



Published in final edited form as:

J Proteome Res. 2008 August ; 7(8): 3114–3126. doi:10.1021/pr800205b.

Characterization of the Mouse Pancreatic Islet Proteome and Comparative Analysis with Other Mouse Tissues

Vladislav A. Petyuk[‡], Wei-Jun Qian[‡], Charlotte Hinault[§], Marina A. Gritsenko[‡], Mudita Singhal[†], Matthew E. Monroe[‡], David G. Camp II[‡], Rohit N. Kulkarni^{§,*}, and Richard D. Smith^{‡,*}

[‡]Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352

[§]Division of Cell and Molecular Physiology, Joslin Diabetes Center, Department of Medicine, Harvard Medical School, Boston, MA 02215

[†]Computational Biology and Bioinformatics Group, Pacific Northwest National Laboratory, Richland, WA 99352

Abstract

The pancreatic islets of Langerhans, and especially the insulin-producing beta cells, play a central role in the maintenance of glucose homeostasis. Alterations in the expression of multiple proteins in the islets that contribute to the maintenance of islet function are likely to underlie the pathogenesis of type 2 diabetes. To identify proteins that constitute the islet proteome, we provide the first comprehensive proteomic characterization of pancreatic islets for mouse, the most commonly used animal model in diabetes research. Using strong cation exchange fractionation coupled with reversed phase LC-MS/MS we report the confident identification of 17,350 different tryptic peptides covering 2,612 proteins having at least two unique peptides per protein. The dataset also identified ~60 post-translationally modified peptides including oxidative modifications and phosphorylation. While many of the identified phosphorylation sites corroborate those previously known, the oxidative modifications observed on cysteinyl residues reveal potentially novel information suggesting a role for oxidative stress in islet function. Comparative analysis with 15 available proteomic datasets from other mouse tissues and cells revealed a set of 133 proteins predominantly expressed in pancreatic islets. This unique set of proteins, in addition to those with known functions such as peptide hormones secreted from the islets, contains several proteins with as yet unknown functions. The mouse islet protein and peptide database accessible at <http://ncrr.pnl.gov>, provides an important reference resource for the research community to facilitate research in the diabetes and metabolism fields.

Keywords

LC-MS/MS; Pancreatic islets; proteomics; diabetes; mouse model; mass spectrometry; interaction network

*Authors of correspondence: Dr. Richard D. Smith, Biological Sciences Division, Pacific Northwest National Laboratory, P.O. Box 999, MSIN: K8-98, Richland, WA 99352, USA. Email: rds@pnl.gov; Fax: (509) 376-7722, Rohit N. Kulkarni, Joslin Diabetes and Department of Medicine, Harvard Medical School, Boston, MA 02215, USA. E-mail: Rohit.Kulkarni@joslin.harvard.edu, Fax: 617-713-3476.

Supporting Information Available

Supplemental tables (1–5) in Excel format are available. This material is available free of charge via the Internet at <http://pubs.acs.org>.

INTRODUCTION

The pancreatic islets of Langerhans plays a critical role in the regulation of glucose homeostasis by secreting insulin and several peptide hormones including glucagon, somatostatin, pancreatic polypeptide, amylin, peptide YY, prodynorphin, urocortin 3, and ghrelin. Besides insulin-producing beta-cells (65–80%)¹, other major cell types within the islets of Langerhans include the glucagon releasing alpha-cells (15–20%)², somatostatin producing delta-cells (3–10%)³, pancreatic polypeptide containing PP cells (1%)⁴ and ghrelin containing epsilon cells (<1%)⁵. While each of the secreted peptide hormones affect glucose homeostasis either directly or indirectly, their significance is best exemplified by the deficiency of insulin that leads to the development of diabetes mellitus. Type 1 diabetes results from autoimmune destruction of beta-cells and an absolute deficiency of insulin, while type 2 diabetes is commonly characterized by a failure of the beta-cells to produce sufficient amounts of bioactive insulin to compensate for insulin resistance in the liver, muscle or adipose tissue⁶.

Due to the importance of islets for metabolism and particularly in the pathogenesis of both types of diabetes, a comprehensive understanding of islet biology is essential for the development of therapeutic strategies to prevent, manage, and/or cure the diseases. Recent advances in proteomic technologies offer approaches to comprehensively characterize the proteome at the global level. Therefore, it is not surprising that comprehensive proteome analyses of a number of mouse tissues related to diabetes have been recently reported^{7–14}. However, the mouse pancreatic islet proteome to date has only been characterized by two-dimensional electrophoresis (2DE) and only 44 proteins have been identified by this approach¹⁵.

In this work we present the first comprehensive profiling of the mouse islet proteome with the aim of establishing an extensive peptide/protein database for the pancreatic islet proteome of rodent models. The extensive coverage was achieved by analyzing a pooled islet sample from two different states (normal and insulin resistant) applying two-dimensional LC-MS/MS profiling. Due to the limited dynamic range of detection of LC-MS/MS profiling, we chose to use a pooled islet sample from both the normal and disease states for this initial profiling so that a better coverage for those proteins with increased expression in the disease can be achieved compared to analyzing the normal sample alone; thus, a more complete coverage of the proteome can be achieved to facilitate future studies using mouse models.

Specifically, strong cation exchange (SCX) fractionation followed by reversed phase LC-MS/MS was applied for this proteome profiling, resulting in the confident identification of ~4,000 protein groups (2,612 proteins identified with two or more unique peptides). The dataset includes qualitative relative protein abundance information based upon MS/MS spectral counts. Additionally, we explored in utilizing the currently reported proteomic datasets on multiple other mouse tissues and cell types for discovering potential novel tissue-specific proteins by comparative analyses. In this case, 133 proteins were identified to be either specifically expressed or predominantly abundant in the pancreatic islets. Finally, the dataset revealed a number of post-translational modifications (PTMs) present in the islet proteome including phosphorylation and oxidative modifications. These data will be made available as a reference resource for the diabetes research community as well as for bioinformatic data mining to facilitate research in the fields of diabetes and metabolism.

MATERIALS AND METHODS

Islets Isolation and Protein Digestion

Islets were isolated by the intraductal enzyme injection technique using collagenase¹⁶. Briefly, the pancreas was inflated with collagenase following anterograde injection via common bile

duct, dissected, and incubated at 37°C for 22 min. Following density-gradient centrifugation using HISTOPAQUE-1077 (Sigma), islets were then washed and hand-picked under a stereomicroscope (Stereozoom GZ7, Leica). All islets were cultured overnight at physiological glucose levels (7 mM glucose, 10% FBS and antibiotics) to allow the islets to recover from the effects of collagenase digestion. Islets were then be transferred to nuclease- and pyrogen-free tubes and washed with phosphate buffer. Following removal of the buffer, pellets were frozen at –80°C prior to proteomic analyses. Islets were isolated from four male control mice and four littermates with liver-specific insulin receptor knockout at age 6-months. The details regarding the creation of animal models have been described elsewhere¹⁷. All mice have been back-crossed to the C57BL/6 background for at least 9 generations.

Islet samples from individual mice were homogenized and digested using a 2,2,2-trifluoroethanol (TFE)-based protocol¹⁸. Briefly, islets were resuspended in 50 µL of 50% TFE (Sigma-Aldrich, St. Louis, MO) in 50 mM NH₄HCO₃ (pH 7.8) with 5 mM tributylphosphine (Sigma-Aldrich) and homogenized in a 5510 Branson ultrasonic water bath (Branson Ultrasonics, Danbury, CT), followed by incubation at 60°C for 2 h to reduce disulfide bonds. For tryptic digestion, samples were diluted 5-fold with 50 mM NH₄HCO₃, supplied with 1 mM CaCl₂ and 2 µg of trypsin per sample, and incubated overnight at 37°C with gentle shaking. After lyophilization, samples were re-dissolved in 70 µL of 25 mM NH₄HCO₃. Peptide concentrations were determined with BCA assay (Pierce, Rockford, IL). On average islets from each mouse yielded 30–60 µg of tryptic peptides. Aliquots of 15 µg peptides were used from each mouse to form a pooled sample for this initial profiling experiment.

Strong Cation Exchange (SCX) Fractionation

The pooled sample (~120 µg of total peptides pooled from 8 mice) were subjected to LC fractionation by strong cation exchange (SCX) chromatography on a 200 mm × 2.1 mm Polysulfoethyl A column (PolyLC, Columbia, MD) preceded by a 10 mm × 2.1 mm guard column, using a flow rate of 0.2 mL/min. LC separations were performed using an Agilent 1100 series HPLC system (Agilent, Palo Alto, CA). Mobile phase solvents consisted of (A) 10 mM ammonium formate, 25% acetonitrile, pH 3.0 and (B) 500 mM ammonium formate, 25% acetonitrile, pH 6.8. Once loaded, isocratic conditions at 100% A were maintained for 10 min. Peptides were separated by using a gradient from 0–50% B over 40 min, followed by a gradient of 50–100% B over 10 min. The gradient was then held at 100% solvent B for another 10 min. Following lyophilization, all thirty fractions were dissolved in 25 mM NH₄HCO₃ and stored at –80 °C.

Capillary LC-MS/MS analysis

Each SCX fraction was analyzed with an automated custom-built capillary HPLC system coupled online to an LTQ ion trap mass spectrometer (ThermoElectron, San Jose, CA) by using an electrospray ionization interface. The reversed phase capillary column was prepared by slurry packing 3-µm Jupiter C18 particles (Phenomenex, Torrance, CA) into a 150 µm i.d. × 65 cm fused silica capillary (Polymicro Technologies, Phoenix, AZ). The mobile phase solvents consisted of (A) 0.2% acetic acid and 0.05% TFA in water and (B) 0.1% TFA in 90% acetonitrile. An exponential gradient was used for the separation, which started with 100% A, and gradually increased to 60% B over 100 min. The instrument was operated in a data-dependent mode with an *m/z* range of 400–2000. Ten most abundant ions from each MS scan were selected for further MS/MS analysis by using a normalized collision energy setting of 35%. Dynamic exclusion was applied to avoid repeat analyses of the same abundant precursor ion.

Database Search and Data Analysis

The SEQUEST software (ThermoElectron) was used to search the MS/MS data against the mouse International Protein Index (IPI) database (version 3.19 <http://www.ebi.ac.uk/IPI>). Human keratins and porcine trypsin were added into the database as expected contaminants. No cleavage specificity was defined for database searching. We also considered methionine oxidation as a dynamic modification. The following criteria were used to filter raw SEQUEST results: 1) $X_{\text{corr}} \geq 1.6$ for charge state +1 full tryptic peptides; 2) $X_{\text{corr}} \geq 2.4$ for charge state +2 full tryptic peptides and $X_{\text{corr}} \geq 4.3$ for +2 partial tryptic peptides; and 3) $X_{\text{corr}} \geq 3.2$ for charge state +3 full tryptic peptides and $X_{\text{corr}} \geq 4.7$ for +3 partial tryptic peptides. The delta correlation value (ΔC_n) > 0.1 was used in all cases.

To estimate the false discovery rate (FDR) of peptide identifications we searched against a reversed database as previously described¹⁹. In an attempt to remove redundant protein entries, the software tool ProteinProphet was applied as a clustering tool to group similar or related protein entries into a “Protein Group”²⁰. All identified peptides that passed the filtering criteria were assigned an identical probability score of 1.0, and then entered into the ProteinProphet program solely for clustering analyses to generate the final non-redundant list of proteins or protein groups. To further increase the confidence in protein identifications we considered only proteins identified with at least two peptides.

To find peptides with post-translational modifications (PTM) we used X!Tandem software²¹ and applied the following strategy. For the first pass we searched only for fully tryptic peptides with no dynamic modifications. Proteins that were identified with a peptide expectation value less than -2 were carried over for the second round of database searching where we considered only fully tryptic peptides and a set of dynamic PTMs including: oxidation of cysteine to sulfinic acid +31.9898 Da (Csulfinic), oxidation of cysteine to sulfonic acid +47.9847 Da (Csulfonic), serine and threonine phosphorylation +79.966331 Da (STphos). Each modification was searched independently. To estimate the confidence of those PTM peptide identifications we searched for similar modifications but with masses shifted by ± 10 Da (pseudo-PTMs). The ratio of the number of peptides identified carrying pseudo-PTM to the number of peptides carrying normal PTM was used as an estimation of the FDR for PTM peptide identifications. In particular the FDR estimate for a given PTM modification with mass M at amino acid X was calculated as a ratio of the average of the number of peptides with modification mass $M-10$ and $M+10$ at the same amino acid type X to the number of peptides carrying modification M with original un-shifted mass. To achieve acceptable FDR rates, we require the peptides with E -values less than -3 for Csulfonic PTM and less than -4 for Csulfinic and STphospho PTMs and the non-modified peptide having the E -value passing the same threshold.

In addition to oxidation of cysteines and phosphorylation of serine and threonine residues we also searched for acetylation, methylation and ubiquitination of lysine, nitration and phosphorylation of tyrosine, carbonylation of arginine and proline and S-(2-succinyl) cysteine. However, we failed to identify a significant number of peptides with acceptable FDRs for these PTMs.

Compiling the Set of Proteomic Databases from Other Mouse Tissues

To compare the murine islet proteome with other mouse tissue proteomes, we considered the following mouse tissues and derived cell cultures characterized by LC-MS/MS and currently available proteomic datasets: mouse brain^{7, 8}, cortical neurons cell culture⁹, heart^{8, 22}, muscle²², kidney⁹, lung^{9, 10}, placenta⁸, liver^{8, 11, 12}, adipocyte cell culture¹³, and islet alpha-cell culture¹⁴. If the information on individual peptides was available the peptides were remapped to the mouse International Protein Index database v3.19. The probabilities of correct peptide

assignment were then set to 1 and the remapped peptide lists were analyzed by Protein Prophet to derive a likely set of proteins and homologous protein groups. For subsequent steps we considered IPI annotations from the protein groups having a probability equal to or more than 0.95. Finally, the IPI indexes were mapped to Entrez Gene Symbols using the mouse IPI v 3.19 database. If the information about individual peptides was not readily available, the identifiers (e.g. UniProt⁸ or IPI v3.07 annotations¹³) were mapped directly to IPI v 3.19 to obtain the corresponding Entrez gene symbols.

As the peptide spectral count information was not readily available for a number of datasets, we used the number of unique peptides of the protein normalized for protein length^{23, 24} as a measure of protein abundance. To compare protein abundance levels between the organs and cell-types, we used their assigned ranks within the datasets. The most abundant protein was ranked as No. 1. Missing values, in cases where a protein was not detected in a given dataset, were assigned with an arbitrarily high rank value, which was considered to have lower abundance than the detected least abundant protein within the given dataset. For example, if we have a total 9000 proteins for this tissue expression comparison, the least abundant protein for a given dataset is ranked as 2500, and we will have 5500 proteins are not being detected. To be consistency, we assign a high rank value of these missing proteins as the least abundant rank plus half of the total number of missing proteins, i.e., $2500 + 5500/2 = 5250$ for the rank. When two or more proteins had exactly the same abundance value, they were resolved by assigning the average rank to those proteins. The highest abundance value within the dataset corresponds to rank number one.

Annotation of the database

The InterPro protein family (<http://www.ebi.ac.uk/interpro/>), GO gene ontology (<http://www.geneontology.org>), KEGG pathways (<http://www.genome.ad.jp/kegg/pathway.html>), PIR protein information resource (<http://pir.georgetown.edu/>) and SMART simple modular architecture research tool (<http://smart.embl-heidelberg.de/>) annotations for the entire mouse genome were obtained using the DAVID web-based tool and parsed with *ad hoc* written Python script prior to importing into relational Microsoft Access database. The significance of over- or under-representation of a certain annotation term was computed using Fisher exact test using hypergeometric distribution with *ad hoc* R script utilizing **phyper()** function. The P-values were adjusted for multiplicity of testing using Benjamini-Hochberg method.

RESULTS

Islet Proteome Coverage

To obtain extensive proteomic characterization of mouse islet tissue, we utilized a bottom-up proteomics approach, which first entails detecting and identifying peptide sequences via tandem mass spectrometry and subsequently linking those peptide sequences to their respective proteins during downstream data analysis. Since our aim is to establish an islet proteome database as a reference resource for future diabetes and metabolism research using normal and diseased mouse models, we chose to analyze a pooled pancreatic islet sample isolated from both control mice and mice from a insulin resistance model that exhibits marked islet hyperplasia¹⁷ to achieve a more complete coverage of the proteome based on the understanding that proteins with increased expression in the insulin resistance model will be more detectable in the pooled sample compared to the normal. In the LC-MS/MS analyses of a total of 30 SCX fractions and 3 replicated analyses of the unfractionated global sample, 519,992 MS/MS spectra were collected. The results led to a total of 43,654 MS/MS spectra being confidently identified as peptides based on the SEQUEST custom filtering criteria. These spectra correspond to 17,350 unique peptide identifications (Supplementary Table 1) with a FDR of

0.9% based on reversed database searching. Following ProteinProphet analysis, this dataset contains 4,024 protein groups overall with 2,612 protein groups having two or more unique peptide identifications (Supplementary Table 2). Although we report the complete list of identifications, we considered only proteins with two or more peptide identifications for downstream comparative analyses.

The LC-MS/MS profiling can also provide qualitative estimates of the relative protein abundance based on the spectral count information (Supplementary Table 2)^{25, 26}. To account for the protein length difference, the observed spectral counts were normalized by the number of amino acid residues per protein for estimating the relative abundances within the islet proteome. The dynamic range of estimated abundances spanned approximately four orders of magnitude.

The extent of islet proteome coverage was also examined by mapping the dataset to different canonical signaling pathways. 72 KEGG and 62 Ingenuity Pathway Analysis pathways were revealed to be covered with at least 10 genes in this dataset. Insulin receptor signaling pathway is shown as an example of the receptor tyrosine kinase signaling since this pathway plays a key regulating role in islet function and compensatory islet growth response to insulin resistance (reviewed in²⁷⁻³¹). Figure 1 shows coverage of the canonical insulin receptor signaling pathway. Out of 45 proteins in this pathway, 20 proteins were identified with at least two peptides and 7 more with one peptide, together accounting for approximately 60% of the known proteins in this pathway.

Post-translational Protein Modifications

We have also explored the LC-MS/MS datasets for the presence of post-translational protein modifications including oxidative modifications and phosphorylation. Oxidative stress has been suggested to be linked with beta-cell dysfunction and insulin resistance³². Thus, identification of a list of oxidative protein modifications may be useful in revealing primary hot spots of oxidation and for future quantitative proteomic studies regarding the roles of oxidative stress in islet biology.

The methionine (Met) oxidation is known to be a frequent modification and quite commonly included as a dynamic modification in routine peptide identification searches for MS/MS data. Indeed, for this islet dataset the number of peptides containing oxidized Met (1093) constitutes approximately 20% of all Met-containing peptides (1093 out of 5784). It has been controversial whether the detected Met oxidation in LC-MS experiments reflects endogenous oxidation events induced by reactive oxygen species or biologically irrelevant artifacts such as oxidation during sample preparation or electrospray ionization. However, we observed that peptides with oxidized forms of methionine clearly elute earlier during LC separation compared to their unmodified counterparts for a majority of the identified peptides (Figure 2). This suggests that the majority of oxidation is not occurring during the electrospray ionization, which would otherwise produce identical elution times. The observation in elution time differences is in good agreement with the notion that oxidized methionine is less hydrophobic than normal methionine³³. On the average, oxidation caused peptides to elute earlier by 4.3% on the normalized elution time scale. Although this observation suggests that the majority of methionine oxidation events occur prior to electrospray ionization, the data do not conclusively prove the biological origin since oxidation resulting from sample processing remains as a potential source of artifact. Regardless of the origin, it may be informative to track the abundances between different biological conditions quantitatively for these modified peptides, which would aid in identification of the major protein targets in oxidative stress.

Unlike oxidized methionine, other PTMs were observed in fewer number and the estimated FDR values were quite high after applying the initial filtering criteria optimized for regular

peptide identifications. To improve the FDR, we applied an additional filtering criterion that requires the presence of unmodified form of the peptide in addition to the modified form in order to be considered as a 'true' identification. However, such filtering criterion is not compatible with the common approach for assessing FDR, which use reversed or scrambled protein sequences. The peptide identifications from reversed database are random matches in nature, and it is unlikely to include modified peptides and unmodified peptides with the same sequences identified from the reversed database at the same time. Although they may have some cases with non-modified counterparts among peptides from reversed sequence search, this will significantly underestimate the FDR since the percentage of peptides existing in both modified and non-modified peptides among reversed database searches are significantly fewer than that from forward searches. To address this issue, we introduced an alternative strategy to assess the FDR for peptide identifications with PTMs. We propose that the estimate of the number of false hits for a given PTM can be made based on a search for non-existing PTMs with similar properties such as the same amino acid specificity and a similar, but distinct mass. In practice we performed searches for modification on the same amino acid residue with the intended modification mass shifted with ± 10 Da. For example, we found 23 peptides with cysteines oxidized into sulfonic acid (Supplemental Table 3) which corresponds to +47.9847 Da. Searching for dynamic cysteine modification with the masses +37.9847 Da and +57.9847 Da gave zero and one peptide, respectively. Thus, our FDR estimate of the identification of the peptides having sulfonic acid PTM is 2%. We also identified 5 peptides bearing sulfinic acid, another oxidative modification of cysteine and no peptides with shifted PTM masses were detected, suggesting a relatively low FDR. Interestingly, all 5 sites identified as cysteine sulfinic acid were also identified as cysteine sulfonic acid, which agrees well with the notion that sulfinic acid is an intermediate product of oxidation of cysteine residues to sulfonic acid. All of the proteins bearing oxidized cysteine residues seem to be quite abundant: aspartate aminotransferase, actin, glyceraldehyde-3-phosphate dehydrogenase, elongation factors 2 and acetyl-CoA acetyltransferase with all ranked within top 300 out of the 4024 proteins.

In addition, we identified 26 peptides having phosphorylated serine or threonine with 6% estimated FDR. Interestingly, 11 of these sites have been previously reported or predicted based on homology, and are listed in the Swiss-Prot database. Because no specific enrichment of phosphopeptides was performed, this list of the phosphopeptides most likely reflects only very top abundant phosphoproteins. Since an important function of islets is hormone secretion, it is not surprising that we identified relatively abundant secretion-regulatory proteins including chromogranin A (rank 54) and secretogranin-2 (rank 29) proteins³⁴ that were detected with four phosphopeptides and two phosphopeptides, respectively.

Over and Under-Represented Annotations of the Islet Protein Database

The relatively extensive coverage of the islet proteome led us to examine what pathways or biologically important entities are enriched in the pancreatic islets. The most common approach for such analyses has been comparing the obtained proteomic dataset to the entire genome as a reference. However, such analysis using the annotations of the entire mouse genome as a reference often captures the biases of the experimental approach. For example, the global bottom-up LC-MS/MS proteomic profiling is biased towards the detection of high-abundant proteins. Thus the comparison of such a dataset against the entire genome usually will indicate overrepresentation of GO terms that involves mostly high-abundant proteins (e.g. mitochondria, ribosome and/or main metabolic pathways). Such overrepresentations usually are not resulted from the biology, but rather from the biases of the experimental approach. To overcome this issue, we generated an "average" proteomic database for the mouse based on the available proteomic data from different tissues or cells created with the same or a similar LC-MS/MS experimental approach. A number of studies describing the proteomes of different mouse organs, tissues and derived cell cultures, including mouse brain^{7, 8}, cortical neuron cell

culture⁹, heart^{8, 22}, muscle²², kidney⁸, lung^{8, 10}, placenta⁸, liver^{8, 11, 12}, adipocyte cell culture¹³, and islet alpha-cell culture¹⁴ have been included in compiling such reference database. To approximate the protein set of an “average” LC-MS/MS mouse tissue analysis, we assembled available datasets (including the current one derived from pancreatic islets), but only those derived from adult mouse samples were considered in the final combined database. If a gene was observed in multiple different studies, we retained redundant entries in the combined dataset. This approach has the advantage of maintaining the approximate distribution of genes belonging to a given biological annotation between the data obtained from a single profiling study and the combined dataset of multiple profiling studies. As expected, some annotations are no longer statistically significantly over- or under-represented when analyzed against the combined proteomic database as compared to the entire genome as a reference. For example, it is typical to achieve significant coverage of the proteins involved in oxidative phosphorylation, so it appeared significantly over-represented when compared against the entire mouse genome. However, because other proteomic datasets also have an extensive coverage of oxidative phosphorylation pathways, when compared with the pooled proteomic dataset, it is not any longer significantly over-represented, as the p-value equals 0.084 even before the correction for multiplicity of testing (Table 1). Overall, when compared to the pooled proteomic dataset instead of the entire genome, only 9, instead of 255 GO “biological processes” terms, appeared to be significantly over-represented, and none (instead of 76) significantly under-represented (i.e., having adjusted p-values < 0.05). Even so, all the 9 over-represented terms (Table 2) relate to protein transport and secretion, thus likely reflecting insulin and other peptides related to hormone secretion as the main constituents of the pancreatic islets. To determine which protein complexes or sub-networks related to protein transport and exocytosis might contribute to over-representation of corresponding GO terms (Table 2), we collected evidence for protein-protein interactions from multiple sources for the protein list covered by those GO terms and analyzed them using Cytoscape³⁵ plug-ins CABIN³⁶ and MCODE³⁷. We found a number of protein complexes involved in vesicular trafficking and exocytosis, such as the SNARE complex that is involved in fusing the vesicular membrane with endosomes, ARF proteins which are G-proteins responsible for regulation of trafficking, adaptor proteins involved in formation of clathrin-coated vesicles, proteins in the exocytotic complex responsible for fusion of protein-carrying vesicles to the plasma membrane to enable exocytosis, components of oligomeric Golgi complex and others (Figure 3). Notably 29 proteins related to vesicular secretion are regulated by XBP-1 transcription factor, indicating XBP-1 as the dominant regulator. Indeed, it has been shown that XBP-1 is a crucial transcription factor involved in the development and function of exocrine glands³⁸.

None of the KEGG, InterPro, GO molecular function and SMART annotations appeared to be statistically significantly over- or under-represented. Nonetheless, we detected two GO “cellular component” terms, one PIR super family and five PIR keywords as over-represented annotations in the pancreatic islet dataset. The two GO “cellular component” terms and one PIR keyword highlight the over-representation of the Golgi apparatus, which is indeed extensively involved in secretion of insulin and other peptide hormones. PIR super family SF001135:trypsin are often typical contaminants from exocrine pancreatic tissue of the islet isolates^{39, 40}. Secreted peptide hormones and their precursors produced by alpha- (*Gcg*, *Ppy*), beta- (*Chga*, *Iapp*, *Ucn3*) or PP cells (*Ppy*) constitute the over-represented PIR keyword annotation “amidation”, referring to C-terminal amidation, which is essential for the biological activity of many peptide hormones. Both hormones from endocrine islet tissue (*Gcg*, *Ins1*, *Ins2*, *Ppy*) and contaminants from adjacent exocrine pancreatic tissue (*Amy2*, *Cell*) annotated PIR keyword as exclusively expressed in the pancreas. Cleavage on pair of basic residues is a common post-translational modification in the maturation process of peptide hormones from precursors (*Chgb*, *Gcg*, *Iapp*, *Pdyn*, *Ppy*, *Ppy*, *Scg2*, *Scg3*, *Sst*). Notably, all enriched annotations, except trypsin-like proteases potentially from adjacent exocrine tissue, precisely

point toward the intracellular transport and secretion of peptide hormones as the major function of pancreatic islets.

Proteins specifically expressed in pancreatic islets

Although analyses of over-representation of annotations expectedly highlighted the biological role of the pancreatic islets in the secretion of peptide hormones, it is unlikely that this information will provide novel information regarding the biological role of the islet and the functions of individual proteins. An alternative approach would be to identify the set of proteins specifically expressed in the islets. To this end, we compared our dataset with all the proteomic datasets obtained from different mouse tissues as described above. As a measure of protein abundance, we used the number of unique peptides or spectral count (if available) belonging to the protein normalized by the protein length. However, due to the qualitative nature of LC-MS/MS profiling, along with different experimental and instrumental setups and different types of search engines used to interpret the MS/MS spectra, direct comparison of the estimated protein abundances among the datasets obtained from different laboratories should be interpreted with caution. We reasoned that the rank of the abundance rather than the estimated abundance itself should be a more robust measure for comparison among such datasets (Figure 4), since highly abundant proteins should always have low rank values while low abundant proteins should always have high rank values, although their absolute abundance estimates may significantly differ between different datasets. The complete dataset used for analysis is available in Supplementary Table 4. With the qualitative abundance rank information available, we focused on identifying relatively islet-specific proteins, i.e., proteins expressed at high levels (with low ranking values) in pancreatic islets but not present or present at very low levels in other tissues. Figure 4 shows results of hierarchical clustering as a heatmap with a color gradient for ranks of protein abundance levels. In this analysis, it is important to note that to ensure confidence in the proteins identified in all datasets we considered only those having at least two peptide hits. To compare the islet proteome with proteomes in other tissues we considered all the datasets except the pancreatic islet alpha-cell dataset because the alpha-cells are part of pancreatic islet tissue. We found a cluster of 133 proteins that are almost exclusively present in the pancreatic islet dataset having two or more peptides per protein (Supplementary Table 5).

As shown in Figure 5, most enzymes involved in the citric acid cycle are present in most of the samples with relatively high abundance. Conversely, proteins involved in regulation of secretion (*Chga*, *Chgb*, *Scg2*, *Scg3*, *Scg5*), and moreover islet-specific peptide hormones (*Gcg*, *Iapp*, *Ins1*, *Ins2*, *Pdyn*, *Ppy*, *Pyy*, *Sst*, *Ucn3*) are among the top abundant proteins in the pancreatic islets, but with very low abundance in other organs and tissues. A notable exception is glucagon, which is expressed high in the pancreatic islet dataset as well as in the alpha cell dataset as expected.

This subset of islet-cell specific proteins covers the entire range of abundances with a slight shift towards low abundance proteins (Figure 6). Of these 133 proteins, 68 were not identified in other measurements even by a single peptide, and thus are highly likely specific to pancreatic islets. These 133 proteins were classified into the following annotation groups: (1) secreted protein hormones, (2) proteases and protease inhibitors, (3) proteins involved in transport, secretion and associated with the Golgi apparatus, (4) ribosome and translation, (5) regulation of transcription, (6) proteasome and ubiquitin, (7) glycolysis and oxidative phosphorylation, (8) lipases, (9) lysosome, (10) helicases, (11) proteins with other functional annotations and (12) proteins without any functional annotations. Table 3 lists several novel unannotated islet-specific proteins that do not contain any putative domains with known biological functions (Table 3). These unknown proteins may play an important role in islet function and are potential candidates for further detailed investigation.

DISCUSSION

The availability of complete genome sequences has greatly accelerated the establishment of genomic and proteomic technologies as powerful tool for studying tissue or cell-specific gene expressions at the system-level and for delineating novel pathways involved in metabolic diseases such as diabetes^{41, 42}. In particular, mass spectrometry-based proteomics has become an important tool for molecular and cellular biology research and for systems biology studies by providing large scale measurements of relative protein abundances including post-translational protein modifications⁴³. The importance of studying biological systems at the protein level is further emphasized by recent studies that clearly indicate that mRNA levels do not necessarily correlate with protein abundances^{44–46}.

Mass spectrometry-based proteomic tissue profiling has been extensively applied recently for establishing the proteome composition and protein expression patterns, in different mouse tissues, organs, and cell lines^{7–14}. Such proteome profiling of mammalian tissues or organs is especially valuable for elucidating the diversity in protein composition and expression patterns among mammalian tissues. The proteome database will also serve as a reference resource for more focused hypothesis-driven biological studies and/or for more detailed systems biology studies. For example, one or more of the proteins can serve as potential biomarkers for tissue-specific pathologies. The present study represents the first extensive proteomic characterization of mouse pancreatic islets of Langerhans with the aim of establishing a reference database for mouse islet proteome for future metabolic research using rodent models. We performed this initial survey experiment using a pooled sample from both normal (control) mice and an insulin resistant model with the aim of gaining increased coverage for those proteins that are potentially expressed at higher levels in either of the two conditions. This strategy is advantageous compared to the analysis of islets from the normal state alone because many detectable proteins with increased expression in the disease state could be below the limit of detection if only the normal sample is analyzed. Also, since the database is a qualitative catalog, nearly all proteins identified from the pooled sample will be present in both the normal and disease states but at different abundance levels in the two states. Therefore, the increased coverage achievable for this database should serve as a more useful resource for future studies using both normal and disease mouse models⁴⁷.

The resulting islet proteome database from this study covers ~4,000 proteins. One utility of the database will be mapping different canonical pathways and functional processes to identify which islet proteins are linked with specific metabolic and signaling pathways (Figure 1) since traditional pathway knowledgebases are often non-tissue specific. This database will also be a suitable complement other proteomes that have been characterized for liver, adipocyte, muscle, and brain^{7–14}. Furthermore, we have provided estimated protein-abundances within the islet proteome based on the normalized spectral counts. While it has been reported that spectral counts can provide an estimate on relative protein abundances within the proteome^{25, 48}, such estimates should only be used as a qualitative measure to query whether the protein is either highly abundant or of relatively low abundance. This is because several other factors can influence the spectral count including protein solubility, protein digestion efficiency and peptide ionization efficiency for a given protein.

The extensiveness of the islet proteomic datasets enabled us to compare the results with other available datasets from a number of mouse organs and tissues. We were able to identify a set of 133 proteins that were specific to islets but not detectable or detected with very low abundances in other tissues (Supplemental Table 5). Indeed the subset of 133 proteins contains well-known islet specific secreted hormones, including: glucagon, islet amyloid polypeptide, insulin, prodynorpin, pancreatic polypeptide, peptide YY and urocortin 3. Besides secreted peptide hormones, the subset includes proteins known to be specific to islets, for example:

G6pc2, Reg1 and Sytl4, which are islet specific glucose-6-phosphatase⁴⁹, regenerating islet derived 1⁵⁰ and synaptotagmin-like 4 also known as granuphilin⁵¹, respectively. Although some proteins are indeed known and expected to be islet-specific, the majority of proteins were not known to be restricted to islets. In particular, several hypothetical or unknown proteins were confidently identified as islet-specific i.e. proteins not containing domains with known or reasonably specific functions. These islet-specific proteins, including the unknown proteins, may be important for islet function and are suitable candidates for future studies. An interesting example is the novel transmembrane protein, TMEM27, that was recently demonstrated to stimulate pancreatic beta-cell proliferation⁵². We should note that those 133 proteins are relatively specific to islets based on our data because we only used 8 other organ and tissue types for this comparative analysis: placenta, muscle, heart, kidney, lung, adipocytes, liver and brain. Potentially, those proteins could be expressed in the tissues not yet profiled by LC-MS/MS proteomics.

In addition to protein identification, there is a significant need in identifying potential post-translational protein modifications in a global proteome profiling study since many modifications are known to regulate cell signaling and can also serve as markers of disease progression. Unfortunately, due to the low-abundant nature of most modifications, it has been a challenge to identify protein modifications in global profiling experiments without enrichment. In this work we explored the use of an alternative informatics strategy for identification of modified peptides from LC-MS/MS analyses of a global non-enriched sample. We based our analyses on the notion that post-translational modifications are usually substoichiometric, therefore, the true modified peptides should be present along with unmodified forms. By applying this criterion and an alternative FDR estimation approach using shifted-mass approach (details described in methods), we identified a total of 54 modified peptides including oxidative modifications on cysteine and phosphorylation on serine and threonine with the FDR <5%. Our approach for controlling the FDR of the peptides with PTMs is important for identifying modified peptides within global profiling data, especially for those PTMs that cannot be specifically enriched such as cysteinyl oxidation.

While we realize that the number of identified modified peptides is very limited due to the nature of the global profiling experiment without specific enrichment, the identification of these modifications provides additional value to protein identities or abundances for this initial characterization of the islet proteome. One example is oxidative modifications since oxidative stress has been linked with diabetes^{32, 53}. To our knowledge, the specific oxidative modification in islets have never been identified, presumably due to the limitations in technology. The observed oxidative modifications in this study could potentially provide a list of novel targets that may play a role in oxidative stress response and may also serve as markers of disease progression. An specific example is the oxidation of Cys-244 in glyceraldehyde-3-phosphate dehydrogenase (G3PDH). There are reports showing that Cys-244 is one of the strongest nucleophilic residues and is susceptible to modifications by 4-hydroxy-2-nonenal⁵⁴, a major lipid peroxidation-derived reactive aldehyde, or by normal endogenous metabolites like acyl-CoA⁵⁵ and fumarate⁵⁶. All three modifications result in strong inhibition of the G3PDH enzyme activity. The fact that the oxidation on Cys-244 was detected in both sulfonic and intermediate sulfinic acid forms further supports the confidence of the identifications. However, it remains to be proven that the oxidation of Cys-244 residue indeed inhibits the enzymatic activity of G3PDH.

In summary, the resulting mouse islet proteome database contains the identified peptide sequences, the protein identifications and spectral count information for each protein as information reflecting their relative abundances, and the identified PTMs. The database represents an important reference resource for further data mining and for islet biological studies focused on diabetes. For example, this database will provide a foundation for future

quantitative proteomic studies applying the accurate mass and time tag approach where both accurately measured masses and elution times are utilized for peptide identifications⁵⁷. The available peptide sequences and islet-specific proteins will also be useful for selecting and devising specific targeted proteomic experiments. The database is included as Supplemental Material and available at the NCCR Center for Integrative Biology website (<http://nccr.pnl.gov>) for access by the research community.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank the NIH NCCR grant (RR018522) to Richard D. Smith and the Pacific Northwest National Laboratory LDRD program (W.-J.Q.) and RO1 DK67536 (R.N.K.) and in part by the Harvard Stem Cell Institute (R.N.K.) for support and the Environmental Molecular Sciences Laboratory (EMSL) for use of the instrumentation applied in this research. EMSL is a U.S. Department of Energy (DOE) national scientific user facility located at the Pacific Northwest National Laboratory in Richland, Washington. PNNL is a multi-program national laboratory operated by Battelle Memorial Institute for the DOE under Contract DE-AC05-76RL01830.

References

1. Lacy PE. Electron microscopy of the beta cell of the pancreas. *Am J Med* 1961;31:851–859. [PubMed: 14461448]
2. Kieffer TJ, Habener JF. The glucagon-like peptides. *Endocr Rev* 1999;20(6):876–913. [PubMed: 10605628]
3. Luft R, Efendic S, Hokfelt T, Johansson O, Arimura A. Immunohistochemical evidence for the localization of somatostatin-like immunoreactivity in a cell population of the pancreatic islets. *Med Biol* 1974;52(6):428–430. [PubMed: 4613970]
4. Ekblad E, Sundler F. Distribution of pancreatic polypeptide and peptide YY. *Peptides* 2002;23(2):251–261. [PubMed: 11825640]
5. Prado CL, Pugh-Bernard AE, Elghazi L, Sosa-Pineda B, Sussel L. Ghrelin cells replace insulin-producing beta cells in two mouse models of pancreas development. *Proc Natl Acad Sci U S A* 2004;101(9):2924–2929. [PubMed: 14970313]
6. Kahn SE. The importance of the beta-cell in the pathogenesis of type 2 diabetes mellitus. *Am J Med* 2000;108:2S–8S. [PubMed: 10764844]
7. Wang H, Qian WJ, Chin MH, Petyuk VA, Barry RC, Liu T, Gritsenko MA, Mottaz HM, Moore RJ, Camp II DG, Khan AH, Smith DJ, Smith RD. Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. *J Proteome Res* 2006;5(2):361–369. [PubMed: 16457602]
8. Kislinger T, Cox B, Kannan A, Chung C, Hu P, Ignatchenko A, Scott MS, Gramolini AO, Morris Q, Hallett MT, Rossant J, Hughes TR, Frey B, Emili A. Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* 2006;125(1):173–186. [PubMed: 16615898]
9. Yu LR, Conrads TP, Uo T, Kinoshita Y, Morrison RS, Lucas DA, Chan KC, Blonder J, Issaq HJ, Veenstra TD. Global analysis of the cortical neuron proteome. *Mol Cell Proteomics* 2004;3(9):896–907. [PubMed: 15231876]
10. Cox B, Kislinger T, Wigle DA, Kannan A, Brown K, Okubo T, Hogan B, Jurisica I, Frey B, Rossant J, Emili A. Integrated proteomic and transcriptomic profiling of mouse lung development and Nmyc target genes. *Mol Syst Biol* 2007;3:109. [PubMed: 17486137]
11. Foster LJ, de Hoog CL, Zhang Y, Zhang Y, Xie X, Mootha VK, Mann M. A mammalian organelle map by protein correlation profiling. *Cell* 2006;125(1):187–199. [PubMed: 16615899]
12. Shi R, Kumar C, Zougman A, Zhang Y, Podtelejnikov A, Cox J, Wisniewski JR, Mann M. Analysis of the mouse liver proteome using advanced mass spectrometry. *J Proteome Res* 2007;6(8):2963–2972. [PubMed: 17608399]

13. Adachi J, Kumar C, Zhang Y, Mann M. In-depth analysis of the adipocyte proteome by mass spectrometry and bioinformatics. *Mol Cell Proteomics* 2007;6(7):1257–1273. [PubMed: 17409382]
14. Maziarz M, Chung C, Drucker DJ, Emili A. Integrating global proteomic and genomic expression profiles generated from islet alpha cells: opportunities and challenges to deriving reliable biological inferences. *Mol Cell Proteomics* 2005;4(4):458–474. [PubMed: 15741311]
15. Sanchez JC, Chiappe D, Converset V, Hoogland C, Binz PA, Paesano S, Appel RD, Wang S, Sennitt M, Nolan A, Cawthorne MA, Hochstrasser DF. The mouse SWISS-2D PAGE database: a tool for proteomics study of diabetes and obesity. *Proteomics* 2001;1(1):136–163. [PubMed: 11680894]
16. Pralong WF, Bartley C, Wollheim CB. Single islet beta-cell stimulation by nutrients: relationship between pyridine nucleotides, cytosolic Ca²⁺ and secretion. *EMBO J* 1990;9:53–60. [PubMed: 2403930]
17. Michael MD, Kulkarni RN, Postic C, Previs SF, Shulman GI, Magnuson MA, Kahn CR. Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol Cell* 2000;6(1):87–97. [PubMed: 10949030]
18. Wang H, Qian WJ, Mottaz HM, Clauss TR, Anderson DJ, Moore RJ, Camp DG, 2nd Khan AH, Sforza DM, Pallavicini M, Smith DJ, Smith RD. Development and evaluation of a micro- and nanoscale proteomic sample preparation method. *J Proteome Res* 2005;4(6):2397–2403. [PubMed: 16335993]
19. Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, Petritis K, Camp DG, Smith RD. Probability-Based Evaluation of Peptide and Protein Identifications from Tandem Mass Spectrometry and SEQUEST Analysis: The Human Proteome. *J. Proteome Res* 2005;4:53–62. [PubMed: 15707357]
20. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* 2003;75(17):4646–4658. [PubMed: 14632076]
21. Fenyo D, Beavis RC. A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. *Anal Chem* 2003;75(4):768–774. [PubMed: 12622365]
22. Knyushko TV, Kowalska MI, Gritsenko MA, Clauss TR, Camp DG, Smith RD, Qian W-J, Bigelow DJ. Preferential nitration of cytoskeletal and mitochondrial proteins in heart and skeletal muscle. submitted.
23. Zhang B, VerBerkmoes NC, Langston MA, Uberbacher E, Hettich RL, Samatova NF. Detecting differential and correlated protein expression in label-free shotgun proteomics. *J Proteome Res* 2006;5(11):2909–2918. [PubMed: 17081042]
24. Rappsilber J, Ryder U, Lamond AI, Mann M. Large-scale proteomic analysis of the human spliceosome. *Genome Res* 2002;12(8):1231–1245. [PubMed: 12176931]
25. Qian WJ, Jacobs JM, Camp DG 2nd, Monroe ME, Moore RJ, Gritsenko MA, Calvano SE, Lowry SF, Xiao W, Moldawer LL, Davis RW, Tompkins RG, Smith RD. Comparative proteome analyses of human plasma following in vivo lipopolysaccharide administration using multidimensional separations coupled with tandem mass spectrometry. *Proteomics* 2005;5(2):572–584. [PubMed: 15627965]
26. Zybailov B, Coleman MK, Florens L, Washburn MP. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal Chem* 2005;77(19):6218–6224. [PubMed: 16194081]
27. Kulkarni RN. Receptors for insulin and insulin-like growth factor-I and insulin receptor substrate-1 mediate pathways that regulate islet function. *Biochem Soc Trans* 2002;30(2):317–322. [PubMed: 12023872]
28. Okada T, Liew CW, Hu J, Hinault C, Michael MD, Krtzfeldt J, Yin C, Holzenberger M, Stoffel M, Kulkarni RN. Insulin receptors in beta-cells are critical for islet compensatory growth response to insulin resistance. *Proc Natl Acad Sci U S A* 2007;104(21):8977–8982. [PubMed: 17416680]
29. Vasavada RC, Gonzalez-Pertusa JA, Fujinaka Y, Fiaschi-Taesch N, Cozar-Castellano I, Garcia-Ocana A. Growth factors and beta cell replication. *Int J Biochem Cell Biol* 2006;38(5–6):931–950. [PubMed: 16168703]
30. Kulkarni RN. New insights into the roles of insulin/IGF-I in the development and maintenance of beta-cell mass. *Rev Endocr Metab Disord* 2005;6(3):199–210. [PubMed: 16151624]

31. Kahn BB. Type 2 diabetes: when insulin secretion fails to compensate for insulin resistance. *Cell* 1998;92:593–596. [PubMed: 9506512]
32. Kaneto H, Katakami N, Kawamori D, Miyatsuka T, Sakamoto K, Matsuoka TA, Matsuhisa M, Yamasaki Y. Involvement of oxidative stress in the pathogenesis of diabetes. *Antioxid Redox Signal* 2007;9(3):355–366. [PubMed: 17184181]
33. Black SD, Mould DR. Development of hydrophobicity parameters to analyze proteins which bear post-or cotranslational modifications. *Anal Biochem* 1991;193(1):72–82. [PubMed: 2042744]
34. Karlsson E. The role of pancreatic chromogranins in islet physiology. *Curr Mol Med* 2001;1(6):727–732. [PubMed: 11899258]
35. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13(11):2498–2504. [PubMed: 14597658]
36. Singhal M, Domico K. CABIN: collective analysis of biological interaction networks. *Comput Biol Chem* 2007;31(3):222–225. [PubMed: 17500038]
37. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 2003;4:2. [PubMed: 12525261]
38. Lee AH, Chu GC, Iwakoshi NN, Glimcher LH. XBP-1 is required for biogenesis of cellular secretory machinery of exocrine glands. *Embo J* 2005;24(24):4368–4380. [PubMed: 16362047]
39. Takeda J, Yano H, Eng S, Zeng Y, Bell GI. A molecular inventory of human pancreatic islets: sequence analysis of 1000 cDNA clones. *Hum Mol Genet* 1993;2(11):1793–1798. [PubMed: 7506601]
40. Wang H, Horikawa Y, Jin L, Narita T, Yamada S, Shihara N, Tatemoto K, Muramatsu M, Mune T, Takeda J. Gene expression profile in rat pancreatic islet and RINm5F cells. *J Mol Endocrinol* 2005;35(1):1–12. [PubMed: 16087717]
41. Almind K, Doria A, Kahn CR. Putting the genes for type II diabetes on the map. *Nat. Med* 2001;7(3):277–279. [PubMed: 11231616]
42. Wang J, Webb G, Cao Y, Steiner DF. Contrasting patterns of expression of transcription factors in pancreatic alpha and beta cells. *Proc. Natl. Acad. Sci. U. S. A* 2003;100(22):12660–12665. [PubMed: 14557546]
43. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198–207. [PubMed: 12634793]
44. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 1997;18(3–4):533–537. [PubMed: 9150937]
45. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation Between Protein and mRNA Abundance in Yeast. *Mol. Cell Biol* 1999;19:1720–1730. [PubMed: 10022859]
46. Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA* 2002;99(9):5860–5865. [PubMed: 11972065]
47. Michael MD, Kulkarni RN, Postic C, Previs SF, Shulman GI, Magnuson MA, Kahn CR. Loss of insulin signaling in hepatocytes leads to severe insulin resistance and progressive hepatic dysfunction. *Mol. Cell* 2000;6:87–97. [PubMed: 10949030]
48. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 2004;76(14):4193–4201. [PubMed: 15253663]
49. Arden SD, Zahn T, Steegers S, Webb S, Bergman B, O'Brien RM, Hutton JC. Molecular cloning of a pancreatic islet-specific glucose-6-phosphatase catalytic subunit-related protein. *Diabetes* 1999;48(3):531–542. [PubMed: 10078553]
50. Okamoto H. The Reg gene family and Reg proteins: with special attention to the regeneration of pancreatic beta-cells. *J Hepatobiliary Pancreat Surg* 1999;6(3):254–262. [PubMed: 10526060]
51. Wang J, Takeuchi T, Yokota H, Izumi T. Novel rabphilin-3-like protein associates with insulin-containing granules in pancreatic beta cells. *J Biol Chem* 1999;274(40):28542–28548. [PubMed: 10497219]
52. Akpinar P, Kuwajima S, Krutzfeldt J, Stoffel M. Tmem27: a cleaved and shed plasma membrane protein that stimulates pancreatic beta cell proliferation. *Cell Metab* 2005;2(6):385–397. [PubMed: 16330324]

53. Robertson RP, Harmon JS. Diabetes, glucose toxicity, and oxidative stress: A case of double jeopardy for the pancreatic islet beta cell. *Free Radic Biol Med* 2006;41(2):177–184. [PubMed: 16814095]
54. Ishii T, Tatsuda E, Kumazawa S, Nakayama T, Uchida K. Molecular basis of enzyme inactivation by an endogenous electrophile 4-hydroxy-2-nonenal: identification of modification sites in glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry* 2003;42(12):3474–3480. [PubMed: 12653551]
55. Yang J, Gibson B, Snider J, Jenkins CM, Han X, Gross RW. Submicromolar concentrations of palmitoyl-CoA specifically thioesterify cysteine 244 in glyceraldehyde-3-phosphate dehydrogenase inhibiting enzyme activity: a novel mechanism potentially underlying fatty acid induced insulin resistance. *Biochemistry* 2005;44(35):11903–11912. [PubMed: 16128592]
56. Blatnik M, Frizzell N, Thorpe SR, Baynes JW. Inactivation of Glyceraldehyde-3-Phosphate Dehydrogenase by Fumarate in Diabetes: Formation of S-(2-succinyl)cysteine, a Novel Chemical Modification of Protein and Possible Biomarker of Mitochondrial Stress. *Diabetes*. 2007
57. Smith RD, Anderson GA, Lipton MS, Pasa-Tolic L, Shen Y, Conrads TP, Veenstra TD, Udseth HR. An Accurate Mass Tag Strategy for Quantitative and High Throughput Proteome Measurements. *Proteomics* 2002;2:513–523. [PubMed: 11987125]

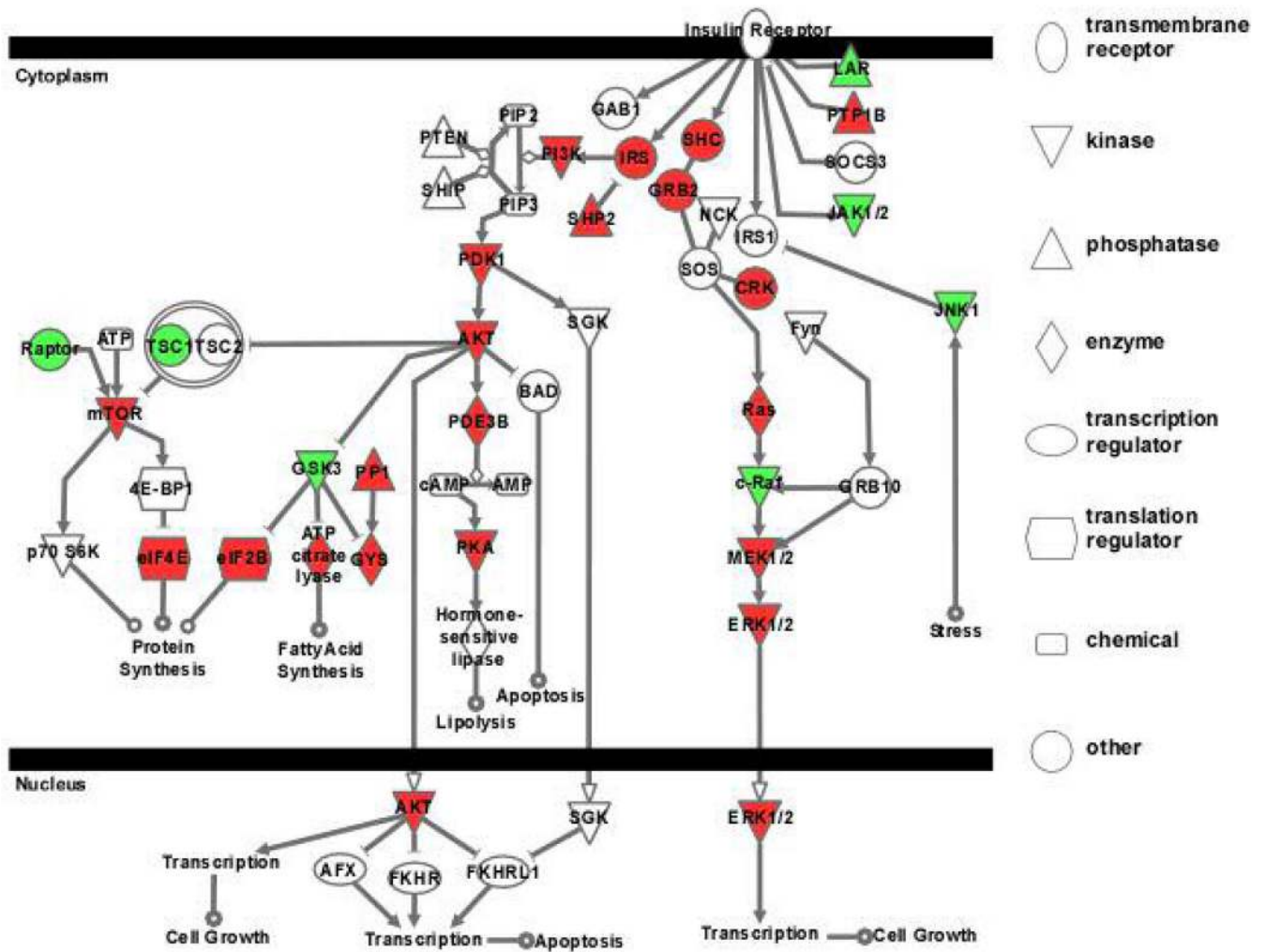


Figure 1. Coverage of the insulin receptor signaling pathway. Proteins detected with one and at least two peptides are colored with green and red, respectively. Undetected proteins and low molecular weight chemical compounds are shown in white.

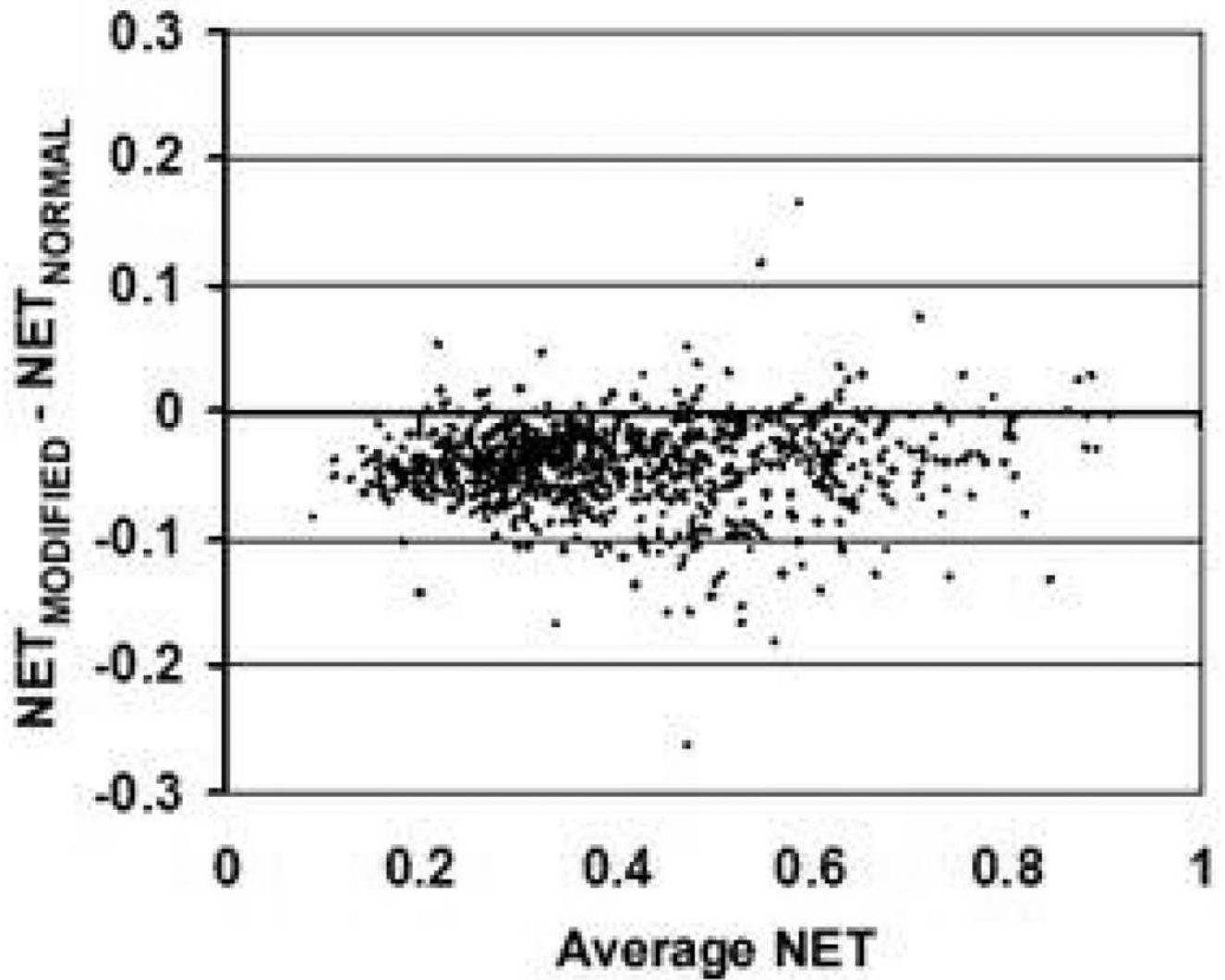


Figure 2. Elution time of peptides containing oxidized methionine compared to peptides with the same sequences with normal non-oxidized methionine. Peptides bearing oxidized methionine tended to elute earlier by 4.3% on the average than their non-modified counterparts.

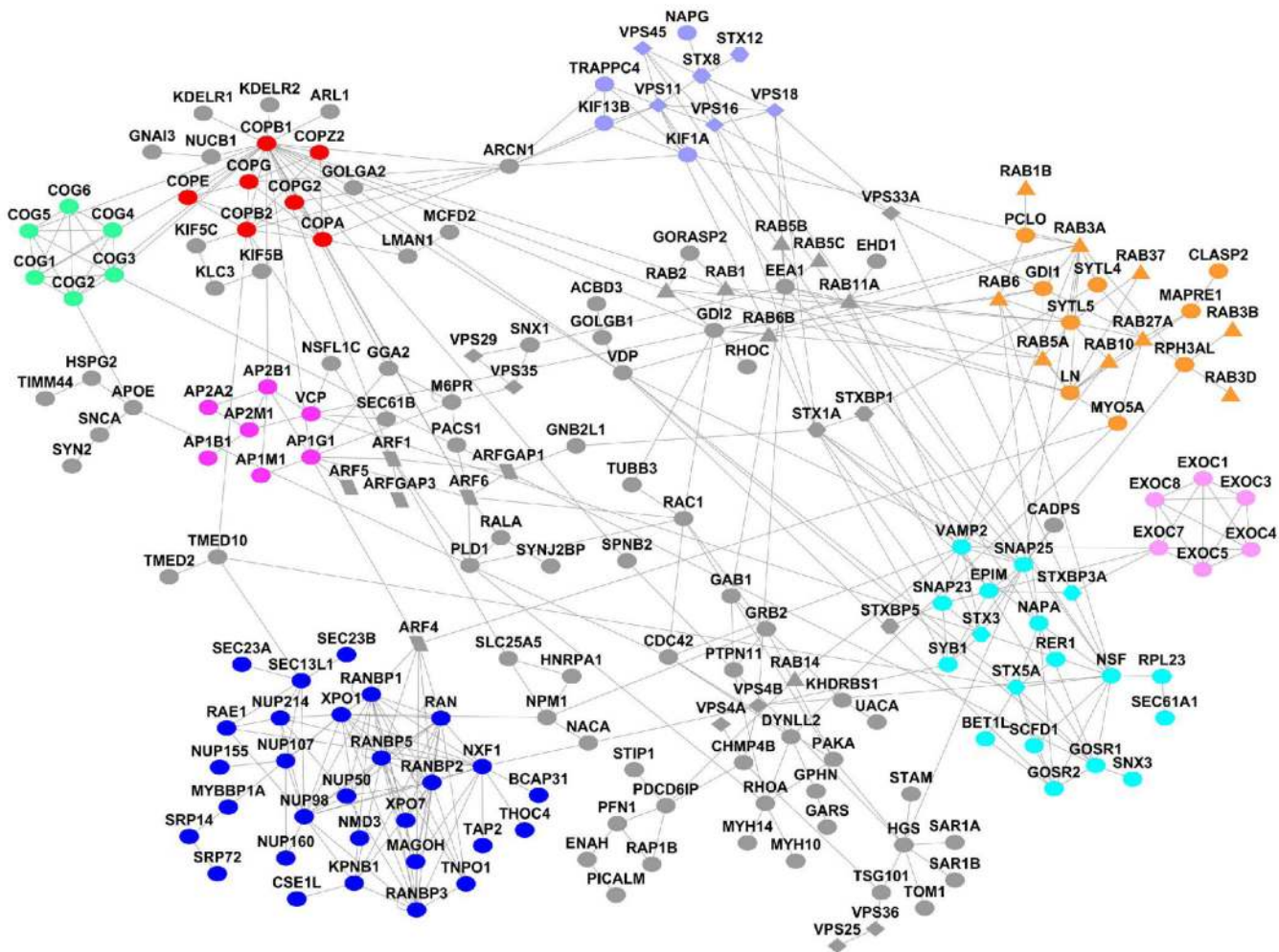


Figure 3. Interaction network of the proteins and protein complexes involved into protein transport and exocytosis in the islet cells. Proteins from overrepresented GO terms (Table 2) were analyzed using Cytoscape plug-in CABIN to build the protein-protein interaction network, followed by analysis with MCODE plug-in to find clusters of highly interconnected sub-networks. The eight clusters are oligomeric Golgi complex (green), coatomer complex (red), cluster containing vacuolar protein sorting proteins and syntaxins 8 and 12 (light blue), cluster of Rab GTPases (orange), exocyst complex (light pink), SNARE complexes (turquoise), cluster mostly consisting of proteins involved in the transport via the nuclear membrane (blue) and cluster of adaptor proteins such as VCP (bright pink). Several of the highly representative functional categories were labeled with distinct shapes other than the default round node shape. GTPases such as Rab and Arf involved in the regulation of trafficking are shown with triangles and parallelograms, respectively. Syntaxins involved in fusion of synaptic vesicles with plasma membrane are shown as hexagon shape. Vacuolar protein sorting proteins are shown as diamonds.

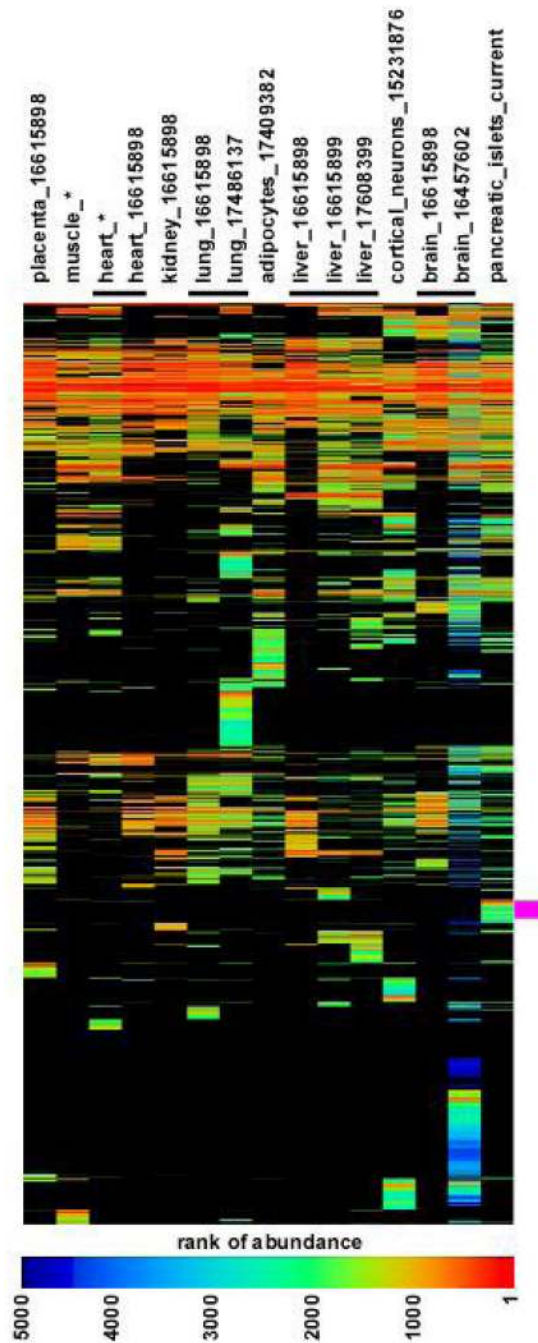


Figure 4. Hierarchical clustering results showing heatmap display with color gradient for gene abundance ranks within the corresponding datasets. Ranks above 5000, that is genes not in top 5000 most abundant, colored with black. The distance function is based on Euclidean distance between the ranks and computed as average linkage. The cluster of the genes predominantly abundant in the pancreatic islets highlighted with magenta. The dataset names refer to the organs or cell types followed by the PubMed unique identifier of the corresponding publication. Asterisk refers to the submitted publication of Knyushko et al. ²².

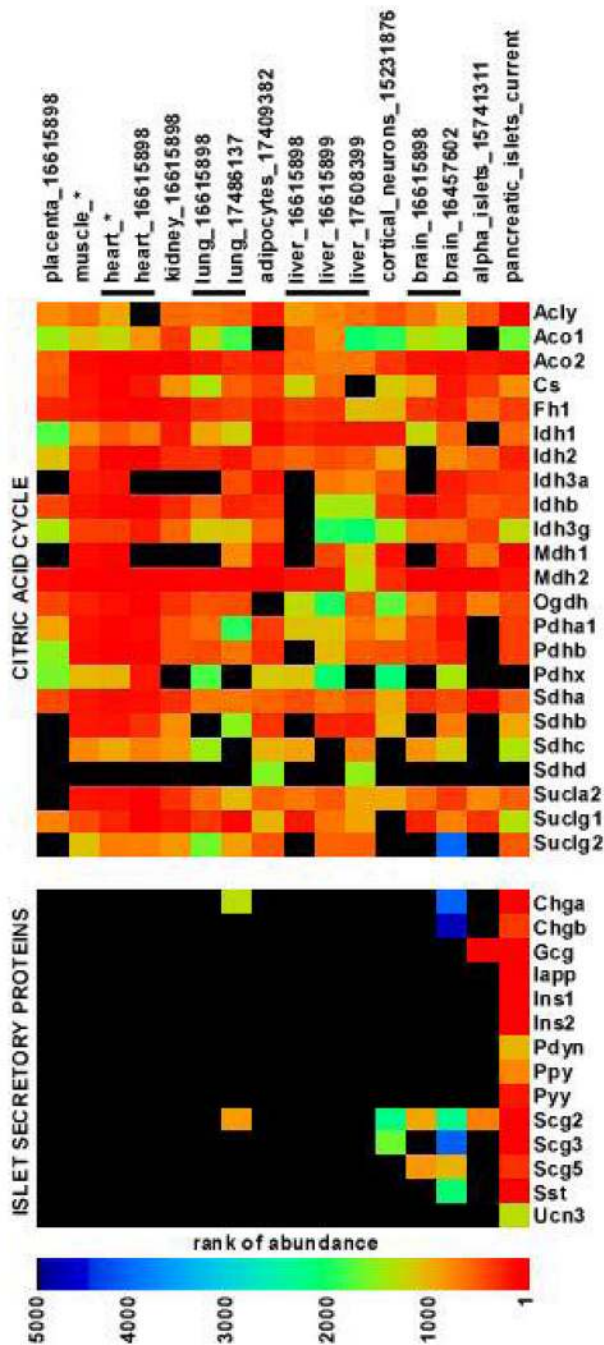


Figure 5. Examples of heatmaps with color gradient for protein abundance ranks for citric acid cycle and islet secreted proteins. The members of citric acid cycle present across most of the datasets and at high abundance ranks. Islet secretory peptide hormones and co-secreted granin proteins are predominantly abundant in the pancreatic islets. The datasets names refer to the organs or cell types followed by the PubMed unique identifier of the corresponding publication. Asterisk refers to a submitted publication (Knyushko et al. ²².)

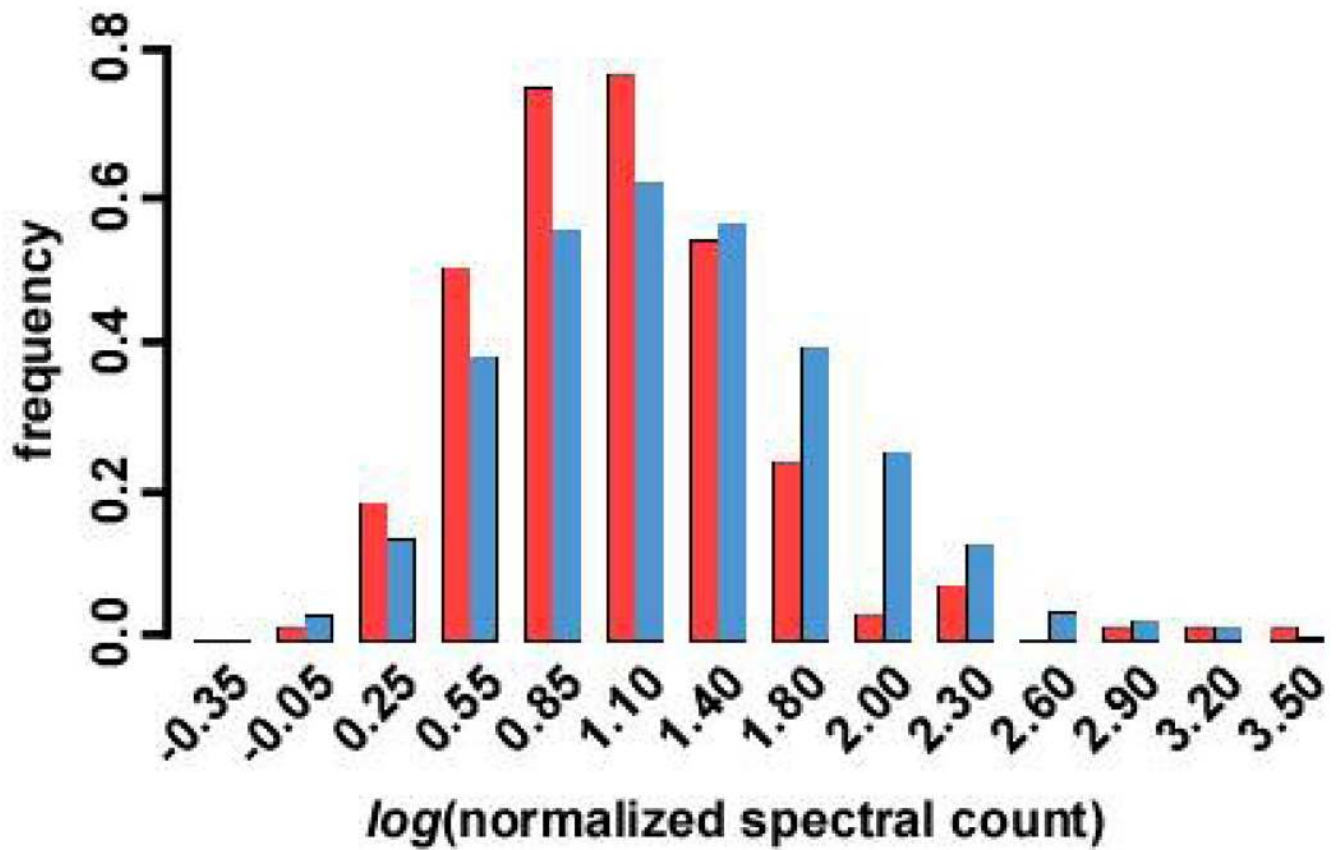


Figure 6. Distribution of log transformed normalized spectral count as a measure of abundance for the entire pancreatic islet dataset (blue) and the subset of proteins unique to the pancreatic islets (red). The abundances of proteins unique to pancreatic islets span across the entire range and on the average slightly shifted to the left part of the histogram.

Table 1

Contingency tables used for testing of over- or under-representation of annotations using the example of oxidative phosphorylation GO:0006119.

A) The reference is the entire mouse genome			
	observed in pancreatic islets	Remained*	annotations in the entire genome
Oxidative phosphorylation	43	24	67
Other	2010	11792	13802
Total	2053	11816	13869

B) The reference is the pooled proteomic datasets of multiple mouse tissues and derivative cell cultures			
	observed in pancreatic islets	Remained	annotations in the combined dataset
oxidative phosphorylation	43	421	464
other	2010	25015	27025
total	2053	25436	27489

* The remained column indicates the total number in the entire genome minus the number observed in pancreatic islets. The observed 4.34-fold enrichment compared to the entire genome is statistically significant. The p-value is 3.5E-20 before and 2.4E-18 after Benjamini-Hochberg correction.

The observed 1.24-fold enrichment compared to the combined dataset is not statistically significant. The p-value is 0.084 before and 1 after Benjamini-Hochberg correction.

Overrepresented annotations of the pancreatic islet protein database compared to the combined database of multiple mouse tissues.

Table 2

Annotation ^a	Term/Family/Keyword	Genes ^b	Share ^c	Fold Enrichment	Adjusted P-value ^d
GO Biological Process	GO:0008104;protein localization, Level 3	237	10%	1.34	1.1E-3
	GO:0015031;protein transport, Level 5	226	10%	1.36	1.4E-3
	GO:0045184;establishment of protein localization, Level 4	228	10%	1.36	2.0E-3
	GO:0046903;secretion, Level 4	93	12%	1.61	2.6E-3
	GO:0051649;establishment of cellular localization, Level 5	235	9.7%	1.30	5.0E-3
	GO:0051641;cellular localization, Level 4	236	9.7%	1.30	5.2E-3
	GO:0046907;intracellular transport, Level 6	235	9.7%	1.30	5.7E-3
	GO:0045045;secretory pathway, Level 5	87	12%	1.59	6.4E-3
	GO:0006886;intracellular protein transport, Level 7	155	10%	1.34	2.7E-2
	GO:0005794;Golgi Apparatus, Level 8	142	11%	1.49	6.6E-4
GO:0005795;Golgi stack, Level 10	105	12%	1.54	2.0E-3	
PIR Super Family	SF001135;trypsin	1810049H19Rik/Try4, 2210010C04Rik, Ctrb1, Ctrl, Ela1, Ela2/Ela2a, Ela3, Kik1/Kik1b5, Prss2	32%	4.17	2.2E-2
PIR Keyword	Golgi stack	110	12%	1.55	2.1E-3
	protein transport	149	10%	1.37	1.2E-2
	Amidation	Chga, Cgc, Iapp, Ppy, Pyy, Ucn3	50%	6.58	1.6E-2
	Pancreas	Amy2, Cel, Cgc, Ins1, Ins2, Ppy	46%	6.07	2.1E-2
	Cleavage on pair of basic residues	Chgb, Cgc, Iapp, Insrr, Pdyn, Ppy, Pyy, Scg2, Scg3, Sst	27%	3.56	2.9E-2

^{a)}The annotation were obtained using <http://david.abcc.ncifcrf.gov/> functional annotation tool. GO gene ontology and PIR protein information resource refer to resources available at <http://www.geneontology.org/>, <http://pir.georgetown.edu/>.

^{b)}Number of the genes observed in islets and covered by the given annotation. If the number of annotated genes is relatively small they are listed explicitly. In the cases when proteins have multiple Entrez gene annotations, which can not be resolved at peptide level, the gene symbols are separated with slash (/) symbol.

^{c)}Share for a given annotation is the percentage of the genes observed in the pancreatic islets sample in the entire pooled database of multiple mouse tissue samples.

^{d)}P-values were computed for Fisher exact test using hypergeometric distribution. Multiplicity of testing was accounted using Benjamini-Hochberg adjustment method.

Table 3
 que to pancreatic islets with unassigned or hypothetical functions.

description	InterPro Domain	Trans-membrane Helices	Signaling Peptide	Protein entry_ID	IPs	Unique Peptides	Rank of Normalized Spectra Count
DNA 1810010M01 gene	PR001229 Jacalin-related lectin	None	+	2411_0	IP100169543	3	424
DNA 1810073N04 gene	IPR013057 Amino acid transporter, transmembrane	8	-	687_0	IP100228647, IP100649374	19	791
sequence A1662250	None	None	-	1281_0	IP100223392	3	834
otological and al expressed 1-like	None	None	-	1680_0	IP100107958	3	897
gene, EG328644	None	None	-	2143_0	IP100224056	4	1268
quence BC023892	IPR012937 Protein of unknown function DUFI693	None	-	1320_0	IP100115303, IP100749568, IP100761403	3	1843
domain containing 134	None	None	+	2920_0	IP100226901	2	1859
asdermin 1	IPR007677 DFN A5 protein	None	-	56_0	IP100113529	3	2108
brane protein 87A	IPR009637 Lung seven transmembrane receptor	7	+	499_0	IP100279051, IP100625262, IP100626341	3	2305
del 447, (NCBI)	IPR010978 tRNA binding arm	None	-	3690_0	IP100752659	3	2288
olog 2 (Drosophila)	IPR008636 HOOK	None	None	3643_0	IP100381165, IP100471449, IP100652864	2	3058