Identification of a Micropeptide on Lipid Droplet That Mediates Lipid Storage and Insulin Sensitivity

Ting Huang[1]*, Adekunle T. Bamigbade[2]*, Shimeng Xu[2]*, Yaqin Deng[2], Kang Xie[2], Ololade O. Ogunsade[2], Ahmed Hammad Mirza[2], Jifeng Wang[2], Pingsheng Liu[1,2,3#], and Shuyan Zhang[2#]

[1]School of Basic Medical Sciences, Southwest Medical University, Luzhou 646000, China

[2]National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

[3]University of Chinese Academy of Sciences, Beijing 100049, China

*Those authors contribute equally.

[#]To whom correspondence should be addressed.

Address: 15 Datun Road, Beijing 100101, China

E-mail: pliu@ibp.ac.cn and E-mail: syzhang@ibp.ac.cn

Running title: Identification of a functional micropeptide on lipid droplet

Abbreviations

LD, lipid droplet; LDAMP, LD-associated micropeptide; sORFs, small open reading frames; OA, oleate; TAG, triacylglycerol; PA, palmitate; TM, total membrane; Cyto, cytosol; PNS, post-nuclear supernatant; LDAM, LD-anchored mitochondria; GF, cells expressing GFP-FLAG; LGF, cells stably expressing LDAMP1-GFP-FLAG; CRISPR/Cas9, clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9

Abstract

Great progress in sequencing technologies and computational biology suggests that the noncoding sequence in mammalian genome is rich in functional elements to produce proteins. Many RNA molecules, mis-annotated as noncoding, actually harbor small open reading frames that are predicted to code proteins called micropeptides. Some micropeptides have been discovered and verified to play critical roles in multiple biological processes. Lipid droplet (LD) is a unique cellular organelle, conserved from bacteria to humans, and closely associated with cellular lipid metabolism and metabolic disorders. However, no micropeptide has been identified on LDs. Here we for the first time explore this possibility. Proteins with small molecular weight from isolated LDs of myoblasts were enriched by electrophoresis and subjected to proteomic analysis, and LD-associated micropeptides (LDAMP) were identified using mass spectrometry coupled with a home-made protein database. A total of 15 micropeptides were identified. One of them was studied further and termed LDAMP1. Similar to LD resident proteins, LDAMP1 was degraded by proteasome. LDAMP1 was localized on LDs by confocal imaging, cell fractionation, as well as immunogold labeling. Using CRISPR/Cas9-mediated genome editing technique, endogenous expression of LDAMP1 was validated by FLAG tag knock-in. Moreover, oleate-induced triacylglycerol accumulation was significantly decreased when myoblasts expressing LDAMP1 and, in turn, the rescue of palmitate-inhibited insulin sensitivity was obviously reduced. Taken together, we report for the first time that translatable noncoding RNA-derived peptides were identified on LDs and a novel LD protein LDAMP1was discovered, which paves a new way for exploring LD dynamics and related metabolic functions.

Keywords:

Micropeptide; lipid droplet; sORF-encoded peptides; insulin sensitivity; LDAMP1; noncoding

Introduction

In past two decades, only 1.5% of the human genome sequence have been identified to actually contain coding elements for making proteins(1). But due to the rapid development of high-throughput sequencing technologies recently, transcriptomics studies have shown that eukaryotic genome is extensively transcribed(1-4). Large-scale projects for the systematic annotation and functional characterization of genes, such as GENCODE and FANTOM, demonstrate that at least 80% of mammalian genomic DNA is actively transcribed and have led to the prediction of many nonclassical protein-coding genes that are previously thought to be noncoding(5-7). It has become more clear that a number of noncoding regions actually harbor small open reading frames (sORFs) less than 300 nucleotides to code for functional small proteins or micropeptides, which has historically been the minimum length for detection. These small proteins always evade detection due to their small size, low abundance, instability, and loss during sample preparation(7). However, the currently developed computational and experimental techniques, especially omics, have facilitated to mine those peptides, such as intricate computational techniques, ribosome profiling and mass spectrometry(7, 8). To date, most micropeptides are identified in prokaryotic organisms(8). A number of micropeptides have also been discovered in eukaryotic organisms, such as plant, zebra fish, insect, mouse, and humans(4, 7, 8). Those functional micropeptides play important roles in many fundamental biological processes. For instance, Toddler has been proved to be an embryonic signal that promotes cell movement in zebrafish(9). Myoregulin (MLN) co-localizes with and inhibits SERCA, a Ca^{2+} pump found in the sarcoplasmic reticulum of skeletal muscle, to regulate calcium homeostasis in mouse(10). MOTS-c, a mitochondrial-derived peptide, is reported to

promote metabolic homeostasis and reduce diet-induced obesity and insulin resistance in mouse(11). The peptides from noncoding transcripts have been reported to localize to some cellular compartments such as mitochondrion(11), late endosome/lysosome(12), and endoplasmic reticulum(10). However, given the putative large number of micropeptides, much more should be mined, especially based on subcellular localization, which would gain the insight into their possible biological functions.

Lipid droplet (LD) is a ubiquitous intracellular organelle containing a neutral lipid core bounded by phospholipid monolayer membrane and some integral and peripheral membrane proteins(13, 14). LDs play critical roles in lipid storage, transport and metabolism(15-17). Abnormal LD dynamics is related to the cause and development of metabolic diseases, such as obesity, atherosclerosis, fatty liver disease, and type 2 diabetes(17-19). The LD-associated proteins are the main executors of LD function. For instance, Perilipin 2 (PLIN2), the most abundant LD resident protein in cells except adipocytes, has been proved to function in hepatic lipid sequestration(20) and help to improve insulin sensitivity in skeletal muscle(21). The most prevalent proteins reported to be associated with LDs are enzymes involved in lipid hydrolysis and synthesis(17). Therefore, it is crucial to study the proteins associated with LDs to unveil the role of LD in biological processes as well as human metabolic diseases. To date, over 100 proteins have been found on LDs and some of them have been proved to be associated with human diseases(17, 18). However, thus far, there has been no report on identification of LD-associated proteins derived from noncoding transcripts.

Here for the first time, the existence of micropeptides on LDs was explored. We adopted an experimental approach to identify micropeptides associated with LDs in myoblast cells. LDs

were isolated and the enriched small proteins less than 17 kDa were subjected to mass spectrometric study. A home-made database that we setup previously for identification of micropeptides was used and a total of 15 novel micropeptides were identified in C2C12 LDs. The top one micropeptide, LD-associated micropeptide 1, LDAMP1, was further studied in detail. This novel micropeptide was shown to localize on LDs and degraded though proteasome. When loaded with oleate (OA), LDAMP1-overexpressing cells displayed a significant reduction in triacylglycerol (TAG) and rescue of palmitate (PA)-reduced insulin sensitivity was inhibited.

Materials and methods

Materials

The Colloidal Blue Staining Kit, Hoechst 33342 and LipidTOX Red were from Invitrogen. Sodium oleate and sodium palmitate were obtained from Sigma-Aldrich. Western lightning plus-ECL reagent was from PerkinElmer. 8% paraformaldehyde solution (EM grade) and LR White resin were purchased from Electron Microscopy Sciences. The synthesized PCR primers used are listed in Supplemental Table S2. Antibodies used in this study are listed in Supplemental Table S3.

Methods

Cell culture

Mouse C2C12 myoblasts (American Type Culture Collections, Manassas, VA) were maintained in DMEM (Macgene Biotech., Beijing) supplemented with 10% FBS (Hyclone), 100 U/mL penicillin and 100 μg/mL streptomycin at 37°C with 5% CO_2.

Preparation of fatty acid

Sodium oleate and sodium palmitate were prepared according to our previously established method(22). Briefly, fatty acids were mixed with ethanol to a concentration of 100 mM and sonicated on ice at 200 W, 10 s on and 3 s off pulses until the solution was milky and homogenous. The prepared fatty acid stocks were kept at 4°C and kept from light. The fatty acid was incubated in 60°C preheated medium before treatment of cells.

Isolation of lipid droplets from C2C12 cells

LDs were isolated from C2C12 cells using the method described previously with modifications(23, 24). Briefly, C2C12 myoblasts treated with 100 μM OA for 12 h were scraped and collected after 3 rinses with ice-cold PBS. Then all the cells were transferred to 50 mL Buffer A (25 mM tricine, 250 mM sucrose, pH 7.8) containing 0.5 mM PMSF. After centrifugation at 3,000*g* for 10 min the cell pellets were re-suspended in 10 mL Buffer A containing 0.5 mM PMSF and incubated on ice for 20 min. The cells were then homogenized by N_2 bomb (750 psi for 15 min on ice). The cell lysate was centrifuged at 3,000*g* for 10 min, the post-nuclear supernatant (PNS) fraction was collected and loaded into a SW40 tube and the sample was overlain with 2 mL Buffer B (20 mM HEPES, 100 mM KCl, and 2 mM $MgCl_2$, pH 7.4). The sample was centrifuged at 182,000*g* for 1 h at 4°C. The white band containing LDs at the top of gradient was collected into a 1.5 mL Eppendorf tube. The LD sample was centrifuged at 20,000*g* for 10 min at 4°C and then the underlying solution was carefully removed. The LDs were gently re-suspended in 200 μL Buffer B. This procedure was repeated four times. The lipid extraction and the protein precipitation were carried out using chloroform/acetone (1:1, v/v) treatment followed by centrifuging the sample at 20,000*g* for 10

min at 4°C. The protein pellet was then dissolved in 2 × SDS sample buffer (125 mM Tris Base, 20% glycerol, 4% SDS, 4% β-mercaptoethanol and 0.04% bromophenol blue, pH 6.8).

Western blot analysis

Briefly, whole cell lysates were suspended in 2 × sample buffer, sonicated and denatured for 5 min at 95°C. 10 μL sample was subjected to gel electrophoresis followed by silver staining or protein blotting onto 0.2 μm PVDF membrane. In order to detect proteins of interest, membranes were incubated with indicated antibodies and subjected to ECL system.

Immunoprecipitation

Cells were cultured to 100% confluence and culture medium was removed by aspiration. Cells were washed with cold PBS three times and then treated with 1 mL cold TETN (25 mM Tris-HCl, 5 mM EDTA, 1% Triton X-100, pH 7.5) solution containing 150 mM NaCl supplemented with 0.5 mM PMSF. The mixture was left on ice for 30 min. Cells were scraped into a protease-free Eppendorf tube and pipetted 8 times followed by rotation at 4°C for 30 min. Meanwhile, FLAG beads (SIGMA M2 Affinity Gel) were blocked with 20 mg/mL BSA in TETN solution containing 150 mM NaCl. Then, beads were rotated at 4°C for 10 min, followed by centrifugation at 1,000*g* for 30 s. Supernatant was removed and this process was repeated three times. After 30 min whole cell lysate was centrifuged at 17,000*g* for 10 min at 4°C. An aliquot of the supernatant was kept as 'Input' mixed with an equal volume of 2 × sample buffer. The remaining supernatant was incubated with FLAG beads, previously blocked with BSA, at 4°C for 3 h. Supernatant-bead complex was centrifuged at 1,000*g* for 30 s at 4°C. An aliquot of the supernatant was kept as 'Supernatant' while the remaining was discarded. Beads were washed with TETN containing 500 mM NaCl followed by 250 mM NaCl-containing TETN

and finally 150 mM NaCl containing TETN, twice for each buffer. Supernatant was discarded, and beads were suspended in 50 μL 2 × sample buffer and vortexed vigorously for 5 min followed by denaturation at 95°C for 5 min (IP sample). The samples which include 'Input', 'Supernatant', and 'IP' were kept until further analyses.

LC-MS/MS Analysis

The LD proteins were separated on a 12% Bis-Tris gel with a 4% stacking gel. The gel was then stained with Colloidal Blue according to the manufacturer's protocol (Invitrogen). The protein band less than 17 kDa was cut into slices and protein sample were manually excised, cut into small plugs, and washed twice in 200 μL of ddH_2O for 10 min each time. The gel bands were dehydrated in 100% acetonitrile for 10 min and dried in a Speedvac for approximately 15 min. Reduction (10 mM DTT in 25 mM NH_4HCO_3 for 45 min at 56°C) and alkylation (40 mM iodoacetamide in 25 mM NH_4HCO_3 for 45 min at room temperature in the dark) were performed, followed by washing of the gel plugs with 50% acetonitrile in 25 mM ammonium bicarbonate twice. The gel plugs were then dried using a Speedvac and digested with sequence-grade modified trypsin (40 ng for each band) in 25 mM NH_4HCO_3 overnight at 37°C. The enzymatic reaction was stopped by adding formic acid to a 1% final concentration. The solution was then transferred to a sample vial for LC-MS/MS analysis. Besides, half of the extracted LD proteins were not stained after electrophoresis and cut directly for proteomic study.

All nano LC-MS/MS experiments were performed on a Q Exactive (Thermo Scientific) equipped with an Easyn-LC 1000 HPLC system (Thermo Scientific). The peptides were loaded onto a 100-μm id × 2-cm fused silica trap column packed in-house with reversed phase silica (Reprosil-Pur C18-AQ, 5 μm, Dr. Maisch GmbH) and then separated on an a 75-μm id × 20-

cm C18 column packed with reversed phase silica (Reprosil-Pur C18-AQ, 3 μm, Dr. Maisch GmbH). The peptides bound on the column were eluted with a 78-min linear gradient. The solvent A consisted of 0.1% formic acid in water solution and the solvent B consisted of 0.1% formic acid in acetonitrile solution. The segmented gradient was 4–12% B, 5 min; 12–22% B, 53 min; 22–32% B, 12 min; 32-90% B, 1 min; 90% B, 7 min at a flow rate of 300 nL/min.

The MS analysis was performed with Q Exactive mass spectrometer (Thermo Scientific). In a data-dependent acquisition mode, the MS data were acquired at a high resolution 70,000 (*m/z* 200) across the mass range of 300–1,600 *m/z*. The target value was 3.00E+06 with a maximum injection time of 60 ms. The top 20 precursor ions were selected from each MS full scan with isolation width of 2 *m/z* for fragmentation in the HCD collision cell with normalized collision energy of 27%. Subsequently, MS/MS spectra were acquired at resolution 17,500 at *m/z* 200. The target value was 5.00E+04 with a maximum injection time of 80 ms. The dynamic exclusion time was 40 s. For nano electrospray ion source setting, the spray voltage was 2.0 kV; no sheath gas flow; the heated capillary temperature was 320°C.

Protein identification

The raw data from Q Exactive were analyzed with Proteome Discovery (version 1.4.1.14, Thermo Scientific) using Sequest HT search engine for protein identification and Percolator for FDR (false discovery rate) analysis against our home-made protein database constructed from noncoding transcripts(25). Some important searching parameters were set as following: two missed cleavages were allowed for searching; the mass tolerance of precursor was set as 10 ppm and the product ions tolerance was 0.02 Da.; the methionine oxidation was selected as variable modifications; the cysteine carbamidomethylation was selected as a fixed modification.

FDR analysis was performed with Percolator and FDR <1% was set for protein identification. The peptides confidence was set as high for peptide filter. For the stained sample, the raw MS data was searched with parameter of trypsin selected as enzyme as well as no-enzyme as parameter while for unstained sample only using no-enzyme parameter.

Plasmid construction

LDAMP1 ORF was amplified from exon 4 in the USPL1 (UniProtKB entry Q3ULM6) coding sequence. LDAMP1 was cloned into pEGFP-N1 with XhoI and HindIII restriction endonucleases. The recombinant plasmid (LDAMP1-GFP cDNA) was in turn amplified with primer pair containing FLAG epitope in the reverse primer and cloned into pQCXIP with NotI and EcoRI restriction endonucleases where GFP-FLAG served as control. Point mutation at positions 7, 49, 61 and 80 to change Lys to Ala in LDAMP1 was carried out using QuikChange mutagenesis (Agilent) according to manufacturer's protocol. Mutant LDAMP1-GFP fusion proteins were subcloned into pQCXIP plasmid. List of primers is indicated in Supplemental Table S2.

Constitutive expression of LDAMP1 in C2C12 Cells

C2C12 cells stably expressing LDAMP1were constructed according to Clontech retroviral packaging system. For virus packaging, VSVG (vesicular stomatitis virus G) envelop plasmid and pHIT packaging plasmid were mixed together in Lipofectamine 2000 according to instruction manual along with LDAMP1-GFP-FLAG recombinant gene in pQCXIP plasmid. The plasmid mix was transfected into Plat-ET cells. Medium from packaged virus was filtered with 0.45-μm filter to remove dead floating cells but leave infective virus viable. C2C12 wild type (WT) cells were cultured into 6-well dishes to a density of 40% followed by infection with

medium containing LDAMP1 expressing virus supplemented with 8 μg/mL polybrene to increase infection efficiency. After 48 h post-infection, cells were sub-cultured into 100-mm dish and screened with 1 μg/mL puromycin for two weeks. Immediately after the screening process, cells were dispensed into 96-well dishes to select for clones.

CRISPR/Cas9-mediated genome editing of LDAMP1

To construct homologous arms for LDAMP1, nucleotide sequence spanning 1,000 bp upstream and downstream of reference stop codon was amplified and assembled in front and behind 3×FLAG of pKI-FLAG-N1 as previously established(26). The 3×FLAG was encrypted in a 34-codon sequence with an additional linker sequence (7 codons). To induce double strand break in the target genes in C2C12 cell genome, CRISPR/Cas9 technique was employed. Briefly, last exon/intron that includes about 50 bp upstream or downstream of stop codon on the left and right was copied from the genomic sequence and pasted on the http://crispr.mit.edu/ interactive webpage to design sgRNA most suitable to induce double strand break in the genome. The short sequence containing 20 nucleotides was synthesized with CACC and AAAC flanking the forward and reverse primers respectively in the 5′ directions. The primers were annealed and ligated into px260a plasmid previously restricted with BbsI.

Knock in plasmids were co-transfected with respective sgRNA plasmids into C2C12 myoblast by electroporation. Forty-eight hours (48 h) post-transfection, cells were screened with 1 μg/mL puromycin for two weeks with a medium change every 48 h. Subsisting cells were screened for clones expressing FLAG fusion protein both by genotyping and Western blot.

Fluorescence microscopy

C2C12 myoblast was transfected with 0.5 μg LDAMP1 with C-terminal fused GFP

(pEGFP-N1). Transfected cells were cultured on glass cover slips in a 24-well dish with empty plasmid as transfection control (pEGFP-N1). Cells were treated with OA-containing medium for 12 h prior to sample preparation for image capture. Twenty-four hours (24 h) post-transfection, cells were rinsed with cold PBS three times and treated with LipidTOX Red (1:1,000) for 30 min, and Hoechst 33342 (1:1,000) for 10 min. Cells were rinsed three times with PBS and the cover slips were placed on confocal dish for image capture using Olympus FV1200 confocal microscope (Olympus Corp, Lake Success, NY).

Fluorescence immunostaining

After incubation of OA for 12 h, C2C12 cells stably expressing GFP-FLAG and LDAMP1-GFP-FLAG on glass cover slips were rinsed with cold PBS three times and fixed with 4% (v/v) paraformaldehyde for 30 min and 50 mM NH_4Cl in PBS was used to quench residual aldehydes. The cells were then permeabilized with 0.1% Triton X-100 for 15 min. 3% BSA was used to block non-specific antibody binding. Cells were then incubated with primary antibody (anti-GFP or anti-FLAG, 1:100) for 2 h and subsequently with secondary antibody with Alexa Fluor 488-conjugated for 1 h. The cells were further stained by LipidTOX Red (1:1,000) for 30 min, and Hoechst 33342 (1:1,000) for 10 min. Images were captured with Olympus FV1200 confocal microscope.

Immunogold labeling

The localization of LDAMP1 was visualized by immunoelectron microscopy. In brief, cells stably expressing LDAMP1-GFP-FLAG were treated with 100 μM OA plus 1 μM MG132 for 12 h. Cells were collected by trypsin digestion and were fixed in 4% (v/v) paraformaldehyde overnight at 4°C. The cells were subsequently dehydrated in an ascending concentration series

of ethanol at room temperature, including 50%，70% and 80%, for 15 min each. The samples were incubated in a 2:1 mixture of LR White resin to 70% ethanol for 1 h, and in three changes of 100% LR White, for 1 h, 12 h and 1 h respectively. Cells in LR White resin were placed in a gelatin capsule to polymerize at 50 °C for 24 h. The resin block with cells were cut into 70 nm sections on Leica EM UC6 Ultramicrotome. The ultrathin sections were mounted on formvar-coated copper grid. For labeling, the grids were incubated on drop of PBS with 1% BSA and 0.15% glycine for 10 min, then transferred to drop of GFP polyclonal antibody (TaKaRa, Cat. No. 632592) at 1:50 for 1 h and subsequently in second antibody conjugated with 18-nm gold particles (Jackson, Cat. No. 111-215-144) at 1:100 for 1 h. After being stained with 1% uranyl acetate for 10 min, the sections were observed with Tecnai Spirit electron microscope (FEI, Netherlands).

Total cellular triacylglycerol quantitation

C2C12 cells were cultured in 6-well dishes until 75% cell density. Subsequently, cells were refreshed with or without 100 μM OA-supplemented DMEM medium for 12 h. Cells in each well were washed with PBS three times by using 2 mL PBS each time. To each well, 0.3 mL of 1% (v/v) Triton X-100 in PBS was dispensed and incubated at 37°C for 10 min. Cells in each well were collected and sonicated using 200 W, 9 s pulse on: 4 s pulse off setting for 5 times. Ten microliter (10 μL) volume of whole cell lysate was used for triacylglycerol analysis while protein served as internal control. Meanwhile, equal aliquot of whole cell lysate was used for protein analysis. Briefly, protein standard supplied along with BCA (Bicinchoninic acid Assay) Protein Kit was diluted according to instruction manual with maximum concentration of 2 mg/mL. Each concentration ranging from 0 mg/mL to 2 mg/mL was prepared in triplicate. Ten

microliters (10 μL) of each standard concentration was dispensed into 96-well dish while 190 μL of BCA chromogenic agent was added. For sample of unknown protein concentration, similar procedure was followed. Samples were incubated at 37ºC for 45 min and the optical density (OD) was read at 562 nm for protein. In the same vein, whole cell lysate sample was equally treated for triacylglycerol, however, the standard and chromogenic agent differ. Triacylglycerol standard ranged from 0 mg/dL to 40 mg/dL. Upon addition of chromogenic agent mix to sample, assay mixture was incubated at 37ºC for 10 min while the OD was read at 505 nm.

Statistical analyses

Data were presented as mean ± SEM unless specifically indicated. The statistical analyses were performed using GraphPad Prism 6. Comparisons of significance between groups were performed using unpaired Student *t*-test by Quickcalc.

Results

Enrichment of lipid droplet small molecular weight proteins and identification of the micropeptides

Previously we established a method for isolation and proteomic study of LDs and collectively identified many proteins of LDs from bacteria to humans(17, 24). But the proteins coded from sORFs have not been identified in LDs. In this study, we explored a proteomic-based strategy to discover the micropeptides in isolated LDs of myoblasts (Fig. 1A). To do so, a database was constituted for the non-annotated proteins and has been used to identify sORF-encoded peptides in mouse serum in our previous study, in which 54 novel micropeptides were

discovered and two were further verified with raised antibodies, suggesting the effectivity and reliability of this database(25). In brief, mouse GENCODE (vM4) and Ensembl transcripts (release 73) were used to annotate the noncoding transcripts in mouse. Then the resulting noncoding transcripts were translated to small ORF-encoded peptides, ranging from 8- to 100-amino acids, by the ORF Finder program and an in-house program. This database, containing a total of 525,792 putative ORFs, was then merged with mouse Uniprot database and Contamination database. The combined database was then trained and verified by current known micropeptides. This database was used in this study for proteomic MS data searching.

In order to study the novel tiny proteins on LDs of myoblasts, we carried out proteomic study on the isolated organelle. Thus, we firstly isolated LDs from C2C12 cells according to our well-established method(23, 24). Then the quality of the isolated LDs was analyzed using silver staining and Western blot methods. In the silver staining result, isolated LDs exhibited unique protein composition compared to other cellular fractions, such as total membrane (TM), cytosol (Cyto), and post-nuclear supernatant (PNS) (Fig. 1Ba, upper panel). The Western blot result showed the enrichment of the LD marker proteins, PLIN2, PLIN3, ACSL4, UBXD8, and CGI-58, on the isolated LDs (Fig. 1Ba, lower panel). Meanwhile, the cytosolic protein, GAPDH, was not detectable in the LD fraction. The endoplasmic reticulum protein, BIP, was mainly enriched in the membrane fraction. Some of the mitochondrial proteins were found on LDs, which is consistent with our previous result that there exist LD-anchored mitochondria (LDAM) in skeletal muscle cells(23, 27). Together, we obtained LD fraction from myoblasts with high quality for further analysis.

After collection of isolated LDs with high quality, we used 12% Bis-Tris gel to separate

and enrich the proteins with molecular weight less than 17 kDa and the gel was stained by Colloidal Blue (Fig. 1Bb). We carried out the isolation of LDs for 4 times and those LDs were applied for the enrichment and the subsequent proteomic study using our database.

The similarity between the three independent LD fractions showed the reproducibility of our system to isolate and enrich the tiny proteins (Fig. 1Bb). More important, the large area below 17 kDa showed small proteins or degraded proteins in isolated LDs and this was the first time to analyze them specifically in detail. The proteins coded by classical genes were listed in Supplemental Table S1 according to their sub-cellular location/function. The total number of proteins identified were 136, 44 of which had more than two unique peptides. 52% of the identified proteins have been reported previously in LD proteomes or have been confirmed to be localized on LDs, indicating the reliability of our system to purify LD and enrich small proteins as well as the proteomic study. Among the identified proteins, the most abundant proteins were ribosome proteins (around 23%, Supplemental Fig. S1). Mitochondrial proteins constituted about 20% of the proteins, which is consistent with both our previous and recent study and suggests a tight anchoring between LDs and mitochondria in the oxidative tissues(27). Besides, another major group was consisted of proteins involved in signaling pathway (up to 14%), supporting the hypothesis that LD functions in signal transduction(17). 7% of the proteins were involved in protein degradation in the ubiquitin-proteasome system, suggesting the role of LDs in protein degradation, which is consistent with previously report(17).

More important, by using our database, a total amount of 15 micropeptides were identified, which arose from six transcript classes, including retained intron, processed transcript, antisense, TEC (to be experimentally confirmed), lincRNA, and nonsense mediated decay (Fig.

1C). Among those peptides, those arose from retained intron and processed transcript occupied 60%, which indicated that LD-associated micropeptides derived from those two classes may be more than from other transcript classes annotated to be noncoding. Interestingly, a transcript category, TEC (to be experimentally confirmed), which has been specifically created to highlight the coding potential of region that require experimentally validation, was found to code two of the identified peptides.

The putative peptides were named as LD-associated micropeptides, LDAMP, and listed in Table 1 in a descending order of their XCorr scores, an important search algorithm to assess the quality of a peptide-spectrum match, which indicates the peptide-identification reliability in mass spectrometry.

Degradation of LDAMP1 is regulated by proteasome

Among those peptides, LDAMP1, with highest XCorr score, was examined further in this study (Table 1). The murine *LDAMP1* is located to the exon 4 of ubiquitin specific peptidase-like 1 (*USPL1*) on chromosome 5, containing 261 nucleotides (starting from 149,194,024 to 149,914,284) (Fig. 2A). However, it is translated in the +1 reading frame, resulting in the 87-amino acid peptide, LDAMP1. To gain insights into the regulation of LDAMP1, this peptide was overexpressed as a GFP-FLAG fusion protein. It is known that micropeptides are always expressed in low abundance and unstable(7). Besides, LD resident proteins, such as PLIN1 and PLIN2, are degraded through the proteasome-dependent pathway(28, 29). As described above, LDAMP1 was identified in isolated LD fraction. Thus, we wondered whether LDAMP1 was also degraded by proteasome. MG132, a well-known specific proteasome inhibitor, was used.

Culture medium was supplemented with or without MG132 in both control cells expressing GFP-FLAG (GF) and cells stably expressing LDAMP1-GFP-FLAG (LGF), and then FLAG immunoprecipitation was performed followed by GFP Western blot. After addition of MG132, as indicated, a band around 40 kDa, representing the molecular weight of LDAMP1 fused with GFP and FLAG, was detected in whole cell lysate (Fig. 2B, upper panel, lane 8). This finding was further confirmed by FLAG immunoprecipitation result. After FLAG immunoprecipitation, the intensity of the 40 kDa band was increased dramatically after addition of MG132 shown by GFP Western blot result (Fig. 2B, upper panel, lane 4) and silver staining result (Fig. 2B, lower panel, lane 4, red star). Together, LDAMP1 was degraded through proteasome.

Therefore, in order to identify the potential ubiquitin acceptor site(s) on LDAMP1, all four lysine residues were mutated to alanine independently. Then we evaluated respective contribution of each residue to peptide stability (Fig. 2C). LDAMP1 and LDAMP1 mutant fused with GFP was stably expressed in C2C12 cells and evaluated for peptide level by Western blot. Lysine residue 7, 49, and 61 appeared to be, in part, important for LDAMP1 stability, shown by the increase of GFP signal compared to cells expressing WT LDAMP1-GFP (Fig. 2C, lanes 3-5 vs. lane 2). Meanwhile, mutant of lysine residue 80 appeared to be unstable and the GFP signal was not detectable similar to WT (Fig. 2C, lane 6 vs. lane 2), suggesting lysine residue 80 may be not involved in ubiquitination. However, when treated with MG132, all GFP signals were detected (Fig. 2C, lanes 9-12). Taken together, these results further support that LDAMP1 is regulated by proteasome protein degradation and suggest that multiple lysine residues on LDAMP1 are potential sites of regulation by the ubiquitin system.

LDAMP1 is localized on lipid droplets in C2C12

LDAMP1 was identified in the LD proteome and behaved as LD resident proteins in the sense of degradation. We next set out to determine whether it indeed localizes on LDs using biochemical and morphological methods. Firstly, we transiently expressed it with GFP tag on C terminus. Figure 3A shows very obvious GFP ring structures around LDs presenting by LipidTOX Red staining, indicating that it was co-localized with LDs.

The localization of the LDAMP1 in the LDAMP1-GFP-FLAG stable cells was further determined by immunofluorescence staining of GFP or FLAG (green signal, Fig. 3B). Consistent with the transiently transfected cells, the LDAMP1 was also localized to LDs in stable cells.

More so, the localization of LDAMP1 was confirmed through cell fractionation. Western blot GFP of cellular fractions from C2C12 cells expressing LDAMP1-GFP-FLAG showed that GFP was co-fractionated with known LD-associated proteins, PLIN3 and ACSL3 (Fig. 3C, upper panel, lane 2), while GFP in control cells expressing GFP-FLAG co-fractioned with cytosolic protein Actin (Fig. 3C, upper panel, lane 5). Meanwhile, almost no signal of GFP could be detected in the membrane traction, which contains endoplasmic reticulum, indicated by marker BIP and PDI, and mitochondrion, indicated by OPA1 and TIM23 (Fig. 3C, upper panel, lane 4). This biochemical result also revealed that LDAMP1 was enriched in LD fraction.

Furthermore, the localization was also displayed using immuno-gold labeling method. Figure 3D showed that the 18-nm colloidal gold particles labeling GFP in LDAMP1-GFP-FLAG stable cells mainly surrounded LDs (black dots).

All these results demonstrated that LDAMP1 localizes to LDs in C2C12 cells.

LDAMP1 is endogenous expressed in C2C12 cells

The next essential question to be addressed is whether LDAMP1 is expressed endogenously. First, LDAMP1 was analyzed computationally for conservation. Alignment of LDAMP1 amino acid sequence against in silico-translated short ORF in USPL1 sequence from 11 different mammalian species (Fig. 4A) revealed that this peptide is conserved and 48% of amino acid residues appeared to be invariant between human and mouse. This result demonstrated that this peptide is conserved across mammals. Second, part of amino acids in the peptide was identified experimentally by mass spectrometry identification (listed in Table 1). Third, a database for analysis of alternatively translated proteins in human RefSeq transcripts also identified LDAMP1 homolog in human within the USPL1 coding sequence and later also in several other mammals(30). These data suggested the possibility of endogenous translation of LDAMP1. Therefore, we used CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas9 (CRISPR-associated protein 9)-mediated gene editing strategy to validate its endogenous expression. 3×FLAG-HA sequence was integrated at the C-terminus of and in-frame with LDAMP1 ORF located within USPL1 exon 4 in C2C12 cells (Fig. 4B).

Correct insertion of 3×FLAG-HA (123 bp) was verified by polymerase chain reaction (PCR) using a pair of primers targeting nucleotide sequence upstream and downstream of LDAMP1 stop codon. Agarose gel electrophoresis of amplicon from the target genomic region indicated presence of two bands in LDAMP1 knock-in cell (Fig. 4C, lane 2). The lower band, adjacent to WT, corresponded to endogenous non-edited allele and the upper band corresponded to precisely edited allele containing 3×FLAG-HA, indicating its endogenous expression at mRNA level. Endogenous expression at protein level was verified by Western blot subsequently.

Since the stability of overexpressed LDAMP1 is regulated by proteasome shown above (Fig. 2), 3×FLAG-HA tagged LDAMP1 knock-in C2C12 cells were cultured with or without 1 μM MG132 for 12 h. Whole cell lysate from WT and LDAMP1 knock-in cell lines were subjected to FLAG immunoprecipitation in order to concentrate the peptide. FLAG immunoprecipitation sample was further analyzed with anti-FLAG and the result showed a distinct band in the neighborhood of 17 kDa in MG132-treated LDAMP1 knock-in cells (Fig. 4D, upper panel, lane 4). This unique band lies in the predicted molecular weight region for 3×FLAG-HA knock-in LDAMP1 (10.2 kDa of LDAMP1 and 4.8 kDa of 3×FLAG and HA tag, encoded in 34 amino acids, plus a 7-amino acid linker). Thus, LDAMP1 is translated endogenously.

LDAMP1 reduces triacylglycerol storage and mediates insulin sensitivity in C2C12 cells

LDAMP1 was the first peptide derived from noncoding transcript found on LDs. We then tried to determine its biological function on LD dynamics. To do so, C2C12 cells stably expressing GFP-FLAG and LDAMP1-GFP-FLAG were incubated in the presence or absence of 100 μM oleate (OA) for 12 h. Total triacylglycerol (TAG) level in GFP-FLAG control and LDAMP1-GFP-FLAG cell lines treated with vehicle (Fig. 5A) showed no significant differences. However, TAG level was declined significantly in OA-treated LDAMP1 stably expressed cells relative to GFP-FLAG control, suggesting that LDAMP1 negatively regulates OA-mediated TAG storage. OA is mainly incorporated into TAG in C2C12 cell and is proved in our and other groups' studies to reverse the inhibitory effect of saturated fatty acid on insulin signaling by converting the active lipids produced by saturated fatty acid to TAG(22, 31, 32). Thus, we wondered whether LDAMP1 affects insulin signaling. In agreement with our previous

report, PA treatment reduced the insulin signal represented by Akt phosphorylation (p-Akt) in WT C2C12 cells (Fig. 5B, lane 6 vs lane 2), while OA itself had no significant effect (Fig. 5B, lane 4 vs lane 2). However, when the WT cells were incubated with both PA and OA, the PA-reduced p-Akt signal was recovered (Fig. 5B, lane 8 vs lanes 6 and 2). In LDAMP1 stably expressed cells, PA still reduced p-Akt signal similarly as in WT (Fig. 5B, lane 14 vs lane 10), but OA could not rescue PA-reduced insulin signal, Akt phosphorylation (Fig. 5B, lane 16 vs lane 10). Therefore, LDAMP1 functioned to reduce the ability of TAG storage and, in turn, affect the insulin sensitivity in C2C12 cells.

Discussion

Besides sequences encoding classical long proteins, much progress has been made in recent years in identifying the full set of sequences that can be translated in the cells. Many of them are found to be small ORFs that potentially encode functional micropeptides less than 100 amino acids. Some of those peptides have been found to play significant biological roles in some cellular processes. LD, the prominent reservoir for lipid energy and membrane building blocks in cells, plays a crucial role in lipid metabolism and storage, and is actively involved in metabolic diseases. However, no micropeptide encoded by noncoding sORFs has been reported on LDs.

We aimed to find out whether there is micropeptide on LDs. A bottom-up strategy is adopted here, that is, the study starts with the isolated organelle directly other than from the whole cell. In combination with mass spectrometry approach, possible LD-associated micropeptides are searched. Here, 15 novel micropeptieds were detected in isolated LDs from

C2C12 cells and one of them had been validated to express endogenously. This is the first report to study the tiny proteins from noncoding sORFs on LDs directly. Similarly, this strategy may be applicable in searching micropeptides from other cellular organelles.

This is also the first systematic study of LD-associated micropeptides. Several traits are displayed by the identified micropeptides in isolated LDs. First, distribution pattern of those peptides (Fig. 1C) across different transcript classes indicated that the micropeptides associated with LDs may be more common in retained_intron and processed_transcript. Second, those micropeptides are in very low abundance. Unlike the classically translated proteins, such as ribosome proteins, half of which were detected more than 3 fragments in mass spectrometry, all of the identified micropeptides matched only one fragment (Table 1). On the other hand, even stably expressed, LDAMP1 could not be easily detected in the whole cell lysate (Fig. 2B, upper panel, lane 7). Only if we enriched the fusion protein its expression could be detected (Fig. 2B, upper panel, lane 3).

In this study, LDAMP1 was chosen to studied further due to its highest XCorr score. This protein was verified to localize on LDs in C2C12 cells through several approaches as described. Endogenous expression of this protein is finally verified by CRISPR/Cas9-mediated gene editing approach. Therefore, this is an authentic and novel small protein found on LDs. To date, more than 100 proteins have been reported to be associated with LDs. However, many of fundamental questions in LD biology remain unclear since LDs have been recognized as an important organelle for only the last two decades(33). Therefore, the proteins derived from noncoding transcripts would be a brand-new territory in study of LDs.

Interestingly, the degradation of LDAMP1 was found to be proteasome-dependent, similar

to the degradation of LD resident proteins, PLIN1 and PLIN2(28, 29). This further supported the proposal that proteasomal regulation is a common way to modulate LD-associated proteins(29). We have identified multiple lysine residues on LDAMP1 important for ubiquitination in proteasome system. However, besides targeting proteins for degradation, ubiquitination plays other important roles in DNA repair, signal transduction and endocytosis(34). Thus, it remains unclear whether multiple ubiquitination of LDAMP1 serves other purposes besides degradation or such signature is common to peptides derived from small ORFs to facilitate their rapid turnover.

Even though no specific LD targeting signal has been found yet, LD-associated proteins that directly target to LDs usually contain a hydrophobic domain or amphipathic α-helix(17, 18). For instance, LD resident protein in plants, oleosin, is proposed to comprise an N-terminal amphipathic domain, a central hydrophobic domain, and a C-terminal amphipathic α-helical domain. The central domain anchors oleosin into the hydrophobic core of LD and N- and C-terminal domains are proposed to reside on the monolayer phospholipid membrane of LDs(35, 36). Using bioinformatic tools, we analyzed the possible secondary structure of LDAMP1 (Supplemental Information). LDAMP1 was predicted to have one amphipathic α-helix on each terminal, spanned by a hydrophobic α-helix. We proposed that LDAMP1 targeting LDs may be in a similar way to oleosin, with the hydrophobic helix anchoring LDAMP1 in LD hydrophobic core and two amphipathic α-helices binding the monolayer phospholipid membrane of LDs (Supplemental Fig. S2). We are working on its exact targeting mechanism.

LDAMP1 is identified as an alternative protein derived as an out-of-frame ORF in the USPL1 coding sequence. A previous work suggests that dual coding is more prevalent in low

expressed genes, because genes which are highly expressed are optimized for efficient translation and their coding sequences are too restrained to accommodate additional coding information(37). Previous work on USPL1 indicated that the protein is translated in low abundance in all cell lines examined(38), which further supports the fact that genes encoding low abundant proteins might be a reservoir of alternative proteins. Interestingly, while USPL1 is located to the nucleus and affects coilin localization(38), LDAMP1 is shown to localize on LDs and influence lipid metabolism, suggesting a variety of functions for alternatively translated proteins.

Skeletal muscle consumes the majority of glucose in humans and plays crucial role in maintaining glucose homeostasis(39). Excessive lipid storage in myocytes is closely related to insulin resistance(19). To determine the mechanism linking TAG storage and insulin sensitivity in skeletal muscle is essential in study of insulin resistance. In our and others previous studies, PA is found to inhibit insulin sensitivity through converting into active lipid molecules and OA can protect insulin signaling from the inhibitory effects of PA in myoblasts by switching lipids into TAG(22, 31, 32). As a LD-associated protein, LDAMP1 was found to reduce the storage of TAG when myocytes were loaded with OA. In turn, PA-reduced insulin sensitivity could not be rescued well by OA. Hence, we propose that this protein could block OA incorporating into TAG thus affect insulin sensitivity in myoblast. However, how it works is unclear. One should work on identifying its potential interacting partner(s), since several known micropeptides are found to exert functions through engaging with and modulating larger regulatory proteins(7). Meanwhile, its effect on insulin sensitivity of skeletal muscle in mice will also be investigated. Taken together, this study may shed light on brand-new factors, LD-associated micropeptides

and/or their functional machineries, which can affect LD dynamics in general and insulin signaling regulation in skeletal muscle.

Summary

In this study, we have shown the first time proteins derived from noncoding transcripts on LDs and one of these, LDAMP1, has been verified to localize on LDs. This protein reduces the TAG storage and the storage-related insulin sensitivity in myoblasts. Those findings suggest the vast pool of putative proteins derived from noncoding transcripts serves as candidates for screening new LD-associated proteins as well as factors modulating LD dynamics and lipid metabolism.

Data availability

All data are contained within the manuscript. Nucleotide sequence of ldamp1 has also been submitted to GenBank with accession number of MT318098.

Acknowledgments

The authors thank Dr. John Zehmer for his critical reading and useful suggestions. The authors also thank Mr. Lianwan Chen for his help in immunogold labeling experiment and Ms. Chen Ye for her critical reading. This work was supported by the National Key R&D Program of China (Grant No. 2018YFA0800900, 2016YFA0500100 and 2018YFA0800700), National Natural Science Foundation of China (Grant No. 31671402, 91954108, 91857201, 31671233, 31701018 and U1702288). This work was also supported by the "Personalized Medicines—

—Molecular Signature-based Drug Discovery and Development", Strategic Priority Research Program of the Chinese Academy of Sciences, Grant No. XDA12040218.

Author contributions

S.Z. and P.L. designed the project. T.H., A.B., and S.X. performed the experiments. Y.D. helped with the immunoprecipitation analysis. K.X. and O.O. helped with triacylglycerol assay. A.H.M. helped with bioinformatic analysis. J.W. helped with MS study. S.Z., P.L., and A.B. wrote the manuscript. All authors have read and approved the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

1. Lander ES. Initial impact of the sequencing of the human genome. Nature. 2011;470(7333):187-97.

2. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. Science. 2005;309(5740):1559-63.

3. Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science. 2007;316(5830):1484-8.

4. Ruiz-Orera J, Alba MM. Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. Trends Genet. 2019;35(3):186-98.

5. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

6. Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, et al. An atlas of human long non-coding RNAs with accurate 5' ends. Nature. 2017;543(7644):199-204.

7. Makarewich CA, Olson EN. Mining for Micropeptides. Trends Cell Biol. 2017;27(9):685-96.

8. Sousa ME, Farkas MH. Micropeptide. PLoS Genet. 2018;14(12):e1007764.

9. Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, et al. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. Science. 2014;343(6172):1248636.

10. Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell. 2015;160(4):595-606.

11. Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, et al. The mitochondrial-derived peptide MOTS-c promotes metabolic homeostasis and reduces obesity and insulin resistance. Cell Metab. 2015;21(3):443-54.

12. Matsumoto A, Pasut A, Matsumoto M, Yamashita R, Fung J, Monteleone E, et al. mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. Nature. 2017;541(7636):228-32.

13. Martin S, Parton RG. Lipid droplets: a unified view of a dynamic organelle. Nat Rev Mol Cell Biol. 2006;7(5):373-8.

14. Zhang C, Liu P. The lipid droplet: A conserved cellular organelle. Protein Cell.

2017;8(11):796-800.

15. Zehmer JK, Huang Y, Peng G, Pu J, Anderson RG, Liu P. A role for lipid droplets in inter-membrane lipid traffic. Proteomics. 2009;9(4):914-21.

16. Farese RV, Jr., Walther TC. Lipid droplets finally get a little R-E-S-P-E-C-T. Cell. 2009;139(5):855-60.

17. Zhang C, Liu P. The New Face of the Lipid Droplet: Lipid Droplet Proteins. Proteomics. 2019;19(10):e1700223.

18. Xu S, Zhang X, Liu P. Lipid droplet proteins and metabolic diseases. Biochim Biophys Acta. 2018;1864(5 Pt B):1968-83.

19. Samuel VT, Shulman GI. Mechanisms for insulin resistance: common threads and missing links. Cell. 2012;148(5):852-71.

20. Crunk AE, Monks J, Murakami A, Jackman M, Maclean PS, Ladinsky M, et al. Dynamic regulation of hepatic lipid droplet properties by diet. PLoS ONE. 2013;8(7):e67631.

21. Bosma M, Hesselink MK, Sparks LM, Timmers S, Ferraz MJ, Mattijssen F, et al. Perilipin 2 improves insulin sensitivity in skeletal muscle despite elevated intramuscular lipid levels. Diabetes. 2012;61(11):2679-90.

22. Peng G, Li L, Liu Y, Pu J, Zhang S, Yu J, et al. Oleate blocks palmitate-induced abnormal lipid distribution, endoplasmic reticulum expansion and stress, and insulin resistance in skeletal muscle. Endocrinology. 2011;152(6):2206-18.

23. Zhang H, Wang Y, Li J, Yu J, Pu J, Li L, et al. Proteome of skeletal muscle lipid droplet reveals association with mitochondria and apolipoprotein a-I. J Proteome Res. 2011;10(10):4757-68.

24. Ding Y, Zhang S, Yang L, Na H, Zhang P, Zhang H, et al. Isolating lipid droplets from multiple species. Nat Protoc. 2013;8(1):43-51.

25. Deng Y, Bamigbade AT, Hammad MA, Xu S, Liu P. Identification of small ORF-encoded peptides in mouse serum. Biophys Rep. 2018;4(1):39-49.

26. Wang Z. Epitope tagging of endogenous proteins for genome-wide chromatin immunoprecipitation analysis. Methods Mol Biol. 2009;567:87-98.

27. Cui L, Mirza AH, Zhang S, Liang B, Liu P. Lipid droplets and mitochondria are anchored during brown adipocyte differentiation. Protein Cell. 2019;10(12):921-6.

28. Masuda Y, Itabe H, Odaki M, Hama K, Fujimoto Y, Mori M, et al. ADRP/adipophilin is degraded through the proteasome-dependent pathway during regression of lipid-storing cells. J Lipid Res. 2006;47(1):87-98.

29. Xu G, Sztalryd C, Londos C. Degradation of perilipin is mediated through ubiquitination-proteasome pathway. Biochim Biophys Acta. 2006;1761(1):83-90.

30. Vanderperre B, Lucier JF, Roucou X. HAltORF: a database of predicted out-of-frame alternative open reading frames in human. Database (Oxford). 2012;2012:bas025.

31. Yu J, Li Y, Zou F, Xu S, Liu P. Phosphorylation and function of DGAT1 in skeletal muscle cells. Biophys Rep. 2015;1:41-50.

32. Palomer X, Pizarro-Delgado J, Barroso E, Vazquez-Carrera M. Palmitic and Oleic Acid: The Yin and Yang of Fatty Acids in Type 2 Diabetes Mellitus. Trends Endocrinol Metab. 2018;29(3):178-90.

33. Welte MA. Expanding roles for lipid droplets. Curr Biol. 2015;25(11):R470-81.

34. Passmore LA, Barford D. Getting into position: the catalytic mechanisms of protein

ubiquitylation. Biochem J. 2004;379(Pt 3):513-25.

35. Huang AH. Oleosins and oil bodies in seeds and other organs. Plant Physiol. 1996;110(4):1055-61.

36. Shao Q, Liu X, Su T, Ma C, Wang P. New Insights Into the Role of Seed Oil Body Proteins in Metabolism and Plant Development. Front Plant Sci. 2019;10:1568.

37. Michel AM, Choudhury KR, Firth AE, Ingolia NT, Atkins JF, Baranov PV. Observation of dually decoded regions of the human genome using ribosome profiling data. Genome Res. 2012;22(11):2219-29.

38. Schulz S, Chachami G, Kozaczkiewicz L, Winter U, Stankovic-Valentin N, Haas P, et al. Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. EMBO Rep. 2012;13(10):930-8.

39. Shulman GI, Rothman DL, Jue T, Stein P, DeFronzo RA, Shulman RG. Quantitation of muscle glycogen synthesis in normal subjects and subjects with non-insulin-dependent diabetes by 13C nuclear magnetic resonance spectroscopy. N Engl J Med. 1990;322(4):223-8.

Table 1. Micropeptides on C2C12 lipid droplets.

Name	Accession	Annotation	XCorr	Charge	Peptide sequence (with detected peptide in red)	M.W. (kDa)
LDAMP1	ENSMUSG00000041264.12	Uspl1|nonsense_mediated_decay	1.84	2	MVNIMEKCVMTSQQASQIPVLTRILPAQLLLWSRVRLWRLMMWLWPLQKILLQLVSHLSWKCQLRADVYHFARLCVSSGKTPRPCAG	10.2
LDAMP2	ENSMUSG00000093765.3	Gm20658|processed_transcript	1.47	3	MPLPSKMEERALLAPLRLDGDSLQAELAMVLMTGAKRMRSTTVHSHCFKSDEACRHKAFVHCVKVVLVLFKCLFPCPHFLFQSGYGIVMLRSINYPQP	11.1
LDAMP3	ENSMUSG00000032595.6	Cdhr4|retained_intron	1.31	3	MLRWTVTHLGLASSIQLPSWWSMVVNPK	3.2
LDAMP4	ENSMUSG00000039716.8	Dock3|processed_transcript	1.17	2	MKVILSRRNQKSFQVWIS	2.2
LDAMP5	ENSMUSG00000102700.1	TEC RP24-396E16.1|TEC	1.08	2	MSLITPRFVGKIGAGSRFSSSLFFLQPIELRTPCSNSKGEKPEVS	4.9
LDAMP6	ENSMUSG00000085034.1	Gm11240|processed_transcript	1.08	2	MLEERHTTQTQMACGNSGLRLTVTSVTSLLQLQQTSKLNVFGRGHGYKKMMECTLRTIWF	6.9
LDAMP7	ENSMUSG00000025909.12	Sntg1|retained_intron	0.81	2	MKKVHLNYLCIFIFRLTCIIRCF	2.9
LDAMP8	ENSMUSG00000051396.10	Hspa14|retained_intron	0.68	2	MQSLVVNVQFFFRWLDWQQNKVE	2.9
LDAMP9	ENSMUSG00000060212.9	Pcnxl2|retained_intron	0.68	2	MEVYSMVQNIMCARLLSGISSLYVDLQSHPGPRRDC	4.1
LDAMP10	ENSMUSG00000036564.12	Ndrg4|processed_transcript	0.63	2	MHLLRKGCWLLVPRLSVTPNWTQPLRPS	3.3
LDAMP11	ENSMUSG00000057359.6	antisense Gm17494|antisense	0.55	3	MRSSCKLFSQLVISKGGTTMDGATPSWWSWVL	3.6
LDAMP12	ENSMUSG00000103146.1	TEC RP23-274C16.4|TEC	0.45	3	MLEFVFYIYMHCMYIHYFEKSLF	3.1
LDAMP13	ENSMUSG00000085162.3	lincRNA Gm12295|lincRNA	0.39	2	MVDLLKSNT	1
LDAMP14	ENSMUSG00000037847.9	Nmrk1|retained_intron	0.25	2	MLQMARWFFKWTWERKLWGAPSPRQSAASEEGQDFPGKEEVGSAVFPFIFRS	6.1
LDAMP15	ENSMUSG00000085782.1	antisense Gm15624|antisense	0.19	2	MCTGTHTQTHMCRYHTTFTHT	2.5

Figures and figure legends

Figure 1

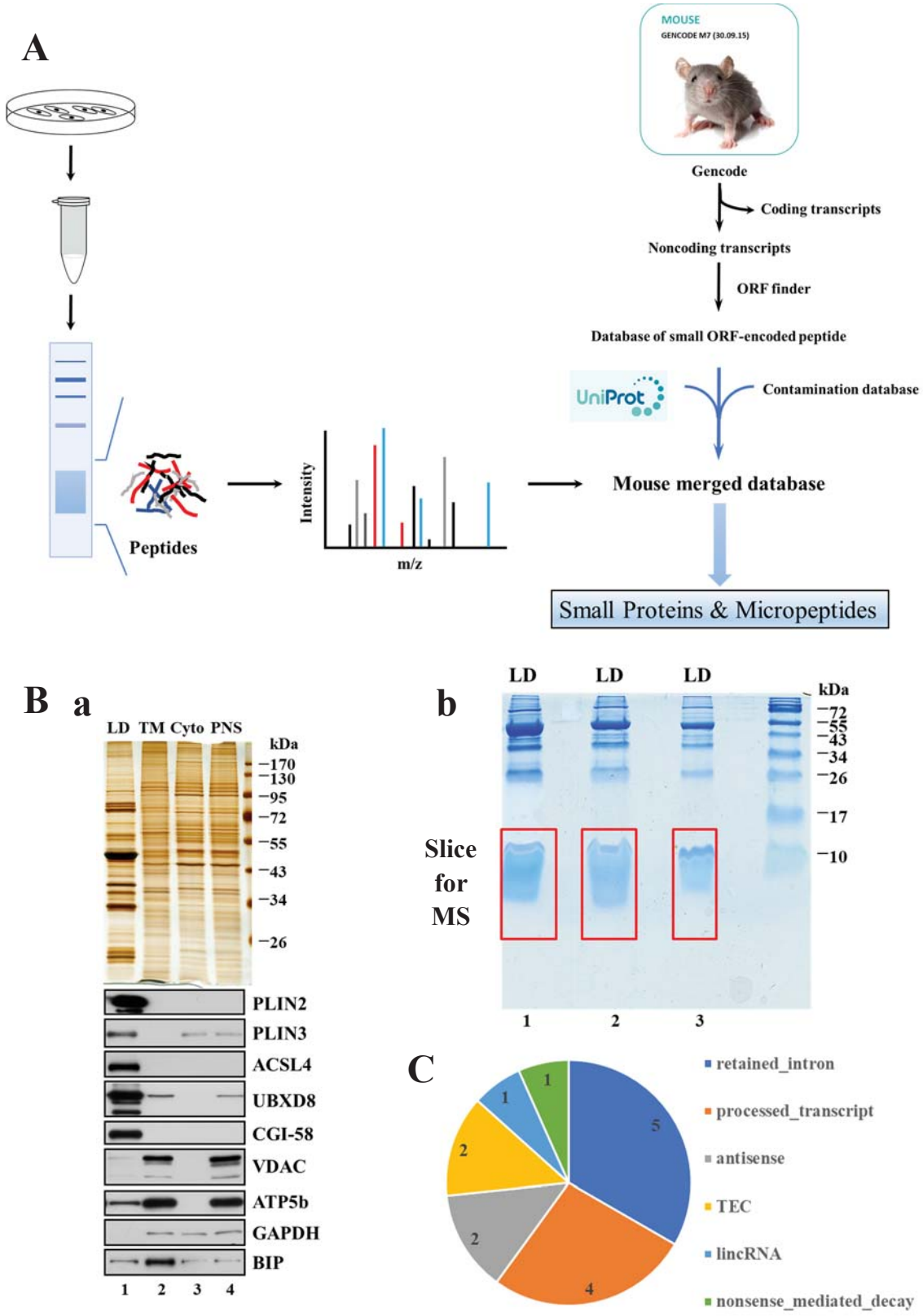


Figure 1. Identification of micropeptides from C2C12 lipid droplets.

A Workflow for the study. Briefly, after isolation of LDs, the small proteins were enriched and the digested peptides were subjected to mass spectrometry analysis. The obtained peptide mass spectra were searched against our home-made database previously reported. The flowchart for construction of database is adopted and modified from Figure 1 in our previous work(25). **B a** Quality verification of the isolated LDs. The quality of the isolated LDs from C2C12 were verified by silver staining and Western blot analysis. After isolation, the LD proteins were extracted and subjected to silver staining. Meanwhile, the marker proteins of LD (PLIN2, PLIN3, ACSL4, UBXD8 and CGI-58), endoplasmic reticulum (BIP), mitochondrion (VDAC, ATP5b), as well as cytosol (GAPDH) were analyzed by Western blot. TM, total membrane; Cyto, cytosol; PNS, post-nuclear supernatant. **b** Enrichment of the LD proteins less than 17 kDa. The LD proteins were separated by 12% Bis-Tris gel and the proteins less than 17 kDa were carefully cut for MS study. **C** Classification of LD-associated micropeptides derived from noncoding RNA. The 15 putative LD-associated micropeptides were grouped according to transcript of origin. TEC, to be experimentally confirmed.

Figure 2

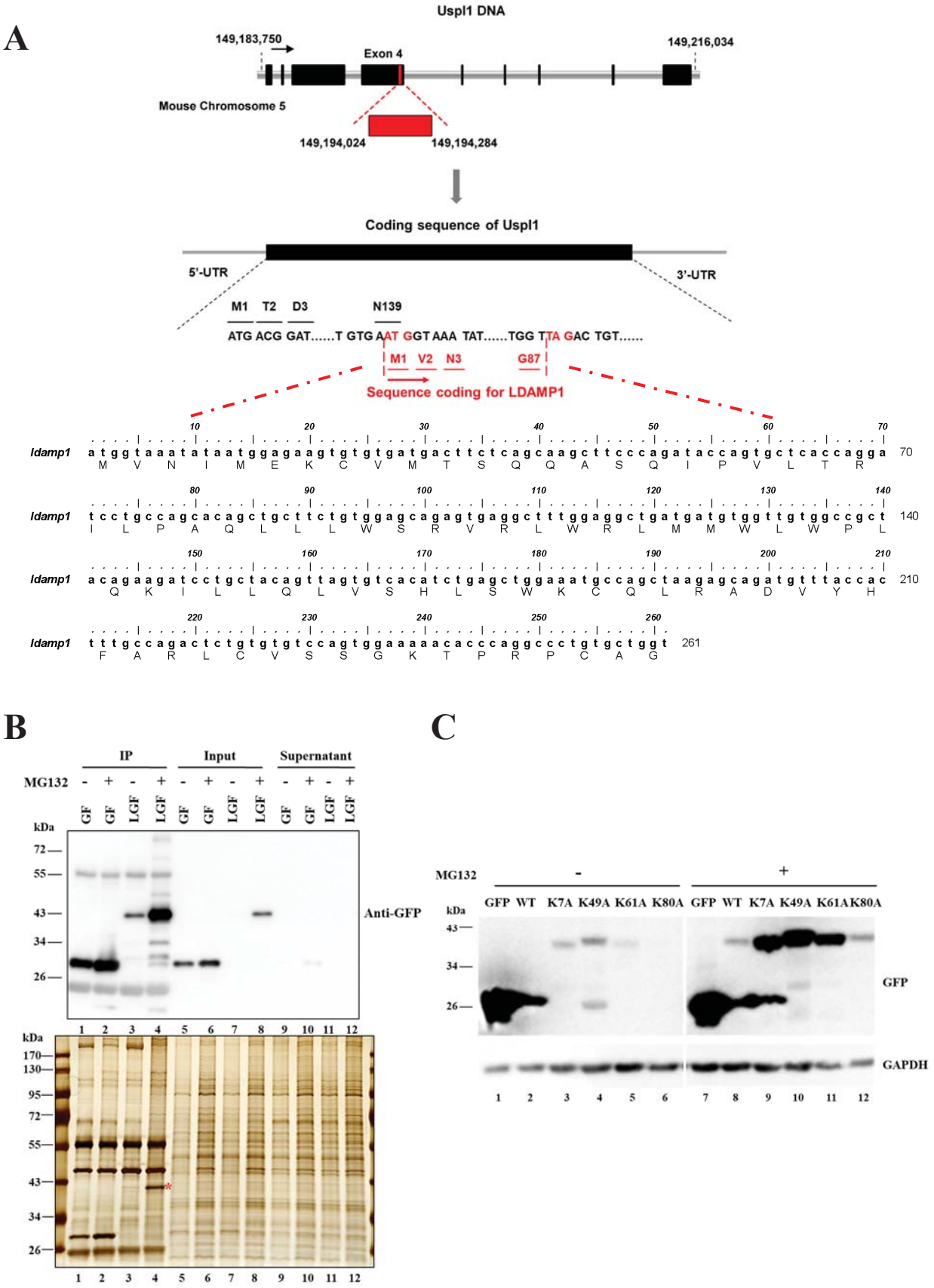

A
Uspl1 DNA
149,183,750
Exon 4
149,216,034
Mouse Chromosome 5
149,194,024
149,194,284
Coding sequence of Uspl1
5'-UTR
3'-UTR
M1 T2 D3
N139
ATG ACG GAT.....T GTG AAT GGT AAA TAT.....TGG TTA GAC TGT.....
M1 V2 N3
G87
Sequence coding for LDAMP1
ldamp1 atggtaaatataatggagaagtgtgtgatgacttctcagcaagcttcccagataccagtgctcaccagga 70
M V N I M E K C V M T S Q Q A S Q I P V L T R
ldamp1 tcctgccagcacagctgcttctgtggagcagagtgaggctttggaggctgatgatgtggttgtggccgct 140
I L P A Q L L L W S R V R L W R L M M W L W P L
ldamp1 acagaagatcctgctacagttagtgtcacatctgagctggaaatgccagctaagagcagatgtttaccac 210
Q K I L L Q L V S H L S W K C Q L R A D V Y H
ldamp1 tttgccagactctgtgtgtccagtggaaaaacacccaggccctgtgctggt 261
F A R L C V S S G K T P R P C A G
B
IP
Input
Supernatant
MG132
GF
LGF
kDa
72
55
43
34
26
Anti-GFP
170
130
95
C
MG132
GFP WT K7A K49A K61A K80A
GFP
GAPDH

Figure 2. LDAMP1 is degraded by proteasome.

A Schematic representation of the locus and translation of LDAMP1. In mice, *LDAMP1* (red) is located in the exon 4 of *USPL1* (black) on chromosome 5 and LDAMP1 was translated in the +1 frame. The nucleotide and amino acid sequence are shown (using BioEdit). **B** Effect of MG132 on the expression of LDAMP1. Cells stably expressing LDAMP1-GFP-FLAG (LGF) and GFP-FLAG control cells (GF) were treated with or without 1 μM MG132 for 12 h. Cells were washed with PBS, lysed with 1 mL TETN solution and probed with FLAG beads to precipitate FLAG tagged protein. Precipitate was eluted in sample buffer and separated on 10% SDS-PAGE. Silver stained gel (lower panel) indicated protein profile in each sample and served as protein loading control. Separated protein was blotted and probed with GFP antibody (upper panel). GF, GFP-FLAG cells; LGF, LDAMP1-GFP-FLAG cells; IP, immunoprecipitation. **C** Analysis of potential ubiquitin acceptor site(s) on LDAMP1. Lysine residues in LDAMP1 were mutated to alanine using LDAMP1-GFP as template. Cells stably expressing mutated LDAMP1 and control cells were treated with or without 1 μM MG132 for 12 h. Whole cell lysate were probed with GFP and GAPDH, the loading control. WT, cells expressing LDAMP1-GFP.

Figure 3

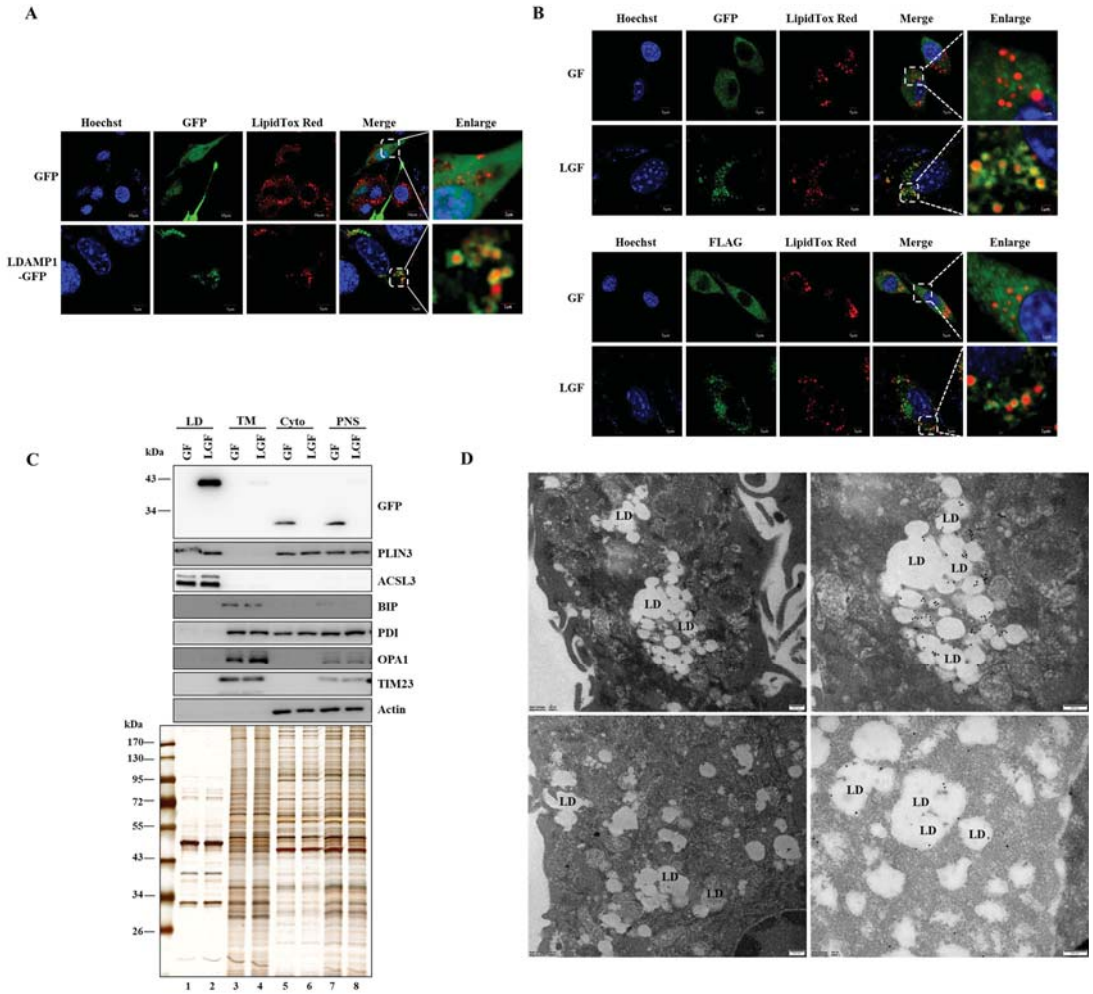

A
Hoechst
GFP
LipidTox Red
Merge
Enlarge
GFP
LDAMP1
-GFP
B
Hoechst
GFP
LipidTox Red
Merge
Enlarge
GF
LGF
Hoechst
FLAG
LipidTox Red
Merge
Enlarge
GF
LGF
C
LD
TM
Cyto
PNS
GF
LGF
kDa
43
34
GFP
PLIN3
ACSL3
BIP
PDI
OPA1
TIM23
Actin
170
130
95
72
55
43
34
26
1 2 3 4 5 6 7 8
D
LD

Figure 3. LDAMP1 localizes to lipid droplets in C2C12 cells.

The localization of LDAMP1 in C2C12 cells was studied using several methods. **A** The localization of transiently expressed-LDAMP1. LDAMP1-GFP was transiently expressed in C2C12 cells and treated with 100 μM OA for 12 h. Twenty-four hours post-transfection, cells were treated with LipidTOX Red and Hoechst (blue) to stain LDs and nucleus respectively. The cells were viewed using confocal microscope. In upper panel, Bar = 10 μm, except for Bar = 2 μm in enlarge image. In lower panel, Bar = 5 μm, except for Bar = 1 μm in enlarge image. **B** The localization of stably expressed-LDAMP1. C2C12 cells stably expressing GFP-FLAG and LDAMP1-GFP-FLAG were treated with 100 μM OA for 12 h. Then the cells were analyzed by immunofluorescence staining for GFP or FLAG with green-fluorescent signal to show the localization of stably expressed-LDAMP1. Subsequently, the cells were stained with LipidTOX Red for LD and Hoechst for nucleus. The cells were viewed using confocal microscope. Bar = 5 μm, except for Bar = 1 μm in enlarge image. **C** Localization of LDAMP1 demonstrated by cell fractionation. C2C12 cells stably expressing LDAMP1-GFP-FLAG and GFP-FLAG control cells were cultured in 100 μM OA supplemented medium for 12 h. Cells were harvested for cell fractionation. The proteins in different cellular fractions were separated by SDS-PAGE and analyzed by Western blot with indicated antibodies. GF, GFP-FLAG cells; LGF, LDAMP1-GFP-FLAG cells. TM, total membrane; Cyto, cytosol; PNS, post-nuclear supernatant. **D** Immuno-gold labeling for LDAMP1. C2C12 cells stably expressing LDAMP1-GFP-FLAG were treated with 100 μM OA plus 1 μM MG132 for 12 h. Then the cells were prepared for immuno-gold labeling of GFP to display the localization of LDAMP1. Briefly, cells were fixed with 4% (v/v) paraformaldehyde, dehydrated in an ascending concentration series of ethanol

and embedded in LR White resin. After ultra-thin sectioning, sections were stained for GFP with secondary antibody conjugated to 18 nm colloidal gold. The sections were observed with transmission electron microscope. Bar = 200 nm.

Figure 4

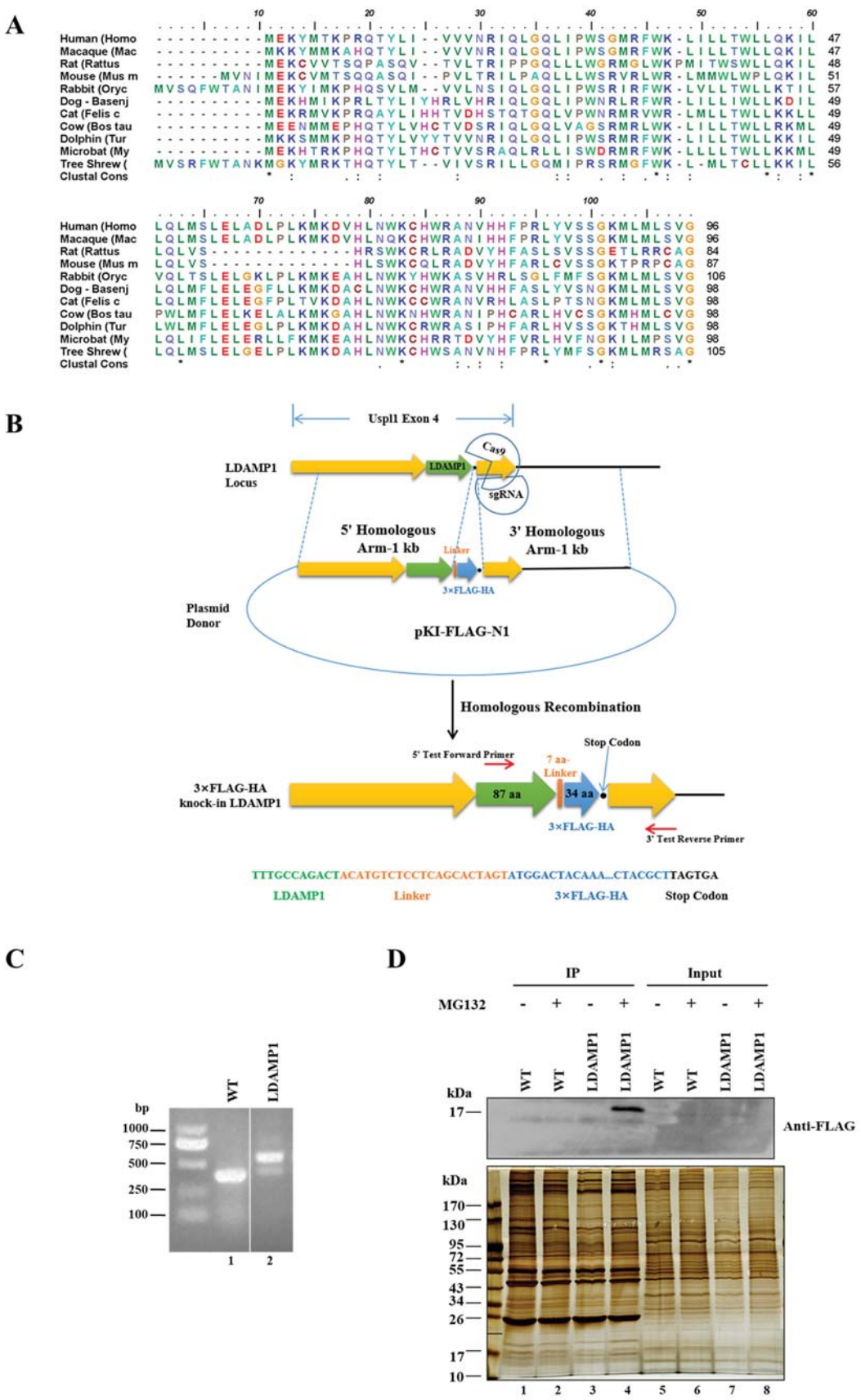
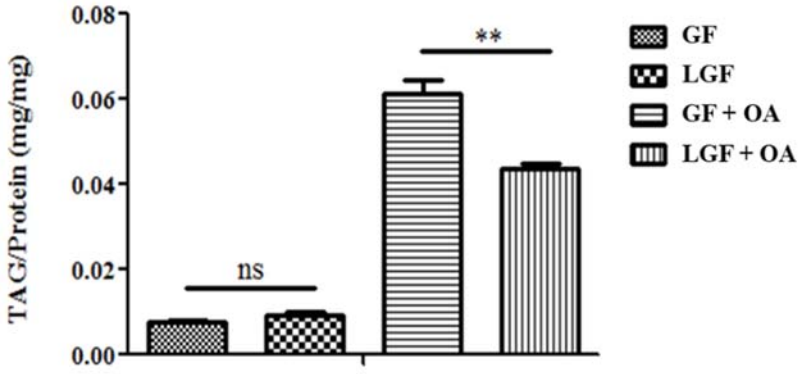

A
Human (Homo
Macaque (Mac
Rat (Rattus
Mouse (Mus m
Rabbit (Oryc
Dog - Basenj
Cat (Felis c
Cow (Bos tau
Dolphin (Tur
Microbat (My
Tree Shrew (
Clustal Cons
B
Uspl1 Exon 4
LDAMP1 Locus
LDAMP1
Cas9
sgRNA
5' Homologous Arm-1 kb
3' Homologous Arm-1 kb
Linker
3×FLAG-HA
Plasmid Donor
pKI-FLAG-N1
Homologous Recombination
5' Test Forward Primer
7 aa-Linker
Stop Codon
3×FLAG-HA knock-in LDAMP1
87 aa
34 aa
3×FLAG-HA
3' Test Reverse Primer
TTTGCCAGACTACATGTCTCCTCAGCACTAGTATGGACTACAAA...CTACGCTTAGTGA
LDAMP1
Linker
3×FLAG-HA
Stop Codon
C
WT
LDAMP1
bp
1000
750
500
250
100
1
2
D
IP
Input
MG132
WT
LDAMP1
kDa
17
Anti-FLAG
170
130
95
72
55
43
34
26
17
10
1
2
3
4
5
6
7
8

Figure 4. Verification of LDAMP1 expression endogenously.

A Alignment of putative LDAMP1 peptide sequences from a variety of mammals. LDAMP1 amino acid sequences were translated from short ORF in USPL1 sequence in 11 different mammalian species. Alignment of those sequences were analyzed using BioEdit. Amino acid identity is indicated. **B** A schematic cartoon to show CRISPR/Cas9-mediated LDAMP1 knock-in strategy. Cells carrying 3×FLAG-HA tag in the C-terminus of LDAMP1 allele was generated by CRISPR/Cas9-mediated knock-in technique. **C** Verification of the insertion of 3×FLAG-HA tag by PCR. Genomic DNA from selected clones was probed with primer pairs designed to amplify 200 bp upstream and 200 bp downstream of the LDAMP1 stop codon. PCR amplicon was subjected to 2% (w/v) agarose gel electrophoresis. **D** Verification of endogenous translation of LDAMP1. LDAMP1 knock-in cell line and WT C2C12 cells were treated with or without 1 μM MG132 for 12 h. Cells were washed with PBS, lysed with 1 mL TETN solution and probed with anti-FLAG beads to precipitate FLAG-tagged protein. Precipitate was eluted in sample buffer and separated on 5-15% gradient SDS-PAGE and further analyzed by FLAG Western blot (upper panel). Silver stained gel (lower panel) indicated protein profile in each sample and served as protein loading control. IP, immunoprecipitation.

Figure 5

A


0.08
0.06
0.04
0.02
0.00
TAG/Protein (mg/mg)
ns
**
GF
LGF
GF + OA
LGF + OA

B

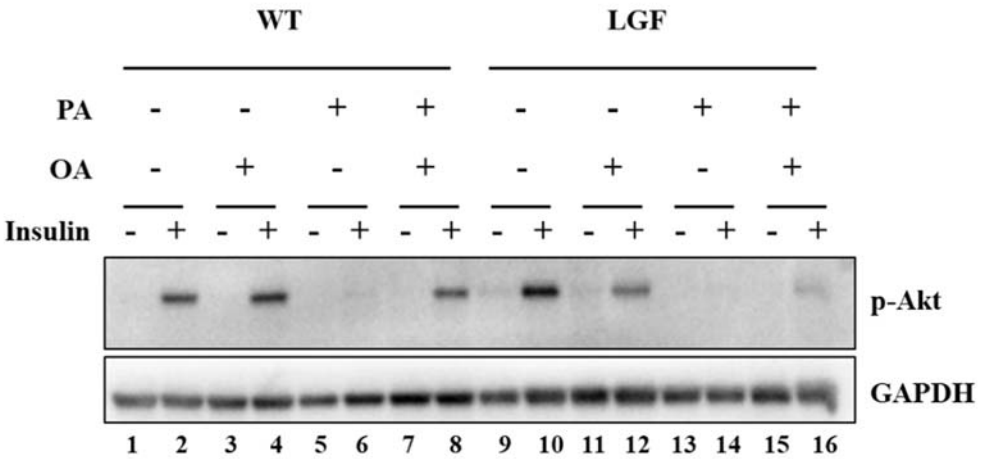

WT
LGF
PA - - + + - - + +
OA - + - + - + - +
Insulin - + - + - + - + - + - + - + - +
p-Akt
GAPDH
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Figure 5. LDAMP1 reduces the storage of triacylglycerol and mediates insulin sensitivity in C2C12 cells.

A Reduction of triacylglycerol level by LDAMP1. Stable cells expressing LDAMP1-GFP-FLAG and GFP-FLAG control cells were cultured to confluence in 6-well dishes and treated with or without 100 μM oleate (OA) for 12 h. Cells were collected in PBS (pH 7.4) containing 1% (v/v) Triton X-100 and sonicated. Aliquot of cell lysate was subjected to triacylglycerol (TAG) analysis with protein as internal control. Data represent mean ± SEM (n = 4), analyzed by unpaired Student *t*-test. **, P<0.001. GF, GFP-FLAG cells; LGF, LDAMP1-GFP-FLAG cells. **B** Inhibitory effect of LDAMP1 on insulin sensitivity in C2C12 cells. Cells were incubated in the presence or absence of 100 μM OA with or without 500 μM palmitate (PA) for 12 h before insulin stimulation (200 nM insulin for 10 min at 37°C). Cells were then lysed with sample buffer and the proteins were subjected to Western blot analysis of p-Akt and GAPDH. LGF, LDAMP1-GFP-FLAG cells.