Title

Profiling mouse brown and white adipocytes to identify metabolically relevant small ORFs and functional microproteins

Journal

Cell Metabolism, 35(1)

ISSN

1550-4131

Authors

Martinez, Thomas F
Lyons-Abbott, Sally
Bookout, Angie L
et al.

Publication Date

2023

DOI

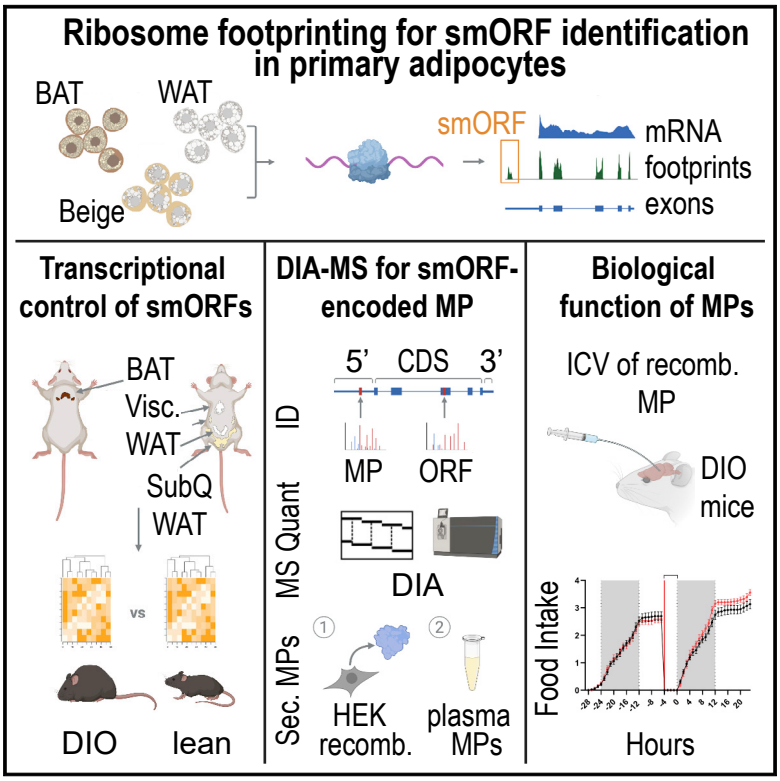
10.1016/j.cmet.2022.12.004

Cell Metabolism

Resource

Profiling mouse brown and white adipocytes to identify metabolically relevant small ORFs and functional microproteins

Graphical abstract



Authors

Thomas F. Martinez, Sally Lyons-Abbott, Angie L. Bookout, ..., Michael J. MacCoss, Alan Saghatelian, Christopher A. Barnes

Correspondence

asaghatelian@salk.edu (A.S.), chrisb@rnes.pro (C.A.B.)

In brief

Martinez et al. identify thousands of unannotated smORF-encoded microproteins in primary mouse adipocytes. Hundreds of microproteins are differentially expressed in tissues from obese versus lean mice. Dozens of secreted microproteins are also detected by mass spectrometry. A predicted FAM237B homolog is shown to be an orexigenic hormone in obese mice.

Highlights

- Thousands of unannotated smORFs encode microproteins in primary mouse adipocytes
- Hundreds of microproteins are differentially expressed in obese versus lean mice
- Dozens of microproteins are found to be secreted by mass spectrometry
- Identified protein homolog of mouse FAM237B that is an orexigenic hormone in obese mice

Martinez et al., 2023, Cell Metabolism *35*, 166–183
January 3, 2023
https://doi.org/10.1016/j.cmet.2022.12.004

CellPress

CellPress
OPEN ACCESS

Cell Metabolism

Resource

Profiling mouse brown and white adipocytes to identify metabolically relevant small ORFs and functional microproteins

Thomas F. Martinez,[1] Sally Lyons-Abbott,[2] Angie L. Bookout,[2] Eduardo V. De Souza,[3,4,5] Cynthia Donaldson,[5] Joan M. Vaughan,[5] Calvin Lau,[5] Ariel Abramov,[2] Arian F. Baquero,[2] Karalee Baquero,[2] Dave Friedrich,[2] Justin Huard,[2] Ray Davis,[2] Bong Kim,[2] Ty Koch,[2] Aaron J. Mercer,[2] Ayesha Misquith,[2] Sara A. Murray,[2] Sakara Perry,[2] Lindsay K. Pino,[7] Christina Sanford,[2] Alex Simon,[2] Yu Zhang,[2] Garrett Zipp,[2] Cristiano V. Bizarro,[3,4] Maxim N. Shokhirev,[6] Andrew J. Whittle,[2] Brian C. Searle,[8] Michael J. MacCoss,[7] Alan Saghatelian,[5,*] and Christopher A. Barnes[2,9,10,*]

[1]Department of Pharmaceutical Sciences, Department of Biological Chemistry, Chao Family Comprehensive Cancer Center, University of California, Irvine, Irvine, CA, USA

[2]Novo Nordisk Research Center Seattle, Inc., Seattle, WA, USA

[3]Centro de Pesquisas em Biologia Molecular e Funcional (CPBMF) and Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Brazil

[4]Programa de Pós-Graduação em Biologia Celular e Molecular, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul 90616-900, Brazil

[5]Clayton Foundation Laboratories for Peptide Biology, Salk Institute for Biological Studies, La Jolla, CA, USA

[6]Razavi Newman Integrative Genomics and Bioinformatics Core, Salk Institute for Biological Studies, La Jolla, CA, USA

[7]Department of Genome Sciences, University of Washington, Seattle, WA, USA

[8]Department of Biomedical Informatics, The Ohio State University, Columbus, OH, USA

[9]Velia Therapeutics, Inc., San Diego, CA, USA

[10]Lead contact

*Correspondence: asaghatelian@salk.edu (A.S.), chrisb@rnes.pro (C.A.B.)

https://doi.org/10.1016/j.cmet.2022.12.004

SUMMARY

Microproteins (MPs) are a potentially rich source of uncharacterized metabolic regulators. Here, we use ribosome profiling (Ribo-seq) to curate 3,877 unannotated MP-encoding small ORFs (smORFs) in primary brown, white, and beige mouse adipocytes. Of these, we validated 85 MPs by proteomics, including 33 circulating MPs in mouse plasma. Analyses of MP-encoding mRNAs under different physiological conditions (high-fat diet) revealed that numerous MPs are regulated in adipose tissue *in vivo* and are co-expressed with established metabolic genes. Furthermore, Ribo-seq provided evidence for the translation of *Gm8773*, which encodes a secreted MP that is homologous to human and chicken FAM237B. *Gm8773* is highly expressed in the arcuate nucleus of the hypothalamus, and intracerebroventricular administration of recombinant mFAM237B showed orexigenic activity in obese mice. Together, these data highlight the value of this adipocyte MP database in identifying MPs with roles in fundamental metabolic and physiological processes such as feeding.

INTRODUCTION

Recent advances in ribosome profiling (Ribo-seq) and proteogenomics have identified thousands of unannotated peptides and small proteins, microproteins (MPs), encoded by small open reading frames (smORFs) in mammalian genomes.[1–6] Among these, several MPs have been characterized with biologically significant activities including cell stress and survival[1,7] and muscle development and function.[8–11] Furthermore, some MPs with therapeutically beneficial activities are being looked at preclinically. For example, the DWORF (dwarf open reading frame) MP binds and enhances SR-Ca^{2+}-ATPase (SERCA) activity and was shown to mitigate cardiomyopathy in mice using a gene therapy approach.[12] Though only a small fraction of identified MPs have been characterized, these early examples suggest that MPs can play important roles in as diverse pathways as larger annotated proteins.

Obesity and related metabolic diseases impact a large proportion of the population.[13] Therefore, it is vital to have a better understanding of the processes that regulate obesity and metabolic health to provide additional, improved therapies in the future. Peptides and small proteins, such as insulin and leptin, have an outsized role in metabolism, suggesting that MPs, which are of similar size and biophysical properties, might be a rich source of metabolic regulators. In support of this hypothesis, several MPs have recently been discovered with metabolic roles, including mitoregulin[12,14–17] and MOTS-c.[18] Adipose tissue is an endocrine organ capable of secreting several peptide and

CellPress
OPEN ACCESS

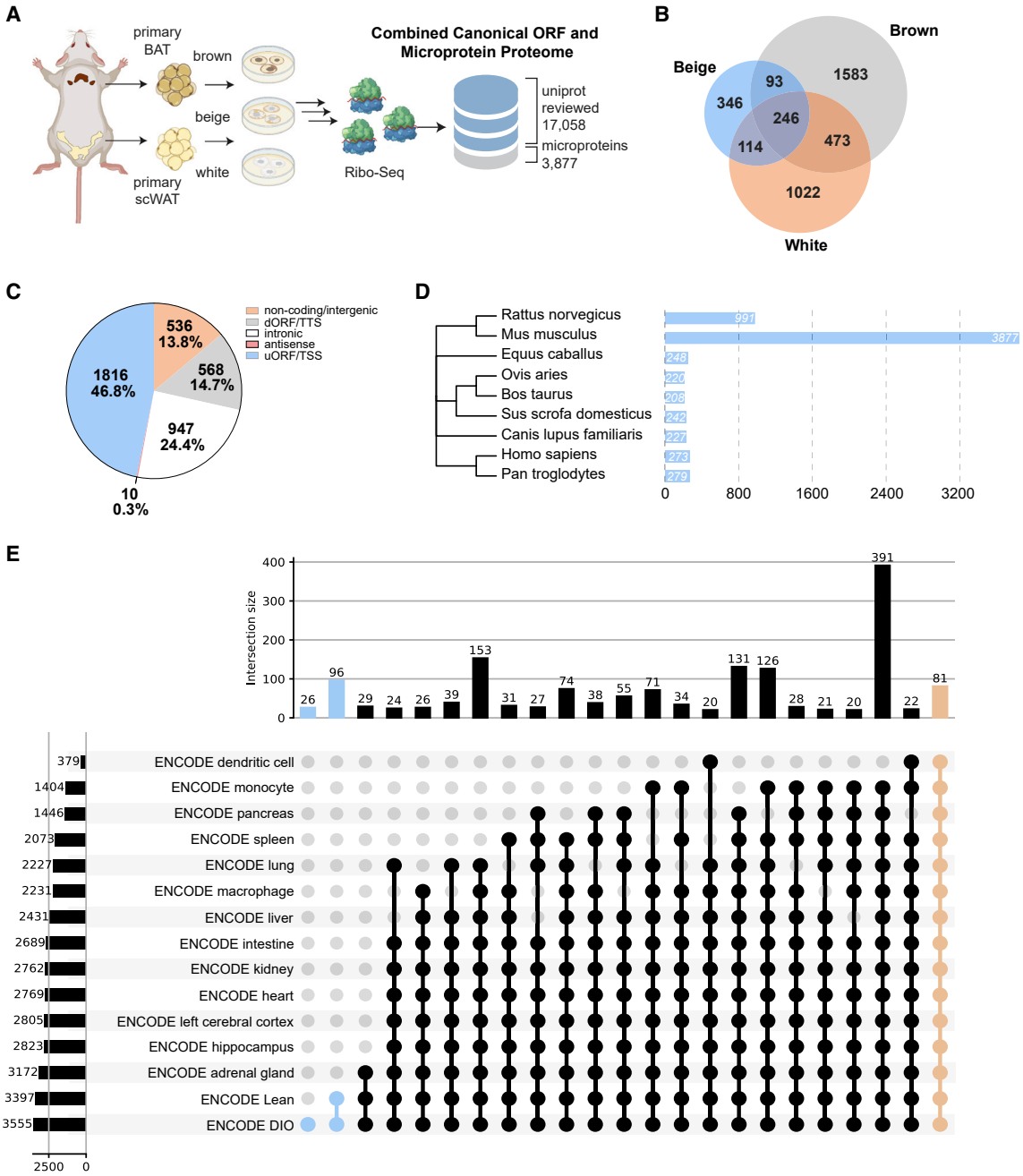


Figure 1. Ribosome profiling to define primary differentiated white, brown, and beige mouse adipocyte microproteins

(A) Primary subcutaneous white, beige (from subcutaneous WAT), and brown adipocytes were generated from freshly isolated subcutaneous white adipose tissue (WAT) and brown adipose tissue (BAT). Primary differentiated adipocytes were then analyzed by ribosome profiling (Ribo-seq), leading to the identification of thousands of MP-encoding small open reading frames (smORFs) not found in curated databases (UniProt, Refseq, Ensembl).

(B) A Venn diagram of the identified smORF distributions between primary white, beige, and brown adipocytes.

(C) Homer analysis of the MP-encoding smORF positions was used to estimate numbers of smORFs in non-coding/intergenic regions, those downstream of a CDS (downstream ORF [dORF]/translational termination site [TTS]), intronic regions, antisense RNAs, and upstream of a CDS (upstream ORF [uORF]/translational start site [TSS]).

(D) Unrooted phylogenetic tree inferred from the mitochondrial DNA sequences of closely related species. The bars on the side represent the number of homologous sequences for the novel smORFs that could be found in the transcriptome of the species in the tree with a tBLASTn search.

(E) UpSet plot showing the overall tissue distribution of the smORF-containing transcripts (SCTs) in mouse ENCODE samples. Each row in the lower part of the plot represents a different tissue or cell type. The linked dots in these rows represent an intersection of the SCTs for a given sample. For instance, the linked dots in the third column represent SCTs that are expressed only in the lean, DIO, and adrenal gland samples. Blue dots denote a subset of SCTs that are unique to the

(legend continued on next page)

protein factors that regulate feeding, energy balance, and thermogenesis.[19–21] In this work, we performed Ribo-seq on primary mouse differentiated brown, white, and beige adipocytes, which led to the annotation of 3,877 new MP-coding smORFs, indicating a large reservoir of uncharacterized MPs.

We demonstrate several different strategies for prioritizing MP-coding smORFs in our dataset that are most likely involved in metabolism. First, we identify several important metabolic genes that are likely post-transcriptionally regulated by translated smORFs in the genes' 5′ untranslated regions (UTRs). We also highlight many smORFs encoded on RNAs that are regulated by high-fat diet (HFD) and co-expressed with established metabolic genes, suggesting the encoded MPs might also function in metabolism. Because many metabolically relevant peptides are circulating factors, we also used mass spectrometry (MS)-based proteomics to identify and quantify dozens of MPs in primary adipose cell lysates, conditioned media, and mouse plasma. Conservation analysis identified over 200 MPs that are likely to be functional given their sequence conservation across multiple mammalian species. Conservation in particular has been a strong predictor for functional MPs, and therefore we followed up on one such hit found in our dataset. We show that an smORF derived from our Ribo-seq results on the predicted gene *Gm8773* encodes a conserved MP that is homologous to human and chicken FAM237B. In addition, we show that this MP is likely secreted and that central administration of recombinant Gm8773 microprotein leads to changes in feeding behavior in mice. While these initial analyses highlight many likely metabolically relevant smORFs, this database's power will come from investigators performing similar analyses with transcriptomic datasets under different physiological conditions or from genetically engineered/drug-treated mice to identify MPs of further interest. In aggregate, these examples provide a roadmap on how to best use these data to begin to connect and eventually characterize MPs to fundamentally important metabolic biology.

RESULTS

Ribo-seq of primary mouse brown, white, and beige adipocytes identifies thousands of novel smORFs

Adipose tissue is a complex mixture of cell types that includes adipocytes, fibroblasts, vascular cells, and immune cells.[22] Using primary mouse adipocytes avoids the complexity from profiling different cell types to generate a more reliable adipocyte dataset. We generated primary brown (BAT), white (WAT), and beige adipocytes using established protocols[23–27] (Figure 1A). Ribo-seq of these cells provided a dataset that was then processed using the translated ORF classifier RibORF[28,29] to annotate MP-coding smORFs in these cells[4] (Figure 1A). We previously showed that Ribo-seq-based identification of translated smORFs is inherently noisy, but confidence in smORF annotations could be improved by collecting replicates (Table S1).[4] Therefore, we collected two biological replicates for each cell type and generated Ribo-seq data with clear 3-nt periodicity (Figure S1).

Ribo-seq identified a total of 3,877 previously unannotated smORFs in total from all three cell types (Figure 1B; Table S1). Of these, 246 smORFs were called translated in all three cell types, 926 were identified in at least two cell types, and 1,050 were detected in at least two replicates from any cell type (Figure 1B; Table S1). These numbers are consistent with the number of unannotated smORFs identified in several human cell lines.[1,4] The length distribution of mouse adipose MPs also follows a similar curve as human MPs where the 33-amino-acid median length is close to the 32-amino-acid median length identified for novel human MPs[4] (Figure S2A). Next, an analysis of the amino acid frequencies in predicted mouse adipose MPs showed higher usages of alanine, glycine, proline, arginine, and tryptophan, as well as lower levels of aspartate, glutamate, asparagine, glutamine, and tyrosine, which is also identical to what we had observed in human MPs[4] (Figure S2B). While smORFs and encoded MPs are not as well conserved as canonical genes on the whole (*vide infra*), several features such as median length and amino acid composition seem to be shared by smORFs across mammals.

smORF classes

One of the biggest insights to come from the discovery of thousands of smORFs is that many mammalian mRNAs are polycistronic (encode for more than one protein),[1–4,30] which contradicts a dogma in molecular biology.[31] In this dataset, there are 1,816 smORFs in the 5′ UTR of mRNAs, or upstream ORFs (uORFs), which is the largest category of smORFs in these data (Figure 1C) and is consistent with other datasets.[4] Because mRNAs were long thought to produce only a single protein, uORFs were rarely studied for their role in producing functional proteins. Instead, the focus of most uORF research has been on their role as genetic elements that can repress the translation of downstream main ORFs (mORFs) dynamically.[30]

For instance, the uORFs in the *Atf4* and *Chop* mRNAs repress the translation of *Atf4* and *Chop* under homeostatic conditions, but under stress conditions, the uORFs are bypassed, leading to the expression of *Atf4* and *Chop* and triggering the cellular stress response.[32] Other uORFs have similar roles, and as a consequence the discovery of a uORF on an mRNA indicates the possibility that the mORF is regulated post-transcriptionally. Thus, one use of this dataset is the identification of several key metabolic genes that contain uORFs, including *Insr*, *Igf1*, *Cd36*, *Adra1a*, *Adrb1*, *Ppara*, *Plin1*, *Irs2*, *Ucp1*, and *Ucp2*. To further determine whether translated uORFs identified in our data could potentially regulate metabolic pathways, we employed gene ontology (GO) analysis on the annotated genes containing uORFs.[33] The top 10 redundancy-filtered pathways regulated by genes containing uORFs included lipid metabolic processes, positive regulation of catabolic processes, mitochondrion organization, and glycoprotein metabolic process (Figure S2C; Table S3) Indeed, the *Cd36* uORFs were previously

RNA-seq datasets generated in this study. Black dots represent subsets of SCTs that are also expressed in the analyzed datasets from the mouse ENCODE project. Orange dots represent a subset of SCTs that are present in every group. The bar plots on the upper part of the plot represent the number of expressed SCTs unique to that subset. Bar plots on the left side of the lower part of the plot denote the number of SCTs with at least 1 TPM that are expressed in a given sample.

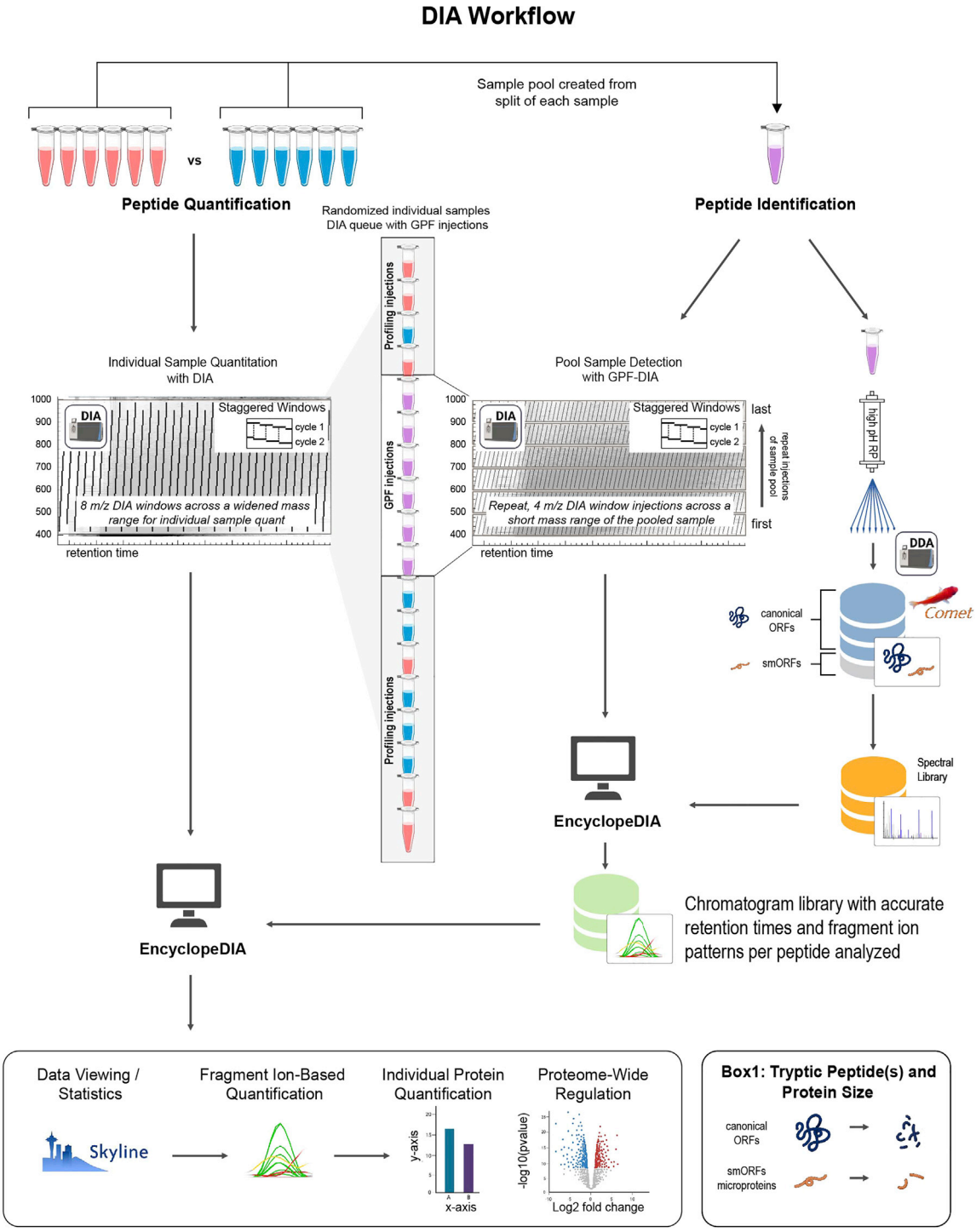


Figure 2. Schematic of data-independent acquisition-mass spectrometry (DIA-MS) method that employs the chromatogram library approach

An idealized experiment with two groups of six (red and blue) is depicted in the upper left. A small volume of each sample is split and pooled for peptide identification analysis (purple). First, the pool is subjected to high-pH reversed-phase (RP) fractionation where peptide identifications are generated from DDA injections of each RP fraction. The Ribo-seq-generated smORF-encoded MP sequences are amended to the canonical reviewed UniProt proteome database (blue/gray library icon) where each RP fraction DDA injection is searched (Comet + PeptideProphet) with the results used to make a spectral library of peptide identifications (orange library icon). The pool is also subjected to a small (4 *m/z*) DIA gas-phase fractionation (GPF) window. The GPF chromatogram library method utilizes 6 replicate injections of the sample pool with small, staggered windows (4 *m/z*) across a short mass range (100 *m/z*). Each short mass range injection covers one-sixth of the total mass range of the profiling method used on each sample. The profiling method for quantifying each sample uses larger

(legend continued on next page)

characterized and shown to be necessary for glucose regulation of *Cd36* translation.[34]

There are also 568 smORFs located in the 3′ UTR and downstream of annotated coding mRNAs, which are referred to as downstream ORFs (dORFs) (Figure 1C). In contrast to uORFs, which repress translation, dORFs have been shown to enhance the translation of the mORF in some cases.[35] The remaining classes of smORFs include 947 smORFs that overlap annotated intronic regions and 536 smORFs that are in non-coding RNAs (ncRNAs) or within intergenic regions (Figure 1C). The data show protein-coding smORFs in mice are found throughout the transcriptome in regions previously considered untranslated (example of uORF and dORF in Figure S2D) and are consistent with distributions observed in other eukaryotic systems.[1–4,36]

Microprotein classes

Though only a small percentage of all MPs have been functionally characterized, several of these contain single-pass transmembrane domains,[7–11] and in some of these cases the transmembrane domain is also the functional domain of the microprotein. For example, several single-pass transmembrane MPs in mammalian and fly muscle utilize the transmembrane domain to activate or inhibit the sarcoplasmic reticulum calcium channel to regulate muscle contraction. These examples suggest that MPs with a transmembrane domain warrant additional scrutiny when looking for functional MPs. The predictive software tool transmembrane hidden Markov model (TMHMM)[37] identified 84 MPs likely to contain at least one transmembrane domain (Table S2; Figure S2).

Secreted MPs are also of interest since many physiologically active peptides and small proteins are secreted. Secreted peptides and proteins can be predicted by identifying signal peptides that target MPs into the secretory pathway. Indeed, this approach was used to identify a handful of MPs that were secreted upon overexpression.[38] We used SignalP and Phobius to predict secreted MPs in our dataset[39,40] and identified a total of 183 potentially secreted MPs (Table S2; Figure S2E). Of these, 27 MPs had positive scores in both the TMHMM and SignalP/Phobius analyses, and 3,583 MPs had no signal peptide or transmembrane domain.

Microprotein conservation

There are several approaches to look for evolutionary conservation. The most rigorous is PhyloCSF, which accounts for phylogeny and codon substitution frequency to assess evolutionary protein-coding potential and has been used to identify protein-coding ORFs from sequence alone.[41] Using PhyloCSF, we find that 204 of the smORFs have a positive average score and thus are likely functional coding regions (Table S1). In addition to PhyloCSF, we also employed tBLASTn[42] to find MP sequence similarity between mouse and several other species (Figure 1D; Table S1). Not surprisingly, mouse and rat shared the largest number of high-similarity smORFs, with 991 passing the scoring threshold (Figure 1D). Overlap with other species ranged between 208 and 279 smORFs, with 273 smORFs between mouse and human showing high similarity (Figure 1D). An explanation for the high number of microproteins with no homologous sequences in closely related organisms is the prevalence of small coding sequences arising from *de novo* gene birth events, which are common among smORFs[43] and represent a subset of novel sequences that are unique to mouse in this case. An exemplar of a highly conserved MP comes from an smORF in the 5′ UTR of the creatinine transporter (Slc6a8), which is identical in several species and has a positive average PhyloCSF score, suggesting that this peptide sequence is functional (Figure S2F). Conservation has proven to be a frequently common feature of functional MPs, and therefore this set of smORFs is likely to be of greatest interest to biologists for further validation studies.

Tissue distribution of smORF mRNAs

Canonical genes that show a restricted expression pattern often have a role in biology related to that tissue. For instance, adipose-enriched genes identified in the tissue map proteome include *Adipoq*, *Lep*, *Pparg*, and *Pnpla2 (Atgl)*.[44] These genes, along with 212 others, are part of the current list of adipose-enhanced genes on the human protein atlas database (https://www.proteinatlas.org/[44]). To identify smORF-containing transcripts that are adipose specific, we analyzed mouse tissue RNA sequencing (RNA-seq) data from the ENCODE mouse project.[45] We then selected the smORF-containing transcripts with a TPM value of at least 1 and included these as a subset in an UpSet plot, only retaining groups with at least 20 genes (Figure 1E). The analysis identified 122 smORF-containing transcripts that are specific to adipose tissue (Figure 1E). As with the canonical genes enriched in adipose tissue, consequently MPs encoded on these adipose-enriched smORFs should be studied further for potential roles in adipose tissue biology.

Proteomics validation of smORF translation in primary differentiated BAT and WAT cells

Proteomics serves to validate that some smORFs are translated and produce stable, detectable MPs, but the absence of an MP from proteomics should not be interpreted as that MP not existing. To try to improve confidence and sensitivity of our proteomics analysis, we utilized a combined data-dependent acquisition (DDA) and data-independent acquisition (DIA) MS workflow. We chose to use DIA-MS because this method excels in the reliable quantification of individual peptides, while also

staggered windows (8 *m/z*) from a mass range of 400–1,000 that covers the entire mass range of the GPF injections. The GPF injections are inserted into the randomized queue of individual samples to be quantified, allowing very accurate retention time realignment and improved accuracy in the peptide extractions from the DIA-MS profiling data. EncyclopeDIA is first used to create the chromatogram library (green library icon) that contains all of the DIA-based peptide identifications and accurate retention times. EncyclopeDIA is then used again to extract fragment ion-based peptide information from the profiling injections using the chromatogram library generated from the GPF injections. Extracted ion chromatograms for each peptide are then post-processed in Skyline,[51] allowing for viewing the data, proteome regulation analysis (generation of protein or peptide volcano plots), and exporting the summed peptide fragment ion intensities for each peptide. Multiple peptides per protein can be summed to quantify individual proteins, but the method is also highly accurate on a single peptide level. Box 1 depicts how this methodology is important for smORF-encoded MP discovery. Each smORF sequence may only generate a small number of analytical peptides in a tryptic digest (or any given sample prep methodology). In contrast, a canonical ORF may generate many analytical peptides per protein in a tryptic digest.

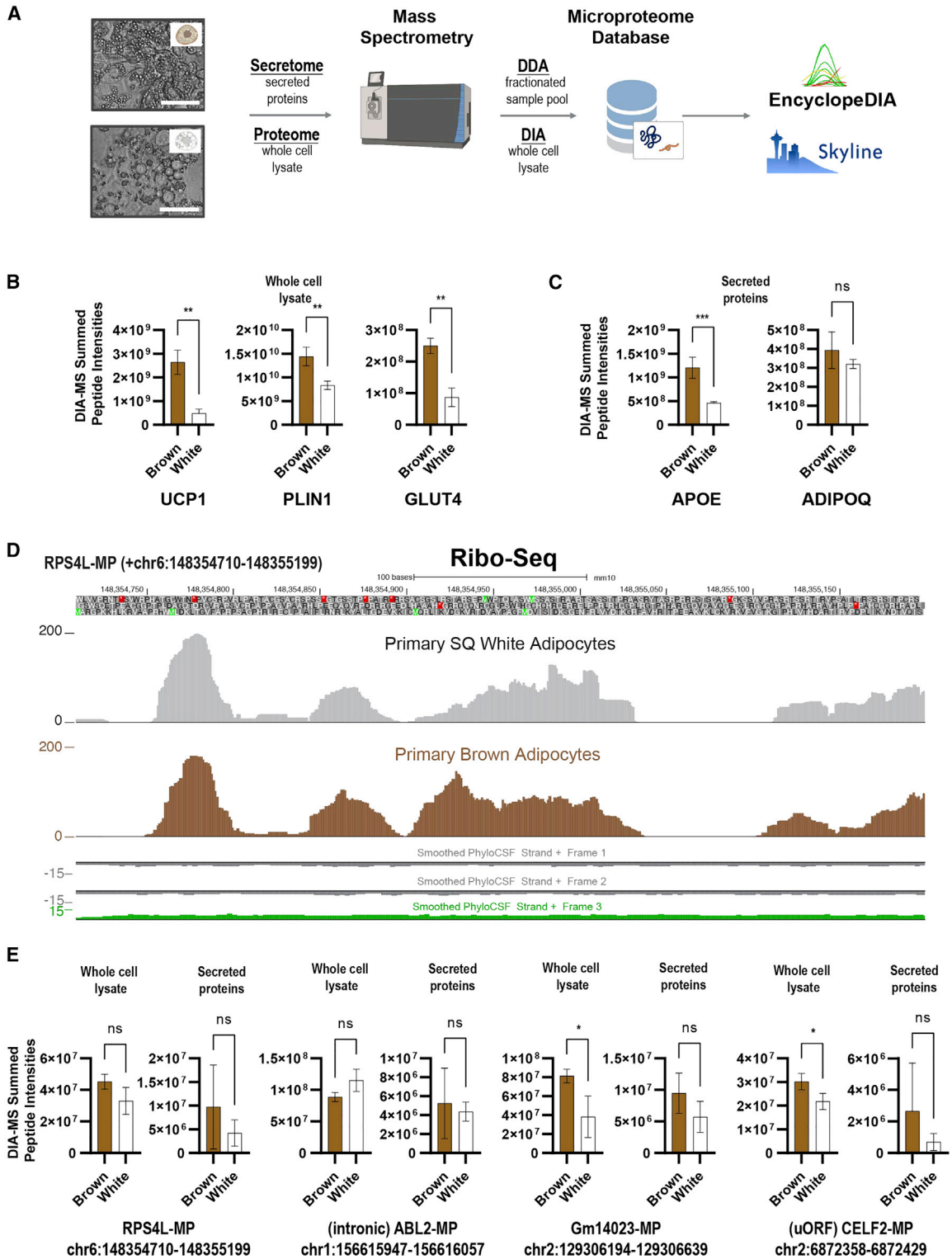


Figure 3. DIA-MS quantitation of canonical ORF proteins and small ORF microproteins in primary differentiated brown and subcutaneous white adipocytes

(A) A schematic of the experimental design quantifies the proteomes of primary differentiated brown and subcutaneous white adipocytes, allowing for identification with DDA and quantification with DIA-MS of both canonical ORF proteins and smORF-encoded MPs in whole-cell lysates and conditioned media (secretomes).

(legend continued on next page)

being fast enough for proteome-wide quantitation.[46–50] Our DIA-MS approach uses a complex pooling and fractionation strategy (outlined in Figure 2) to allow for precise measurements of individual peptides without missing values between replicates. Briefly, peptide sequence identifications are derived from non-quantitative DDA-MS injections of high pH reversed phase fractionations of representative sample pools. The same pools are additionally used in gas phase fractionation small-window DIA-MS injections to determine relative fragment ion intensities, retention times, and DIA-specific peak shapes of the individual peptides. We perform quantitation using wide-window DIA at the individual peptide level, represented as relative intensity values across a panel of samples (Figure 2).

Figure 3A outlines the DIA-MS strategy for quantifying proteins of primary differentiated BAT and WAT cell lysates and conditioned media (“secretome”). We observed expected changes such as higher UCP1, PLIN1, and GLUT4 protein levels from brown adipocyte cell lysates[24,52,53] (Figure 3B) and higher APOE and ADIPOQ levels in conditioned media[54] (Figure 3C). Analysis of MPs identified a total of 55 from cell lysates and conditioned media (Table S4; Figure 3). We quantified several MPs with DIA-MS, including an MP encoded on the *Rps4 l* pseudogene (RPS4L-MP) as well as an MP on an annotated intronic region of *Abl2* (“(intronic) ABL2-MP”) (Figures 3D and 3E), and found no difference in five of the MPs we quantified, but did show a modest change in the Gm14023-MP and CELF2-MP between brown and white cells (Figure 3E). Together, these examples validate the translation of some of the newly discovered MPs and demonstrate a sensitive workflow for the quantitation of peptides.

Upstream ORFs in mouse and human mRNAs

As stated previously, uORFs are often *cis* regulators of translation that can dynamically regulate translation of downstream mORFs.[55] Using previously identified smORFs from HEK293T, HeLa-S3, and K562 human cell lines,[4] we identified genes with established roles in adipose tissue biology that contain a uORF in both mouse and human. For example, *Pten*, a well-known tumor suppressor, improves adipose tissue metabolism and overall metabolic health when expressed at higher levels in mice, via up-regulation of *Ucp1* in BAT.[56] Both mouse and human *Pten* have a conserved uORF that encode a 45-amino-acid MP (Figure 4A). While these uORFs might have roles in regulating the translation of their downstream mORFs, sequence conservation at the protein level suggests that these MPs may have functions besides post-transcriptional regulation of the *Pten* mORF. Interestingly, a different *Pten* uORF-encoded MP was recently shown to regulate lactate metabolism in a mouse glioblastoma model.[57]

By contrast, mouse and human PPAR-δ both contain a translated uORF that differ substantially. PPAR-δ is an important nuclear receptor with broad roles in maintaining energy balance, and *Ppar-δ* knockout mice are metabolically less active and glucose intolerant.[58] The mouse *Ppar-δ* (mPPAR-δ) uORF detected by Ribo-seq encodes a 48-amino-acid MP that starts with MGK (Figure 4A). By contrast, the homologous human *PPAR-δ* (hPPAR-δ) uORF is truncated and starts with the sequence MHV due to the presence of a stop codon not found in mPPAR-δ. Because this truncated uORF is still translated in hPPAR-δ,[4] it is likely that the most important feature is the presence of the uORF and not the polypeptide sequence of any resultant MP. This suggests that post-transcriptional regulation of the PPAR-δ mORF is important. Both examples of functional uORFs from *Pten* and *Ppar-δ* provide interesting starting points for further investigation of these genes.

smORF-containing transcripts are regulated during diet-induced obesity

None of these newly discovered smORF-encoded MPs have been previously characterized. Therefore, we sought to highlight informatics approaches that can be used to begin to identify subsets of smORFs that are regulated in a manner that is similar to biologically active ORFs, and thus could serve as good starting points for functional studies.

First, we searched for evidence of smORF regulation during broad multi-organ metabolic changes triggered by diet-induced obesity (DIO). RNA-seq data were collected from BAT, epididymal WAT (eWAT), liver, subcutaneous WAT (scWAT), retroperitoneal fat (Retro Fat), and mesenteric fat (Mesen Fat) from cohorts of 27-week-old male DIO mice fed HFD for 21 weeks and healthy age-matched chow-fed controls (Tables S5 and S6). Analysis of this dataset showed consistent obesity-induced alterations system-wide as demonstrated by GO term enrichment analysis of up- and downregulated processes in BAT and eWAT (Figure S3B) as well as hundreds of smORF-containing transcripts that are significantly changed in adipose tissue depots of DIO mice (Figures 4B and S4; Table S6). This included marked changes to the *Irs1* and *Ucp2* smORFs (Figure S3A). By contrast, the liver had fewer changes (Figures 4B and S4C; Tables S5, and S6). We also wanted to determine whether smORF expression is capable of separating different tissue depots and metabolic states based on diet. Because the majority of smORFs are uORFs, which are confounded by potential translational regulation of the annotated mORF, we analyzed uORF and non-uORF smORFs separately. Principal component analysis (PCA) of non-uORF smORF expression levels was able to distinguish BAT from non-BAT fat depots and liver, and within

(B) DIA-MS quantification of canonical ORF proteins UCP1, PLIN1, and GLUT4 from the whole-cell lysates of the same primary differentiated brown and subcutaneous white adipocytes analyzed with Ribo-seq in Figure 1.
(C) DIA-MS quantification of known secreted proteins APOE and ADIPOQ (Adiponectin).
(D) Ribo-seq coverage of a conserved smORF (chr6:148,354,710–148355199) identified in both primary differentiated subcutaneous white and brown adipocytes.
(E) DIA-MS quantification of MPs in the whole-cell lysates and secretomes of both the differentiated brown and subcutaneous white adipocyte cultures showing 4 MPs quantified in both the whole-cell lysates and secretomes that includes MPs from the smORFs: chr6:148,354,710–148355199 (Ribo-Seq); chr1:156,615,947–156616057; chr2:129,306,194–129306639; and chr2:6,872,358–6872429.
Scale bar, 25 μm (A); bar plots represent the DIA-MS-based summed fragment ion intensities of each peptide for each protein/microprotein entry with at least one peptide per ORF/smORF where statistical comparisons were performed with a two-tailed Student’s t test with values representing the means ± SD; asterisks indicate *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

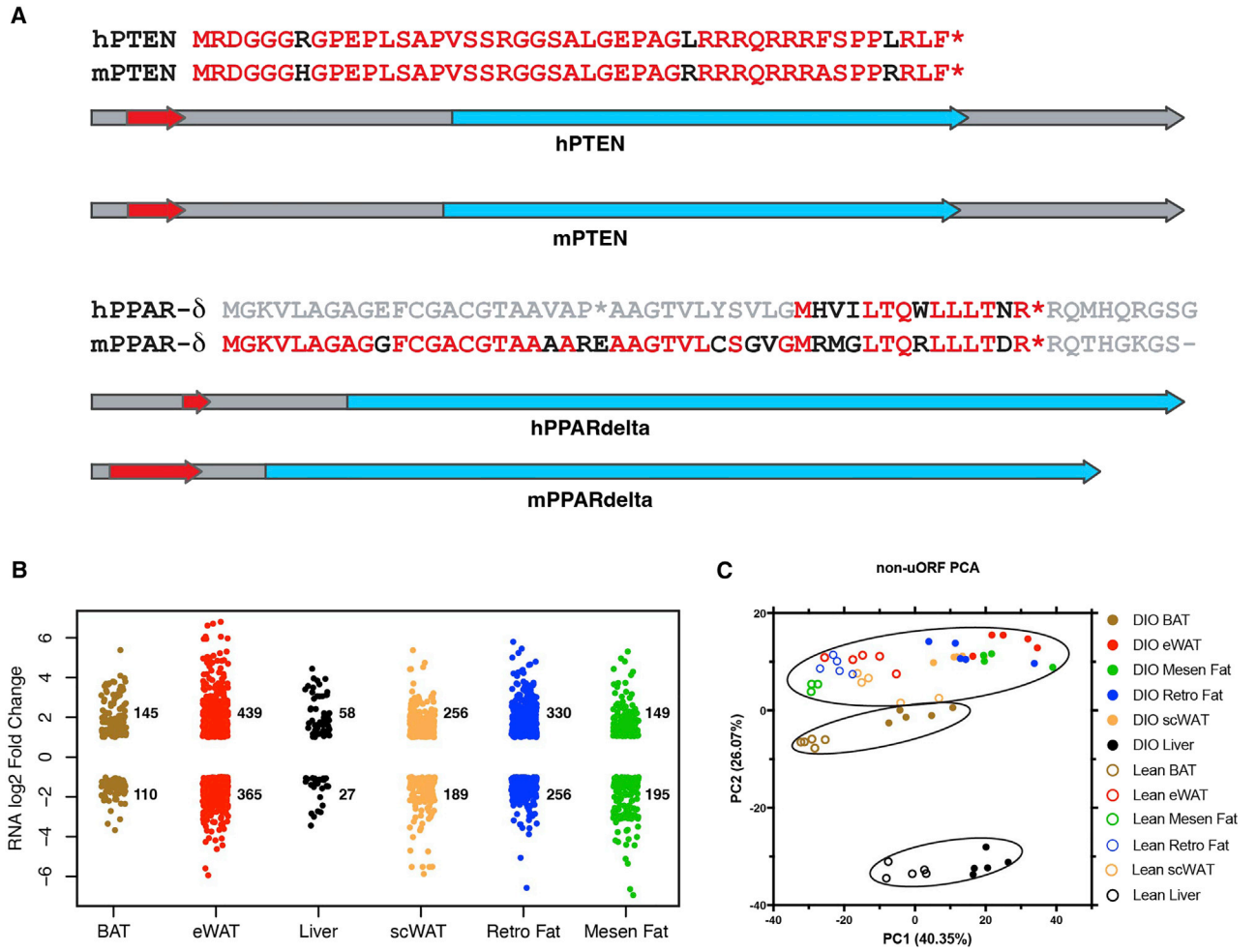


Figure 4. Adipose protein-coding smORFs are differentially transcribed in diet-induced obese mice

(A) Representation of human and mouse PTEN and PPAR-δ mRNAs, with uORFs in red and mORFs in blue, as well as the translation of the 5′ UTRs to reveal MPs above the mRNAs. With PPAR-δ, the presence of non-conserved uORFs in mice and humans supports a role for uORFs in translational regulation of PPAR-δ, but the lack of conservation makes it unlikely that the MPs from these uORFs are functional. By contrast, the uORFs from PTEN can regulate PTEN translation, but are also very likely to be functional in their own right because of the strong conservation between mouse and human PTEN uORF MPs. Thus, non-sequence conserved uORFs could still reveal post-transcriptional regulation across evolution.

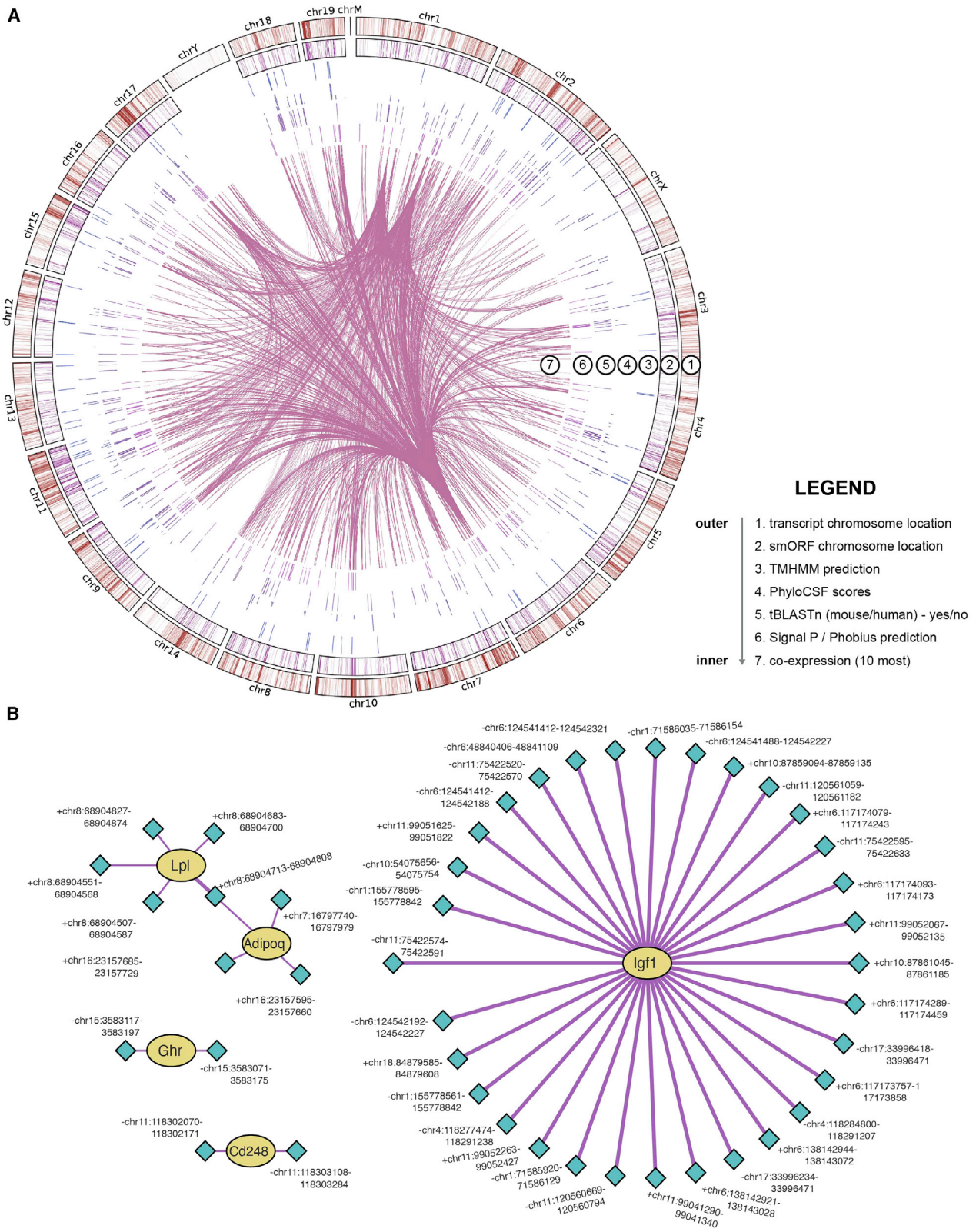
(B) Changes in RNA expression for adipose protein-coding smORFs induced by DIO in various tissues (padj < 0.05 and |log2 fold change| ≥ 1).

(C) PCA of non-uORF protein-coding smORF RNA expression levels in tissues derived from DIO and lean mice.

tissue groupings separation is also observed for DIO versus lean mice (Figure 4C). A similar observation was observed for uORFs; however, separation of groups is less between DIO and lean mice (Figure S3C). These differentially expressed smORFs potentially encode MPs that regulate adipose tissue biology changes under different nutrient conditions.

Specific genes might already offer a clue of the encoded MP's function. The long non-coding (lnc)RNA gene brown adipose tissue enriched long non-coding RNA 1 (*Lncbate1*) contains five predicted smORFs (Figure S3D). *Lncbate1* has decreased expression in several fat depots of obese mice (Table S5) and has been shown to be required for development and maintenance of brown adipocytes.[59] Since most methods that are used to characterize lncRNAs would change MP levels as well (e.g., siRNA, shRNA), these data now call into question whether this gene's functions are actually driven by the encoded MPs instead of, or in conjunction with, the transcribed RNA. Thus, MPs and smORFs from previously characterized lncRNAs represent a high-priority pool of MPs that warrant further investigation.

Next, we sought to determine which smORF mRNAs are co-expressed with annotated genes so that we could use the known functions of the annotated genes to connect smORFs to different biological pathways. These data are represented using a circos plot that shows canonical gene transcripts adjacent to our Ribo-seq-derived smORFs along with additional rings that show conservation (PhyloCSF) and mouse-to-human tBLASTn (yes/no) as well as signal peptide predictions and transmembrane domain predictions (Figure 5A). The center of the ring represents the 10 most interconnected smORFs in the network that are co-expressed with annotated genes from the DIO treatment (i.e., same datasets as Figure 4B), where co-expressed genes are most likely to function in similar pathways[60] (Table S7). Instead of



(legend on next page)

looking at the smORFs themselves to infer their functions, we selected genes known to be important for lipid metabolism and identified which smORFs have correlated expression levels in the co-expression network. This way we can infer that they might function in the same pathways as these genes. For example, *Lpl*, *Adipoq*, *Ghr*, *Cd248*, and *Igf1* have several to many smORFs that are co-regulated, including smORFs that are evolutionarily conserved. It would be reasonable to hypothesize that these co-regulated, conserved smORFs might have roles in biology related to these genes, and this could be tested in cellular studies.

DIA proteomics of mouse plasma microproteins

Even though proteomics offers the ability to detect far fewer MPs, the ability to detect secreted proteins is valuable and out of reach of genomics approaches. Furthermore, not all proteins are secreted via the canonical secretory pathway and would be missed when restricting searches to signal peptide-containing proteins. To identify and quantify circulating MPs, we analyzed the plasma proteome from DIO and lean mice at 26 and 41 weeks of age (Figure 6A). We used a low MW protein fractionation method in conjunction with the workflow outlined in Figure 2 to provide deep MP detections from within the challenging large dynamic range of plasma (Figures 6B and 6C). DIA-MS was then used to quantify circulating MPs across all the samples in the cohort.

We detected 33 circulating MPs each with a single tryptic peptide (Table S8), highlighting the advantage of using DIA-MS. Only two of these circulating peptides have a predicted signal peptide when analyzed using SignalP 5.0 (Figure S5), underscoring the utility of proteomics to identify secreted MPs that lack a signal peptide. The two MPs with signal peptides are encoded by smORFs on the putative ncRNA A530053G22Rik (Figure 7A) and lymphocyte antigen 6 complex pseudogene 9030619P08Rik (Figure S5B), the latter of which MP shares 50%–60% sequence identity to several other mouse lymphocyte antigen 6 complex members, a family with emerging roles in immunity and cancer.[61] In addition, an MP encoded on AW112010 scored just under the threshold for signal peptide prediction by SignalP 5.0 (Figure 7C), but passed when using SignalP 4.1 (data not shown). AW112010-MP was recently characterized with a role in mucosal immunity,[62,63] but this is the first evidence that AW112010-MP is a circulating factor (Figure 6D). Relative quantitation of the proteomes between the mice groups in Figure 6A could detect known biomarkers (ALDOB, APOC2, etc.)[64] as well as one MP that was changed (Figures 6E and 6F). Specifically, we found increased levels in aged/obese mice of a 24-amino-acid MP encoded by a uORF from the fibroblast growth factor receptor substrate 2 (FRS2; Figures 6F–6H). FRS2 is an intracellular protein that links ligand-activated fibroblast growth factor receptor 1 (FGFR1) to downstream ERK and PI3K/AKT signaling pathways[65–67] (Figure 6G). Given that fibroblast growth factor signaling is an important regulator of metabolism,[67] this secreted FRS2-encoded MP could also function as a metabolic regulator in aged/obese mice. These data provide a preliminary example that some MPs may eventually find uses as biomarkers.

Gm8773 is translated and encodes the mouse orexigenic FAM237B neurosecretory MP

In our Ribo-seq data, we identified a translated smORF from the predicted gene *Gm8773* in primary differentiated BAT- and scWAT-derived adipocyte cultures (Figure 7A), providing the first evidence for translation of this mRNA in mice. We became interested in *Gm8773* because this 132-amino-acid MP contains a signal peptide that would generate a 108-amino-acid secreted MP in its final predicted processed form and is also predicted to be glycosylated[68] (Figures 7B and 7C). Furthermore, the Gm8773-MP is a homolog of the human, rat, and chicken FAM237B (also called neurosecretory protein GM or NPGM).[70–72] FAM237B has orexigenic effects in chickens[73] and hFAM237B has reported activity against the feeding receptor GPR83.[74] Until recently, *Gm8773* was annotated as an ncRNA. Having validated the translation of mouse FAM237B, we wanted to test whether this MP has activity in mice. Previous studies have identified *Gm8773* to be expressed in the NPY-labeled neurons in the arcuate nucleus (ARH) (Figure 7D)—the brain's feeding center.[75,69] We further located *Gm8773* mRNA expression in the ARH with qPCR on a broad panel of tissue RNAs showing high expression in the hypothalamus followed by *in situ* hybridization locating the transcript in the ARH (Figures 7E and 7F). Together, combined with the *Gm8773* high degree of conservation, predicted glycosylation, positive SignalP score, and negative TMHMM scores, we hypothesized that mFAM237B might act as a secreted hormone with a role in mouse feeding.

To test this hypothesis, we generated recombinant mFAM237B for *in vivo* testing. Recombinant expression of Gm8773/FAM237B in mammalian cells (HEK 293T) generated a glycosylated MP of the correct molecular weight that ran as a monomer in SDS-PAGE (Figure 7G), and this recombinant MP was used for *in vivo* intracerebroventricular (i.c.v.) injections into DIO mice 30 min before the nocturnal cycle. We observed an induction of feeding in the mFAM237B-treated mice (Figure 7H), consistent with the effects of FAM237B in chickens.[70,73] Thus, the validation of mFAM237B translation from *Gm8773* helped

Figure 5. Adipose protein-coding smORFs are located throughout the genome and can be co-expressed with important lipid metabolism regulators

(A) Circular genome plot showing the genomic landscape of the novel smORFs. The annotated genes are represented as red bars in the outermost section, while the smORFs are depicted as purple bars in the second section. The third ring depicts the smORFs with a positive transmembrane prediction from TMHMM. The fourth ring summarizes the PhyloCSF information, with peaks representing positive PhyloCSF scores. The bars in the fifth ring represent the presence of homologous sequences in human identified with tBLASTn for the smORF in the same coordinates. The sixth ring represents smORFS with a positive signal peptide from either SignalP 5.0 or Phobius or both. The center of the plot contains links illustrating the co-expression between one of the top 10 ranked SCTs in the network and another gene.

(B) Network showing co-expression of smORFs with genes whose functions are related to lipid metabolism. Cyan diamonds represent a smORF, and yellow circles represent an annotated gene. Edges correspond to a correlation between the expression levels of two different nodes across multiple conditions.

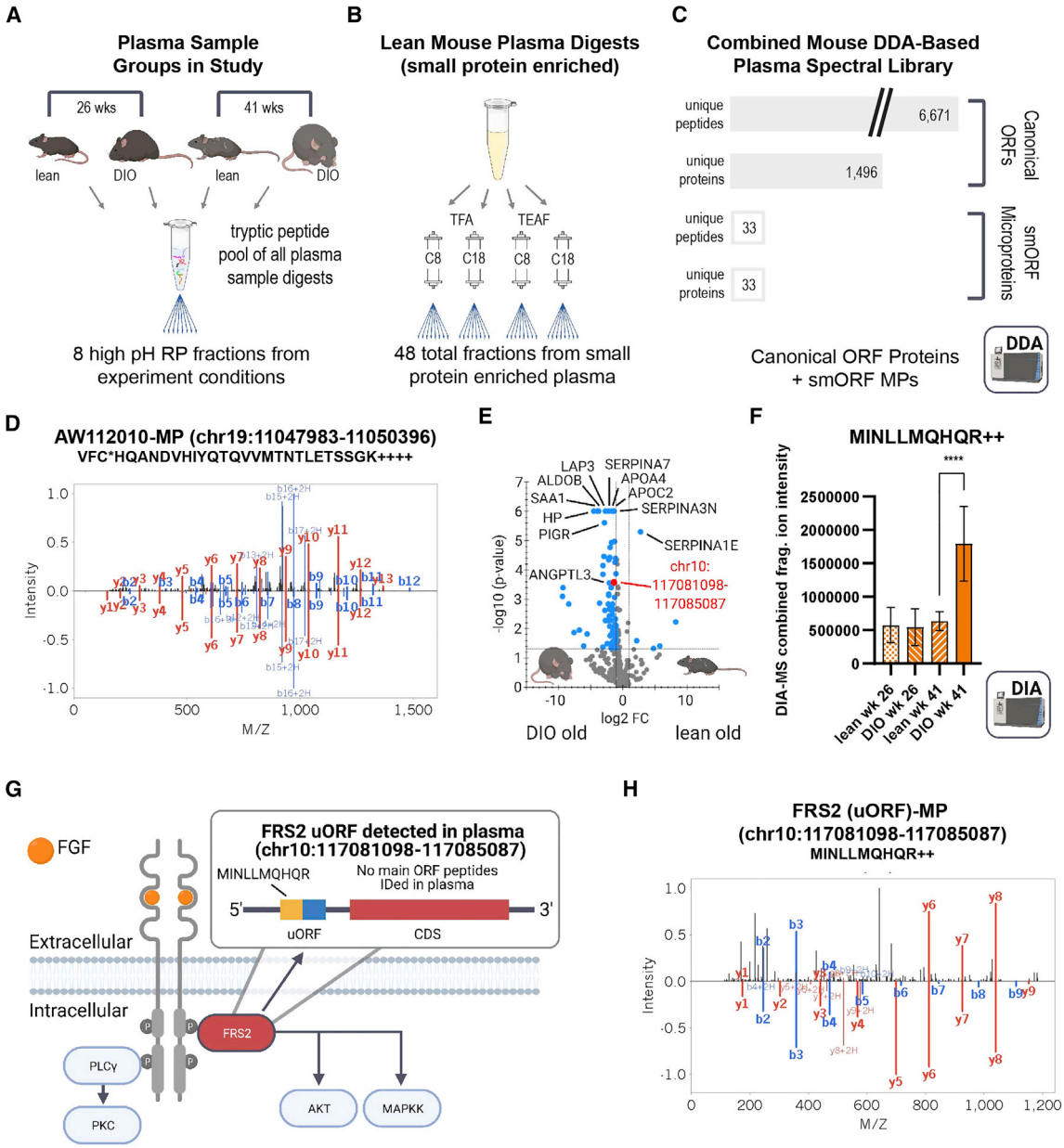


Figure 6. Plasma proteomics of aged obese mice identifies evidence for multiple smORF-encoded microproteins with one FRS2 uORF circulating at a higher level in the aged obese state

(A) Experimental design for plasma proteomics of 26- and 41-week-old mice with both DIO and lean groups in both ages (n = 12 per group).

(B) Strategy for non-quantitative DDA-based deep identification of canonical ORF proteins and smORF-encoded MPs in lean mouse plasma shows a combinatorial fractionation strategy designed to enrich small proteins using both C8- and C18-based fractionation with both trifluoroacetic acid (TFA) and triethylammonium formate (TEAF) ion-pairing agents. Additionally, plasma from the experiment depicted in (A) was fractionated (not pictured) with high-pH RP fractionation as depicted in Figure 2.

(C) Summary of all proteins identified with the fractionation strategies outlined in (B) showing peptides and proteins identified from both the canonical ORFs and smORF-encoded MPs.

(D) Annotated MS2 fragment ion spectrum of a peptide (VFC*HQANDVHIYQTQVVMTNTLETSSGK++++, * = reduced and alkylated cysteine) that maps to an MP generated from the lncRNA AW112010 (chr19:11,047,983–11050396) discovered in the circulation via the plasma fractionation depicted in (B). The peptide is depicted in a butterfly plot with the measured MS2 spectrum above the x axis and the Prosit-predicted fragment ion pattern below the x axis.

(E) Volcano plot of DIA-MS quantification of the experiments depicted in (A) comparing the DIO old condition to the lean old condition depicting regulated canonical ORF proteins along with a regulated smORF-encoded MP that correlates with chr10:117,081,098–117085087.

(F) Quantification with DIA-MS of a tryptic peptide (sequence: MINLLMQHQR++) for the smORF-encoded MP from chr10:117,081,098–117085087 across all of the biological conditions in (A) where quantification represents the DIA-MS-based summed fragment ion intensity depicting the mean ± SD with statistics performed using a one-way ANOVA where ****$p < 0.0001$.

(legend continued on next page)

establish a role for this MP as a novel orexigenic factor in mice and suggests that this MP has a role in the brain-adipose axis that controls feeding. The advantage of having established this system in mice is that we can use the large repository of knockout mice and genetic engineering tools to better study the role of this MP in ways that would have not been feasible in chickens or other organisms, and thus may accelerate the development of novel medicines based on this MP.

DISCUSSION

Adipose depots have well-documented endocrine functions with leptin, a potent circulating adipokine, being a long-standing exemplar.[76] This piqued our interest in the possibility of identifying novel secreted MPs in adipocytes. While tissue ribosome profiling is feasible, low overall yields of RNA in adipocytes create challenges in tissue-based Ribo-seq, and profiling endogenous adipose tissue would result in a mixture of cell types. For these reasons, we focused our Ribo-seq efforts on primary white, beige, and brown adipocytes. This led to the identification of 3,877 smORFs that all potentially encode MPs that are absent from the SwissProt UniProt reviewed proteome database, providing an entirely new set of adipose proteins to study and understand. We imagine that these genes can and should be included in large-scale genomic screening efforts like using CRISPR to identify novel functions as translational regulators[77] and/or functional MPs,[1,7,78] while focused follow-up experiments on MPs with MS evidence are also warranted as the protein starting point provided here accelerates the justification for more in-depth experimentation for each smORF.

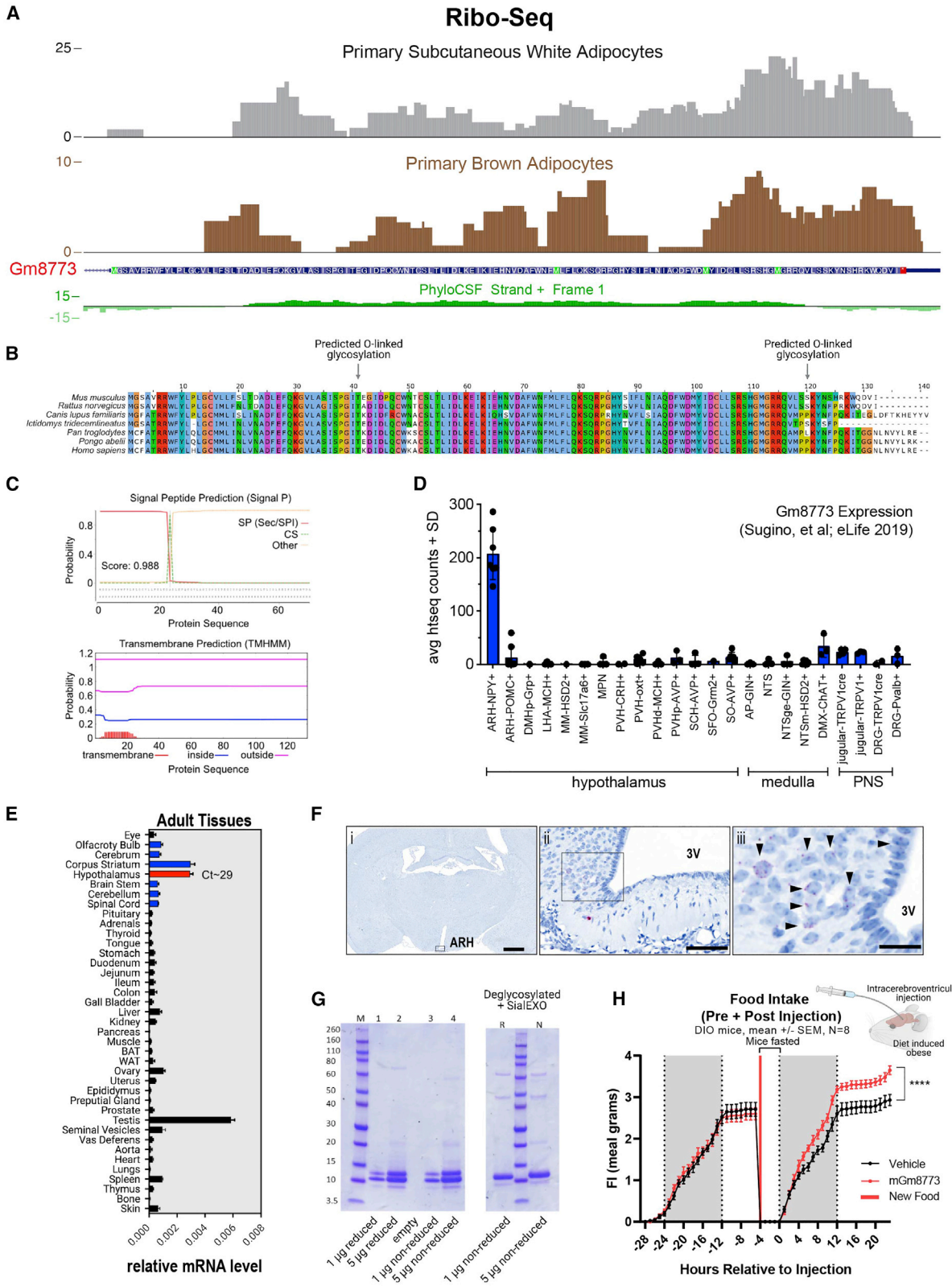
To understand the potential impact of large groups of MPs on metabolism, we examined smORF regulation and found hundreds that are transcriptionally regulated in various adipose tissue depots in mice fed HFD. These putative MPs along with the evolutionarily conserved smORFs represent a subset that should be at the top of any list for subsequent functional studies. More generally, with thousands of available RNA-seq datasets, researchers can identify smORFs regulated under different metabolic conditions or genetic models where changes in MP expression levels can lead to new testable hypotheses.

In general, MPs are found on transcript regions thought to be non-coding—i.e., the 5′ and 3′ UTRs of known coding mRNAs and on lncRNAs. For mRNAs with multiple ORFs, the smORFs are thought to act as translational regulators of the mORFs.[35,79,80] The uORF regulator of PGC1α is one example of an important translational regulator where the uORF generates a stable MP that in part controls translation of the canonical ORF,[77] and a recent study showed a translated uORF MP that can inhibit PKC phosphorylation.[81] Of course, many of these smORFs might regulate mORF translation via ribosome stalling *and* also produce a functional MP.[1] Therefore, it is interesting to find smORFs on the mRNAs of several metabolic genes (e.g., *Insr*, *Plin1*, *Ucp1*), as these underappreciated genetic elements might have a key role in the positive or negative regulation of the translation of these genes, or indeed produce bioactive MPs. Furthermore, the presence of non-conserved uORFs in genes like PPAR-δ might be a way to find functional uORFs that do not encode functional MPs but are instead translational regulators. Similarly, the discovery that some functional "non-coding" RNAs, such as *Lncbate1*, may in fact be translated[59] may lead to the expansion of our understanding of their biology as potential protein-coding genes and not just ncRNAs, although the possibility of both biological functions co-occurring also exists.[82,83] For metabolism, certainly the Ribo-seq-based discovery of translation of an ncRNA whose amino acid sequence has a predictable secretion signal peptide is of high interest for follow-up studies such as those shown here with the observed central feeding control associated with our recombinant Gm8773.

As new genes, antibodies to MPs need to be developed, which can be difficult and time consuming.[7] Proteomics offers a solution to MP detection and complements Ribo-seq by reporting on the stability and location of an MP. However, owing to their length, MPs produce few tryptic peptides for detection, and the tryptic peptides they generate are often not unique to the MP, or none of the tryptic peptides are detectable because of their biophysical properties.[84–86] These limitations hamper the validation of MPs and have also limited previous proteomics-based quantification efforts. DIA-MS provides a more reliable detection and quantitation platform for dealing with single peptides by MS, resulting in an ideal strategy for MPs. Indeed, using DIA-MS, we were able to detect and quantify MPs in our primary adipocyte cultures and also in plasma. With the ability of proteomics to inform on the location of an MP, we revealed the existence of several circulating MPs in mouse plasma, including a well-annotated spectrum for AW112010-MP, a recently discovered mouse MP with a role in gut immunity. Revealing AW112010-MP as a circulating MP suggests a broader role in immunity that complements the original discovery of the protein.[62]

At the outset of these studies, Gm8773 was listed as an ncRNA and Ribo-seq provided the first empirical evidence that Gm8773 produces a protein in mice (this misannotation was recently revised and Gm8773 is currently listed as a protein-coding gene in RefSeq, but a "Predicted gene" in UniProt: Q3UQ24). As a result of its misannotation, there are no reports of Gm8773 translation or function, but fortunately, homology to a characterized chicken protein FAM237B led us to show that this gene makes a stable, folded, secretable, and glycosylated recombinant MP that regulates food intake when administered centrally to DIO mice, which is also consistent with published chicken and rat experiments.[70,73] To complement the central food intake increase effects seen here, the Gm8773 gene was previously reported to be co-expressed with AgRP, a known orexigenic peptide, in the hypothalamus in an scRNA-seq atlasing effort.[75] Given the availability of readily available genetic mouse models and experimental paradigms in mice, research into the potential impact of this gene can now accelerate with more confidence aided by the verification of its protein function in mice. More

(G) The amino acid sequence of the smORF-encoded MP from chr10:117,081,098–117085087 that maps to the uORF region of fibroblast receptor substrate 2 (FRS2) with the identified tryptic peptide (sequence: MINLLMQHQR) depicted in yellow and the whole smORF in blue/yellow.
(H) Annotated MS2 fragment ion spectrum of the tryptic peptide from the FRS2 uORF (sequence: MINLLMQHQR++) was used to quantify this MP in (E) and (F) with DIA-MS.



(legend on next page)

generally, this example highlights the value of MP annotation to uncover these novel proteins and facilitate their characterization to provide new biological insights and therapeutic opportunities. Our workflow outlining the approach to validating the extracellular function of the Gm8773-MP can be generalized to the discovery of function of any of the smORFs in the dataset presented here. For any new smORF, we recommend focusing first on conservation, regulation under biological conditions, presence on an ncRNA, and secretion (or presence in the extracellular space) as the top criteria for generating hypotheses to test as follow-up. As with any new biology, however, all of the evidence must be considered in aggregate especially where gaps exist, which will be common with any newly discovered gene.

Altogether, the smORF database generated in this study will find value in cross-referencing with other -omics datasets including genomics data, bulk RNA-seq data from disease models, proteomics data by amending protein databases with MP sequences, and the ever-expanding single-cell datasets. Moreover, with DIA-MS adoption increasing in proteomics due to superior quantitation, we imagine that the integration of sequences will reveal new MPs that are increased or decreased under different conditions. These -omics studies will lead to targeted hypothesis-driven experiments to look at the function of specific smORFs or MPs based on their location or regulation. For instance, upstream and downstream smORFs on genes such as *Ucp1* and *Irs1* suggest a layer of translational regulation that is currently not well understood. Thus, the impact of this work will be the identification of new genes with roles in murine metabolism, and eventually some of these genes will provide insights that can generate therapeutic opportunities in obesity and diabetes and potentially other diseases.

Limitations of study

In this study, we provide evidence for the translation of thousands of smORFs in primary brown and white adipocytes, and the proteomic detection of MPs from cell proteomes, conditioned media, and plasma. MS underestimates MP numbers because these polypeptides are short and generate few detectable tryptic peptides. Furthermore, without trypsinization, MPs do not ionize well, which partially explains the difference in numbers between MS and Ribo-seq. That said, smORF annotation by Ribo-seq does not guarantee that the encoded MPs are stable, long-lived molecules capable of regulating biology, though our hypothesis is that there are enough biologically active MPs within this pool to make it worthwhile to interrogate these genes. Indeed, the Ribo-seq data provided the first empirical evidence for the translation of the biologically active Gm8773 MP. Pharmacological experiments with the recombinant Gm8773 MP and i.c.v. injections reveal an intriguing feeding phenotype. To determine whether these pharmacological experiments define a physiological role of Gm8773, subsequent studies should employ mouse genetics that knockout/knockin Gm8773 in cells and tissues of a living animal.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Creation of diet-induced obese (DIO) mouse tissue atlas for RNA-Seq analysis
  - Creation of DIO/lean young/old mouse plasma samples
  - Intracerebroventricular (ICV) administration of Gm8773 protein product for assessing food intake and body weight effects in the diet-induced obese mouse
- METHOD DETAILS
  - Isolation and differentiation of primary mouse white, beige, and brown adipocytes for both ribosome profiling and mass spectrometry
  - Ribo-Seq and total RNA-Seq sample preparation from primary mouse white, beige, and brown adipocytes
  - Ribo-Seq and bioinformatics analysis
  - Whole cell lysates (“proteomes”) and secreted protein (“secretome”) sample preparation from primary mouse white, beige, and brown adipocytes
  - Trypsin digestion and desalting of differentiated brown and white adipocyte proteome and secretome samples
  - Bulk mRNA-Seq analysis with next-generation sequencing in DIO/lean tissue

Figure 7. Translation, expression, and activity of Gm8773

(A) Ribo-seq evidence for the translation of Gm8773 in both differentiated subcutaneous white and brown adipocytes.

(B) Amino acid level conservation of the predicted protein sequence of Gm8773 along with the two residues that have a positive O-linked glycosylation prediction score (NetOGlyc 4.0[68]).

(C) Signal peptide (SignalP 5.0) and transmembrane domain (TMHMM 2.0) predictions for Gm8773 predicted protein sequence.

(D) Replotting of published transcriptional co-expression data of Gm8773 expression within specific nuclei of the hypothalamus and other regions of the brain; note Gm8773 co-expressed with NPY-containing neurons.[69]

(E) Relative comparative tissue level mRNA expression with qPCR of Gm8773 across a panel of mouse tissues.

(F) *In situ* hybridization of Gm8773 mRNA localization to the arcuate nucleus of the hypothalamus, showing (i) the arcuate nucleus and (ii) median eminence, and (iii) arrows denoting mGm8773+ cells in the mediobasal hypothalamus. Scale bars, 500 μm (i), 250 μm (ii), 50 μm (iii).

(G) Expression of recombinant Gm8773 in HEK cells showing a double band at the molecular weight of a protein monomer in both reducing and non-reducing conditions (left side) along with the collapsing of the Gm8773 doublet protein band down to a single band upon treatment with an O-linked glycosylation deglycosylating enzyme (right side).

(H) Increased food intake is observed in mice following the intracerebroventricular (i.c.v. injection) administration of the recombinant Gm8773 protein from (G). Statistics were performed with two-way ANOVA with significant time by treatment interaction, ****$p < 0.0001$.

- Bioinformatics quantitation of NGS transcriptome data of bulk tissue (adipose depots and liver) mRNA DIO/lean
- Circular genome visualization and co-expression networks
- Tissue distribution of smORFs
- Sequence conservation of smORFs across different species
- Sample preparation and protein digestions including cleanup and desalting of tryptic peptides for DIO/lean young/old plasma
- Small protein sub-fractionation in mouse plasma
- Sample pooling and fractionation for mass spectrometry
- Data-dependent acquisition mass spectrometry (DDA-MS)
- Data-independent acquisition mass spectrometry (DIA-MS)
- Bioinformatics for proteomics
- qPCR for Gm8773 mRNA in Mouse Tissue extracts
- *In situ* hybridization and visualization of Gm8773 transcript in the mouse brain
- Recombinant expression and purification of Gm8773 protein product
- Acute food intake effects of ICV Gm8773 injections in DIO mice

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cmet.2022.12.004.

ACKNOWLEDGMENTS

The authors would like to thank Erin Whalen, Fiona McMurray, and Nick Cox for critical feedback throughout the planning and writing of this work. The authors also thank Kevin Grove and Mads Tang-Christensen for support. B.C.S. acknowledges financial support from NIH grant R01 GM133981. M.J.M. acknowledges financial support from NIH grants P41 GM103533, R24 GM141156, and U19 AG065156, and a sponsored research agreement with Novo Nordisk Research Center Seattle, Inc. A. Saghatelian acknowledges financial support from NIH grants P30CA014195, R01GM102491, and RC2DK129961; Frederick Paulsen and the Ferring Foundation; and a sponsored research agreement with Novo Nordisk Research Center Seattle, Inc. T.F.M. acknowledges financial support from NIH grant K01CA249038. C.V.B. acknowledges financial support from CNPq/FAPERGS/CAPES/BNDES to the National Institute of Science and Technology on Tuberculosis (INCT-TB), Brazil (grant numbers 421703-2017-2/17-1265-8/14.2.0914.1). C.V.B. (310344/2016-6) is also a research career awardee of the National Council for Scientific and Technological Development of Brazil (CNPq). Data analyses in this study were financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES).

AUTHOR CONTRIBUTIONS

Study conceptualization, C.A.B., T.F.M., A.J.W., M.J.M., and A. Saghatelian. C.A.B., A.J.W., A.L.B., S.A.M., D.F., Y.Z., and K.B. conceptualized the animal studies. T.F.M., M.N.S., C.D., J.M.V., Y.Z., A.L.B., and C.L. performed the RNA experiments and/or Ribo-seq analysis methodologies. J.M.V. performed the plasma fractionations. T.F.M., M.N.S., A.L.B., S.A.M., E.V.D.S., C.V.B., and C.A.B. performed the bioinformatics methodologies, visualization, and data curation work. C.A.B. and B.K. performed the MS. C.A.B., L.K.P., B.C.S., and M.J.M. performed the MS bioinformatics methodologies, visualization, and data curation. S.L.-A., A.A., T.K., A.M., and A. Simon performed the protein production methodologies. A.J.M. performed the *in situ* hybridization methodologies. A.F.B., K.B., D.F., J.H., R.D., S.P., C.S., and G.Z. performed the *in vivo* animal study methodologies. C.A.B., M.J.M., and A. Saghatelian performed supervision of the research. T.F.M., C.A.B., E.V.D., and A. Saghatelian wrote the original draft.

DECLARATION OF INTERESTS

All authors affiliated with the Novo Nordisk Research Center Seattle, Inc. have worked for a for-profit commercial pharmaceuticals company that produces and sells medicines for the treatment of obesity and diabetes. B.C.S. is a founder and shareholder of Proteome Software. M.J.M. has a sponsored research agreement with and is a paid consultant for Thermo Fisher Scientific. A. Saghatelian is a paid consultant and shareholder for and cofounder of Exo Therapeutics and Velia Therapeutics. T.F.M. is a paid consultant and shareholder of Velia Therapeutics. C.A.B. is a current employee of Velia Therapeutics.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location.

Received: March 15, 2022
Revised: September 19, 2022
Accepted: December 6, 2022
Published: January 3, 2023

REFERENCES

1. Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D., and Weissman, J.S. (2020). Pervasive functional translation of noncanonical human open reading frames. Science *367*, 1140–1146.
2. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.
3. Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.
4. Martinez, T.F., Chu, Q., Donaldson, C., Tan, D., Shokhirev, M.N., and Saghatelian, A. (2020). Accurate annotation of human protein-coding small open reading frames. Nat. Chem. Biol. *16*, 458–468.
5. Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L., and Saghatelian, A. (2013). Peptidomic discovery of short open reading frame–encoded peptides in human cells. Nat. Chem. Biol. *9*, 59–64.
6. van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J.F., Adami, E., Faber, A.B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.-L., et al. (2019). The translational landscape of the human heart. Cell *178*, 242–260.e29.
7. Chu, Q., Martinez, T.F., Novak, S.W., Donaldson, C.J., Tan, D., Vaughan, J.M., Chang, T., Diedrich, J.K., Andrade, L., Kim, A., et al. (2019). Regulation of the ER stress response by a mitochondrial microprotein. Nat. Commun. *10*, 4883.
8. Anderson, D.M., Anderson, K.M., Chang, C.L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J., Bassel-Duby, R., and Olson, E.N. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell *160*, 595–606.
9. Bi, P., Ramirez-Martinez, A., Li, H., Cannavino, J., McAnally, J.R., Shelton, J.M., Sánchez-Ortiz, E., Bassel-Duby, R., and Olson, E.N. (2017). Control of muscle formation by the fusogenic micropeptide myomixer. Science *356*, 323–327.
10. Nelson, B.R., Makarewich, C.A., Anderson, D.M., Winders, B.R., Troupes, C.D., Wu, F., Reese, A.L., McAnally, J.R., Chen, X., Kavalali,

E.T., et al. (2016). A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science *351*, 271–275.

11. Zhang, Q., Vashisht, A.A., O'Rourke, J., Corbel, S.Y., Moran, R., Romero, A., Miraglia, L., Zhang, J., Durrant, E., Schmedt, C., et al. (2017). The microprotein Minion controls cell fusion and muscle formation. Nat. Commun. *8*, 15664.
12. Makarewich, C.A., Baskin, K.K., Munir, A.Z., Bezprozvannaya, S., Sharma, G., Khemtong, C., Shah, A.M., McAnally, J.R., Malloy, C.R., Szweda, L.I., et al. (2018). MOXI is a mitochondrial micropeptide that enhances fatty acid beta-oxidation. Cell Rep. *23*, 3701–3709.
13. GBD 2015 Obesity Collaborators, Afshin, A., Forouzanfar, M.H., Reitsma, M.B., Sur, P., Estep, K., Lee, A., Marczak, L., Mokdad, A.H., Moradi-Lakeh, M., et al. (2017). Health effects of overweight and obesity in 195 countries over 25 years. N. Engl. J. Med. *377*, 13–27.
14. Chugunova, A., Loseva, E., Mazin, P., Mitina, A., Navalayeu, T., Bilan, D., Vishnyakova, P., Marey, M., Golovina, A., Serebryakova, M., et al. (2019). LINC00116 codes for a mitochondrial peptide linking respiration and lipid metabolism. Proc. Natl. Acad. Sci. USA *116*, 4940–4945.
15. Friesen, M., Warren, C.R., Yu, H., Toyohara, T., Ding, Q., Florido, M.H.C., Sayre, C., Pope, B.D., Goff, L.A., Rinn, J.L., and Cowan, C.A. (2020). Mitoregulin controls beta-oxidation in human and mouse adipocytes. Stem Cell Rep. *14*, 590–602.
16. Lin, Y.F., Xiao, M.H., Chen, H.X., Meng, Y., Zhao, N., Yang, L., Tang, H., Wang, J.L., Liu, X., Zhu, Y., and Zhuang, S.M. (2019). A novel mitochondrial micropeptide MPM enhances mitochondrial respiratory activity and promotes myogenic differentiation. Cell Death Dis. *10*, 528.
17. Stein, C.S., Jadiya, P., Zhang, X., McLendon, J.M., Abouassaly, G.M., Witmer, N.H., Anderson, E.J., Elrod, J.W., and Boudreau, R.L. (2018). Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency. Cell Rep. *23*, 3710–3720.e8.
18. Lee, C., Zeng, J., Drew, B.G., Sallam, T., Martin-Montalvo, A., Wan, J., Kim, S.J., Mehta, H., Hevener, A.L., de Cabo, R., and Cohen, P. (2015). The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. Cell Metab. *21*, 443–454.
19. Kusminski, C.M., Bickel, P.E., and Scherer, P.E. (2016). Targeting adipose tissue in the treatment of obesity-associated diabetes. Nat. Rev. Drug Discov. *15*, 639–660.
20. Reilly, S.M., and Saltiel, A.R. (2017). Adapting to obesity with adipose tissue inflammation. Nat. Rev. Endocrinol. *13*, 633–643.
21. Smith, U., and Kahn, B.B. (2016). Adipose tissue regulates insulin sensitivity: role of adipogenesis, de novo lipogenesis and novel lipids. J. Intern. Med. *280*, 465–475.
22. Vijay, J., Gauthier, M.-F., Biswell, R.L., Louiselle, D.A., Johnston, J.J., Cheung, W.A., Belden, B., Pramatarova, A., Biertho, L., Gibson, M., et al. (2020). Single-cell analysis of human adipose tissue identifies depot- and disease-specific cell types. Nat. Metab. *2*, 97–109.
23. Cannon, B., and Nedergaard, J. (2001). Cultures of adipose precursor cells from brown adipose tissue and of clonal brown-adipocyte-like cell lines. Methods Mol. Biol. *155*, 213–224.
24. Cannon, B., and Nedergaard, J. (2004). Brown adipose tissue: function and physiological significance. Physiol. Rev. *84*, 277–359.
25. Rosen, E.D., and Spiegelman, B.M. (2014). What we talk about when we talk about fat. Cell *156*, 20–44.
26. Roberts, L.D., Boström, P., O'Sullivan, J.F., Schinzel, R.T., Lewis, G.D., Dejam, A., Lee, Y.K., Palma, M.J., Calhoun, S., Georgiadi, A., et al. (2014). beta-Aminoisobutyric acid induces browning of white fat and hepatic beta-oxidation and is inversely correlated with cardiometabolic risk factors. Cell Metab. *19*, 96–108.
27. Hausman, D.B., Park, H.J., and Hausman, G.J. (2008). Isolation and culture of preadipocytes from rodent white adipose tissue. Methods Mol. Biol. *456*, 201–219.
28. Ji, Z. (2018). RibORF: identifying genome-wide translated open reading frames using ribosome profiling. Curr. Protoc. Mol. Biol. *124*, e67.
29. Ji, Z., Song, R., Regev, A., and Struhl, K. (2015). Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. eLife *4*, e08890.
30. Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc. Natl. Acad. Sci. USA *106*, 7507–7512.
31. Mouilleron, H., Delcourt, V., and Roucou, X. (2016). Death of a dogma: eukaryotic mRNAs can code for more than one protein. Nucleic Acids Res. *44*, 14–23.
32. Vattem, K.M., and Wek, R.C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc. Natl. Acad. Sci. USA *101*, 11269–11274.
33. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.
34. Griffin, E., Re, A., Hamel, N., Fu, C., Bush, H., McCaffrey, T., and Asch, A.S. (2001). A link between diabetes and atherosclerosis: glucose regulates expression of CD36 at the level of translation. Nat. Med. *7*, 840–846.
35. Wu, Q., Wright, M., Gogol, M.M., Bradford, W.D., Zhang, N., and Bazzini, A.A. (2020). Translation of small downstream ORFs enhances translation of canonical main open reading frames. EMBO J. *39*, e104763.
36. Fields, A.P., Rodriguez, E.H., Jovanovic, M., Stern-Ginossar, N., Haas, B.J., Mertins, P., Raychowdhury, R., Hacohen, N., Carr, S.A., Ingolia, N.T., et al. (2015). A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. Mol. Cell *60*, 816–827.
37. Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. *6*, 175–182.
38. Hu, F., Lu, J., Matheson, L.S., Díaz-Muñoz, M.D., Saveliev, A., Turner, M., and Turner, M. (2021). ORFLine: a bioinformatic pipeline to prioritize small open reading frames identifies candidate secreted small proteins from lymphocytes. Bioinformatics *37*, 3152–3159.
39. Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. *37*, 420–423.
40. Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. *338*, 1027–1036.
41. Mudge, J.M., Jungreis, I., Hunt, T., Gonzalez, J.M., Wright, J.C., Kay, M., Davidson, C., Fitzgerald, S., Seal, R., Tweedie, S., et al. (2019). Discovery of high-confidence human protein-coding genes and exons by whole-genome PhyloCSF helps elucidate 118 GWAS loci. Genome Res. *29*, 2073–2087.
42. Gertz, E.M., Yu, Y.-K., Agarwala, R., Schäffer, A.A., and Altschul, S.F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol. *4*, 41.
43. Fesenko, I., Shabalina, S.A., Mamaeva, A., Knyazev, A., Glushkevich, A., Lyapina, I., Ziganshin, R., Kovalchuk, S., Kharlampieva, D., Lazarev, V., et al. (2021). A vast pool of lineage-specific microproteins encoded by long non-coding RNAs in plants. Nucleic Acids Res. *49*, 10328–10346.
44. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. Science *347*, 1260419.
45. Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

46. Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics *11*. O111.016717.

47. Navarro, P., Kuharev, J., Gillet, L.C., Bernhardt, O.M., MacLean, B., Röst, H.L., Tate, S.A., Tsou, C.-C., Reiter, L., Distler, U., et al. (2016). A multicenter study benchmarks software tools for label-free proteome quantification. Nat. Biotechnol. *34*, 1130–1136.

48. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol. Syst. Biol. *14*, e8126.

49. Pino, L.K., Just, S.C., MacCoss, M.J., and Searle, B.C. (2020). Acquiring and analyzing data independent acquisition proteomics experiments without spectrum libraries. Mol. Cell. Proteomics *19*, 1088–1103.

50. Searle, B.C., Pino, L.K., Egertson, J.D., Ting, Y.S., Lawrence, R.T., MacLean, B.X., Villén, J., and MacCoss, M.J. (2018). Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. Nat. Commun. *9*, 5128.

51. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968.

52. Orava, J., Nuutila, P., Lidell, M.E., Oikonen, V., Noponen, T., Viljanen, T., Scheinin, M., Taittonen, M., Niemi, T., Enerbäck, S., and Virtanen, K.A. (2011). Different metabolic responses of human brown adipose tissue to activation by cold and insulin. Cell Metab. *14*, 272–279.

53. Shimizu, Y., Nikami, H., Tsukazaki, K., Machado, U.F., Yano, H., Seino, Y., and Saito, M. (1993). Increased expression of glucose transporter GLUT-4 in brown adipose tissue of fasted rats after cold exposure. Am. J. Physiol. *264*, E890–E895.

54. Bartelt, A., John, C., Schaltenberg, N., Berbée, J.F.P., Worthmann, A., Cherradi, M.L., Schlein, C., Piepenburg, J., Boon, M.R., Rinninger, F., et al. (2017). Thermogenic adipocytes promote HDL turnover and reverse cholesterol transport. Nat. Commun. *8*, 15010.

55. Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. Gene *299*, 1–34.

56. Huang, W., Queen, N.J., McMurphy, T.B., Ali, S., and Cao, L. (2019). Adipose PTEN regulates adult adipose tissue homeostasis and redistribution via a PTEN-leptin-sympathetic loop. Mol. Metab. *30*, 48–60.

57. Huang, N., Li, F., Zhang, M., Zhou, H., Chen, Z., Ma, X., Yang, L., Wu, X., Zhong, J., Xiao, F., et al. (2021). An upstream open reading frame in phosphatase and tensin homolog encodes a circuit breaker of lactate metabolism. Cell Metab. *33* 454–144.e9.

58. Lee, C.-H., Olson, P., Hevener, A., Mehl, I., Chong, L.-W., Olefsky, J.M., Gonzalez, F.J., Ham, J., Kang, H., Peters, J.M., and Evans, R.M. (2006). PPARδ regulates glucose metabolism and insulin sensitivity. Proc. Natl. Acad. Sci. USA *103*, 3444–3449.

59. Alvarez-Dominguez, J.R., Bai, Z., Xu, D., Yuan, B., Lo, K.A., Yoon, M.J., Lim, Y.C., Knoll, M., Slavov, N., Chen, S., et al. (2015). De novo reconstruction of adipose tissue transcriptomes reveals long non-coding RNA regulators of brown adipocyte development. Cell Metab. *21*, 764–776.

60. Contreras-López, O., Moyano, T.C., Soto, D.C., and Gutiérrez, R.A. (2018). Step-by-step construction of gene co-expression networks from high-throughput Arabidopsis RNA sequencing data. Methods Mol. Biol. *1761*, 275–301.

61. Upadhyay, G. (2019). Emerging role of lymphocyte antigen-6 family of genes in cancer and immune cells. Front. Immunol. *10*, 819.

62. Jackson, R., Kroehling, L., Khitun, A., Bailis, W., Jarret, A., York, A.G., Khan, O.M., Brewer, J.R., Skadow, M.H., Duizer, C., et al. (2018). The translation of non-canonical open reading frames controls mucosal immunity. Nature *564*, 434–438.

63. Yang, X., Bam, M., Becker, W., Nagarkatti, P.S., and Nagarkatti, M. (2020). Long noncoding RNA AW112010 promotes the differentiation of inflammatory T cells by suppressing IL-10 expression through histone demethylation. J. Immunol. *205*, 987–993.

64. Niu, L., Geyer, P.E., Wewer Albrechtsen, N.J., Gluud, L.L., Santos, A., Doll, S., Treit, P.V., Holst, J.J., Knop, F.K., Vilsbøll, T., et al. (2019). Plasma proteome profiling discovers novel proteins associated with non-alcoholic fatty liver disease. Mol. Syst. Biol. *15*, e8793.

65. Rabin, S.J., Cleghon, V., and Kaplan, D.R. (1993). SNT, a differentiation-specific target of neurotrophic factor-induced tyrosine kinase activity in neurons and PC12 cells. Mol. Cell Biol. *13*, 2203–2213.

66. Xu, H., Lee, K.W., and Goldfarb, M. (1998). Novel recognition motif on fibroblast growth factor receptor mediates direct association and activation of SNT adapter proteins. J. Biol. Chem. *273*, 17987–17990.

67. Nies, V.J.M., Sancar, G., Liu, W., van Zutphen, T., Struik, D., Yu, R.T., Atkins, A.R., Evans, R.M., Jonker, J.W., and Downes, M.R. (2015). Fibroblast growth factor signaling in metabolic regulation. Front. Endocrinol. *6*, 193.

68. Hansen, J.E., Lund, O., Tolstrup, N., Gooley, A.A., Williams, K.L., and Brunak, S. (1998). NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj. J. *15*, 115–130.

69. Sugino, K., Clark, E., Schulmann, A., Shima, Y., Wang, L., Hunt, D.L., Hooks, B.M., Tränkner, D., Chandrashekar, J., Picard, S., et al. (2019). Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain. Elife *8*, e38619.

70. Iwakoshi-Ukena, E., Shikano, K., Kondo, K., Taniuchi, S., Furumitsu, M., Ochi, Y., Sasaki, T., Okamoto, S., Bentley, G.E., Kriegsfeld, L.J., et al. (2017). Neurosecretory protein GL stimulates food intake, de novo lipogenesis, and onset of obesity. eLife *6*, e28527.

71. Ukena, K., Iwakoshi-Ukena, E., Taniuchi, S., Bessho, Y., Maejima, S., Masuda, K., Shikano, K., Kondo, K., Furumitsu, M., and Tachibana, T. (2014). Identification of a cDNA encoding a novel small secretory protein, neurosecretory protein GL, in the chicken hypothalamic infundibulum. Biochem. Biophys. Res. Commun. *446*, 298–303.

72. Matsuura, D., Shikano, K., Saito, T., Iwakoshi-Ukena, E., Furumitsu, M., Ochi, Y., Sato, M., Bentley, G.E., Kriegsfeld, L.J., and Ukena, K. (2017). Neurosecretory protein GL, a hypothalamic small secretory protein, participates in energy homeostasis in male mice. Endocrinology *158*, 1120–1129.

73. Kato, M., Iwakoshi-Ukena, E., Furumitsu, M., and Ukena, K. (2021). A novel hypothalamic factor, neurosecretory protein GM, causes fat deposition in chicks. Front. Physiol. *12*, 747473.

74. Sallee, N.A., Lee, E., Leffert, A., Ramirez, S., Brace, A.D., Halenbeck, R., Kavanaugh, W.M., and Sullivan, K.M.C. (2020). A pilot screen of a novel peptide hormone library identified candidate GPR83 ligands. SLAS Discov. *25*, 1047–1063.

75. Campbell, J.N., Macosko, E.Z., Fenselau, H., Pers, T.H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A.M.J., Resch, J.M., McCarroll, S.A., et al. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. Nat. Neurosci. *20*, 484–496.

76. Ahima, R.S., and Flier, J.S. (2000). Leptin. Annu. Rev. Physiol. *62*, 413–437.

77. Dumesic, P.A., Egan, D.F., Gut, P., Tran, M.T., Parisi, A., Chatterjee, N., Jedrychowski, M., Paschini, M., Kazak, L., Wilensky, S.E., et al. (2019). An evolutionarily conserved uORF regulates PGC1α and oxidative metabolism in mice, flies, and bluefin tuna. Cell Metab. *30*, 190–200.e6.

78. Prensner, J.R., Enache, O.M., Luria, V., Krug, K., Clauser, K.R., Dempster, J.M., Karger, A., Wang, L., Stumbraite, K., Wang, V.M., et al. (2021). Noncanonical open reading frames encode functional proteins essential for cancer cell survival. Nat. Biotechnol. *39*, 697–704.

79. Chen, Y.-J., Tan, B.C.-M., Cheng, Y.-Y., Chen, J.-S., and Lee, S.-C. (2010). Differential regulation of CHOP translation by phosphorylated eIF4E under stress conditions. Nucleic Acids Res. *38*, 764–777.

80. Lee, Y.-Y., Cevallos, R.C., and Jan, E. (2009). An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2α phosphorylation. J. Biol. Chem. *284*, 6661–6673.

81. Jayaram, D.R., Frost, S., Argov, C., Liju, V.B., Anto, N.P., Muraleedharan, A., Ben-Ari, A., Sinay, R., Smoly, I., Novoplansky, O., et al. (2021). Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of PKCs. Proc. Natl. Acad. Sci. USA *118*. e2018899118.

82. Lee, C.Q.E., Kerouanton, B., Chothani, S., Zhang, S., Chen, Y., Mantri, C.K., Hock, D.H., Lim, R., Nadkarni, R., Huynh, V.T., et al. (2021). Coding and non-coding roles of MOCCI (C15ORF48) coordinate to regulate host inflammation and immunity. Nat. Commun. *12*, 2130.

83. Niu, L., Lou, F., Sun, Y., Sun, L., Cai, X., Liu, Z., Zhou, H., Wang, H., Wang, Z., Bai, J., et al. (2020). A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation. Sci. Adv. *6*, eaaz2059.

84. Bogaert, A., Fijalkowska, D., Staes, A., Van de Steene, T., Demol, H., and Gevaert, K. (2022). Limited evidence for protein products of noncoding transcripts in the HEK293T cellular cytosol. Mol. Cell. Proteomics *21*, 100264.

85. Deutsch, E.W., Lane, L., Overall, C.M., Bandeira, N., Baker, M.S., Pineau, C., Moritz, R.L., Corrales, F., Orchard, S., Van Eyk, J.E., et al. (2019). Human proteome project mass spectrometry data interpretation guidelines 3.0. J. Proteome Res. *18*, 4108–4116.

86. Ma, J., Diedrich, J.K., Jungreis, I., Donaldson, C., Vaughan, J., Kellis, M., Yates, J.R., and Saghatelian, A. (2016). Improved identification and analysis of small open reading frame encoded polypeptides. Anal. Chem. *88*, 3967–3975.

87. Bookout, A.L., Jeong, Y., Downes, M., Yu, R.T., Evans, R.M., and Mangelsdorf, D.J. (2006). Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. Cell *126*, 789–799.

88. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

89. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

90. Ma, J., Saghatelian, A., and Shokhirev, M.N. (2018). The influence of transcript assembly on the proteogenomics discovery of microproteins. PLoS One *13*, e0194518.

91. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

92. Lin, M.F., Jungreis, I., and Kellis, M. (2011). PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics *27*, i275–i282.

93. Wingett, S.W., and Andrews, S. (2018). FastQ Screen: a tool for multi-genome mapping and quality control. F1000Res. *7*, 1338.

94. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

95. Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., and Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. *47*, W199–W205.

96. Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. *30*, 918–920.

97. Eng, J.K., Jahan, T.A., and Hoopmann, M.R. (2013). Comet: an open-source MS/MS sequence database search tool. Proteomics *13*, 22–24.

98. Deutsch, E.W., Mendoza, L., Shteynberg, D., Slagel, J., Sun, Z., and Moritz, R.L. (2015). Trans-proteomic pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteomics. Clin. Appl. *9*, 745–754.

99. Gessulat, S., Schmidt, T., Zolg, D.P., Samaras, P., Schnatbaum, K., Zerweck, J., Knaute, T., Rechenberger, J., Delanghe, B., Huhmer, A., et al. (2019). Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. Nat. Methods *16*, 509–518.

100. Searle, B.C., Swearingen, K.E., Barnes, C.A., Schmidt, T., Gessulat, S., Küster, B., and Wilhelm, M. (2020). Generating high quality libraries for DIA MS with empirically corrected peptide predictions. Nat. Commun. *11*, 1548.

101. Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat. Methods *4*, 923–925.

102. The, M., MacCoss, M.J., Noble, W.S., and Käll, L. (2016). Fast and accurate protein false discovery rates on large-scale proteomics data sets with percolator 3.0. J. Am. Soc. Mass Spectrom. *27*, 1719–1727.

103. Dalbøge, L.S., Jacobsen, J.M., Mehrotra, S., Mercer, A.J., Cox, N., Liu, F., Bennett, C.M., Said, M., Tang-Christensen, M., Raun, K., et al. (2021). Evaluation of VGF peptides as potential anti-obesity candidates in pre-clinical animal models. Peptides *136*, 170444.

104. McGlincy, N.J., and Ingolia, N.T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. Methods *126*, 112–129.

105. Bairoch, A., and Apweiler, R. (1997). The SWISS-PROT protein sequence data bank and its supplement TrEMBL. Nucleic Acids Res. *25*, 31–36.

106. Revelle, W.R. (2017). psych: procedures for personality and psychological research. https://cran.r-project.org/web/packages/psych/index.html.

107. Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. Int. J. Complex Syst. *1695*, 1–9.

108. Yoon, J., Blumer, A., and Lee, K. (2006). An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. Bioinformatics *22*, 3106–3108.

109. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

110. Pearson, W.R. (2013). An introduction to sequence similarity ("homology") searching. Curr. Protoc. Bioinform. *Chapter 3*, Unit3.1.

111. Zhang, H., Gao, S., Lercher, M.J., Hu, S., and Chen, W.-H. (2012). EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. Nucleic Acids Res. *40*, W569–W572.

112. Amodei, D., Egertson, J., MacLean, B.X., Johnson, R., Merrihew, G.E., Keller, A., Marsh, D., Vitek, O., Mallick, P., and MacCoss, M.J. (2019). Improving precursor selectivity in data-independent acquisition using overlapping windows. J. Am. Soc. Mass Spectrom. *30*, 669–684.

113. Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. *74*, 5383–5392.

114. Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods *4*, 207–214.

115. Bookout, A.L., Cummins, C.L., Mangelsdorf, D.J., Pesola, J.M., and Kramer, M.F. (2006). High-throughput real-time quantitative reverse transcription PCR. Curr. Protoc. Mol. Biol. *Chapter 15*. Unit 15.8.

116. Hultman, K., Scarlett, J.M., Baquero, A.F., Cornea, A., Zhang, Y., Salinas, C.B.G., Brown, J., Morton, G.J., Whalen, E.J., Grove, K.L., et al. (2019). The central fibroblast growth factor receptor/beta klotho system: comprehensive mapping in mus musculus and comparisons to non-human primate and human samples using an automated in situ hybridization platform. J. Comp. Neurol. *527*, 2069–2085.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Healthy C57BL/6J female mouse brown and subcutaneous white adipose tissue (7 week old)	The Jackson Laboratory	JAX: 000664
Diet-induced obese 27-week old C57BL/6J male mouse liver, brown adipose tissue (BAT), epididymal white adipose tissue ("eWAT"), subcutaneous white adipose tissue (scWAT), retroperitoneal fat ("Retro Fat") and mesenteric fat ("Mesen Fat") fed high-fat diet or 21 weeks along with age matched counterparts fed chow	The Jackson Laboratory	JAX: 000664
Aged (41-weeks) diet-induced obese (fed *ad lib* 45% HFD from age 10 weeks) C57BL/6J male mouse plasma	The Jackson Laboratory	JAX: 000664
Aged (41-weeks) lean C57BL/6J male mouse plasma	The Jackson Laboratory	JAX: 000664
Young (26-week) diet-induced obese (fed *ad lib* 45% HFD from age 10 weeks) C57BL/6J male mouse plasma	The Jackson Laboratory	JAX: 000664
Young (26-week) lean mouse plasma	The Jackson Laboratory	JAX: 000664
Male and female C57BL/6J mouse tissues RNA extract bank for qPCR	The Jackson Laboratory	JAX: 000664
Chemicals, peptides, and recombinant proteins		
DMEM	ThermoFisher/Gibco	Cat# 10566016
HBSS	ThermoFisher/Gibco	Cat# 14025134
Collagenase Type I	ThermoFisher/Gibco	Cat# 17100017
Collagenase Type II	ThermoFisher/Gibco	Cat# 17101015
BSA	Millipore Sigma	Cat# A9418
Newborn calf serum	ThermoFisher/Gibco	Cat# 16010167
Glutamax	ThermoFisher/Gibco	Cat# 35050079
Sodium ascorbate	Millipore Sigma	Cat# A7631
Insulin	ThermoFisher/Gibco	Cat# 12585014
Rosiglitazone	Millipore Sigma	Cat# R2408
T3	Millipore Sigma	Cat# T6397
Cycloheximide	Fisher Scientific	Cat# AAJ66004X
Trizol	ThermoFisher	Cat# 15596026
Triton X-100	ThermoFisher	Cat# A16046.AP
Turbo DNase	ThermoFisher	Cat# AM2238
RNase I	Lucigen	Cat# N6901K
Superase-In RNase I inhibitor	ThermoFisher	Cat# AM2694
Episcript RT	Lucigen	Cat# ERT12910K
Exonuclease I	Lucigen	Cat# X40520K
Hybridase	Lucigen	Cat# H39500
CircLigase I	Lucigen	Cat# CL4111K
Phusion Hot Start II High-Fidelity Master Mix	ThermoFisher	Cat# F565L
Urea	Sigma-Aldrich	Cat# U5128
ammonium bicarbonate	Sigma-Aldrich	Cat# A6141
DMEM, high glucose, no glutamine, no phenol red	ThermoFisher/Gibco	Cat# 31053028
SDS	Sigma-Aldrich	Cat# 436143
Triethylammonium bicarbonate buffer	Sigma-Aldrich	Cat# T7408
MgCl2	Sigma-Aldrich	Cat# M8266

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HALT protease & phosphatase inhibitors	ThermoFisher	Cat# 78440
Tris(2-carboxyethyl)phosphine hydrochloride	Sigma-Aldrich	Cat# C4706
Iodoacetamide	Sigma-Aldrich	Cat# I1149
porcine trypsin	Promega	Cat# V5113
Pierce retention time calibration (PRTC) peptides	Pierce	Cat# 88321
Rodent diet with 60 kcal% from fat	Research Diets	Cat# D12492i
Rodent diet with 10 kcal% from fat	Research Diets	Cat# D12450Bi
PicoLab Rodent Diet 20	LabDiet	Cat# 5053
Rodent diet with 45 kcal% from fat	Research Diets	Cat# D12451i
PPS silent surfactant	Agilent	Cat# 400500
8M guanidine-Cl	Pierce	Cat# 24115
Trifluoracetic acid	Pierce	Cat# 28904
β-mercaptoethanol	Fischer Scientific	Cat# PI35602
Triethylamine	Pierce	Cat# 25108
Standard rodent diet	Harlan Teklad	Cat# 7001
RNAStat60	Tel-Test, Inc.	Cat# RNA STAT-60
RNAscope rodent-specific probe for mouse Gm8773 mRNA	Advanced Cell Diagnostics	Cat# 542179
RNAscope RED kit	Advanced Cell Diagnostics	Cat# 322350
Expi293 medium	ThermoFisher/Gibco	Cat# A14351-01
Expifectamine 293 Transfection Reagent	ThermoFisher/Gibco	Cat# A14524
SialExo enzyme	Genovis	Cat# G1-SM1-020
gel filtration standards	BioRad	Cat# 1511901
Critical commercial assays		
Ribo-Zero Mammalian Kit	Illumina	Cat# 20040526
TruSeq Ribo Profile (Mammalian) Library Prep Kit index primers 1–12.	Illumina	Cat# RPYSC12116
BCA Protein Assay Kit	ThermoFisher/Pierce	Cat# 23225
RNeasy Mini Kit	Qiagen	Cat# 74004
TruSeq stranded mRNA kit	Illumina	Cat# 20020594
Deposited data		
GEO Superseries ID of all raw and analyzed data	This paper	GEO: GSE198109
Raw and analyzed Ribo-Seq data	This paper	GEO: GSE197909
Raw and analyzed mRNA-Seq data of primary metabolic cells	This paper	GEO: GSE198107
Tissue mRNA-Seq data of DIO and lean mice	https://doi.org/10.1101/2021.06.25.449953	GEO: GSE185466
Mouse Primary brown, beige, and white whole cell lysate proteome data (released to ProteomeXchange upon publication)	This paper	MassIVE: MSV000089022
Mouse Primary brown, beige, and white secretome proteome data (released to ProteomeXchange upon publication)	This paper	MassIVE: MSV000089023
Mouse plasma proteome data from DIO/lean young/old mice (released to ProteomeXchange upon publication)	This paper	MassIVE: MSV000089021
Data S1 – Source Data.zip	This paper	N/A
Experimental models: Organisms/strains		
Mouse: C57Bl6/J 7 week old female	The Jackson Laboratory	JAX: 000664
Mouse: C57Bl6/J 7 week old male	The Jackson Laboratory	JAX: 000664
Oligonucleotides		
Ribo-Seq Library Construction - 3′ adapter – 5'-/5phos/AGATCGGAAGAGCACACGTCTGAA/3ddC/-3′	N/A	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ribo-Seq Library Construction - RT primer – 5'-/5Phos/AGATCGGAAGAGCGTCGTGTAGGGAAAGAG/iSp18/GTGACTGGAGTTCAGACGTGTGCTC-3′	Martinez et al.[4]	N/A
Ribo-Seq Library Construction - PCR forward primer – 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC-3′	Martinez et al.[4]	N/A
18s General SYBR Green qPCR primer – 5′–ACCGCAGCTAGGAATAATGGA-3′	Bookout et al.[87]	N/A
18s General SYBR Green qPCR primer – 5′–GCCTCAGTTCCGAAAACCA-3′	Bookout et al.[87]	N/A
Gm8773 (NR_033499) SYBR Green qPCR primer – 5′-GCGTGGCCACCCACTCT	This paper	N/A
Gm8773 (NR_033499) SYBR Green qPCR primer – 5' – GCAGGACCTCGCTCCTTTTC-3′	This paper	N/A
Software and algorithms		
STAR v2.53b	Dobin et al.[88]	https://github.com/alexdobin/STAR
Cufflinks	Trapnell et al.[89]	https://github.com/cole-trapnell-labcufflinks
Stringtie v2.1.4	Trapnell et al.[89]	https://github.com/skovaka/stringtie2
MAPS v1.0	Ma et al.[90]	https://bitbucket.org/shokhirev/maps/src/master/
GTFtoFasta	Martinez et al.[4]	N/A
RibORF	Ji et al.[28]	https://github.com/zhejilab/RibORF
HOMER	Heinz et al.[91]	http://homer.ucsd.edu/homer/
PhyloCSF	Lin et al.[92]	https://github.com/mlin/PhyloCSF
fastqc	Wingett et al.[93]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
DESeq2	Love et al.[94]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
WebGestaltR	Liao et al.[95]	https://github.com/bzhanglab/WebGestaltR
ProteoWizard version 3.0.9974	Chambers et al.[96]	https://proteowizard.sourceforge.io/
Comet (version 2018.01 rev. 0)	Eng et al.[97]	http://comet-ms.sourceforge.net/
Trans-Proteomic Pipeline (TPP version 5.1.0)	Deutsch et al.[98]	https://sourceforge.net/projects/sashimi/files/Trans-Proteomic%20Pipeline%20%28TPP%29/
Skyline (version 21.1.0.146)	MacLean et al.[51]	https://skyline.ms/project/home/software/Skyline/begin.view
Prosit	Gessulat et al.[99]	https://github.com/kusterlab/prosit/
EncyclopeDIA (version 0.9.0)	[50,100]	https://bitbucket.org/searleb/encyclopedia/wiki/Home
Percolator 3.1	Käll et al. and The et al.[101,102]	https://github.com/percolator
Other		
100 μm cell strainer	Corning	Cat# 431752
Micro-Spin S-400 HR columns	Millipore Sigma	Cat# GE27-5140-01
S-trap	Protifi	N/A
Li Heparin Microtainer tubes	Fisher Scientific	Cat# 365965
MCX columns in 96-well format	Waters	Cat# 186001830BA
BondElut C18 SPE cartridges	Agilent/Varian	Cat# 12102028
High pH Reversed-Phase Peptide Fractionation Kit	ThermoFisher	Cat# 84868
1.9 μm ReproSil-Pur C18 silica beads	Dr. Maisch	Cat# r119.b9
75 μm inner diameter fused silica capillary, Self-Pack PicoFrit	New Objective	Cat# PF360

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Ambion Turbo DNA-free kit (Austin, TX	Fisher Scientific	Cat# AM1907
Sigma GenElute Endo-free maxi prep kits (prod# NA0410-1KT	Millipore Sigma	Cat# NA0410-1KT
ExpiFectamine 293 Transfection Kit	ThermoFisher	Cat# A14525
Superdex 75	Millipore Sigma	Cat# GE17-5174-01
eStain L1 Protein Staining System	Genscript	Cat# L00753
Bolt 4–12% Bis-Tris gels	ThermoFisher	Cat# NW04122BOX
Zorbax 300-Diphenyl column (1.8 μm, C8)	Agilent	Cat# 863750-944
Waters BEH column, 1.7 μm, 200 Å, 4.6 mm ID x 150 mm L	Waters	Cat# 186005225
26-gauge-guided cannula	PlasticsOne	Cat# C315GS-4/SPC
G-ænial Bond and G-ænial Universal Flo: G-Bond Unit Dose Kit, G-ænial Universal Flo B1 Refill	GC America	Cat# 002302, Cat# 004207
Vetbond	3M	Cat# 70200742529

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Christopher A. Barnes (chrisb@rnes.pro)

Materials availability

This study did not generate new unique reagents in sufficient quantities to share broadly.

Data and code availability

- RNA-seq and Ribo-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Proteomics data have been deposited at MassIVE and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. The raw tabular data used to create figure panels throughout this work is compiled in Data S1- Source Data.zip
- This paper does not report original code
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All of the animal experiments with the exclusion of the tissue bank RNA extracts that were used for qPCR of Gm8773 were performed in accordance with internationally accepted principles for the use of laboratory animals and were approved by the Novo Nordisk Research Center Seattle Institutional Animal Care and Use Committee and the Novo Nordisk Ethical Review Committee. All functional animal studies were performed on male mice.

The mouse tissue bank RNA extracts that were used for qPCR of Gm8773 are detailed below. All protocols were approved by the University of Texas Southwestern Medical Center Institutional Animal Care and Use Committee.

Creation of diet-induced obese (DIO) mouse tissue atlas for RNA-Seq analysis

30 Male, C57Bl/6J mice were obtained from The Jackson Laboratory (Stock #000664 Bar Harbor, ME, USA) and were assigned to high fat (DIO) or low fat (Control) diet groups at 6 weeks of age. The diets were 60% energy from fat (D12492i, Research Diets; N = 15) and 10% energy from fat (D12450Bi, Research Diets; N = 15). Sanitized water and the test diets were available *ad libitum*. Animals were transported to Novo Nordisk Research Center Seattle at 21 weeks of age and were maintained on the same test diets for an additional 6 weeks in a light and temperature-controlled room (12/12 light-dark cycle, 22 ± 2°C). At 27 weeks of age the group mean body weights were 47.8 g +/− 2.8 g (DIO) and 31.2 g +/− 1.3 g (Control). Animals were euthanized via CO2 inhalation approximately 3–8 h into the start of the light period; with food removal 2 h prior to euthanasia. The order of euthanasia and tissue collection was alternated between groups to control for bias. Adipose and liver tissues (and other tissues not reported here) were removed rapidly, frozen in liquid nitrogen and stored at −80°C until processed for RNA. All animal experiments were performed in accordance with internationally accepted principles for the use of laboratory animals and were approved by the Novo Nordisk Research Center Seattle Institutional Animal Care and Use Committee and the Novo Nordisk Ethical Review Committee.

Creation of DIO/lean young/old mouse plasma samples

Male C57/BL6J mice were purchased from Jackson and grouped housed in a 12 h light/dark cycle with lights on at 6 a.m. and with *ad libitum* access to food and water throughout the experiment. The diet-induced obese (DIO) groups represented 45% of the original animals in the cohort that were kept on chow (PicoLab Rodent Diet 20) diet for their first 4 weeks after arrival and subsequently switched to high fat diet (45% HFD Research Diets: D12451i) at week 10 with these DIO animals remaining on the HFD for the duration of the experiment. The chow (PicoLab Rodent Diet 20) fed animals remained on the same diet throughout the entire experiment. Plasma was collected from both DIO and lean (chow) animal groups that were harvested at either 26- or 41-weeks of age representing the "young" and "old" groups, respectively. Each group represents 12 animals (n = 12). For plasma collections, animals were euthanized with asphyxiation with CO_2 followed by cardiac puncture with individual blood collections into Li Heparin tubes (BD Microtainer tubes cat# 365965)). 120μL of plasma was separated for each sample by centrifugation at 10,000 RCF for 4 min 4°C followed by transferring the plasma supernatant into fresh tubes and immediately snap freezing the aliquots in liquid nitrogen.

Intracerebroventricular (ICV) administration of Gm8773 protein product for assessing food intake and body weight effects in the diet-induced obese mouse

ICV experiments were performed as described previously.[103] Briefly, a permanent 26-gauge-guided cannula (C315GS-4/SPC, PlasticsOne, Roanoke, VA, USA) projecting to the lateral ventricle was implanted into DIO mice which were male C57BL/6J mice purchased from Jackson Laboratories with obesity induced in-house via continuous feeding with high fat diet (45% HFD Research Diets: D12451i) from age 10 weeks on. DIO mice were cannulated for ICV injections at age 51 weeks. The surgeries were performed as follows: Isoflurane anesthesia was administered to Mice who were placed in a stereotaxic instrument that was maintained at 36–37°C throughout surgery on a heating pad. Continuous isoflurane was administered using a specialized mask. A midline cranial incision was then made and the skull was exposed and cleaned. A single hole (1 mm in diameter) was drilled in the skull in the lower left quadrant relative to bregma. After the horizontal skull position was confirmed, the cannula was placed according to bregma (−0.7 mm posterior, −1.2 mm lateral [left], and −2.0 mm ventral). G-ænial Bond and G-ænial Universal Flo (GC America, Inc., Alsip, IL, USA, G-Bond Unit Dose Kit Cat# 002,302 and G-ænial Universal Flo B1 Refill Cat# 004,207) were used to fix the cannula in place which was cured with LED light. Vetbond (3 M, St. Paul, MN, USA) was used to cloes the surgical incision. Gm8773 was injected (2 μL) using a Hamilton syringe connected to an infusion cannula with a 0.5 mm projection. Cannula placement was confirmed using NPY at a dose of 0.3 nmol/mouse (via 2 μL injections of 2.5 μg/μL NPY solution) injected in the early light phase. A cumulative food intake of ≥0.5 g in 3 h following ICV NPY injection indicated successful cannula placement.

METHOD DETAILS

Isolation and differentiation of primary mouse white, beige, and brown adipocytes for both ribosome profiling and mass spectrometry

Twenty female C57Bl6/J mice aged 7 weeks old were sacrificed by CO_2 asphyxiation followed by cervical dislocation. Interscapular and subscapular BAT was excised with dissecting scissors, being careful to avoid surrounding muscle and white adipose tissue. For white adipose tissue, subcutaneous (scWAT) depots were excised bilaterally. Fat pads were placed collectively in a 50 mL falcon containing ice-cold DMEM. DMEM was carefully decanted through a mesh filter, and the pads washed briefly in HBSS. Most of the HBSS was then decanted, before transferring each tube of fat pads to a single well of a 6-well culture plate. Dissecting scissors were then used to thoroughly mince each collection of pooled fat explants. Digests were performed in the following solutions (pre-warmed to 37°C): 1) for scWAT, 40 mg Collagenase Type I with 80 mg Collagenase Type II in 28 mL HBSS and 12 mL 7.5% w/v BSA (Sigma); 2) for BAT, 80 mg Collagenase Type I with 80 mg Collagenase Type II in 28 mL HBSS and 12 mL 7.5% w/v BSA (Sigma). Digests were incubated at 37°C for 15–20 min with shaking (225 rpm), resulting in complete dissociation of adipose tissue. Digests were then filtered through a 100 μm mesh (yellow, Corning) into 50 mL falcon tubes and allowed to sit on ice for 20 min. Visible fat floating as a layer was carefully aspirated. Working from the top, the supernatant was pipetted into a clean 50 mL Falcon tube, leaving behind the lowest 10 mL of digested material. This fraction is enriched for pre-adipocytes and was mixed 1:1 with 10% newborn calf serum (Gibco). Cells were spun down at 700 g for 15 min and resuspended in growth media consisting of: 1) for scWAT ("scWAT media"), DMEM plus 10% NCS with Pen/Strep, 4 mM Glutamax (Gibco), 150 μM sodium ascorbate, 30 nM insulin, and 1 μM rosiglitazone (Millipore Sigma, R2408); and 2) for BAT ("BAT media") DMEM plus 10% NCS with Pen/Strep, 4 mM Glutamax (Gibco), 150 μM sodium ascorbate, 4 nM insulin, and 2 nM T3 (Millipore Sigma, T6397). Cells were spun again for 10 min and resuspended ready for plating in the following volumes (based on 2 mL of media for each well seeded): 1) for scWAT, a ratio of 5 mice in 38 mL of plating media with suspended cells, and 2) for BAT, a ratio of 5 mice in 25 mL of plating media with suspended cells. Cells were plated into 6-well dishes with 2 mL of cell suspension per well. Cells cultured as the "Beige" phenotype represent the scWAT cultures that were switched to the BAT media 24 h after plating in the scWAT media. Media was changed daily for days 1–4 and every other day from days 4–8. On day 6, media was changed to phenol red free DMEM and rosiglitazone was removed from the scWAT cultures to allow wash out before harvest. All cultures were treated or harvested or both on day 8.

Ribo-Seq and total RNA-Seq sample preparation from primary mouse white, beige, and brown adipocytes

For sample collection of BAT/WAT/Beige cultures for Ribo-Seq and total mRNA-Seq used for smORF discovery, each culture was washed directly out of the incubators 2x with ice-cold PBS supplemented with 100 μg/mL cycloheximide (CHX; Fisher Scientific,

AAJ66004X). After the last wash, liquid nitrogen was gently ladled onto the surfaces of the cells and plates were stored at −80°C prior to ribosome footprinting and preparation of sequencing libraries. Each biological replicate for Ribo-Seq in our experiments represents 11 wells of a 6-well dish leaving one well to be lysed with Trizol reagent (Thermo #15596026) for the bulk mRNA-Seq used int the *de novo* transcriptome assembly of the cells under analysis. For the Ribo-Seq analysis, two separate primary isolations along with their preps for both RNA-Seq and Ribo-Seq were performed for the discovery work. Cells were lysed with 400 μL of ice-cold lysis buffer (20 mM Tris–HCl, pH 7.4, 150 mM NaCl, 5 mM MgCl2, 1% Triton X-100, with 1 mM DTT, 25 U ml−1 Turbo DNase (Thermo Fisher, catalog no. AM2238) and 100 μg mL−1 CHX added fresh) was dripped onto the plate. Cells were incubated on ice in lysis buffer for 10 min with periodic vortexing and pipetting to disperse the cells. The lysate was then clarified by centrifugation at 15,000 g for 10 min. Cell lysates were flash-frozen in liquid nitrogen and stored at −80°C for up to 7 days before ribosome footprinting. For each cell type, ribosome footprinting was carried out by digesting 40–60 μg of RNA in 200–300 μL lysate with 0.375 U/μg−1 RNase I (Lucigen, N6901K) for 50 min at room temperature. Digestion reactions were quenched with 200 U Superase-In RNase I inhibitor (Thermo Fisher, AM2694) on ice. Following digestion, monosomes were purified from small RNA fragments using Micro-Spin S-400 HR columns (GE Life Sciences), and ribosome protected RNA fragments (RPFs) were extracted by acid phenol chloroform and isopropanol precipitation. Sequencing libraries were prepared as in McGlincy and Ingolia[104] with some modifications. First, the Ribo-Zero Mammalian Kit (Illumina) was used to deplete rRNA after RPF extraction and just prior to RPF size selection by gel extraction. Second, the Zymo clean & concentrator step after adaptor ligation is omitted and the reaction was carried over straight into reverse transcription. For the reverse transcription step to form cDNA, Episcript RT (Lucigen, ERT12910K) was used. Following reverse transcription, excess primer was degraded using Exonuclease I (Lucigen, X40520K) and the RNA templates were degraded using Hybridase (Lucigen, H39500). For the cDNA circularization step, CircLigase I (Lucigen, CL4111K) was used. PCR amplification was then carried out using Phusion Hot Start II High-Fidelity Master Mix (Thermo Fisher, F565L) for 9–12 cycles. The adapters and primers for library construction used were as follows: 3′ adapter – 5'-/5phos/AGATCGGAAGAGCACACGTCTGAA/3ddC/-3'; RT primer – 5'-/5Phos/AGATCGGAAGAGCGTCGTGTAGGGAAAGAG/iSp18/GTGACTGGAGTTCAGACGTGTGCTC-3'; PCR forward primer – 5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC-3′; Illumina TruSeq Ribo Profile (Mammalian) Library Prep Kit index primers 1–12.

Ribo-Seq and bioinformatics analysis

We followed the methodology described in Martinez et al.[4] to generate custom open reading frame databases for microprotein discovery, with some key modifications. After trimming adapters, removal of mm10 rRNA and tRNA sequences, and alignment to the mm10 genome with STAR v2.53b,[88] instead of Cufflinks,[89] Stringtie v2.1.4[89] and MAPS v1.0[90] were run using default parameters on combined alignments from total and mRNA RNA-Seq libraries from BAT, beige, and WAT tissues, followed by 3-frame translated using a custom script GTFtoFasta (see Martinez et al.[4] for details). The resulting 3-frame translated ORF databases were then scored for translation using RibORF[28] with the pipeline described by Martinez et al.[4] which briefly included shifting reads of each length to obtain base-pair resolution, and then scoring each candidate smORF using RibORF, keeping only highly scoring (score $\geq$0.7), short ($\leq$150 aa), and novel (not overlapped with known coding regions in RefSeq, and a maximum blastp alignment evalue to SwissProt.[105] The resulting microproteins were then collated into a non-redundant table and were further annotated using HOMER[91] (genomic location of their ORFs), and using PhyloCSF[92]; conservation from a multi-way alignment of mammals) as described previously.[4] Candidate microproteins were then split into categories based on the genomic location of their ORF with respect to genomic features: uORFs exist upstream of known genes, non-uORFs are all other ORFs, iORFs specifically exist in intergenic regions, while dORFs exist downstream of known genes. These categories of ORFs were then further tested for expression in RNA-Seq datasets and visualized using read pileup tracks.

Whole cell lysates ("proteomes") and secreted protein ("secretome") sample preparation from primary mouse white, beige, and brown adipocytes

For whole cell proteome analysis, 1 mL of 8M urea (Sigma-Aldrich #U5128) in 100 mM ammonium bicarbonate (Sigma-Aldrich # A6141) (Urea lysis buffer, "ULB") was pipetted into each well of a 6-well dish of BAT-derived "brown" cells and scWAT-derived "white" cells that were differentiated with the protocol above where each well represents a single biological replicate. The proteome and secretome samples were prepared in parallel with the samples used for Ribo-Seq and total mRNA-Seq. Lysates were pipetted up and down to aid in lysis before collection. Additionally, parallel secretome samples were washed 3x with pre-warmed serum free phenol red free DMEM (Gibco # 21063029) (3x) followed by the addition of 500 uL of pre-warmed serum free phenol red free DMEM (Gibco #31053028). The cells were allowed to secrete proteins for 90 min before collection of the secretion media which was collected from each well into separate 2 mL LoBind eppendorf tubes and snap frozen in liquid nitrogen.

Trypsin digestion and desalting of differentiated brown and white adipocyte proteome and secretome samples

For tryptic digestions, protein concentration estimates of each of the lysates were determined using BCA (Pierce #23225) and 100 μg of each sample was digested using a modified protocol suitable for S-trap (Protifi) digestion of lipid laden samples. The S-trap solubilization/lysis buffer is 5% SDS(SDS, Sigma-Aldrich # 436143), 50 mM Triethylammonium bicarbonate buffer (TEAB, Sigma-Aldrich #T7408), 2 mM MgCl2 (Sigma-Aldrich #M8266), 1X HALT protease & phosphatase inhibitors (ThermoFisher # 78440) so each sample from the ULB extractions described above were brought to this final concentration prior to proceeding through the protocol. Once in the S-trap solubilization/lysis buffer, samples were reduced with 10 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP,

Sigma-Aldrich #C4706) for 30 min at room temperature with 600 RPM shaking on a thermomixer followed by alkylation with 40 mM iodoacetamide (Sigma-Aldrich #I1149) in the dark at room temperature with 600 RPM shaking on a thermomixer. 12% aqueous phosphoric acid was added at 1:10 ratio giving a concentration of ~1.2% phosphoric acid to neutralize iodoacetamide. Samples were vortexed and then spun down. S-trap binding buffer consisting of 90% methanol in 100 mM TEAB was added at 6x the volume of the sample, vortexed, then spun down. The acidified lysates with S-trap binding buffer were added to the 96-well S-traps columns and the plates were affixed to a plate vacuum manifold. Repeat additions of the sample were performed until all of each sample had passed through the columns. Columns were then washed with 200 μL S-trap binding buffer (3x). To remove lipids, columns were washed with 150 μL of a 50/50 mixture of chloroform and methanol (3x) followed by one more wash with S-trap binding buffer. The S-trap digestion plate was then moved to a clean LoBind collection plate and 5 μg of porcine trypsin (Promega, #V5113) suspended in 50 mM TEAB was added to each sample. The plates were loosely sealed with Teflon plate covers and incubated for 1 h at 47°C. After digestion, 80 μL of 50 mM TEAB was added to all wells and the plates were centrifuged on top of the new collection plates at 1500 g for 2 min. An additional 80 μL 0.2% aqueous formic acid was added to all wells and centrifuged into the same collection plates at 1500 g for 2 min. Finally, 80 μL of 50% acetonitrile (ACN) containing 0.2% formic acid was added and centrifuged at 1500 g for 2 min for a final concentration of 10% v/v ACN. The elution plates were frozen at −80°C overnight with Teflon gaskets and wrapped in parafilm until holes were poked in the gaskets and the samples were lyophilized to dry. Samples were resuspended in 50 μL water with 0.1% formic acid (buffer A) supplemented with 50 fmol/μL Pierce retention time calibration (PRTC) peptides (Pierce #88321) and the plates were floated in a water bath sonicator for 5 min followed by 5 min on a thermomixer at 600 rpm prior to mass spectrometry analysis.

Bulk mRNA-Seq analysis with next-generation sequencing in DIO/lean tissue

Mice were sacrificed and tissues of interest were extracted and kept at −80°C until RNA extraction. Samples were thawed and homogenized in Trizol with one 5 mm stainless steel ball using a TissueLyser II (Qiagen). RNA was extracted using chloroform and a Qiagen RNeasy Mini Kit (#74004) following the manufacturer's instructions. RNA concentration and quality was measured on a Nanodrop and Bioanalyzer. Samples with a 260/280 ratio of >1.8 and RIN >8 were shipped on dry ice for sequencing at Covance Genomics Lab, Redmond, WA. cDNA libraries were prepared with TruSeq stranded mRNA kit (Illumina #20020594) and sequenced on an Illumina HiSeq 2000 using paired-end, 100 nucleotide reads with an average depth of 20 million reads per sample.

Bioinformatics quantitation of NGS transcriptome data of bulk tissue (adipose depots and liver) mRNA DIO/lean

Raw sequencing data was quality assessed using fastqc (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).[93] Reads were aligned to the mm10 reference genome using STAR v2.53b[88] and reads were quantified using HOMER analyzeRepeats[91] using the top expressed isoform as proxy for gene expression. Differential expression between DIO and lean mice was carried out using DESeq2[94] and genes with FDR <0.05 and log2fold >1 were identified as significantly changed. Gene pathway enrichment analysis was carried out using WebGestaltR[95] in ORA mode, using all genes as the background for testing. Candidate ORFs identified from Ribo-Seq were quantified across ORF exons analogously to annotated genes and differential ORF expression was tested with DESeq2 using a threshold of FDR <0.05 and log2fold >1. Principal component analysis (PCA) was carried out with the prcomp function and plotted in R. Ovals, were manually drawn to highlight specific groups of samples.

Circular genome visualization and co-expression networks

To generate the circular genome plot, we used the Python package Pycircos (v0.3.0). The genome coordinates for each annotated gene present in the reference mm10 GTF file from UCSC were plotted in the outermost ring as red bars. We did the same for the novel smORFs identified with Ribo-Seq, which were plotted in the second ring from border to center as purple bars. To improve visualization, we increased the width of each bar by 1.000.000, as smORFs would not be visible otherwise. We then added a bar corresponding to the presence of a transmembrane domain predicted by TMHMM for each smORF to the third ring. Afterward, we added peaks to the fourth section corresponding to the positive PhyloCSF scores for each smORF. For the fifth section, we added a bar for each smORF that had a hit in the tBLASTn search for humans. To the sixth ring, we added a bar for each smORF whose protein sequence contained a signal peptide domain predicted by either Phobius and/or Signal P 5.0.[39,40] The links in the center correspond to the co-expression between a smORF and another gene. To infer the smORF co-expression, we followed the protocol from Contreras et al. with some modifications.[60] First, we performed a differential expression analysis for all DIO and Lean samples, and selected only the ones that had a padj <0.05 and a Log2FoldChange >1. Then, we used a normalized matrix of the counts for this gene subset to perform a pairwise comparison for each gene combination. We tested for a Pearson correlation for each gene pair using the R package Psych,[106] which was developed for psychometric analyses but contains useful functions for correlation analyses in general. Then, we selected genes with a padj <0.05 and an absolute correlation of at least 0.9 to generate the co-expression networks. For that step, we used the package igraph,[107] and calculated network statistics such as 'degree' and 'betweenness centrality' for each gene. Briefly, degree corresponds to the number of connections a specific node (gene) has, i.e. the number of co-expression links, and betweenness centrality measures how many nodes use this node as a path.[108] These two metrics are used to calculate the Rank for each gene, which we used to select the 10 smORFs with the largest gene hubs. Afterward, we added links to the center of the circular plot to show co-expression evidence for each smORF from this subset and another gene. We then selected smORFs that were co-expressed with genes related to lipid metabolism and generated a summarized network to show the smORFs that are the most relevant for subsequent studies.

Tissue distribution of smORFs

To check for the distribution of the novel Ribo-Seq smORFs across different tissues, we downloaded fastq files for RNA-Seq experiments from the ENCODE Mouse atlas[45] for dendritic cells, monocytes, pancreas, spleen, lung, macrophages, liver, intestine, kidney, heart, left cerebral cortex, hippocampus, and adrenal gland. To check for low-quality reads and adapter contamination, we ran TrimGalore (https://github.com/FelixKrueger/TrimGalore, v0.6.6) and aligned the trimmed reads to the mm10 genome with STAR using a GTF file containing both the novel smORFs and the annotated reference genes as guide. Next, we ran analyzeRepeats.pl from the HOMER suit[91] on each set of alignments, and extracted the mean TPM values for the replicates. Then, we applied a Variance Stabilizing Transformation on the TPM values and plotted the smORF-containing transcripts (SCTs) into a heatmap using the R package Complex Heatmap.[109] Similarly, we used a Python implementation of the UpSet plot (https://github.com/jnothman/UpSetPlot) to represent the intersection among the SCTs across multiple different tissues using the same mean TPM values. We included in the UpSet plot only SCTs with at least a mean TPM of 1 and subsets containing the intersection of at least 20 genes. Blue-colored bars and dots represent transcript abundance that is exclusive to the RNA-Seq samples from this study, black-colored features represent the abundance for SCTs present in both fat and ENCODE samples, and orange-colored features correspond to transcripts that are present in every single one of the analyzed tissues.

Sequence conservation of smORFs across different species

To assess the number of homologous sequences for the novel smORFs in the transcriptome of other eukaryotes, we downloaded the reference transcriptome of *Rattus norvegicus*, *Equus caballus*, *Ovis aries*, *Bos taurus*, *Sus scrofa* domesticus, *Canis lupus* familiaris, *Homo sapiens*, and *Pan troglodytes* from NCBI. Afterward, we performed a tBlastn search against each transcriptome using the microproteins sequences of the smORFs as query and, following the guidelines of Pearson, 2013,[110] we classified as homologous sequences any hit that had an alignment with an E-value $\leq$ 0.001 and a score $\geq$ 50. This way, without using an identity threshold, we can identify homologs with low sequence conservation. Next, we downloaded the common tree from NCBI for these species in phylip format and used EvolView[111] to plot bars to the right side of each leaf in the tree corresponding to the number of smORFs that have some level of conservation in these transcriptomes.

Sample preparation and protein digestions including cleanup and desalting of tryptic peptides for DIO/lean young/old plasma

For plasma protein digests for the DIO/lean young/old mouse experiments, 1 μL of plasma was diluted to 100 μL in water. 50 μL of this solution was combined with 50 μL of 0.2% PPS silent surfactant (Agilent # 400500) in 100 mM ammonium bicarbonate (Sigma-Aldrich # A6141) for a final concentration of 0.1% PPS and 50 mM ammonium bicarbonate. The remaining 50 μL was stored for BCA protein concentration estimates. The twice diluted plasma was incubated at 95°C for 5 min to facilitate protein denaturation and then reduced with the addition of 500 mM dithiothreitol (DTT) to a final concentration of 5 mM and incubation at 60°C for 30 min. Reduced samples were then alkylated at 15 mM iodoacetamide at room temperature for 30 min in the dark. Alkylation was quenched by adding an additional aliquot of DTT to bring the final concentration to 10 mM. Each sample was digested with sequencing grade trypsin (Promega #V5113) at a dilution of 1:10 trypsin:protein with starting plasma concentrations estimated at 70 mg/mL such that 3.5 μg of trypsin was added to each sample. Digestion was allowed to proceed for 18 h at 37°C on a thermomixer set to 1300 RPM. The following day, digestion was quenched by the addition of 4.5 μL of 5 M HCL, which were allowed to incubate at room temperature for 1 h to facilitate hydrolysis of the PPS surfactant. Samples were centrifuged at 15,000 g at 4°C for 5 min to pellet insoluble material and precipitated PPS surfactant. Tryptic digests were desalted with MCX columns in 96-well format (Waters # 186001830BA) as follows. An MCX column plate was affixed to a vacuum manifold. Samples were added to the columns careful to avoid the precipitated PPS pellet at the bottom of the tube and then samples were allowed to sit for 10 min. Salts were washed with 1 mL 0.1% formic acid followed by a wash with 1 mL of 90% ACN/10% water. The column plate was placed over the top of a new collection plate 600 μL 10% ammonium hydroxide in methanol was added to each column and allowed to sit for 30 min. Vacuum was used to pull the ammonium hydroxide/methanol mixture through to the new collection plates that were vacuum centrifuged to dryness. Samples were reconstituted in 20 μL buffer A supplemented with PRTC peptides were spiked in to a final concentration of 50 fmol/μL for DIA-MS analysis.

Small protein sub-fractionation in mouse plasma

For the quantification of plasma proteomes from the DIO/lean young/old experiment, an additional plasma protein physical fractionation step was employed with the goal of enriching for the small proteome. Normal lean plasma was enriched for small proteins in an effort to deeply identify microproteins in the circulation using the following four microprotein enrichment strategies. For this, 1 mL of lean mouse plasma (1 mL for each enrichment) was mixed with an equal volume of 8M guanidine-Cl (Pierce #24115), 0.2% Trifluoracetic acid (TFA, Pierce #28904), 1% β-mercaptoethanol (Fischer #PI35602) and filtered through a 0.45 μm filter. Samples were then enriched for peptides and small proteins, both known and novel, with 4 separate methods as follows. Two of the enrichments were performed with Agilent BondElut C18 SPE cartridges (Varian Product #12102028) using either triethylammonium formate buffer at pH 3.0 (TEAF) or 0.1% TFA, (Pierce #28904) as the counter ion. TEAF was made by adding 23 mL of 88% formic acid to 1.9 L of dH_2O and adjusting pH to 3.0 using triethylamine (Pierce #25108). The other two enrichments were performed with Agilent BondElut C8 SPE cartridges (Varian Product #12102029) also using either TEAF at pH 3.0 or 0.1% TFA. Eluted with 75% acetonitrile in counter ion. In all 4 cases, columns were prewet with 3 mL methanol followed by equilibration with 3 mL of either the TEAF at pH 3.0 solution

or the 0.1% TFA solution. Samples were applied to the conditioned columns followed by a wash in either the TEAF at pH 3.0 solution or the 0.1% TFA solution. Samples were eluted in 75% acetonitrile and lyophilized prior to digestion. Each plasma fraction that had been enriched for the small proteome was digested with the plasma digest protocol in the section entitled "Sample preparation and protein digestions including cleanup and desalting of tryptic peptides for DIO/lean young/old plasma" with the peptides fractionated with disposable high pH reversed phase fractionation columns as described by the manufacturer (Thermo, #84868) as described in "Data-independent acquisition (DIA) mass spectrometry for quantitative proteomics including spectral library generation with data-dependent acquisition (DDA)".

Sample pooling and fractionation for mass spectrometry

Sample pooling and queueing across all DDA and DIA experiments reported in this paper followed the strategy described in Figure 2. Briefly, for each experiment, tryptic digested peptides from a representative batch of biological samples were pooled. These pools were used for both GPF-DIA acquisitions and high pH reversed phase fractionation for DDA to generate spectral libraries. For the experiment described in Figure 3, separate pools were generated for proteome samples and for secretome samples. For the experiment described in Figure 5, subaliquots of all biological samples from all conditions were pooled. While an aliquot of each pool was reserved for GPF-DIA, these pools were additionally fractionated with disposable high pH reversed phase fractionation columns as described by the manufacturer (Thermo, #84868). Briefly, 80 μg of a pool of tryptic peptides was loaded onto the hydrophobic resin spin column and fractionally eluted using 8 buffers with increasing concentrations of acetonitrile as follows: 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20.0%, and 50%. The peptide fractions were dried completely with vacuum centrifugation and resuspended in 15 μL of buffer A for further analysis with DDA.

Data-dependent acquisition mass spectrometry (DDA-MS)

Each fraction described in the previous section was injected into a Thermo Fusion Lumos attached to a Thermo EASY-nLC 1200. For DDA, tryptic digest fractions were separated on self-packed 30 cm columns packed with 1.8 μm ReproSil-Pur C18 silica beads (Dr. Maisch) inside of a 75 μm inner diameter fused silica capillary (#PF360 Self-Pack PicoFrit, New Objective). The 30 cm column was coiled inside of a Sonation PRSO-V1 column oven set to 55°C prior to ionization into the MS. The HPLC was performed using 200 nL/min flow with solvent A as 0.1% formic acid in water and solvent B as 0.1% formic acid in 80% acetonitrile. For each injection, 3 μL (approximately 1 μg) was loaded and eluted with a linear gradient from 7% to 38% buffer B over 90 min. Following the linear separation, the system was ramped up to 75% buffer B over 5 min and finally set to 100% buffer B for 15 min, which was followed by re-equilibration to 2% buffer B prior to the subsequent injection. Data were acquired using DDA with a maximum 3 s cycle time configuration between MS1s and with 30 s dynamic exclusion. Precursor spectra were collected from 300 to 1650 *m/z* at 60,000 resolution (AGC target of 5e5, max IIT of 20 ms). MS/MS were collected on +2H to +7H precursors achieving a minimum AGC of 2e3. MS/MS scans were collected at 30,000 resolution (AGC target of 1e5, max IIT of 47 ms) with an isolation width of 1.4 *m/z* with an NCE of 27.

Data-independent acquisition mass spectrometry (DIA-MS)

For DIA-MS analysis, there are two methods involved in this study. There is the small window GPF-DIA-MS method (4 *m/z*) and the larger-window DIA-MS method used for individual sample quantitation (8 *m/z*). For GPF-DIA-MS, the Thermo Fusion Lumos was set to acquire six progressive GPF-DIA-MS acquisitions of the specific biological sample pool used in each comparison using 120,000 precursor resolution and 30,000 fragment resolution. Automatic gain control (AGC) target was set to 4e5 with the fragment maximum ion inject time (IIT) set to 60 ms. The NCE was set to 33 and +2H was assumed as the default charge state. For the GPF-DIA-MS acquisitions, we chose to use 4 *m/z* precursor isolation windows in a staggered-window pattern with the following optimized window placements ranges: 398.4 to 502.5 *m/z*, 498.5 to 602.5 *m/z*, 598.5 to 702.6 *m/z*, 698.6 to 802.6 *m/z*, 798.6 to 902.7 *m/z*, and 898.7 to 1002.7 *m/z*.[112] To quantify the proteomes of individual samples, we used single-injection DIA-MS acquisitions (120,000 precursor resolution, 15,000 fragment resolution, AGC target of 4e5, fragment max IIT of 20 ms) with 8 *m/z* overlapping precursor isolation windows in a staggered-window pattern with optimized window placements over a range of 396.4–1004.7 *m/z*.

Bioinformatics for proteomics

DDA data processing, database search, and spectral library creation for high pH reversed phase fractionated pools

All Thermo RAW files were converted to mzXML format using ProteoWizard (version 3.0.9974).[96] DDA-based peptide identifications were derived by searching mzXML files against a custom mouse proteome database that included the canonical, reviewed *Mus musculus* database downloaded from uniprot.org amended with the 3,877 microprotein sequences compiled from the Ribo-Seq smORF discovery work performed on the primary mouse BAT/WAT/Beige cells (Table S1). Peptide identifications were derived by concatenated target-decoy searching using Comet (version 2018.01 rev. 0)[97] with a modified "high-high" parameters set. In brief, this included variable modifications of up to three methionines per peptide, peptide n-terminal pyro-glutamate formation on glutamine, and protein n-terminal acetylation as well as fixed cysteine carbamidomethylation. Additionally, the "clip_nterm_methionine" parameter was turned on. Fully tryptic searches were performed with a 20 ppm precursor tolerance and a 0.02 Da fragment tolerance permitting up to two missed cleavages. High-pH reversed-phase fractions were combined and search results were filtered to a 1% peptide spectrum match (PSM) FDR using PeptideProphet[113] from the Trans-Proteomic Pipeline (TPP version 5.1.0).[98] For the proteome and secretome experiments in Figure 3, a combined spectral library from both compartments was used for the subsequent DIA-MS. For the plasma proteome work in Figure 5, a combined spectral library was generated from searches against all of

the high pH reveresed phase fractions of the experimental samples themsleves along with all of the small protein enriched lean plasma samples that were also digested and fractionated with high pH reversed phase fractionation. The spectral libraries generated for Figures 3 and 5 were not combined such that the Figure 3 spectral library was used in the Figure 3 DIA-MS and the Figure 5 spectral library was used in the Figure 5 DIA-MS. Search results were imported into Sklyine such that the peptide prophet probability that correlated with a 1% PSM-level FDR was added as the "cut-off score" in Skyline where the search results were converted to.blib format for use in DIA-MS workflows to follow. Fragment spectra from key DDA peptides (VFCHQANDVHIYQTQVVMTNT LETSSGK++++ and MINLLMQHQR++) were visually validated with deep-learning fragmentation pattern predictions generated by Prosit[99] using model version: "Prosit_2020_intensity_hcd".

DIA data processing for DIO/lean adipose tissue atlas, DIO/lean young/old plasma proteome profiling, and BAT/WAT/Beige proteome and secretome analysis

All Thermo RAW files were converted to mzML format and overlap demultiplexed with 10 ppm accuracy after peak picking using ProteoWizard (version 3.0.9974). The GPF-DIA-MS injections for each experimental comparison were searched against the experiment-specific spectral library (.blib from DDA search above) to derive peptide identifications with accurate chromatographic retention time information for each peptide using EncyclopeDIA (version 0.9.0), which was configured with the default settings of 10 ppm precursor, fragment, and library tolerances. Both B and Y ions were allowed to be considered with trypsin digestion assumed in the EncyclopeDIA settings set. Following the protocol outlined in Searle et al.,[50,100] all EncyclopeDIA searches were performed with built in target/decoy[114] peptide-level FDR estimation using Percolator 3.1.[101,102] Proteins were then parsimoniously grouped and filtered to a 1% protein-level FDR. The EncyclopeDIA.elib report files were imported into Skyline, along with the individual DIA-MS raw files for data visualization and group comparisons.

qPCR for Gm8773 mRNA in Mouse Tissue extracts

Animal care and use

Quantitation of Gm8773 mRNA was carried out essentially as described in Bookout et al.[87] Briefly, C57BL/6J mice (6-week old male and female) were purchased from Jackson Laboratory (Bar Harbor, ME) and were allowed to reach 8–9 weeks of age on a 12 h light/dark cycle and fed standard diets (Harlan Teklad #7001; Madison, WI). Females were housed together to ensure a synchronous estrus cycle, although that point was not determined prior to the harvest.

Tissue harvest

Animals were sacrificed by halothane inhalation (Halocarbon Laboratories; River Edge, NJ) at lights on (ZT0). All tissues were from male mice except for female reproductive tissues. Mice were exsanguinated via the vena cava with whole tissues collected and snap frozen in liquid nitrogen as described in Bookout et al.[87] The pancreas was prepared immediately due to its high RNase content by collected directly into the RNA-isolation reagent. The rest of the snap-frozen tissues were stored at −80°C until RNA extraction. All protocols were approved by the University of Texas Southwestern Medical Center Institutional Animal Care and Use Committee.

RNA isolation and cDNA preparation

RNAStat60 (TelTest; Friendswood, TX) was used to extract RNA according to the manufacturer's directions with a few tissue-specific modifications to aid in the mechanical disruption of some organs prior to RNA isolation, described in Bookout et al.[87] Total RNA was pooled in equal quantities for each tissue (n = 6), with RNA pools from male mice used for all tissue analyses except those of the ovary and uterus. DNase treatment and cDNA synthesis were carried out as previously described.[115]

qPCR

The Gm8773 mRNA tissue distribution profile was measured using an ABI 7900HT Sequence Detection System as described previously[87,115] using the efficiency-corrected ΔΔCt method, with 18S rRNA as the endogenous reference gene and SYBR green chemistry.

In situ hybridization and visualization of Gm8773 transcript in the mouse brain

Formalin-fixed, paraffin-embedded (FFPE) blocks containing mouse brain were sectioned at 5 μm onto Fisher Super-Frost Plus glass (Fisher Scientific, Hampton, NH, USA). Sections were hybridized with a rodent-specific probe to detect mouse Gm8773 mRNA (#542179, Advanced Cell Diagnostics, Newark, CA) on a Ventana Discovery ULTRA (Ventana Medical Systems, Tucson, AZ, USA), and amplified/stained using an RNAscope RED kit (Advanced Cell Diagnostics), as described previously.[116] In this assay we used our standard pre-treatment conditions (24 min target retrieval, 12 min protease) to probe Gm8773 while maintaining CNS morphometry. After ISH, all slides were counterstained with hematoxylin and bluing, and were coverslipped using EcoMount (BioCare, Pacheco, CA) All sections were imaged at 20x on an AxioScan.Z1 (Zeiss, Jena, Germany). Post-hoc processing matched brightness/contrast across all slides, and images were compiled in Adobe Illustrator (Adobe Inc., San Jose, CA) for presentation.

Recombinant expression and purification of Gm8773 protein product

Plasmids

For recombinant expression, synthetic sequences encoding mouse Gm8773 (Uniprot Q3UQ24 D25-I132) was cloned into a mammalian expression vector encoding an N-terminal HSA fusion tags with expression under the control of the CMV promoter. The native signal peptide as predicted by SignalP was replaced with the CD33 signal peptide, yielding a final sequence of D25-I132. Sigma GenElute Endo-free maxi prep kits (prod# NA0410-1KT) were used to generate higher amounts of plasmid DNA for scale up production efforts.

Generation of recombinant proteins in transient mammalian host systems

For transient protein expression, the Gibco Expi293 Expression system (prod# A14525) was used, following the vendor's recommended protocol. Cells were grown at 37°C in Expi293 medium (prod# A14351-01) at 5–8%CO2, 85% humidity, and shaking, with a maximum passage number of 30. The day prior to transfection, cells were split to 2e6 cells/ml. The day of transfection, cells were counted and transferred to a 1.6 L flask (Thomson, 931,113) at 2.9e6 cells/ml. Plasmid DNA was combined with Expifectamine 293 Transfection Reagent following the vendor's suggested concentration and ratio. Cultivation was continued for 18–20 h at 37°C, 5–8%CO2, 85% humidity, and shaking, at which point the cultures were supplemented with ExpiFectamine 293 Transfection Enhancer 1 and Enhancer 2. After a total of 5 days post-transfection, the cell supernatants were collected by centrifugation and filtration (0.2 uM Millipore S2GPU11RE) and stored at 4°C.

Purification of Gm8773

900mL of filtered supernatant was captured on 2 × 5mL Ni excel columns at 5 mL/min, washed with 5 column volumes (CV) of PBS pH7.2 and eluted in 2.5CV of PBS pH 7.2, 300 mM Imidazole, 300 mM L-Arginine. Selected fractions were pooled for a total of 256 mg and digested with 2.5 mg HRV 3C for a 1:100 dilution overnight at 4C. After cleavage the pooled fractions were applied to a Superdex 75 (Millipore Sigma GE17-5174-01) in PBS pH 7.2 + 200 mM L-Arginine Buffer to separate the HSA tag from the now untagged protein. Selected pooled fractions were characterized by SDS-PAGE, analytical size exclusion and LC-MS.

Analysis of Gm8773

For analysis of deglycoslated protein 20ug of protein was incubated with 0.5uL of SialExo enzyme (Genovis, #G1-SM1-020) for 1 h at 37C. SDS-PAGE analysis was performed using Bolt 4–12% Bis-Tris gels (Invitrogen NW04122BOX) and stained using the eStain system (Genscript L00753). LC/MS analysis was performed using an Agilent 1290 Infinity II UHPLC system coupled to an Agilent 6545 Q-TOF. LC separations were performed on a Zorbax 300-Diphenyl column (Agilent, 1.8 μm, 300 Å) with a flow rate of 0.4 mL/min used with a solvent gradient of 30%–70% B in 6 min. Solvent A was 0.1% (v/v) formic acid in water and the composition of solvent B was 0.1% (v/v) formic acid in 100% acetonitrile. The mass spectrometer was operated in positive ion mode with a full-scan MS spectra from 400 to 3,200 *m/z*. Analytical size exclusion was performed using an Agilent UPLC HP1290 system using gel filtration standards (BioRad 1,511,901). Separations were performed on a Waters BEH column, 1.7 μm, 200 Å, 4.6 mm ID x 150 mm L (Waters, # 186005225) with a flow rate of 0.35 mL/min in 10 min. The mobile phase was 20 mM Phosphate, 150 mM NaCl, 2% Isopropanol, pH 7.

Acute food intake effects of ICV Gm8773 injections in DIO mice

Acute effects on food intake were monitored in DIO mice with the ICV-cannulations described above. DIO mice were which were male C57BL/6J mice purchased from Jackson Laboratories with obesity induced in-house via continuous feeding with high fat diet (45% HFD Research Diets: D12451i) from age 10 weeks on. DIO mice were cannulated for ICV injections at age 52 weeks. Food intake was monitored using a fully automated human-intereference-free food-intake monitoring system that allows for continuous FI monitoring over time (BioDAQ, Research Diets, Inc., New Brunswick, NJ, USA). Animals were randomized into groups according to body weight and 48 h pre FI prior to the ICV injections (n = 8). Animals each received one ICV injection of vehicle or test compound (0.495 nmol/mouse, dose volume 3μL). Compounds were dosed to 4-h-fasted mice just prior to lights out, and food intake data were collected for 24 h post dosing.

QUANTIFICATION AND STATISTICAL ANALYSIS

All data were analyzed with GraphPad Prism 9.0.1. Bar plots in Figure 3 represent the DIA-MS-based summed fragment ion intensities of each peptide for each protein/microprotein entry with at least one peptide per ORF/smORF where statistical comparisons were calculated using a two-tailed Student's t test where values represent the means ± SD and asterisks indicate * = p value <0.05, ** = p value <0.01, and *** = p value <0.001. Values in Figure 5F represent the same DIA-MS-based summed fragment ion intensity depicting the mean ± SD with statistics performed using a one-way ANOVA where **** = p value <0.0001. Lastly, the ICV-based food intake values in Figure 6H with statistics performed with a two-way ANOVA with significant time by treatment interaction where **** = p value <0.0001.

Cell Metabolism, Volume *35*

Supplemental information

Profiling mouse brown and white adipocytes to identify metabolically relevant small ORFs and functional microproteins

Thomas F. Martinez, Sally Lyons-Abbott, Angie L. Bookout, Eduardo V. De Souza, Cynthia Donaldson, Joan M. Vaughan, Calvin Lau, Ariel Abramov, Arian F. Baquero, Karalee Baquero, Dave Friedrich, Justin Huard, Ray Davis, Bong Kim, Ty Koch, Aaron J. Mercer, Ayesha Misquith, Sara A. Murray, Sakara Perry, Lindsay K. Pino, Christina Sanford, Alex Simon, Yu Zhang, Garrett Zipp, Cristiano V. Bizarro, Maxim N. Shokhirev, Andrew J. Whittle, Brian C. Searle, Michael J. MacCoss, Alan Saghatelian, and Christopher A. Barnes

Supplemental Information

Title: Profiling Mouse Brown and White Adipocytes to Identify Metabolically Relevant Small ORFs and Functional Microproteins

Authors: Thomas F. Martinez[1], Sally Lyons-Abbott[2], Angie L. Bookout[2], Eduardo V. De Souza[3,4,5], Cynthia Donaldson[5], Joan M. Vaughan[5], Calvin Lau[5], Ariel Abramov[2], Arian F. Baquero[2], Karalee Baquero[2], Dave Friedrich[2], Justin Huard[2], Ray Davis[2], Bong Kim[2], Ty Koch[2], Aaron J. Mercer[2], Ayesha Misquith[2], Sara A. Murray[2], Sakara Perry[2], Lindsay K. Pino[7], Christina Sanford[2], Alex Simon[2], Yu Zhang[2], Garrett Zipp[2], Cristiano V. Bizarro[3,4], Maxim N. Shokhirev[6], Andrew J. Whittle[2], Brian C. Searle[8], Michael J. MacCoss[7], Alan Saghatelian[5,*], Christopher A. Barnes[2,9,*]

SUPPLEMENTARY MATERIALS ASSOCIATED WITH THIS DOCUMENT

Supplementary Figures S1-S5

Supplementary Tables S1-S8

Fig. S1

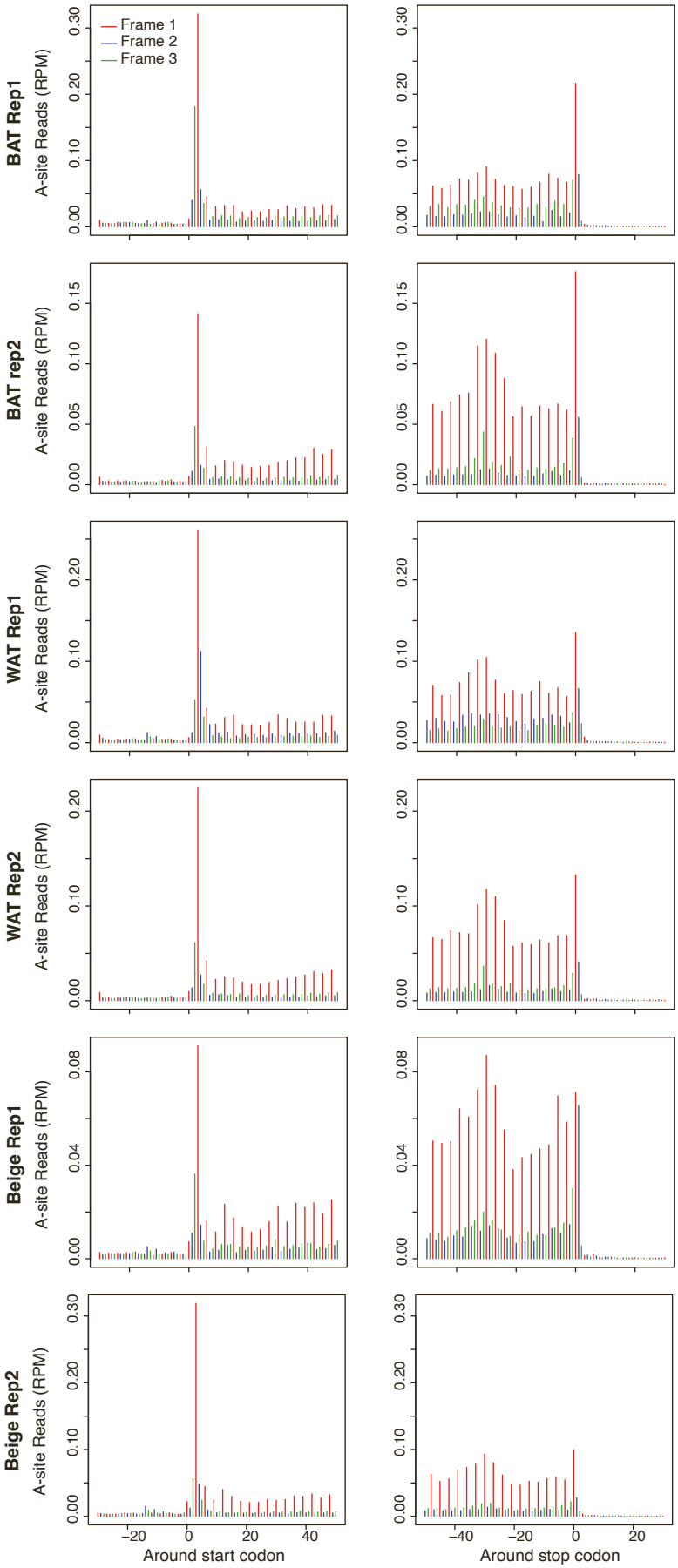

Frame 1
Frame 2
Frame 3
BAT Rep1
BAT rep2
WAT Rep1
WAT Rep2
Beige Rep1
Beige Rep2
A-site Reads (RPM)
Around start codon
Around stop codon

Fig. S1. Metagene plots for BAT, WAT, and Beige Ribo-Seq datasets, related to Figure 1. Metagene plots showing RPF A-site alignment around the start site and stop site for each replicate created using RibORF. 26-31 nt RPF reads were included in the plot for BAT Rep1, 25-29 nt for BAT Rep2, 25-30 nt for WAT Rep1, 25-29 nt for WAT Rep2, 25-29 nt for Beige Rep1, and 25-30 nt for Beige Rep2.

Fig. S2

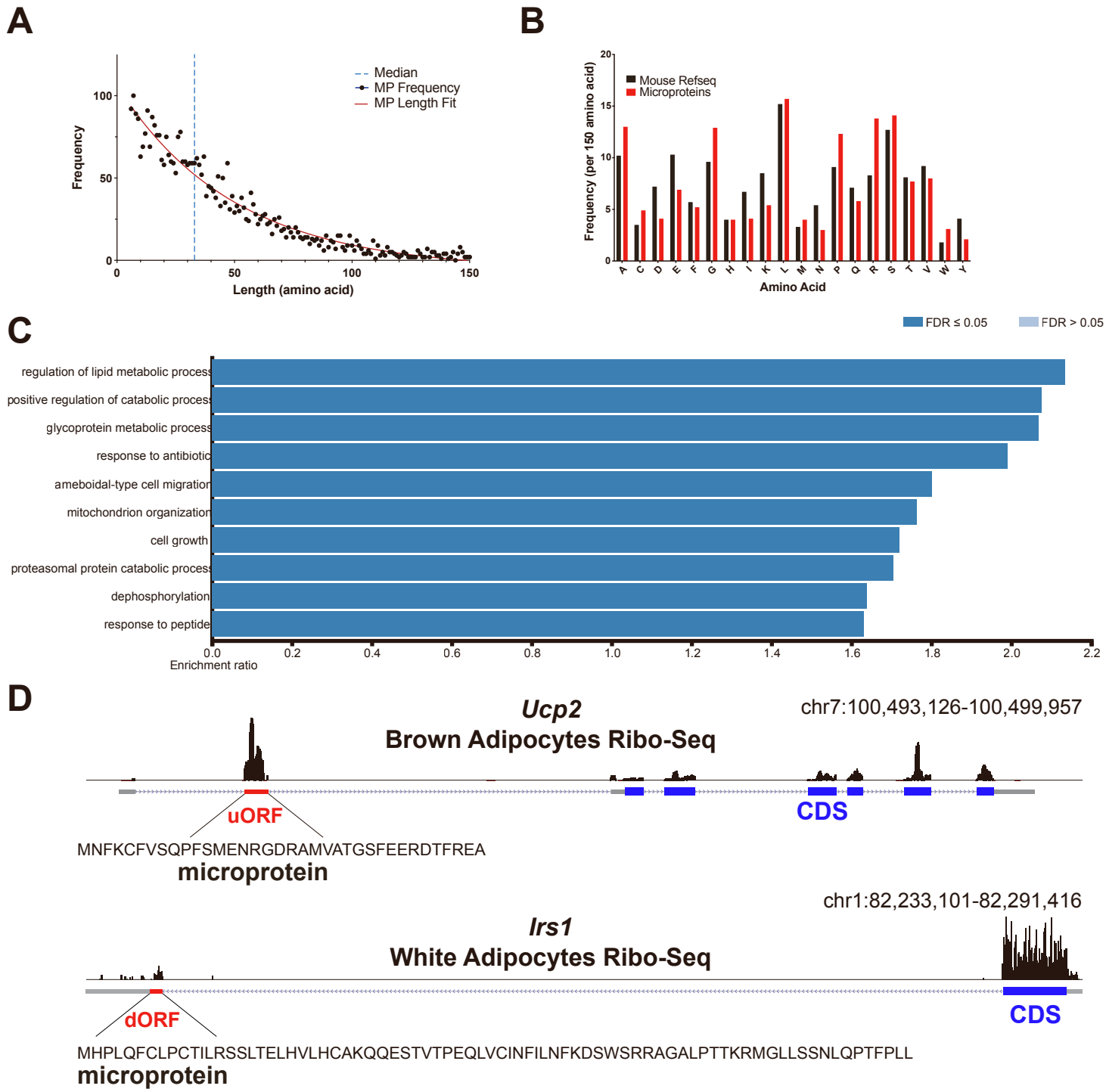

A
Median
MP Frequency
MP Length Fit
Frequency
Length (amino acid)
B
Mouse Refseq
Microproteins
Frequency (per 150 amino acid)
Amino Acid
C
FDR ≤ 0.05
FDR > 0.05
regulation of lipid metabolic process
positive regulation of catabolic proces
glycoprotein metabolic process
response to antibiotic
ameboidal-type cell migration
mitochondrion organization
cell growth
proteasomal protein catabolic proces
dephosphorylation
response to peptide
Enrichment ratio
D
Ucp2
Brown Adipocytes Ribo-Seq
chr7:100,493,126-100,499,957
uORF
CDS
MNFKCFVSQPFSMENRGDRAMVATGSFEERDTFREA
microprotein
chr1:82,233,101-82,291,416
Irs1
White Adipocytes Ribo-Seq
dORF
CDS
MHPLQFCLPCTILRSSLTELHVLHCAKQQESTVTPEQLVCINFILNFKDSWSRRAGALPTTKRMGLLSSNLQPTFPLL
microprotein
E
TRANSMEMBRANE: TMHMM w/o Signal P (84)
SECRETED: All Signal P + Phobius w/o TMHMM (183)
Signal P and TMHMM positive prediction (27)
INTRACELLULAR:
no TMHMM
no Signal P (3583)
F
microprotein from creatine transporter (Slc6a8, uORF)
Mus_musculus
Rattus_norvegicus
Amblyraja_radiata
Ornithorhynchus_anatinus
Echeneis_naucrates
Arvicanthis_niloticus
Pan_troglodytes
Homo_sapiens

Fig. S2. Characterization of smORFs from Ribo-Seq on primary differentiated brown, beige, and white adipocytes, related to Figure 1.

(A) The length distribution of the microproteins encoded by smORFs results in a median length of 33-amino acids (blue dotted line). (B) Comparison of the amino acid composition of microproteins (red bars) to mouse RefSeq proteins (black bars). (C) GO enrichment analysis on the annotated genes containing uORFs listed in Table S1[33]. The top 10 redundancy-filtered pathways are shown. (D) Primary differentiated brown adipocyte Ribo-Seq data shows ribosome occupancy on a uORF of the *Ucp2* gene, and subcutaneous white adipocyte Ribo-Seq data detected a dORF in what was considered the 3'-UTR of *Irs1*. (E) Distribution of transmembrane and signal peptide predictions of smORFs from Ribo-Seq on primary differentiated brown, beige, and white adipocytes. (F) Alignment of microprotein amino acid sequence of Slc6a8 smORF across multiple species showing how some microproteins such as the 14-amino acid microprotein encoded from the creatine transporter Slc6a8 uORF are evolutionarily conserved.

Fig. S3

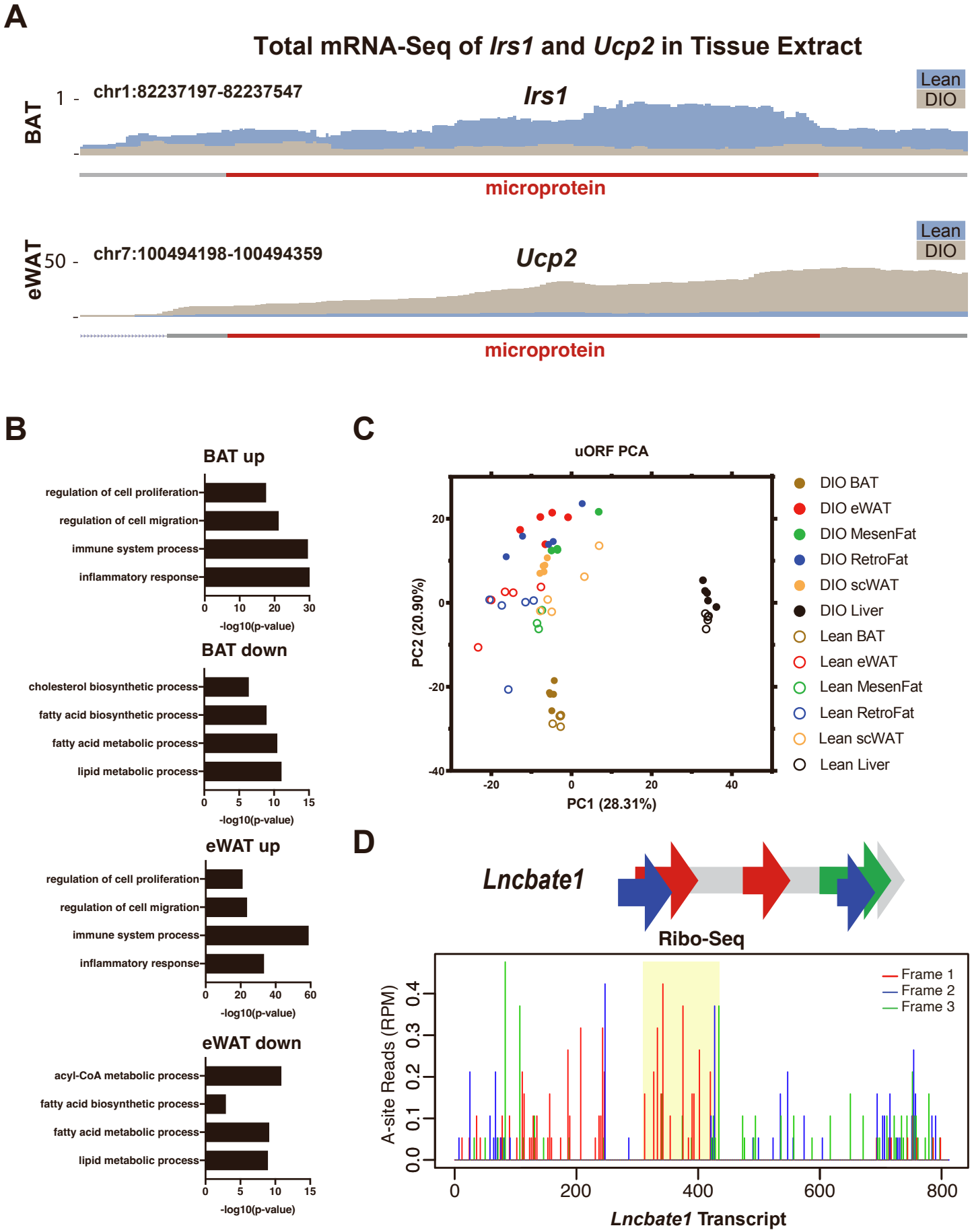
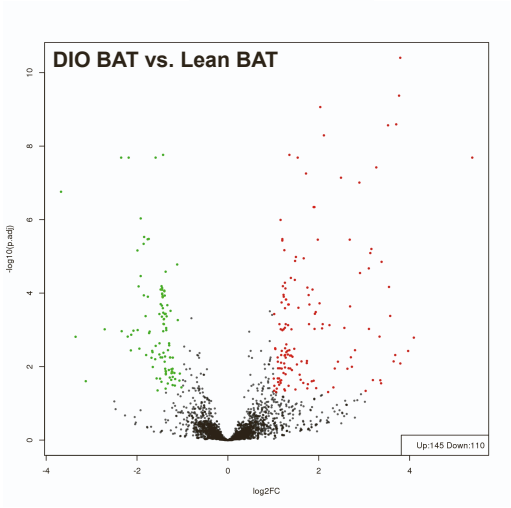

A
Total mRNA-Seq of Irs1 and Ucp2 in Tissue Extract
BAT
chr1:82237197-82237547
Irs1
Lean
DIO
microprotein
eWAT
chr7:100494198-100494359
Ucp2
Lean
DIO
microprotein
B
BAT up
regulation of cell proliferation
regulation of cell migration
immune system process
inflammatory response
-log10(p-value)
BAT down
cholesterol biosynthetic process
fatty acid biosynthetic process
fatty acid metabolic process
lipid metabolic process
-log10(p-value)
eWAT up
regulation of cell proliferation
regulation of cell migration
immune system process
inflammatory response
-log10(p-value)
eWAT down
acyl-CoA metabolic process
fatty acid biosynthetic process
fatty acid metabolic process
lipid metabolic process
-log10(p-value)
C
uORF PCA
PC2 (20.90%)
PC1 (28.31%)
DIO BAT
DIO eWAT
DIO MesenFat
DIO RetroFat
DIO scWAT
DIO Liver
Lean BAT
Lean eWAT
Lean MesenFat
Lean RetroFat
Lean scWAT
Lean Liver
D
Lncbate1
Ribo-Seq
A-site Reads (RPM)
Frame 1
Frame 2
Frame 3
Lncbate1 Transcript

Fig. S3. Adipose protein-coding smORFs are differentially transcribed in diet-induced obese mice, related to Figure 4.

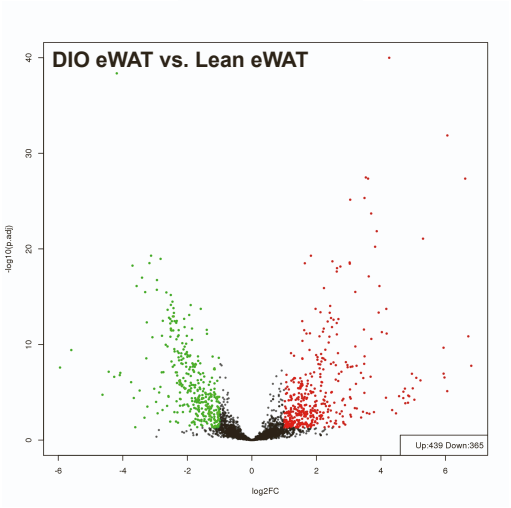
(A) Total mRNA-Seq read maps (equivalently scaled) in lean (blue) and DIO (brown) over the microprotein encoding smORFs of Irs1 (comparison depicted in BAT, top) and Ucp2 (comparison depicted in eWAT, bottom). (B) Gene ontology (GO) biological process analysis of bulk RNA-Seq-based transcriptomics expression of smORF genes up- and down- regulated in both BAT and epididymal WAT ("eWAT") tissue depots from both DIO and lean mice. smORF coordinates listed in Table S1 were used to extract DIO vs. lean expression levels of uORF-containing smORFs. Depicted are up- and down-regulated GO-enriched processes where changes in RNA expression for adipose protein-coding smORFs induced by DIO in various tissues (padj < 0.05 and |log2 fold change| ≥ 1). (C) Changes in RNA expression for adipose protein-coding smORFs induced by DIO in various tissues (padj < 0.05 and |log2 fold change| ≥ 1) showing PCA analysis of uORF-containing protein-coding smORF RNA expression levels in tissues derived from DIO and lean mice. (D) Schematic showing the locations of translated smORFs identified on *Lncbate1*. Five smORFs were identified across the primary differentiated subcutaneous white, brown, and beige adipocytes: two in frame 1, two in frame 2, and one in frame 3 (left). Representative ribosomal A-site plots (Ribo-Seq) from brown adipocytes for *Lncbate1* with one smORF identified as translated in both brown and subcutaneous white adipocytes highlighted in yellow (right).

Fig. S4

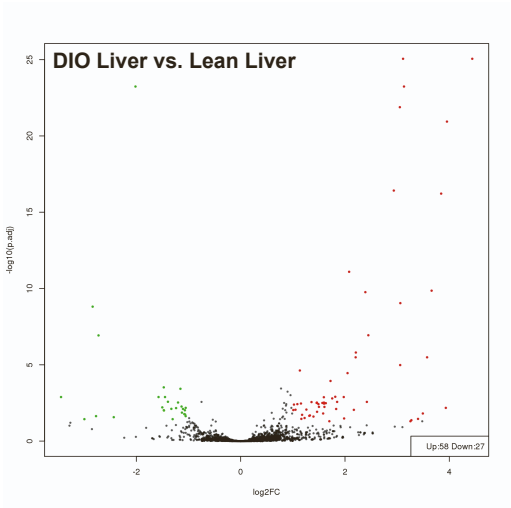
A



B



C



D

E

F

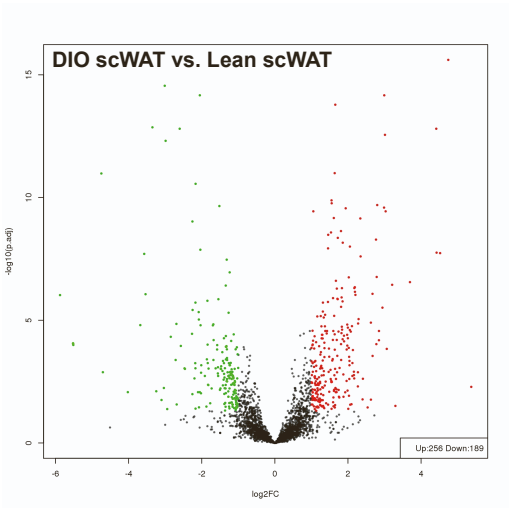
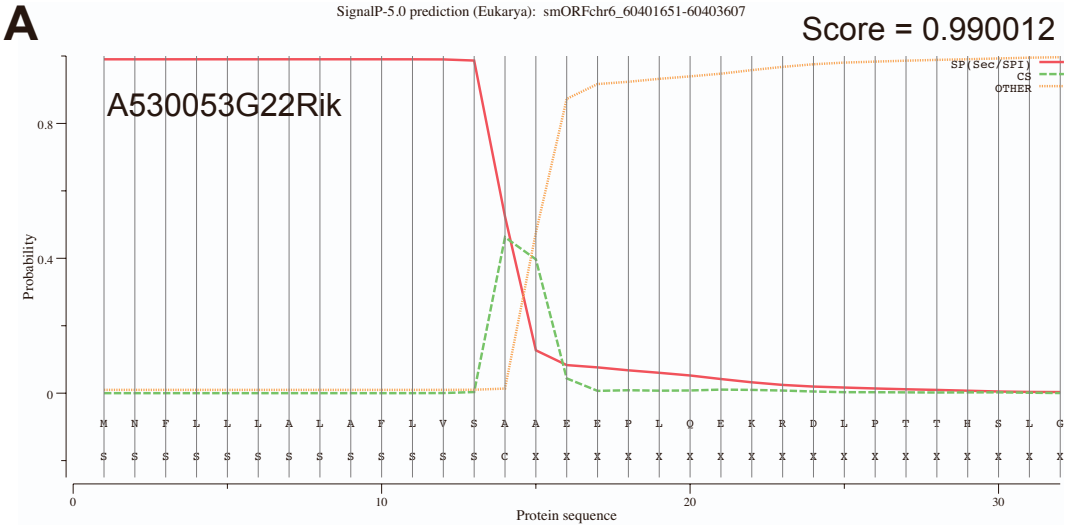
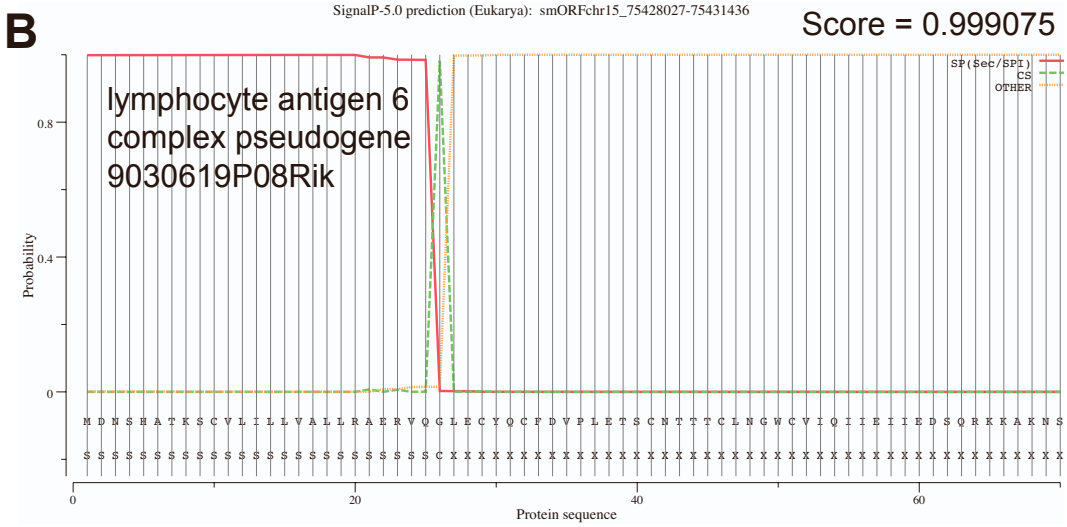


Fig. S4. Differentially expressed smORF genes in multiple adipose tissue depots and liver using bulk RNA-Seq-based transcriptomics, related to Figure 4. smORF coordinates listed in Table S1 were used to extract DIO vs. lean expression levels of all smORF gene coordinates for (A) BAT; (B) epididymal white adipose tissue ("eWAT"); (C) liver; (D) mesentyric adipose tissue ("Mesen Fat"); (E) retroperitoneal adipose tissue ("Retro Fat"); and (F) subcutaneous white adipose tissue ("scWAT"). Changes in RNA expression for adipose protein-coding smORFs induced by DIO in various tissues (padj < 0.05 and |log2 fold change| ≥ 1).

Fig. S5


A
SignalP-5.0 prediction (Eukarya): smORFchr6_60401651-60403607
Score = 0.990012
A530053G22Rik
SP(Sec/SPI)
CS
OTHER
Probability
Protein sequence


B
SignalP-5.0 prediction (Eukarya): smORFchr15_75428027-75431436
Score = 0.999075
lymphocyte antigen 6 complex pseudogene 9030619P08Rik
SP(Sec/SPI)
CS
OTHER
Probability
Protein sequence

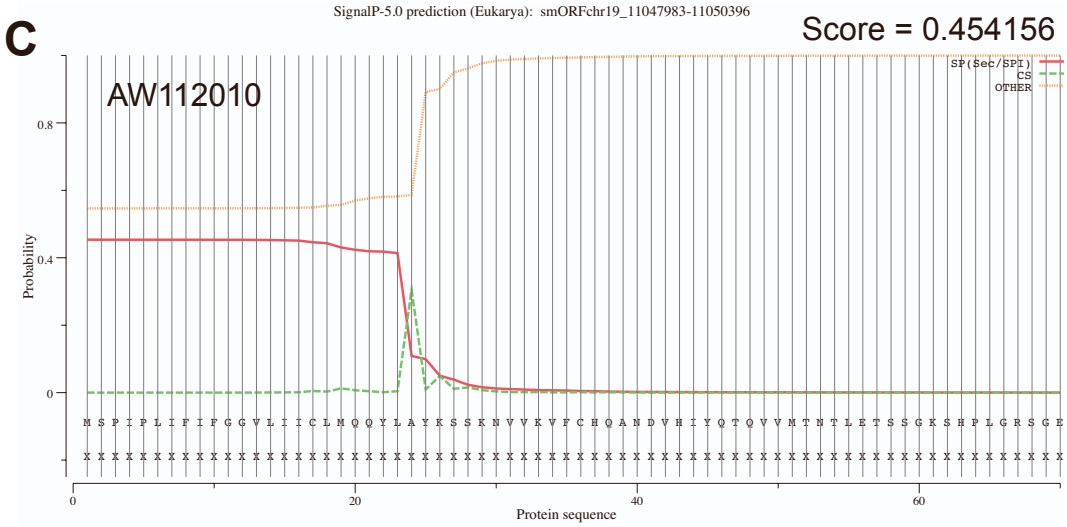

C
SignalP-5.0 prediction (Eukarya): smORFchr19_11047983-11050396
Score = 0.454156
AW112010
SP(Sec/SPI)
CS
OTHER
Probability
Protein sequence

Fig. S5. Signal peptide predictions of microproteins identified in plasma from DIO/lean young/old mouse experiment, related to Figure 6. Microprotein sequences were input into Signal P 5.0[39]. Predictions are depicted above the threshold of 0.4. (A) prediction for microprotein from chr6:60401651-60403607; (B) prediction for microprotein from chr15:75428027-75431436; and (C) prediction for microprotein from chr19:11047983-11050396.