Adipocytes are central players in energy metabolism and the obesity epidemic, yet their protein composition remains largely unexplored. We investigated the adipocyte proteome by combining high accuracy, high sensitivity protein identification technology with subcellular fractionation of nuclei, mitochondria, membrane, and cytosol of 3T3-L1 adipocytes. We identified 3,287 proteins while essentially eliminating false positives, making this one of the largest high confidence proteomes reported to date. Comprehensive bioinformatics analysis revealed that the adipocyte proteome, despite its specialized role, is very complex. Comparison with microarray data showed that the mRNA abundance of detected versus non-detected proteins differed by less than 2-fold and that proteomics covered as large a proportion of the insulin signaling pathway. We used the Endeavour gene prioritization algorithm to associate a number of factors with vesicle transport in response to insulin stimulation, a key function of adipocytes. Our data and analysis can serve as a model for cellular proteomics. The adipocyte proteome is available as supplemental material and from the Max-Planck Unified Proteome database. Molecular & Cellular Proteomics 6:1257–1273, 2007.

Obesity has become a global health epidemic, which leads to an increased population risk for obesity-related complications such as hypertension, dyslipidemias, type II diabetes mellitus, and cardiovascular diseases associated with the onset of insulin resistance (1, 2). Studies in the last few years have transformed our thinking about the function of adipocytes (fat cells) in physiology in general and obesity in particular (3). They are no longer regarded just as a passive depot for storing excess energy in the form of triglyceride but as an endocrine cell that actively regulates the pathways responsible for energy balance by the secretion of various bioactive substances termed adipocytokines. Furthermore recent research has highlighted the lipid drop-let as a dynamic and actively regulated organelle (4–6). To elucidate the pleiotropic function of the adipocyte, several proteomics studies have been conducted. Most of the studies to date focused on specific functions or phenomena of adipocytes such as adipogenesis (7–10), trafficking of glucose transporter (11–13), secretion of adipocytokines (14, 15), mitochondrial biogenesis (16), difference in cell size (17), and lipid metabolism (18, 19). A large scale proteomics study of adipocytes has not been reported before and could serve as a useful resource for fundamental biomedical research. The global approach of coupling large scale proteome studies with bioinformatics analysis has proved to be fruitful in gaining insights into complex hidden biological processes, for example in elucidating transcriptional regulators involved in organelle biogenesis (20–22). Furthermore large scale organellar proteomics studies can contribute tremendously to ongoing functional genomics and systems biology efforts (23, 24). These considerations as well as our interest in insulin signaling and the metabolic syndrome prompted us to perform an in-depth proteomics analysis of the cellular and organellar proteome of adipocytes. We used a combination of one-dimensional gel electrophoresis and on-line electrospray tandem mass spectrometry with biochemical procedures for subfractionation of the cellular proteome (Fig. 1). State of the art protein identification technology recently developed in our laboratory, involving a linear ion trap (LTQ)1-FTICR mass spectrometer with very high mass accuracy (25), allowed us to identify more than 3,000 proteins with extremely stringent identification criteria. Extensive bioinformatics analysis and comparison with transcriptome data gathered by the Diabetes Genome Anatomy Project (DGAP) of which we are a member revealed several layers of information related to the adipocyte proteome that were in turn mapped to an ensemble of biological processes, functions, and pathways. Additionally by using a systemic protein prioritization methodology described recently (26) we predicted candidate proteins hith-

1 The abbreviations used are: LTQ, linear ion trap; DGAP, Diabetes Genome Anatomy Project; IPI, International Protein Index; GO, gene ontology; DMEM, Dulbecco’s modified Eagle’s medium; bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; EST, expressed sequence tag; BLAST, Basic Local Alignment and Search Tool; MGI, Mouse Genome Informatics; SIM, selected ion monitoring; BINGO, Biological Networks Gene Ontology; SNARE, soluble N-ethylmaleimide-sensitive factor attachment protein receptor; t-SNARE, target SNARE; GLUT4, glucose transporter 4.
Large Scale Proteomics of Adipocyte

Cell Culture—3T3-L1 preadipocytes were grown in DMEM with 10% calf serum plus antibiotics in 10% CO₂ at 37 °C. Mouse 3T3-L1 preadipocytes were differentiated essentially as described previously (15, 27). Briefly cells were grown to confluence in DMEM with 10% calf serum plus antibiotics in 10% CO₂. When the cells reached confluence, the medium was replaced with DMEM supplemented with 10% fetal bovine serum. Two days after the cells reached confluence (day 0), they were induced to differentiate by changing the medium to DMEM containing 10% fetal bovine serum, 0.5 mM 3-isobutyl-1-methyxanthine (Sigma), 1 μM dexamethasone (Sigma), and 167 μM insulin (Sigma). After 48 h (day 2), the medium was replaced with DMEM supplemented with 10% fetal bovine serum and 167 μM insulin. After an additional 48 h (day 4), insulin was withdrawn, and the medium was changed every 2nd day. Fat accumulation was measured by Oil Red O staining, and differentiated cells withdrawn, and the medium was changed every 2nd day. Fat accumulation was measured by Oil Red O staining, and differentiated cells were harvested and used for further experiments.

Subcellular Fractionation and Western Blotting—Differential centrifugation was used to fractionate adipocytes as described previously (28, 29). Briefly cells were washed with ice-cold PBS twice and suspended in HES buffer (5 mM HEPES, pH 7.4, 0.5 mM EDTA, 250 μM sucrose, protease inhibitors (Complete tablets; Roche Applied Science). Cells were then sheared by 10 passages through a 25-gauge needle and centrifuged at 1,000 × g for 10 min at 4 °C. The supernatant served as the source of cytosol, mitochondria, and membranes. The nuclear pellet was resuspended in 7 ml of 0.25 M sucrose, TKM buffer (50 mM Tris-HCl, pH 7.4, 25 mM KCl, 5 mM MgCl₂, Complete protease inhibitor). The suspension was overlaid with 1 ml of 2.3 M, 3 ml of 2.1 M, and 6 ml of 1.6 M sucrose layer (sucrose, TKM buffer). The sucrose step gradient was centrifuged at 160,000 × g for 1 h (Sorvall Surespin 630 rotor). The purified nuclei settled at the 2.1–2.3 M interface. The nuclear layer was isolated; diluted in 10 ml of 0.2 M sucrose, TKM buffer; and centrifuged at 2,700 × g for 10 min. The nuclear pellet was resuspended in 200 μl of HES buffer. The mitochondria were isolated from the crude cytoplasm by centrifugation at 16,000 × g for 1 h with the resulting pellet suspended in 0.58 M TES buffer (0.58 M sucrose, 10 mM Tris-HCl, pH 7.4, 0.1 mM EDTA, Complete protease inhibitor). The suspension was centrifuged at 16,000 × g for 20 min, and the pellet was suspended in 3 ml of 1.75 M TES buffer while the supernatant served as the microsomal and cytosolic fraction. On the suspension of 1.75 M TES buffer, 3 ml of 1.55 M, 3 ml of 1.29 M, and 7 ml of 0.58 M sucrose layers were overlaid, and the sucrose step gradient was centrifuged at 160,000 × g for 1 h. The mitochondrial layer at the 1.29–1.55 M interface was isolated, diluted in 15 ml of 0.2 M HES buffer, and centrifuged at 16,000 × g for 20 min. The purified mitochondrial pellet was resuspended in 300 μl of HES buffer. Membrane was isolated by centrifugation of the postmitochondrial cytoplasmic fraction at 160,000 × g for 1 h while the supernatant served as the cytosolic fraction. The membrane pellet was resuspended in HES buffer and centrifuged again at 160,000 × g for 1 h. The purified membrane pellet was resuspended in 200 μl of HES buffer. Proteins in the cytosolic fraction were enriched using Centriprep YM-3 membrane concentrators (Millipore, Billerica, MA). Total protein in all fractions was then quantified using the Coomassie Protein Assay kit (Pierce) and stored at −80 °C. Equal protein amounts (7 μg) from each of the subcellular fractions were loaded onto 10 and 15% SDS-polyacrylamide gels, electrophoresed, transferred to nitrocellulose membranes, and immunoblotted with antibodies against histone H3 (Cell Signaling Technology, Beverly, MA), cytochrome c (BD Pharmingen), insulin receptor β chain (Santa Cruz Biotechnology, Santa Cruz, CA), and MEK1 (Upstate Biotechnology Inc., Lake Placiad, NY) followed by horseradish peroxidase-conjugated secondary antibodies. The membranes were subjected to chemiluminescence detection (ECL) according to the manufacturer’s instructions (GE Healthcare).

One-dimensional SDS-PAGE and In-gel Digest—Proteins (100, 150, 150, and 150 μg) from each of the subcellular fractions (nuclear, mitochondrial, membrane, and cytosol fractions, respectively) were separated by one-dimensional SDS-PAGE using NuPage® Novex bis-Tris gels and NuPage MES-SDS running buffer (Invitrogen) according to the manufacturer’s instructions. The gel was stained with Coomassie using Colloidal Blue Staining kit (Invitrogen). Protein bands were excised and subjected to in-gel tryptic digestion essentially as described previously (30). Briefly the gel pieces were destained and washed, and after dithiothreitol reduction and iodoacetamide alkylation, the proteins were digested with porcine trypsin (modified sequencing grade; Promega, Madison, WI) overnight at 37 °C. The resulting tryptic peptides were extracted from the gel pieces with 30% acetonitrile, 0.3% trifluoroacetic acid and with 100% acetonitrile. The extracts were evaporated in a vacuum centrifuge to remove organic solvent and then desalted and concentrated on reversed-phase C₁₈ StageTips as described previously (31).

NanoFlow LC-MS² or MS³—All nanoflow LC-MS/MS and MS/MS/MS experiments were performed basically as described previously (25, 32). All digested peptide mixtures were separated by on-line reversed-phase nanoscale capillary LC and analyzed by electrospray MS/MS and MS/MS/MS. The experiments were performed on an Agilent 1100 nanoflow system connected to an LTQ-FTICR mass spectrometer (Thermo Electron, Bremen, Germany) equipped with a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark). Binding and chromatographic separation of the peptides took place in a 15-cm fused silica emitter (75-μm inner diameter) in-house packed with reversed-phase ReproSil-Pur C₁₈-AQ 3-μm resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). Peptide mixtures were injected onto the column with a flow of 500 nl/min and subsequently eluted with a flow of 250 nl/min from 10 to 64% acetonitrile in 0.5% acetic acid in a 105-min gradient. Data were acquired in data-dependent mode using Xcalibur software. The precursor ion scan MS spectra (m/z 300–1575) were acquired in the FTICR instrument with resolution R = 25000 at m/z 400 (number of accumulated ions, 5 × 10⁴). The three most intense ions were isolated and fragmented in the linear ion trap by collisionally induced dissociation using 3 × 10⁴ accumulated ions. They were simultaneously scanned by FTICR selected ion monitoring with 10-Da mass range, R = 50,000, and 5 × 10⁴ accumulated ions for even more accurate molecular mass measurements. For MS³, the most intense ions with m/z >300 in each MS² spectra were further isolated and fragmented. In data-dependent LC-MS² experiments dynamic exclusion was used with 30-s exclusion duration.

Proteomics Data Analysis—Proteins were identified via automated database search (Mascot; Matrix Science, London, UK) of all tandem mass spectra against an in-house curated version of the mouse International Protein Index (IPI) protein sequence database (version 3.07, 43,335 protein sequences; European Bioinformatics Institute, www.ebi.ac.uk/IPI/) containing all mouse protein entries from SwissProt, TrEMBL, RefSeq, and Ensembl as well as frequently observed contaminants (porcine trypsin, Achromobacter lyticus lysyl endopeptidase, and human keratins). A “decoy database” was prepared by reverse database. Carbamidomethylcysteine was set as fixed modification, and oxidized methionine, protein N-acetylation, N-pyroglutamate, and deamidation of asparagine and glutamine were searched as variable modifications. Initial mass tolerances for protein identification on MS peaks were 5 ppm and on MS/MS peaks were 10 ppm.
0.5 Da. Two “missed cleavages” were allowed. The instrument setting for the Mascot search was specified as “ESI-Trap.” Peptide identification information was extracted from the Mascot result file into EPICenter (Proxeon Biosystems) (33). Besides the standard search engine results used for peptide assignment (score, expected versus calculated fragment ions, and delta mass), additional empirical information was computed by the EPICenter peptide validation module to assist in the assignment (33). Peptides satisfying the following four criteria were accepted for identification: 1) peptides for which the MS² score were above the 99th percentile of significance (Mascot score >32), 2) fully tryptic peptides with sequence length 7 or longer, 3) peptides for which delta scores (the difference in score between the first and second scoring peptide) were at least 5.0, and 4) peptides for which y ion or b ion score was at least 50.0. The Mascot result file was also imported into MSQuant (75), and the MS³ score was calculated which y ion or b ion score was at least 50.0. The Mascot result file was calculated automatically. Finally the protein identification list was created by accepting the peptides (which passed our criterion mentioned before) and consolidating them according to the following method. Proteins with at least two peptides having a cumulative MS² score (Mascot score) of at least 64 were counted as identified protein. This protein identification criterion corresponds to the confidence of $p = 0.0001$ if both peptide identifications are considered independently. For proteins identified by a single peptide, we required the presence of an MS³ spectrum and a combined score for MS² and MS³ of above 52, which corresponded to a level of false positives of $p = 0.0001$. EPICenter automatically assigns identified peptides to proteins and organizes all proteins with shared peptides into a single group (protein group). EPICenter selects the protein with most of the peptides as an anchor protein and marks proteins that are identified by at least one distinct and separate peptide as conclusively identified proteins. We counted proteins as identified only when a protein conclusively identified as described above or a protein group consisted of only isoforms or overlapping database entries.

**Enrichment Analysis of Gene Ontology (GO) Categories**—Biological Networks Gene Ontology (BinGO) (34), the Cytoscape (35) plugin for finding statistically over- or under-represented GO categories, was used for the enrichment analysis of our adipocyte proteome dataset. The 3T3-L1 proteome dataset was compared against a reference set of complete mouse proteome (IPI mouse) GO annotations. According to the instructions for BinGO (34) the custom GO annotation for the reference set of whole IPI version 3.13 mouse dataset was created by extracting the GO annotations available for mouse IPI identifiers from the EBI Gene Ontology Annotation Mouse 22.0 release (www.ebi.ac.uk/GOA/洌USE-release.html) at the time of release. The Gene Ontology Annotation Mouse 22.0 release contained annotation for 32,776 proteins compiled from different sources. The analysis was done using “Hypergeometric test,” and we selected all GO terms that were significant with $p < 0.0001$ after correcting for multiple term testing by Benjamini and Hochberg false discovery rate. The test of over-represented InterPro terms is called $l_{\text{enrich}}$. For each term $l_{\text{enrich}}[k]$ in set $l_{\text{enrich}}$, the -fold enrichment measure was calculated by the following formula,

$$\text{fold}(l_{\text{enrich}}[k]) = \frac{\% \text{ Adipocyte}(l_{\text{enrich}}[k])}{\% \text{ IPIMouse}(l_{\text{enrich}}[k])}$$  \hspace{1cm} (Eq. 4)

where

$$\% \text{ Adipocyte}(l_{\text{enrich}}[k]) = \frac{\text{Count}(\text{Adipocyte}_{\text{annotated with }} l_{\text{enrich}}[k])}{\text{Count}(\text{Adipocyte}_{\text{annotated in GO Bio process}})}$$  \hspace{1cm} (Eq. 5)

and

$$\% \text{ IPIMouse}(l_{\text{enrich}}[k]) = \frac{\text{Count}(\text{IPIMouse}_{\text{annotated with }} l_{\text{enrich}}[k])}{\text{Count}(\text{IPIMouse}_{\text{annotated in Interpro}})}.$$  \hspace{1cm} (Eq. 6)

Subsequently these enriched InterPro domains were grouped in functional categories according to their representative biological functions.

**Proteome mRNA Concordance Analysis for 3T3-L1 Adipocytes**—To estimate the depth of the proteome we covered in our survey, we compared our identified proteome list with the DGAP microarray dataset for normal 3T3-L1 adipocytes. The available dataset is in triplicates for Affymetrix MG_U74A, B, and C arrays. The MG_U74A, B, and C arrays contain 12,654, 12,636, and 12,728 probe
sets, respectively, which cover known mouse genes and ESTs. The DGAP dataset contains triplicates for each array type. The analysis was carried out in two steps. In the first step we estimated the basal expression of the 3T3-L1 adipocyte transcriptome, and in the second step we mapped our 3T3-L1 adipocyte proteome dataset on the transcriptome data. We analyzed the triplicates of each array type separately and calculated the MAS5 expression values using the “mas5” function implemented in the “affy” package of R software (36). The expression values were then converted to log2 scale. And the log2 expression values were z-transformed to facilitate the comparison of mRNA expression across three array types. The data were further filtered based on the present (P) versus absent (A) call percentage, which is a widely accepted measure of microarray data quality. We used a criterion of 66% P call for accepting a probe set as expressed, i.e. a probe set was accepted if it had a P call in two of three samples. Subsequently the data for the MG_U74A, B, and C arrays were combined in one set. Only 7,656 probe sets of a total 37,886 unique probe sets met this criterion, and they were taken as surrogate for basal 3T3-L1 mRNA expression. Subsequently we mapped our proteome list on the estimated basal expression set. We used BioMart version 0.5.7b (“Mus musculus genes NCBIM36”) dataset to map adipocyte IPI identifiers to the MGU_74A, B and C probe sets. Thus we could map 3,287 IPI identifiers to 3,329 MG_U74(A,B,C) probe sets. Finally the overlap of the adipocyte basal (7,656) probe sets and our survey (3,329) probe sets was calculated. And this gave us a final number of 2,182 probe sets that were found in both the datasets. Hence these 2,182 probe set data were regarded as the genes that we could identify, and the remaining 5,474 probe sets were used for the not-identified set. We used this consolidated information to calculate the average mRNA expression for the identified versus non-identified proteome using the expression levels of 7,656 probe sets.

**Protein Prioritization Analysis**—The recently reported software application for the computational prioritization of genes, Endeavour (26), was used for protein prioritization. 3T3-L1 proteins (IPI identifiers) were mapped to human orthologs (Ensembl Gene identifiers) using BioMart version 0.5.7b (76) and the *Mus musculus* genes NCBIM36 dataset. In total 2,990 IPI protein identifiers were successfully mapped to human Ensembl gene identifiers. The training set (S\textsubscript{Train\_ing}) was created by choosing 29 genes involved in vesicular trafficking in insulin signaling pathway as shown in Fig. 6C. The mapped 3T3-L1 proteome Ensembl list was taken as the candidate test set (S\textsubscript{Test\_ing}). The following data sources were used for ranking: 1) literature (abstracts in EntrezGene), 2) functional annotation (Gene Ontology), 3) microarray expression (Atlas gene expression), 4) EST expression (EST data from Ensembl), 5) protein domains (InterPro), 6) pathway membership from the Kyoto Encyclopedia of Genes and Genomes, 7) cis-regulatory modules (TOUCAN), and 8) sequence similarity (BLAST) data. A model for “vesicular trafficking” genes was created in Endeavour using the above mentioned data sources. Finally the candidate test set (S\textsubscript{Test\_ing}) genes were ranked for their putative role in vesicular trafficking in the insulin pathway by measuring their similarity to genes in training set (S\textsubscript{Train\_ing}).

Annotating the Proteins Annotated as Hypothetical Proteins—To assign putative functions to 335 “hypothetical protein” identifiers we used the Blast2GO (37) tool, which assigns GO annotation to an unknown protein based on its sequence similarity (orthology) to other protein sequences in a preselected database. The GO annotations are assigned based on a four-tier annotation mapping procedure as described in the original study. We used the Swiss-Prot database for this analysis as it serves as the most comprehensive experimentally validated protein database.

Hierarchical Clustering of the Cellular Compartment Profiles of the Adipocyte Proteome—A cellular compartment distribution matrix was created for the 3,287 IPI identifiers that we identified in our analysis. Briefly suppose C is the 3,287 × 4 cellular compartment matrix corresponding to 3,287 proteins and four compartments (nuclei, mitochondria, membrane, and cytosol). For a particular IPI i, i e [1,3287] and a particular compartment column j, j e [1, 4], if the IPI was observed with k unique peptides in the compartment then C[i,j] = k. Otherwise if the IPI i was not observed in compartment column j then C[i,j] = 0. This 3,287 × 4 matrix was called the cellular compartment distribution matrix. The matrix was further converted to a probability distribution matrix, C\textsubscript{prob}, of the same dimensions (3,287 × 4) with each element calculated by the following formula.

\[
C\textsubscript{prob}[i,j] = \frac{C[i,j]}{\sum_{j=1}^{4} C[i,j]} \quad (Eq. 7)
\]

The matrix C\textsubscript{prob} was then used for one-dimensional hierarchical clustering using Genesis (38) software. The distance metric used was “euclidean,” and the clustering was done using the “average linkage clustering” technique. Subsequently the data from earlier large scale studies (20, 22) were overlaid on the clustered dendrogram to ascertain the proteome concordance and depth of our fractionation and subcellular identification. Also the available GO cellular component terms corresponding to the four compartments were extracted for the 3,287 IPI identifiers and overlaid on the clustered dendrogram.

Pathway Mapping of Identified Proteins in Subcellular Compartments—To ascertain the coverage of our dataset with respect to the established key pathways and biological processes, we used the recently developed functional mapping tool GenMapp version 2.1 (39) to map our 3T3-L1 adipocyte dataset on publicly available mouse MAPPs. Because IPI is not currently supported by GenMAPP, we mapped the IPI identifiers to their MGI counterparts using IPI protein cross-reference information as available for IPI mouse version 3.13. Overall 3,124 IPI identifiers (85.6% of the total) could be mapped to their respective MGI identifiers. Subsequently we created a compartment-wide list for the mapped MGI identifiers based on the presence/absence of a particular protein in either of the four compartments, and the data were mapped to the latest available mouse MAPPs.

Adipocyte Proteome Comparison with Previous Human Cell Line and Drosophila Lipid Droplet Proteomics Studies—We used ProteinCenter (Proxome Bioinformatics, Odense, Denmark), which is a proteomics data mining and management software, to compare our datasets with previously published datasets of six human cell lines (40) and the *Drosophila* lipid droplet proteome (5, 6). Briefly for mapping our dataset to any other dataset we loaded the datasets as two groups in ProteinCenter. The datasets were then clustered based on the sequence similarity, and the optimization criterion used was “most homogeneous groups” wherein protein sequences are clustered to make the individual groups as homogeneous as possible. The similarity threshold of 80% was chosen as the cutoff to define clusters of sequences. The adipocyte proteins belonging to the clusters of at least two proteins with at least one identifier from the compared dataset were deemed overlapping with that dataset. The adipocyte proteins belonging to singleton clusters, which could not be clustered with any sequence of the compared dataset, were chosen as adipocyte-specific.

**RESULTS AND DISCUSSION**

**High Confidence Protein Identification of Mouse Adipocyte Organelles**—To reduce the complexity of the proteome and obtain compartment-specific information in adipocytes, we performed differential ultracentrifugation from differentiated 3T3-L1 adipocyte cells. The proteome of four compartments (nuclei, mitochondria, membrane, and cytosol) was examined using the workflow depicted in Fig. 1. As shown in Fig. 2,
Purity of subcellular compartments was excellent as visualized by organellar markers across the four fractions. To reduce sample complexity and dynamic range in protein abundance levels, proteins in each compartment were separated by one-dimensional SDS-PAGE, and 11 or 12 bands were excised and subjected to in-gel tryptic digestion. In total, 45 fractions were analyzed by LC on-line-coupled to electrospray mass spectrometry. We used a hybrid mass spectrometer consisting of an LTQ-FTICR instrument. The mass spectrometer was programmed to perform survey scans of the whole peptide mass range, select the three most abundant peptide signals, and perform narrow range, selected ion mon-
itoring (SIM) scans for high mass accuracy measurements. Simultaneously with the SIM scans, the linear ion trap fragmented the peptide, obtained an MS/MS spectrum, and further isolated and fragmented the most abundant peak in the MS/MS mass spectrum to yield the MS3 spectrum (32). Fig. 3A shows an MS mass spectrum of eluting peptides (see Ref. 41 for an introduction to peptide sequencing). A selected peptide was measured in SIM mode and fragmented (MS2) (Fig. 3B). The most intense fragment in the MS2 spectrum was selected for the second round of fragmentation (Fig. 3C). As can be seen in the figure, high mass accuracy, low background level, and additional peptide sequence information obtained from MS3 spectra yield high confidence peptide identification. Total cycle time for the analysis described above was approximately 5 s.

To obtain the protein “parts list” of adipocytes, high confidence protein identification and reporting were essential. We applied a stringent multistep filter to minimize false-positive identifications while maintaining favorable detection of lower abundance and lower molecular weight proteins (see “Experimental Procedures”). In addition to standard search engine results, i.e. the Mascot probability score (42), MS³ score and additional empirical information (y ion and b ion score, number of sibling peptides, and proline score) (33) were used for peptide and protein identification. Proteins were identified with criteria corresponding to an estimated probability of false positives of \( p = 0.0001 \). We also performed a decoy database search (43) to test the experimental level of false-positive rate in our dataset. In this approach peptides are matched against the database containing forward oriented normal sequences and the same sequences with their amino acid sequences reversed. After applying the stringent criteria mentioned above, we found no false positives for protein identified with two or more peptides and only two false-positive protein hits with one peptide. These results indicate a false-positive identification rate of 2 of 3,287 or 0.06% at the protein level. Thus we conclude that our dataset contains no or very few false-positive identifications. The peptides identified in this study will also be made available in repositories such as PeptideAtlas (44).
Determining the identities of proteins from sequenced peptides is complicated because the same peptide sequence can be present in multiple different proteins or protein isoforms (45). Standard search engines such as Mascot report proteins even when they do not have distinct peptides with their sequence specific to these proteins. Thus, sharing information on identified peptides was checked by EPICenter (33) and manually verified. EPICenter organizes all entries with shared peptides into a single group (protein group). The protein that contains most of the peptides is selected as an anchor, and all peptides into a single group (protein group). The protein that is manually verified. EPICenter organizes all entries with shared sequence specific to these proteins. Thus, sharing information even when they do not have distinct peptides with their sequence specific to these proteins. We counted proteins as identified only when a protein had at least one distinct peptide. If a protein group consisted of only isoforms or overlapping database entries indistinguishable by MS then only the anchor protein was counted; thus the number of identified proteins is a lower boundary of the actual value.

From 45 LC-MS/MS/MS runs, 182,271 MS/MS spectra were submitted to Mascot database search, and 52,585 MS/MS spectra satisfied the criteria for peptide validation. Among these, 22,706 represent unique sequences in the four fractions. A total of 3,287 proteins were identified from four fractions with 8,953 unique peptides (Supplemental Tables 1 and 2). Of all proteins, 20.4% were identified with single peptide identification and two stages of peptide fragmentation. Interestingly 71.3% of the total were identified within only one fraction, whereas 16.7, 7.8, and 4.2% were identified within two, three, and four fractions, respectively. This confirms the relatively high quality of organellar separation as already suggested by marker analysis in the Western blotting experiments in Fig. 2. In two previous studies analyzing several mouse tissues, each similarly separated into four subcellular compartments, protein overlap among compartments was deeper, and fewer than 50% of total identified proteins were specific to one compartment (22, 46).

**Depth and Coverage of the 3T3-L1 Adipocyte Proteome**—We compared our proteome dataset with the recently published mouse liver organelle proteome map (20) and the above mentioned study of six mouse tissues (brain, heart, kidney, liver, lung, and placenta) (22). As shown in Fig. 4A, more than two-thirds of the proteins identified by us in adipocytes overlapped with those other proteomes. These proteins are candidates for the “household proteome,” i.e. proteins performing general cellular functions and therefore present in different cell lines and tissues. However, the proportion of proteins specific to adipocytes in our study (28.7%) was also relatively high. In contrast, in a previous study that compared six human cell lines, specific proteins (proteins that were exclusively found in a single cell line) account for only 6–36% of all identified proteins (40). As shown in Fig. 4B, nearly half of our cytoplasmic proteins overlapped with the combined cytoplasmic proteins from six human cell lines (40). Secreted proteins were enriched ~3.5-fold in identified cytoplasmic proteins from adipocyte compared with cytoplasmic proteins from the six-cell line proteome. Fig. 4

**Fig. 4.** Two-thirds of adipocyte cell line proteins are also found in recent mouse organelle studies. A, mouse proteins reported recently in membrane-enclosed organelles of mouse liver (Foster et al. (20)) and a study of six mouse tissues (but without fat tissue) (Kislinger et al. (22)) were BLASTed against identified proteins in the current study by ProteinCenter (Proxeon Bioinformatics). Only those proteins with at least 95% identity were considered to match. B, the adipocyte cytoplasmic proteins were “BLASTed” against the combined cytoplasmic six-cell line proteome (Schirle et al. (40)). Only those proteins with at least 80% sequence identity were considered to match.

**Fig. 5.** Cumulative probability distribution of the mRNA levels of identified versus not identified proteins. Proteins identified in adipocytes were mapped onto DGAP 3T3-L1 mRNA data. The cumulative probability distributions of mRNA abundance for the genes whose protein products were detected (green) or not detected (pink) by proteomics are shown. The mean expression levels for both groups are indicated.
clearly shows that our proteome dataset contains many proteins that were not identified in previous large scale proteome analysis using both tissues and cell lines. Our result may reflect the depth of high confidence analysis now possible and/or the specificity of adipocytes for their critical role in energy balance and whole body homeostasis.

Because the Drosophila lipid droplet is a common genetic model for adipocyte-related functions, we also compared the adipocyte proteome to the Drosophila lipid droplet proteome reported in two recent publications. However, only 15 (6%) and 17 (13%) of the lipid droplet proteins had homologs at the 80% similarity level in our adipocyte dataset, respectively (5, 6). This number is not higher than expected by chance and may reflect the fact that we are comparing an organelle with a cellular proteome and that Drosophila does not have specific adipose tissue.

**Fig. 6.** Pathway mapping of identified proteins in the four subcellular fractions. Proteins identified in four cellular compartments (cytosol, membrane, mitochondria, and nuclei) were mapped to known pathways using GenMAPP. A, proteasome degradation; B, cytoplasmic ribosomal proteins; C, insulin signaling pathway. The compartments are represented by four different colors for each gene symbol. MHC, major histocompatibility complex; E1, ubiquitin-activating enzyme; E2, ubiquitin carrier protein; E3, ubiquitin-protein isopeptide ligase; PI, phosphatidylinositol; PDK, protein-dependent kinase; MAP, mitogen-activate protein; MAPK, mitogen-activated protein kinase; MEK, mitogen-activated protein kinase/extracellular signal-regulated kinase kinase.
Previously microarray studies have been undertaken to unravel various aspects of 3T3-L1 adipocyte differentiation, development, and function (47–50) and because they serve as a useful resource for gaining insights into mRNA expression and cellular dynamics. Moreover microarray studies provide an estimate of the transcriptome in a particular biological state at any given time, and we wished to use these data as a reference for estimating the coverage and depth of our large scale proteome study. We analyzed the gene expression levels of normal 3T3-L1 adipocyte from Affymetrix microarray data generated in the DGAP. The available dataset contains triplicates for each of the MGU_74A, B, and C Affymetrix array types. We combined the three array type datasets for our analysis. In total they contain 37,886 probe sets of which 7,656 were deemed “present” by using the 66% P call criterion (see “Experimental Procedures”). Of these 7,656 probes, 2,182 could be mapped to our identified proteome.

We then divided the genes judged to be expressed in 3T3-L1 according to the microarray data into two groups: those identified in our study and those not identified (Fig. 5). If proteomics was biased to detect only high abundance proteins, we would expect a large difference in mRNA signal between the two groups. Remarkably the distribution of mRNA expression levels was less than 2-fold higher for the genes whose products were detected in our proteomics analysis as compared with those that were not identified. To further substantiate this finding, we also compared our proteome data with the Gene Atlas V2 mouse microarray data for adipose tissue (51). Again we observed that the distribution of mRNA expression level was less than 2-fold higher for the genes whose products were detected in our proteomics analysis as compared with those that were not (Supplemental Fig. 1). This suggests that proteomics experiments, despite the remaining limitations in complex mixture analysis (52), have
become quite comprehensive and able to detect low abundance proteins in cellular proteomes. Our previous study on mouse tissue mitochondria, in contrast, still showed a substantial tendency of mass spectrometry to preferentially detect products of high abundance messages (21).

We then analyzed the coverage of our proteome in terms of protein complex and pathways known to be present in adipocytes. We mapped identified proteins onto the Kyoto Encyclopedia of Genes and Genomes pathway (53). As shown in Fig. 6A and B, almost all of the known proteins in the ribosome and proteasome complexes were identified. It is interesting that in the membrane fraction only the core 20 S proteasome, and not the 19 S (PA700) and 11 S (Psme1–3) regulatory complexes, was identified (Fig. 6A). The precise distribution of 20 S and 19 S complexes at the cellular organelle level has not been reported in the literature. However, in agreement with our observations, Brooks et al. (54) reported that the 20, 11, and 19 S complexes localized predominantly in the cytosol and also localized in nuclear and membrane fractions prepared from rat liver. The function of the 20 S proteasome at the adipocyte membrane is unknown and would be interesting to elucidate in future studies.

In contrast, less than half of the known proteins in the insulin pathway map were identified (Fig. 6C). Coverage of kinases and transcription factors was low, whereas we detected more than half of the proteins related to vesicular trafficking. Interestingly in the analysis of 3T3-L1 microarray data half of the key components of the insulin signaling path-

![Diagram](image_url)
way were also flagged as not expressed (i.e. filtered out by microarray quality measures when using the P calls >66% criterion). This suggests that the limitation of detection of low abundance proteins/genes such as kinases and transcription factors are no worse for MS-based proteomics than for other high throughput “omics" technologies.

Endocrine function is one of the key functions of adipocytes, and several hormones, called adipokines or adipocytokines, are synthesized by adipose tissue. Among them, adiponectin, adipin, and nicotinamide phosphoribosyltransferase (visfatin) were identified in this study. Furthermore we identified proteins secreted from 3T3-L1 cells (9, 14, 15) including procollagens, complement C3, galectin-3, prohibitin, thioredoxin, peroxiredoxin 1, \( \beta_2 \)-microglobulin, haptoglobin, vimentin, cathepsin B, and CD14 antigen. We also identified NUCB1 (Nucleobindin-1) and NUCB2 (Nucleobindin-2) proteins, which are homologous and are reported to be secreted, although until recently their precise function was unknown (55, 56). Interestingly in a very recent study an N-terminal fragment derivative of NUCB2 named nesfatin-1 (NEFA/Nucleobindin-2-encoded satiety-and fat-influencing protein-1) was shown to act as a satiety molecule in the hypothalamus of rat brain (57). However, we did not identify tumor necrosis factor \( \alpha \), interleukin-6, leptin, and plasminogen activator inhibitor 1 precursor (PAI-1). The absence of leptin in our survey may not only be due to its extremely low abundance but may also be explained by earlier observations that leptin expression at the mRNA level in cultured 3T3-L1 adipocytes is markedly down-regulated or absent compared with the leptin expression in mouse adipose tissue (58). Moreover we classified 60 proteins as secreted (GO:0046903), and our dataset was more than 2.5-fold enriched for secreted proteins when compared with complete IPI mouse proteome dataset version 3.07. Also 482 proteins had a predicted signal peptide, corroborating the endocrine nature of the adipocyte proteome.

Annotation: Subcellular Localization—To compare our subcellular fractions and detection coverage we benchmarked our data against previously reported large scale proteomics analyses of mouse organelles (20, 22) and Gene Ontology annotation. For our proteome, we built a cellular compartment distribution matrix (3,287 proteins \( \times \) 4 fractions) by first counting peptides of each protein in each fraction. Then we normalized the data to arrive at a probability matrix for the distribution of each protein in the four compartments (see “Experimental Procedures”). Hierarchical clustering of this matrix showed that more than 70% of the proteins localized to four clusters with propensity for one specific fraction (Fig. 7). A fifth cluster contained proteins for which there was no clear pattern of distribution. Additionally we overlaid the data of the two large scale mouse tissue experiments as well as cellular compartment annotations from Gene Ontology on the clustered dendrogram. As seen in the figure, most of the adipocyte cytosolic proteome showed high concordance with the experimental studies. Similarly the membrane cluster showed good agreement. For the mitochondrial and nuclear fractions there was excellent correlation; however, the depth of our study was much greater. (Only the top part of our clusters have detected counterparts in the other studies.) The GO annotations agreed well for the mitochondria- and nucleus-specific clusters. No major enrichment was seen for the membrane and cytoplasmic fractions, which seem to be less well annotated in GO.

Annotation: Cellular Process and Molecular Function—The identified proteins were functionally categorized based on Gene Ontology annotation terms using the BiNGO program package (34). In total, 2,437 proteins were linked to at least one annotation term within the GO molecular function category, and 129 GO terms were over-represented at significance value \( p < 0.001 \) when compared with the GO annotation of the entire IPI mouse database. For the GO biological process category 2,135 proteins had an annotation, and 208 GO terms were over-represented. To remove the influence of the household proteome and to gain insights into adipocyte-specific molecular functions and biological processes, we also performed the GO enrichment analysis using proteome datasets obtained from six mouse tissues (22). Common over-represented terms between adipocyte proteome and the large scale tissue study were removed. The resulting GO terms enriched only in the adipocyte proteome are shown in Supplemental Figs. 2 and 3. Most of the terms were related to protein metabolism, catalysis, oxidoreductase activities, binding, sterol metabolism, protein transport, and mitochondrial biogenesis. This result supports the notion that the function of adipocytes is not only limited to be the fat repository of body but that they are also active in protein metabolism and energy metabolism, although most of the cell volume is occupied by lipid droplet.

Annotation: Novel Proteins—Coverage and level of detail of current GO annotation is one of the limitations for the analysis using Gene Ontology. As mentioned above only about two-thirds of the identified proteins were mapped to at least one annotation term within the GO molecular function and biological process categories. Furthermore we found that 335 proteins had no specific protein names and gene symbols; thus we used the Blast2GO (37) tool to assign putative GO annotations to these terms. This allowed us to assign 175 IPI identifiers to GO terms as well as 45 IPI identifiers to Enzyme Classification (EC) identifiers. Hence overall we could assign putative annotation to more than 50% of unannotated proteins in our identified dataset (Supplemental Table 3). The assigned GO biological process terms were mainly related to cellular metabolism, transport, protein biosynthesis, organelle organization and biogenesis, and cellular physiological process (Supplemental Fig. 4). The GO molecular function terms were mainly related to protein binding, nucleic acid binding, RNA binding, catalytic activity, hydrolase activity, oxidoreductase activity, and structural molecule activity (Supplemental Fig. 5). This result aligns well with our enrichment analysis of...
GO annotations for the whole adipocyte proteome and indicates likely roles of these hypothetical/unknown proteins in adipocyte function and biology.

Protein Domain Enrichment for Insights into Protein Function—Classification of proteins based on their amino acid sequence or three-dimensional structure is one of the most
FIG. 8. Significantly over-represented InterPro domains for the set of identified adipocyte proteins. Significantly over-represented InterPro terms with $p < 0.001$ and not also over-represented in the mouse tissue analysis are shown. For each InterPro term the bar shows the enrichment fold ratio for the identified 3T3-L1 proteome in our survey with respect to the InterPro annotations of the entire mouse proteome. The InterPro terms are further grouped by representative biological function shown with text in blue. SH3, Src homology 3; FKBP, FK506-binding protein. OB, oligosaccharide/oligonucleotide-binding.
Molecular & Cellular Proteomics 6.7

Table I

Putative proteins involved in the vesicular trafficking in insulin signaling predicted by the method of gene prioritization

| IPI identifier | Description | Gene symbol | p value | Number of unique peptides identified in nucleolar organelle |
|---------------|-------------|-------------|---------|----------------------------------------------------------|
| IPI00116770   | Rab-18, member Ras oncogene family | Rab18 | 2.57E-06 | 1 | 4 | 9 | 5 |
| IPI00221615   | ADP-ribosylation factor 5 | Arf5 | 3.82E-06 | 0 | 0 | 0 | 11 |
| IPI00271059   | Rab-4B, member Ras oncogene family | Rab4b | 4.77E-06 | 0 | 0 | 4 | 0 |
| IPI00132276   | Vesicle-associated membrane protein 3 | Vamp3 | 6.03E-06 | 0 | 2 | 6 | 2 |
| IPI00113849   | Cell division cycle 42 homolog | Cdc42 | 6.83E-06 | 0 | 0 | 3 | 6 |
| IPI00331663   | Unnamed protein product | Arf4 | 7.24E-06 | 0 | 0 | 0 | 7 |
| IPI00122965   | Rab-3A, member Ras oncogene family | Rab3a | 8.52E-06 | 0 | 0 | 0 | 2 |
| IPI00114560   | Rab-1A, member Ras oncogene family | Rab1 | 9.91E-06 | 2 | 0 | 14 | 10 |
| IPI00137227   | Rab-2A, member Ras oncogene family | Rab2 | 1.07E-05 | 2 | 5 | 14 | 5 |
| IPI00132410   | Rab-5A, member Ras oncogene family | Rab5a | 1.13E-05 | 1 | 4 | 6 | 0 |
| IPI00118217   | Syntaxin-7 | Stx7 | 1.24E-05 | 0 | 0 | 3 | 0 |
| IPI00453589   | Vesicle-associated membrane protein 8 (VAMP-8) | Vamp8 | 1.43E-05 | 0 | 0 | 2 | 3 |
| IPI00126042   | Rab-14, member Ras oncogene family | Rab14 | 1.49E-05 | 3 | 5 | 17 | 7 |
| IPI00230011   | Rab6A protein | Rab6 | 1.61E-05 | 0 | 1 | 10 | 0 |
| IPI00137647   | Synaptobrevin-like 1 | Syb1 | 1.71E-05 | 0 | 0 | 0 | 2 |
| IPI00321581   | GS32 | Snap29 | 1.90E-05 | 0 | 0 | 0 | 1 |
| IPI00124291   | Vacular protein sorting 45 | Vps45 | 2.35E-05 | 0 | 0 | 7 | 0 |
| IPI00331128   | Cell line NK14-derived transforming oncogene | Rab8a | 2.36E-05 | 0 | 3 | 7 | 4 |
| IPI00116729   | Ras-related protein Rab-22A (Rab-22) (Rab-14) | Rab22a | 3.24E-05 | 0 | 0 | 3 | 0 |
| IPI00125880   | Protein kinase C and casein kinase substrate in neurons 2 | Pascin2 | 4.34E-05 | 0 | 0 | 2 | 0 |
| IPI00111416   | Syntaxin-12 | Stx12 | 4.87E-05 | 0 | 0 | 6 | 3 |
| IPI00469799   | Splice isoform 2 of Golgi autoantigen, Golgin subfamily A member 3 | Golga1 | 4.91E-05 | 0 | 0 | 2 | 0 |
| IPI00416303   | Aminopeptidase-like-1 | Npep1 | 5.52E-05 | 0 | 0 | 0 | 2 |
| IPI00224219   | sec1 family domain-containing 1 | Scfd1 | 6.32E-05 | 0 | 0 | 9 | 4 |
| IPI00109506   | Unnamed protein product | Stx6 | 6.79E-05 | 0 | 0 | 9 | 0 |
| IPI00224518   | Rab-related protein Rab-5C | Rab5c | 7.15E-05 | 2 | 0 | 10 | 0 |
| IPI00225581   | Dedicator of cytokinesis protein 1 (fragment) | Dock1 | 7.19E-05 | 0 | 0 | 1 | 0 |
| IPI00134941   | c-K-ras protein | Kras | 9.95E-05 | 0 | 2 | 1 | 0 |
| IPI00116558   | Ras homolog gene family, member G | Rhog | 9.96E-05 | 0 | 0 | 4 | 0 |
| IPI00121335   | Thymoma viral proto-oncogene 2 | Aktr2 | 1.09E-04 | 0 | 0 | 2 | 0 |
| IPI00116688   | Rab-3D, member Ras oncogene family | Rab3d | 1.09E-04 | 0 | 0 | 5 | 0 |
| IPI00378015   | Drebrin-like protein (SH3 domain-containing protein 7) | Dbnl | 1.17E-04 | 0 | 0 | 0 | 2 |
| IPI00130118   | Rab-10, member Ras oncogene family | Rab10 | 1.28E-04 | 1 | 4 | 10 | 5 |
| IPI00378145   | Rab-6B, member Ras oncogene family | Rab6b | 1.30E-04 | 0 | 1 | 0 | 5 |
| IPI00453776   | Early endosome antigen 1 | Eea1 | 1.40E-04 | 0 | 0 | 0 | 16 |
| IPI00433771   | Prenylated SNARe protein Ypt6 | Ypt6 | 1.41E-04 | 0 | 0 | 3 | 4 |
| IPI00132665   | Blocked early in transport 1 homolog | Bet1 | 1.41E-04 | 1 | 2 | 1 | 1 |
| IPI00229483   | SEC24-related gene family, member C | Sec24c | 1.43E-04 | 0 | 0 | 10 | 2 |
| IPI00131445   | Dynamin-2 (dynamin UDNN) | Dnm2 | 1.49E-04 | 0 | 0 | 4 | 0 |
| IPI00331284   | Vesicle transport through interaction with t-SNAREs 1B homolog | Vti1b | 1.50E-04 | 1 | 0 | 6 | 0 |
| IPI00130554   | SNAP-associated protein | Snap2 | 1.94E-04 | 0 | 0 | 0 | 2 |

established practices in protein science and also adopted by current large scale structural genomics endeavors (59). Moreover, knowledge of independently folding protein domains can provide useful pointers into the complex interplay of proteome interactions and regulation by post-translational modifications (60). To obtain an additional perspective of the adipocyte proteome we performed InterPro domain enrichment analysis using our adipocyte proteome dataset and the proteome datasets obtained from six mouse tissues (22) and extracted InterPro domains enriched only in the adipocyte proteome (Fig. 8). The functions of these enriched domains were mainly related to signal transduction, redox system, protein transport, translation, transcription, protein degradation, fatty acid metabolism, and phospholipid biosynthesis in agreement with the enriched GO terms described before. Enrichment of redox-related domains, such as the thioredoxin domain, is interesting because it has been suggested that the reduced redox state encourages triglyceride synthesis, adipocyte differentiation, and the development of adipose tissue (61), whereas an increase in the markers of systemic oxidative stress has been associated with obesity and metabolic syndrome (62). Similarly the domains related to vesicular protein transport, such as Rab small GTPase, Rab type, t-SNARE, Longin-like, and a domain zinc finger, Tim10/DDP-type, which is related to protein import into mitochondrial inner membrane, were substantially enriched.
Domains related to transcription and translation were also enriched; especially three domains of aminoacyl-transfer RNA synthetases were enriched, namely aminoacyl-tRNA synthetase, class 1α, anticodon-binding; aminoacyl-tRNA synthetase, class II; and aminoacyl-tRNA synthetase, class I. Transcription and translation are basic functions of the cell; thus proteins related to such functions are generally thought to be housekeeping proteins. This observed enrichment may not reflect basic adipocyte biology but simply the fact that a rapidly growing cell line needs to express more proteins than the comparatively more inert tissue. Further insights may be obtained by quantitative study of protein expression in different cell lines and tissues and by creating a protein expression atlas similar to a gene expression atlas (63, 64).

Protein Prioritization Analysis on Vesicular Trafficking in Adipocytes—One of the important features of adipocytes is insulin-regulated glucose uptake. In adipocytes, the majority of this glucose uptake results from the translocation of the glucose transporter 4 (GLUT4) to the cell surface membrane. Since the cloning of GLUT4 in 1989 in several laboratories (65–69) numerous studies have attempted to elucidate the molecular basis of the insulin receptor signaling pathway and membrane trafficking processes. One of the unresolved questions is the connection between Akt activation and GLUT4 translocation. GO term enrichment analysis revealed that protein transport was enriched in the adipocyte proteome (see above), and some of the identified vesicular trafficking proteins are known to be involved in the insulin signaling pathway (Fig. 6C). As 26, 35, and 36% of identified proteins in our study were not annotated by Gene Ontology molecular function, biological process, and cellular component categories, respectively, we tried to predict candidate proteins involved in GLUT4 translocation using a bioinformatics approach. Very recently, an algorithm termed Endeavour was developed for gene prioritization to rank genes involved in human diseases and biological processes (26). The concept of prioritization by Endeavour is that candidate test genes are ranked based on their similarity with a set of known training genes. The similarity measure is in turn calculated by integrating functional, process, gene ontology (GO), pathway, and sequence similarity information obtained from diverse data sources. As training genes we chose 29 genes involved in vesicular trafficking that are on the map in Fig. 6C. We used 2,990 proteins that were identified and mapped to human Ensembl gene identifiers in our study as test genes. For 41 gene products we identified and mapped to human Ensembl gene identifiers in our study as test genes. For 41 gene products we identified and mapped to human Ensembl gene identifiers in our study as test genes.

Protein Prioritization Analysis on Vesicular Trafficking in Adipocytes—One of the important features of adipocytes is insulin-regulated glucose uptake. In adipocytes, the majority of this glucose uptake results from the translocation of the glucose transporter 4 (GLUT4) to the cell surface membrane. Since the cloning of GLUT4 in 1989 in several laboratories (65–69) numerous studies have attempted to elucidate the molecular basis of the insulin receptor signaling pathway and membrane trafficking processes. One of the unresolved questions is the connection between Akt activation and GLUT4 translocation. GO term enrichment analysis revealed that protein transport was enriched in the adipocyte proteome (see above), and some of the identified vesicular trafficking proteins are known to be involved in the insulin signaling pathway (Fig. 6C). As 26, 35, and 36% of identified proteins in our study were not annotated by Gene Ontology molecular function, biological process, and cellular component categories, respectively, we tried to predict candidate proteins involved in GLUT4 translocation using a bioinformatics approach. Very recently, an algorithm termed Endeavour was developed for gene prioritization to rank genes involved in human diseases and biological processes (26). The concept of prioritization by Endeavour is that candidate test genes are ranked based on their similarity with a set of known training genes. The similarity measure is in turn calculated by integrating functional, process, gene ontology (GO), pathway, and sequence similarity information obtained from diverse data sources. As training genes we chose 29 genes involved in vesicular trafficking that are on the map in Fig. 6C. We used 2,990 proteins that were identified and mapped to human Ensembl gene identifiers in our study as test genes. For 41 gene products we identified and mapped to human Ensembl gene identifiers in our study as test genes. For 41 gene products we identified and mapped to human Ensembl gene identifiers in our study as test genes. For 41 gene products we identified and mapped to human Ensembl gene identifiers in our study as test genes. For 41 gene products we identified and mapped to human Ensembl gene identifiers in our study as test genes. For 41 gene products we identified and mapped to human Ensembl gene identifiers in our study as test genes. For 41 gene products we identified and mapped to human Ensembl gene identifiers in our study as test genes.
Probabilistic protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567.

43. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* 2, 667–675.

44. Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., Chen, S., Eddes, J., Loevenich, S. N., and Aebersold, R. (2006) The PeptideAtlas project. *Nucleic Acids Res.* 34, D655–D658.

45. Nesvizhskii, A. I., and Aebersold, R. (2005) Interpretation of shotgun proteomic data: the protein identification problem. *Mol. Cell. Proteomics* 4, 1419–1440.

46. Jiang, X. S., Zhou, H., Zhang, L., Sheng, Q. H., Li, S. J., Li, L., Hao, P., Y. X., Xia, Q. C., Wu, J. R., and Zeng, R. (2004) A high-throughput approach for subcellular proteomics: identification of rat liver proteins using subcellular fractionation coupled with two-dimensional liquid chromatography tandem mass spectrometry and bioinformatic analysis. *Mol. Cell. Proteomics* 3, 441–455.

47. Soukas, A., Socci, N. D., Saatkamp, B. D., Novelli, S., and Friedman, J. M. (2001) Distinct transcriptional profiles of adipogenesis in vivo and in vitro. *J. Biol. Chem.* 276, 34167–34174.

48. Burton, G. R., Nagarajan, R., Peterson, C. A., and McGehee, R. E., Jr. (2004) Microarray analysis of differentiation-specific gene expression during 3T3-L1 adipogenesis. *Gene (Amst.)* 329, 167–185.

49. Gerhold, D. L., Liu, F., Jiang, G., Li, Z., Xu, J., Li, L., Han, Z., Li, Y. X., Xia, Q. C., Wu, J. R., and Zeng, R. (2004) A high-throughput approach for subcellular proteomics: identification of rat liver proteins using subcellular fractionation coupled with two-dimensional liquid chromatography tandem mass spectrometry and bioinformatic analysis. *Mol. Cell. Proteomics* 4, 1419–1440.

50. Burton, G. R., Nagarajan, R., Peterson, C. A., and McGehee, R. E., Jr. (2004) Microarray analysis of differentiation-specific gene expression during 3T3-L1 adipogenesis. *Gene (Amst.)* 329, 167–185.

51. Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Somogyi, M., Friedman, A., Holder, D. J., Doebber, T. W., Berger, J., Elbrecht, A., Lin, M. C., Hendil, K. B., Tanaka, K., Dyson, J., and Rivett, J. (2000) Subcellular localization of proteasomes and their regulatory complexes in mammalian cells. *Biochem. J.* 346, 167–185.

52. Kaneko, H., Kurosawa, K., and Kinosita, K. (2002) Oxidative stress is associated with adiposity and insulin resistance in men. *J. Clin. Endocrinol. Metab.* 88, 4673–4676.

53. Siddiqui, A. S., Khattra, J., Delaney, A. D., Zhao, Y., Astell, C., Asano, J., Babakaff, R., Barber, S., Beland, J., Bhohacce, S., Brown-John, M., Chand, S., Charest, D., Charters, A. M., Cullum, R., Dhalla, N., Featherstone, R., Gerhard, D. S., Hoffman, B., Holt, R. A., Hou, J., Kuo, B. Y., Lee, L. L., Lee, S., Leung, D., Ma, K., Matsuoc, S., Mayb, M., McDonald, H., Prabu, A. L., Pandh, P., Riggins, G. J., de Alaga, T. R., Rupert, J. L., Smalious, D., Stott, J., Tsai, M., Varholc, R., Vrijlac, P., Wong, D., Wu, M. K., Xie, Y. Y., Yang, G., Zhang, I., Hirsc, M., Jones, S. J., Helgason, C. D., Simpson, E. M., Hoodles, P. A., and Marra, M. A. (2005) A mouse atlas of gene expression: large-scale digital gene expression profiles from precisely defined developing C57BL/6J mouse tissues and cells. *Proc. Natl. Acad. Sci. U. S. A.* 102, 18485–18490.

54. Bhattacharjee, Y. (2006) Neuroscience. ‘Google of the brain’: atlas maps brain’s genetic activity. *Science* 313, 1879.

55. Birmbaum, M. J. (1989) Identification of a novel gene encoding an insulin-responsive glucose transporter protein. *Cell* 57, 305–315.

56. Charron, M. J., Brosius, F. C., III, Alper, S. L., and Lodish, H. F. (1989) A glucose transport protein expressed predominately in insulin-responsive tissues. *Proc. Natl. Acad. Sci. U. S. A.* 86, 2535–2539.

57. Fukumoto, H., Kayano, T., Buse, J. B., Edwards, Y., Plch, P. F., Bell, G. I., and Seino, S. (1989) Cloning and characterization of the major insulin-responsive glucose transporter expressed in human skeletal muscle and other insulin-responsive tissues. *J. Biol. Chem.* 264, 7776–7777.

58. James, D. E., Strube, M., and Mueckler, M. (1989) Molecular cloning and characterization of an insulin-regulatable glucose transporter. *Nature* 338, 83–87.

59. Kaestner, K. H., Christy, R. J., McLenithin, J. C., Braiterman, L. T., Cornelius, P., Pekala, P. H., and Lane, M. D. (1989) Sequence, tissue distribution, and differential expression of mRNA for a putative insulin-responsive glucose transporter in mouse 3T3-L1 adipocytes. *Proc. Natl. Acad. Sci. U. S. A.* 86, 3150–3154.

60. Minea, C. P., Sano, H., Kane, S., Sano, E., Fukuda, M., Peranan, J., Lane, W. S., and Lienhard, G. E. (2005) AS160, theAkt substrate regulating GLUT4 translocation, has a functional Rab GTPase-activating protein domain. *Biochem. J.* 391, 87–93.

61. Immamura, T., Huang, J., Usui, I., Satoh, H., Bever, J., and Olefsky, J. M. (2003) Insulin-induced GLUT4 translocation involves protein kinase C-lambda-mediated functional coupling between Rab4 and the motor protein kinesin. *Mol. Cell. Biol.* 23, 4792–4900.

62. Millar, C. A., Powell, K. A., Hickson, G. R., Bader, M. F., and Gould, G. W. (2000) Evidence for a role for ADP-ribosylation factor 6 in insulin-stimulated glucose transport in mouse 3T3-L1 adipocytes. *J. Biol. Chem.* 275, 17619–17625.

63. Usui, I., Immamura, T., Huang, J., Satoh, H., and Olefsky, J. M. (2003) Cdc42 is a Rho GTase family member that can mediate insulin signaling to glucose transport in 3T3-L1 adipocytes. *J. Biol. Chem.* 278, 13765–13774.

64. Zhang, Y., Zhang, Y., Adachi, J., Osen, J. V., Shi, R., de Souza, G., Pasini, E., Foster, L. J., Macke, S., Zouman, A., Kumor, C., Wiringa, J. R., Jun, W., and Mann, M. (2007) MAPU: Max-Planck Unified database of organellar, cellular, tissue and body fluid proteomes. *Nucleic Acids Res.* D655–D658.

65. Schulze, W. X., and Mann, M. (2004) A novel proteomic screen for peptide-protein interactions. *J. Biol. Chem.* 279, 10756–10764.

66. Kasprzyk, A., Keefe, D., Smedley, D., London, D., Spooner, W., Melsopp, C., Hammond, M., Rocca-Serra, P., Cox, T., and Birney, E. (2004) EnsMart: a generic system for fast and flexible access to biological data. *Genome Res.* 14, 160–169.