency ^c
Mitochondrion	TNAAGGTCA	7.6	Esrr α	32%
	CNCTCCGGT	3.4	NRF-2	10%
Early endosome	CYSATTGGYY	4.6	NF-Y	19%
	CNCTCCGGT	4.2	NRF-2	17%
Endoplasmic reticulum	CACGCNA	6.7	AhR	4%
	ACCACGTGGT	6.2	c-Myc/Max	6%
	CCACGTG	5.9	N-Myc	4%
	ACCGGAAGNG	5.4	NRF-2	18%
	SCRCRTGGC	4.4	c-Myc/Max	6%
	TCCMAGAA	4.1	STAT	9%
	CNGNRNCAGGTGNNNGNA	3.7	MyoD	13%
Plasma membrane	CGGCCATCT	5.4	NF- μ E1	6%
	GCCATNTT	4.4	YY1	7%
	SCRCRTGGC	4.3	c-Myc/Max	4%
	RCWTCKG	4.2	c-Ets-1(p54)	4%
	GGGGGCGGGGY	4.0	Sp1	15%
Endoplasmic reticulum and Golgi-derived vesicles	CCGGAART	8.1	Elk-1	9%
	ACCGGAAGNG	6.6	NRF-2	6%
	GGGGCGGGG	5.7	Sp-1	10%
	YGCNCTCCGGB	5.5	GABP	33%
	RCCCCGCCCCC	4.9	Sp-1	5%
	CCACGTCA	4.6	ATF6	8%
Recycling endosomes	CCACGTG	8.8	N-Myc	40%
	YGCGCATGCG	6.2	Nrf-1	24%
	TTTSCGCG	5.4	E2F-1/DP-1	11%
	CCKCCBC	5.3	ETF	15%
	AANATGGC	5.0	YY1	11%
	ATCACGTGAY	4.8	SREBP-1	6%
	AYTTCCGG	4.8	Elk-1	13%
	VCCGGAAGNGCR	4.2	GABP	7%

^a Motifs enriched in the promoter regions of organelle genes. Motif sequences use the standard IUPAC alphabet.

^b Z-score indicating significance of motif enrichment in the organelle genes.

^c Frequency with which the promoter sequence was observed in the organelle gene set.

that correlates very well with immunofluorescence and that should be a valuable resource for the scientific community, perhaps in the search for disease genes (Mootha et al., 2003b; Li et al., 2004). Future improvements in instrumentation and software will allow the scaling up of PCP to allow the mapping of more tissues with higher organellar resolution and to a greater depth.

PCP of mammalian systems has unique features compared to epitope-tagging approaches: PCP can be applied in tissue samples to endogenous proteins without

overexpression and potentially disruptive tags; it is very well adapted for identifying multiple localizations for proteins; and it provides abundant information for all localized proteins. It is also complementary to large-scale, antibody-based protein localization projects (Agaton et al., 2003), which offer exquisite temporal and spatial resolution but may be confounded by issues of antibody specificity.

We have demonstrated that the localization data presented here is not only valuable by itself but can also

be combined with other large-scale datasets to gain unanticipated insights. Through integrative genomics, we have shown that subsets of the proteins assigned to these organelles are coregulated, and we have identified several potential transcription factors and promoter motifs through which this could occur. The organelle map provides a scaffold with which the transcriptional regulatory code of mammalian genomes can be further dissected. As more types of large-scale data are generated, the integrated functional genomics studies can then be analyzed at subcellular resolution using these data as a framework.

EXPERIMENTAL PROCEDURES

Subcellular Fractionation

All sucrose solutions contained Protease Inhibitor Cocktail (Roche) and 0.5 mM EDTA/20 mM Tris, pH 7.2. Mouse (C57BL/6) livers were diced and rinsed in ice-cold water for 2 min to lyse erythrocytes. Tissue was then washed once in 0.3 M buffered sucrose and placed in a Dounce homogenizer with 5 ml of the same solution. The tissue was homogenized by five strokes with the loose pestle and thirty-five strokes with the tight pestle, filtered through two layers of cheese-cloth, and centrifuged for 5 min at $600 \times g$. The postnuclear supernatant was layered onto two 14 ml continuous gradients, one from 1.0 M to 2.0 M sucrose and the other from 0.3 to 1.6 M sucrose. Organelles were resolved by rate-zonal centrifugation (110 min, $95,000 \times g$, 4°C) before extracting twenty-four 0.5 ml fractions from each. Sucrose concentrations were calculated from refractive indices, and protein concentrations were measured using Coomassie Plus (Pierce). Samples for the nuclei versus cytoplasm analysis were prepared as described (Andersen et al., 2002).

Liquid Chromatography/Mass Spectrometry

Twenty micrograms of protein from each fraction was precipitated and prepared for LC/MS as described (Foster et al., 2003). Peptides were resolved by reversed phase chromatography and measured in a linear ion-trap Fourier transform mass spectrometer (LTQ-FT, Thermo Finnigan) set to acquire in data-dependent mode (Olsen et al., 2004). For each cycle, 5×10^6 ions between 300 and 1,500 m/z were measured at 100,000 resolution, and tandem mass spectra of the five most abundant multiply-charged ions were collected. In order to generate closely timed survey scans, no Selected Ion Monitoring (SIM) scans were performed. Blank gradients of 25 min were interspersed between 140-min analytical cycles to precondition the system.

Protein Identification

Fragment spectra were searched against the mouse IPI database (Feb. 27, 2006) (<ftp://ftp.ebi.ac.uk/pub/databases/IPI/current/>) using Mascot Server v2.0 with the following parameters: trypsin specificity, one missed cleavage, cysteine carbamidomethylation (fixed), methionine oxidation and protein amino-terminal acetylation (variable), ESI-TRAP fragmentation, 15 ppm mass window for precursor ion mass after recalibration (measured in FT without SIM scans), and 0.5 Da mass tolerance for fragment ions (measured in LTQ). Average absolute mass accuracy was better than 3.5 p.p.m. MSQuant (Schulze and Mann, 2004), open source software developed in our group, was used to parse and recalibrate peptide information from Mascot result files and then send it to a relational database. Acceptance criteria were set for a statistical confidence in protein identifications of 99.988% (see Supplemental Experimental Procedures).

PCP

Peptide ion intensities were extracted as described (Andersen et al., 2003). Profiles for each protein were calculated as the average of the five most intense peptides from that protein (if available), and χ^2 values

for each protein relative to marker proteins were calculated (Andersen et al., 2003). For mitochondria, both a χ^2_{mito} value (relative to F_1-F_0 ATP synthase β subunit) and χ^2_{hemo} value (relative to hemoglobin α chain, a marker for comigrating erythrocytes) were calculated. Other proteins were then considered mitochondrial if their χ^2_{mito} value was less than 0.15 and less than their χ^2_{hemo} value. For other organelles, whose profiles typically spanned only one fraction, we used χ^2 criteria equivalent to that reported in our previous analysis of the centrosome, which was also concentrated in a single fraction. Proteins were considered residents if they peaked in the same fraction as the marker protein and if their χ^2 for the relevant organelle was less than 0.05. For all organelles and the proteasome, a peak was defined as a point at least 5% more intense than either of its neighbors on either side and 25% the height of the highest point in the profile to reduce the effects of random noise. Three pairs of organelles, Golgi/ERGDV, EE/RE, and PM/Proteasome, migrated to consecutive fractions of the gradient, meaning that if the PCP for a protein peaked in one member of a pair, it could not be assigned to the other member even if its χ^2 value for the member was very small. To avoid this problem, we relaxed the peak criteria for these three pairs and instead required the PCP to peak in one of the two members for the protein to be considered for either organelle. For the cytosol, a "peak" was defined as a signal in Fraction 30 that was at least 50% higher than the signal in Fraction 24. The "sliding normalization window" method used here to scan across all gradient fractions was essential for successful analysis of this data. Direct comparison of complete PCPs overlooks most multiple organelle localizations since the contribution of a second peak in the PCP would have had too large a contribution to the χ^2 calculation.

The reproducibility of PCP between experiments using separate animals was evaluated by analyzing the high-density fractions from three independent gradient separations of different mouse livers. Using the same χ^2 criteria described above, 120 proteins found in at least two of the three analyses showed incongruent localizations in ten of 154 location assignments (Table S8). Eight of the ten were proteins localized to mitochondria in the replicates that were assigned to other compartments in the larger analysis, while the remaining two were assigned as mitochondrial proteins in the main analysis but whose χ^2 values in the replicates exceeded the cutoff. Thus we estimate overall reproducibility of organellar assignments at greater than 90%.

RNA Expression Analysis

Expression neighborhood analysis was performed on mRNA expression data using each set of organelle components as described previously (Mootha et al., 2003a). After organelle neighborhoods were calculated, all genes enriched ($p < 0.0001$, based on binomial distribution) in each neighborhood were retrieved from Affymetrix's NetAffx Analysis Center (<http://www.affymetrix.com/analysis/index.affx>) to identify those genes annotated as transcription factors in the Gene Ontology system (GO:0003700).

Motif Discovery

We generated a whole-genome alignment between human, mouse, rat, and dog. Using the annotation of transcription starting sites (TSS) from the Reference Sequence (RefSeq), we extracted a region of $[-2000, +2000]$ bp centered on the TSS of each mouse gene from the whole-genome alignment and constructed an aligned promoter database, which consists of $\sim 17,700$ genes. For details of the consensus search, please see Supplemental Experimental Procedures.

Assessment of Published Datasets

Primary sequences of reported proteins (Da Cruz et al., 2003; Huh et al., 2003; Mootha et al., 2003a; Andreoli et al., 2004; Cotter et al., 2004; Gaucher et al., 2004; Wu et al., 2004; Zhao et al., 2004) were generated from the relevant online databases and BLASTed against the protein sequences identified in the current study. Only those matches with at least 85% identity with human or rat proteins, at least 95% identity with mouse proteins, or at least 30% identity with yeast proteins

were considered. Localization criteria were relaxed by 33% (see above) prior to using PCP-based localizations to evaluate other datasets.

Supplemental Data

Supplemental Data include eight tables and Experimental Procedures and can be found with this article online at <http://www.cell.com/cgi/content/full/125/1/187/DC1/>.

ACKNOWLEDGMENTS

The authors wish to thank Peter Mortensen and Shao-En Ong for programming assistance, Henning Skov for database assistance, Rikke Jakobsen for technical assistance, Mogens Nielsen for mouse dissections, CEBI for fruitful discussions, and the Proteomics Core Facility at UBC and the Protein Centre at the University of Victoria for technical help. Yong Zhang and the Beijing Genome Center contributed to the organellar database. The Danish National Research Foundation funds CEBI, and L.J.F. is the Canada Research Chair in Organelle Proteomics and is funded by the Canadian Institutes for Health Research Operating Grant #MOP-77688. This work contributes to HUPO's liver proteome initiative.

Received: September 10, 2005

Revised: January 20, 2006

Accepted: March 8, 2006

Published: April 6, 2006

REFERENCES

- Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198–207.
- Agaton, C., Galli, J., Hoiden Guthenberg, I., Janzon, L., Hansson, M., Asplund, A., Brundell, E., Lindberg, S., Ruthberg, I., Wester, K., et al. (2003). Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Mol. Cell. Proteomics* 2, 405–414.
- Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. (2002). Directed proteomic analysis of the human nucleolus. *Curr. Biol.* 12, 1–11.
- Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., Nigg, E.A., and Mann, M. (2003). Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* 426, 570–574.
- Andreoli, C., Prokisch, H., Hortnagel, K., Mueller, J.C., Munsterkotter, M., Scharfe, C., and Meitinger, T. (2004). MitoP2, an integrated database on mitochondrial proteins in yeast and man. *Nucleic Acids Res.* 32, D459–D462.
- Azarani, A., Boileau, G., and Crine, P. (1998). Recombinant human endothelin-converting enzyme ECE-1b is located in an intracellular compartment when expressed in polarized Madin-Darby canine kidney cells. *Biochem. J.* 333, 439–448.
- Boon, K., Caron, H.N., van Asperen, R., Valentijn, L., Hermus, M.C., van Sluis, P., Roobeek, I., Weis, I., Voute, P.A., Schwab, M., and Versteeg, R. (2001). N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. *EMBO J.* 20, 1383–1393.
- Breuzer, L., Halbeisen, R., Jenö, P., Otte, S., Barlowe, C., Hong, W., and Hauri, H.P. (2004). Proteomics of ERGIC membranes from brefeldin A-treated HepG2 cells identifies ERGIC-32, a new cycling protein that interacts with human Erv46. *J. Biol. Chem.* 279, 242–253.
- Cotter, D., Guda, P., Fahy, E., and Subramaniam, S. (2004). MitoProteome: mitochondrial protein sequence database and annotation system. *Nucleic Acids Res.* 32, D463–D467.
- 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.
- Danial, N.N., Gramm, C.F., Scorrano, L., Zhang, C.Y., Krauss, S., Ranger, A.M., Datta, S.R., Greenberg, M.E., Licklider, L.J., Lowell, B.B., et al. (2003). BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis. *Nature* 424, 952–956.
- de Hoog, C.L., and Mann, M. (2004). Proteomics. *Annu. Rev. Genomics Hum. Genet.* 5, 267–293.
- Dell'Angelica, E.C., Klumperman, J., Stoorvogel, W., and Bonifacio, J.S. (1998). Association of the AP-3 adaptor complex with clathrin. *Science* 280, 431–434.
- 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.
- Dunkley, T.P., Watson, R., Griffin, J.L., Dupree, P., and Lilley, K.S. (2004). Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell Proteomics* 3, 1128–1134.
- Foster, L.J., de Hoog, C.L., and Mann, M. (2003). Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. *Proc. Natl. Acad. Sci. USA* 100, 5813–5818.
- Gaucher, S.P., Taylor, S.W., Fahy, E., Zhang, B., Warnock, D.E., Ghosh, S.S., and Gibson, B.W. (2004). Expanded coverage of the human heart mitochondrial proteome using multidimensional liquid chromatography coupled with tandem mass spectrometry. *J. Proteome Res.* 3, 495–505.
- Ghaemmghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephore, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. *Nature* 425, 737–741.
- Glaser, P., and Boone, C. (2004). Beyond the genome: from genomics to systems biology. *Curr. Opin. Microbiol.* 7, 489–491.
- Hashimoto, M., and James, D.E. (2000). Characterization of insulin-responsive GLUT4 storage vesicles isolated from 3T3-L1 adipocytes. *Mol. Cell. Biol.* 20, 416–427.
- Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. *Nature* 425, 686–691.
- Jover, R., Bort, R., Gomez-Lechon, M.J., and Castell, J.V. (2001). Cytochrome P450 regulation by hepatocyte nuclear factor 4 in human hepatocytes: a study using adenovirus-mediated antisense targeting. *Hepatology* 33, 668–675.
- 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.
- Li, J.B., Gerdes, J.M., Haycraft, C.J., Fan, Y., Teslovich, T.M., May-Simera, H., Li, H., Blacque, O.E., Li, L., Leitch, C.C., et al. (2004). Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 117, 541–552.
- Mootha, V.K., Bunkenborg, J., Olsen, J.V., Hjerrild, M., Wisniewski, J.R., Stahl, E., Bolouri, M.S., Ray, H.N., Sihag, S., Kamal, M., et al. (2003a). Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. *Cell* 115, 629–640.
- Mootha, V.K., Handschin, C., Arlow, D., Xie, X., St Pierre, J., Sihag, S., Yang, W., Altshuler, D., Puigserver, P., Patterson, N., et al. (2004). Erralpha and Gabpa/b specify PGC-1alpha-dependent oxidative phosphorylation gene expression that is altered in diabetic muscle. *Proc. Natl. Acad. Sci. USA* 101, 6570–6575.
- Mootha, V.K., Lepage, P., Miller, K., Bunkenborg, J., Reich, M., Hjerrild, M., Delmonte, T., Villeneuve, A., Sladek, R., Xu, F., et al. (2003b). Identification of a gene causing human cytochrome c oxidase deficiency by integrative genomics. *Proc. Natl. Acad. Sci. USA* 100, 605–610.

- Olsen, J.V., Ong, S.E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* 3, 608–614.
- Peters, J.M., Franke, W.W., and Kleinschmidt, J.A. (1994). Distinct 19 S and 20 S subcomplexes of the 26 S proteasome and their distribution in the nucleus and the cytoplasm. *J. Biol. Chem.* 269, 7709–7718.
- Rappsilber, J., Ishihama, Y., Foster, L.J., Mittler, G., and Mann, M. (2003). Approximate relative abundance of proteins within a mixture determined from LC-MS data. In *Abstracts of the 51st Conference in Mass Spectrometry and Allied Topics*.
- Reits, E.A., Benham, A.M., Plougastel, B., Neeffjes, J., and Trowsdale, J. (1997). Dynamics of proteasome distribution in living cells. *EMBO J.* 16, 6087–6094.
- Ross-Macdonald, P., Coelho, P.S., Roemer, T., Agarwal, S., Kumar, A., Jansen, R., Cheung, K.H., Sheehan, A., Symoniatis, D., Umansky, L., et al. (1999). Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* 402, 413–418.
- Scarpulla, R.C. (1997). Nuclear control of respiratory chain expression in mammalian cells. *J. Bioenerg. Biomembr.* 29, 109–119.
- Schulze, W.X., and Mann, M. (2004). A novel proteomic screen for peptide-protein interactions. *J. Biol. Chem.* 279, 10756–10764.
- Simpson, J.C., Wellenreuther, R., Poustka, A., Pepperkok, R., and Wiemann, S. (2000). Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* 1, 287–292.
- 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.
- Syka, J.E., Marto, J.A., Bai, D.L., Horning, S., Senko, M.W., Schwartz, J.C., Ueberheide, B., Garcia, B., Busby, S., Muratore, T., et al. (2004). Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *J. Proteome Res.* 3, 621–626.
- Tulp, A., Fernandez-Borja, M., Verwoerd, D., and Neeffjes, J. (1998). High-resolution density gradient electrophoresis of subcellular organelles and proteins under nondenaturing conditions. *Electrophoresis* 19, 1288–1293.
- Wiemann, S., Artl, D., Huber, W., Wellenreuther, R., Schlegler, S., Mehrle, A., Bechtel, S., Saueremann, M., Korf, U., Pepperkok, R., et al. (2004). From ORFeome to biology: a functional genomics pipeline. *Genome Res.* 14, 2136–2144.
- Wu, C.C., MacCoss, M.J., Mardones, G., Finnigan, C., Mogelsvang, S., Yates, J.R., 3rd, and Howell, K.E. (2004). Organellar proteomics reveals golgi arginine dimethylation. *Mol. Biol. Cell* 15, 2907–2919.
- Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434, 338–345.
- Yates, J.R., 3rd, Gilchrist, A., Howell, K.E., and Bergeron, J.J.M. (2005). Proteomics of organelles and large cellular structures. *Nat. Rev. Mol. Cell Biol.* 6, 702–714.
- Zhao, Y., Zhang, W., and Kho, Y. (2004). Proteomic analysis of integral plasma membrane proteins. *Anal. Chem.* 76, 1817–1823.