Article

Long Non-Coding RNAs Associated with Ribosomes in Human Adipose-Derived Stem Cells: From RNAs to Microproteins

Bernardo Bonilauri [1] iD **, Fabiola Barbieri Holetz** [2] **and Bruno Dallagiovanna** [1,*] iD

1   Laboratory of Basic Biology of Stem Cells (LABCET), Carlos Chagas Institute-Fiocruz-Paraná, Curitiba 81350-010, Brazil; bernardobonilauri@gmail.com
2   Laboratory of Gene Expression Regulation (LABREG), Carlos Chagas Institute-Fiocruz-Paraná, Curitiba 81350-010, Brazil; fabiola.holetz@fiocruz.br
*   Correspondence: bruno.dallagiovanna@fiocruz.br

Abstract: Ribosome profiling reveals the translational dynamics of mRNAs by capturing a ribosomal footprint snapshot. Growing evidence shows that several long non-coding RNAs (lncRNAs) contain small open reading frames (smORFs) that are translated into functional peptides. The difficulty in identifying bona-fide translated smORFs is a constant challenge in experimental and bioinformatics fields due to their unconventional characteristics. This motivated us to isolate human adipose-derived stem cells (hASC) from adipose tissue and perform a ribosome profiling followed by bioinformatics analysis of transcriptome, translatome, and ribosome-protected fragments of lncRNAs. Here, we demonstrated that 222 lncRNAs were associated with the translational machinery in hASC, including the already demonstrated lncRNAs coding microproteins. The ribosomal occupancy of some transcripts was consistent with the translation of smORFs. In conclusion, we were able to identify a subset of 15 lncRNAs containing 35 smORFs that likely encode functional microproteins, including four previously demonstrated smORF-derived microproteins, suggesting a possible dual role of these lncRNAs in hASC self-renewal.

Keywords: lncRNA; ribosome; smORF; microprotein; translation; stem cells

1. Introduction

Long non-coding RNAs (lncRNAs) are generically described as sequences longer than 200 nucleotides (nt) without the ability to encode proteins. Many lncRNAs have properties similar to mRNAs in that they are transcribed by RNA polymerase II, capped, and polyadenylated [1–3]. However, in contrast with protein-coding genes, lncRNAs contain few exons/introns, are expressed at low levels, and demonstrate limited phylogenetic conservation [4]. Despite the first studies on lncRNAs being conducted in the early nineteen nineties [5–7], the functions of most of these transcripts still remain to be fully elucidated. Some lncRNAs have demonstrated regulatory functions, being involved in the regulation of cell development and differentiation, pluripotency, telomere length, metastasis, DNA damage repair, X chromosome inactivation, splicing, chromatin, transcription, among others [8–14]. Although many of these functions are nuclear, a large number of lncRNAs have cytoplasmic localization and function. Many of these cytoplasmic lncRNAs have the capacity to form complexes with RNA-binding proteins (RBP) and regulate mRNA degradation, translation, cellular signaling, and decoy targets. Another possible function is the action like that of the competitive endogenous RNA (ceRNA) sponging complementary microRNAs [15,16].

Human adipose-derived stem cells (hASC) have the ability to self-renew and differentiate into adipocytes, osteoblasts, and chondrocytes. These cells appear as an excellent source for the study of stem cell fate and as a robust source for use in regenerative medicine [17–19]. Thereby, understanding the molecular biology of hASC is of vital importance due to their great potential. Recently, some groups have demonstrated the importance of lncRNAs in

the self-renewal and differentiation processes of hASC, showing a wide transcriptional, post-transcriptional, and translational regulation [20,21]. Surprisingly, over the past years, several studies have shown the association of lncRNAs with the translational machinery [22–25], including our group, which demonstrated the association of lncRNAs with the polysomal fraction during adipogenic differentiation of hASC [26]. However, why these non-coding transcripts are associated with ribosomes remains unknown. Using ribosome profiling, Guttman et al. showed that these associated lncRNAs do not have ribosome release behavior (RRB) at the stop codon. This suggests that these RNAs do not have canonical ORFs that encode proteins, ribosome association being a possible pervasive translation [27]. Another study estimated with computational approaches that the lncRNAs are highly unlikely to have a coding potential due to their similarity to 3′UTR regions, GC content, and unconventional small ORFs (smORFs) with few initiation codons [28]. Recently, however, distinct groups have found that the pattern of ribosome occupancy in lncRNAs is consistent with translation of these smORFs (<300 nt), and the coding potential of smORFs is similar to protein-coding genes newly evolved [23,24,29,30]. Corroborating these findings, new experimental studies demonstrated smORFs-derived microproteins from lncRNAs. These microproteins demonstrate important functions and mechanisms of action in cell physiology, in healthy as well as pathological states [31]. For example, the lncRNA HOXB-AS3 encodes a conserved 53-amino acid (aa) microprotein. This peptide inhibits tumorigenesis through the blockage of PKM splicing, mir-18a processing, and PKM2 formation, causing metabolic reprogramming in colon cancer cells [32]. Another example is the lncRNA LINC00948, which encodes the conserved 46-aa microprotein Myoregulin (MLN), which interacts with SERCA, preventing Ca^{2+} uptake into the sarcoplasmic reticulum in skeletal muscle [33]. In opposition to MLN activity, another lncRNA-derived 34-aa microprotein named DWORF enhances SERCA activity, improving muscle performance [34]. Meanwhile, to date, there is no record of lncRNA-derived microproteins in hASC.

Against this background our goal is to identify translated smORFs within lncRNAs in hASC. To address this question, we proposed the identification of smORFs in lncRNAs through ribosome profiling experiments followed by deep sequencing (Ribo-seq) and transcriptomic and translatomic (Total and Polysomal Fractions) analysis. In conclusion, we verified ribosomal occupancy in many lncRNAs consistent with the translation of smORFs; however, further experimental studies are required to provide evidence for the translation and function of these microproteins.

2. Materials and Methods

2.1. Human Primary Samples

Human adipose-derived stem cells were obtained from adipose tissue from lipoaspirate samples of three female donors aged between 20 and 48 years (Table 1). We randomly selected the donors since no differences in the morphology, immunophenotype characteristics, proliferative rates and differentiation potential between hASC isolated from young and old subjects were demonstrated [35]. Tissue collection and cell isolation were performed after donors informed consent, in accordance with guidelines for research involving human subjects and with the approval of the Ethics Committee of Oswaldo Cruz Foundation, Brazil (approval number CAAE: 48374715.8.0000.5248).

Table 1. Subject characteristics.

SUBJECT	DONOR1	DONOR2	DONOR3	Mean $\pm$ SD
Age	46	20	48	38 $\pm$ 12.75
Gender	F	F	F	F
Weight (kg)	74.5	75	90	79.8 $\pm$ 7.19
Height (cm)	166	174	175	171 $\pm$ 4.02
BMI	27.04	24.77	29.39	27.06 $\pm$ 1.88

2.2. Isolation, Cell Culture, and Characterization

The hASC isolation was performed as previously described [36]. In short, 200 mL of adipose tissue were constantly washed with phosphate-buffered saline (PBS) and digested with 1 mg/mL type I collagenase (Gibco, Thermo Fisher Scientific, Waltham, MA, USA) for 30 min at 37 °C, 5% CO_2 under constant shaking. Next, the cell suspension was filtered through a 100 μm then a 40 μm mesh filter (BD Biosciences, Franklin Lakes, NJ, USA) and centrifuged. The pellet was treated with a hemolysis buffer to remove contaminating erythrocytes. The cells obtained were washed and plated at a density of 1×10^5 cells/cm^2 in T75 culture flasks in DMEM supplemented with 10% FBS, penicillin (100 units/mL), and streptomycin (100 μg/mL) in humid incubator at 37 °C and 5% CO_2. The culture medium was changed twice a week, and all of the experiments were performed with cell cultures at passages 4 to 6. Cell characterization was performed according to the minimal criteria established by the International Society of Cellular Therapy [37], i.e., the flow cytometry analysis conducted as previously described [17].

2.3. Ribosome Profiling, Library Preparation, and Sequencing

The ribosome profiling procedure was based on previously described protocols, with some modifications [38]. In short, the hASC culture (approximately 6×10^6 cells) was treated with 100 μg/mL cycloheximide for 10 min at 37 °C. Cells were detached with trypsin, washed with PBS, and incubated in the lysis buffer (15 mM Tris HCl pH 7.4, 15 mM $MgCl_2$, 0.3M NaCl, 100 μg/mL cycloheximide, 1% Triton X-100) for 10 min on ice. Cell lysate was centrifuged (12,000$\times$ *g* at 4 °C, 10 min), and the supernatant was carefully collected into a new tube. Next, the sample was incubated with a nuclease (Benzonase) for 10 min at 25 °C to produce ribosome RNA footprints (RNA-protected fragments). The nuclease digestion was inhibited with RNaseOUT (Invitrogen™), and lysates were loaded and pelleted into an ultracentrifuge tube over 2 mL of 1 M sucrose cushion by ultracentrifugation at 39,000 rpm (P40ST rotor, HIMAC; CP80WX, Hitachi Medical Systems, Japan) at 4 °C for 2 h. The supernatant was removed, and RNA was purified and submitted to electrophoresis (TBE-urea 6% gel) to recover the ribosome-protected RNA fragments of approximately 30 nucleotides. Concentration and size distribution were determined on a Bioanalyzer DNA chip (Agilent Technologies, Santa Clara, CA, USA) (Figure S1A). Ribo-seq libraries were constructed according to the TruSeq Small RNA Prep Kit Manual. After that, libraries were sequenced on an Illumina HiSeq 2500 platform.

2.4. Bioinformatics and Computational Analysis

The Ribo-seq raw data were first subjected to quality control checking with FastQC (v.0.11.2). As previously described [39], the replicate reads were pooled for analysis to achieve high coverage and sensitivity (Figure S1B,C). Too long (>36 bp) and too short (<20 bp) reads, low quality reads, and adaptor sequences (TGGAATTCTCGGGTGC-CAAGG) were removed using Cutadapt (v.1.6) [40]. The remaining Ribo-seq reads were searched for contaminant rRNA, tRNA, snoRNA, and snRNA sequences using Bowtie2 (v.2.2.5) [41]. The unaligned reads were then mapped to genome build hg19 using STAR (v.2.5.3a) [42], with the following parameters: —seedSearchStartLmaxOverLread 0.5, —outFilterMismatchNmax 2, —outMultimapperOrder Random, —outFilterMultimap

Nmax 20 and —outSAMmultNmax 1. An index with known transcripts annotation file from GENCODE (v.31) was provided for the genome alignment, and novel splice junctions were permitted. The Ribo-seq raw data from Marcon et al. [38] were downloaded and submitted to quality control analysis followed by trimming using Cutadapt (v.1.6). As the SOLiD platform are color-space data, we proceeded with the alignment and exclusion of reads from contaminants using the Subread (v.2.0.1) [43] aligner. The unaligned reads were selected and mapped to the hg19 genome also using the Subread aligner; featureCounts (v.1.6.0) [44] was used to count the read numbers mapped to known genes using the same gene annotation GTF file. Only unique mapped reads were selected.

For the analysis of the RNA-seq raw data, fastq files of the Total and Polysomal Fractions of undifferentiated hASC were downloaded and subjected to quality control checking as described above. Low-quality reads and adaptor sequences were removed as previously described. Next, reads were mapped to genome build hg19 using HISAT2 (v.2.2.1) [45], and featureCounts (v.1.6.0) was used to count read numbers. Gene expression levels of Ribo-seq data and RNA-seq data were estimated using the transcript per million (TPM) normalization, with a cut-off for lncRNAs of $\geq$10 reads in exons (replicate average) and TPM $\geq$ 2; and for protein-coding genes $\geq$10 reads in exons (replicate average), TPM $\geq$ 5.

Other statistics and functional analyses were performed with R (v.3.5.2), including graphical representations. Principal Component Analysis (PCA) and raw reads profile plots were performed using deepTools (v.3.2.1) [46].

2.5. Gene Ontology Analysis

The Gene Ontology (GO) analysis of the protein-coding genes was performed with PANTHER (Released 20200728), using a cut-off criterion of *p*-value < 0.01 and FDR < 0.01 (1%).

2.6. RNA Minimum Free-Energy Calculation

Prediction of RNA minimum free-energy of secondary structures formation was calculated with RNAfold (v.2.4.14) [47]. RNAfold uses the nearest-neighbor thermodynamic model for the prediction. We used a FASTA input with Ribo-seq lncRNAs, protein-coding genes, and random 5′UTR/3′UTR sequences with a length of $\geq$200 nt for the analysis.

2.7. Microprotein Features

Predictions of the putative microprotein localization were performed using DeepLoc 1.0. In short, DeepLoc 1.0 uses a recurrent neural network that processes protein sequences and identifies protein regions important for subcellular localization [48].

3. Results

3.1. Study Overview: Searching for translated lncRNAs in hASC

Ribosome profiling followed by next-generation sequencing (Ribo-seq) maps the position of translating ribosomes over the transcriptome by combining ribosome footprint and deep sequencing. This is possible due to the action of nuclease treatment, resulting in ribosome-protected fragments (RPFs) with a size of around 30 nt [22,49]. We performed a ribosome profiling of hASC followed by bioinformatics analysis, combining our previous data of transcriptome (Total fraction RNA-seq) and translatome (Polysomal fraction RNA-seq) for searching for ribosome occupancy in smORFs of lncRNAs (Figures 1 and S1).
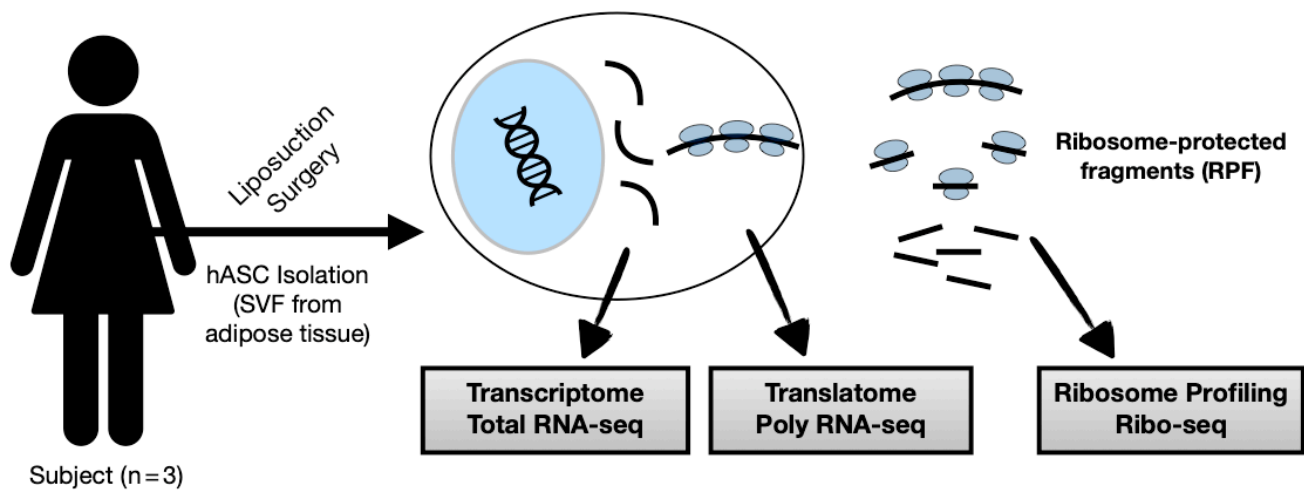
Figure 1. Ribosome profiling was performed in human undifferentiated adipose-derived stem cells (n = 3) followed by massive sequencing (Ribo-seq). For deep analysis, our previous RNA-seq datasets from total fraction (Total RNA-seq) and polysomal fraction (Poly RNA-seq) were collected. Performing bioinformatics analysis, we combined Ribo-seq and RNA-seq datasets to search for the translated smORFs of lncRNAs in hASC. SVF: stromal vascular fraction.

3.2. Identification of Ribosome-Associated lncRNAs in hASC

hASC were first isolated, characterized, and expanded; next, the ribosome profiling assay was performed (see "Materials and Methods"). Our sequenced ribosomal footprints show a characteristic and well-documented length distribution [39,50], where the average length of RPF was 28 nt (Figure S1D). Using a cut-off criterion of more than 10 reads in exons and TPM $\geq$ 2, we verified 222 lncRNAs with ribosome occupancy in our Ribo-seq data (Table S1). As the next step, we decided to systematically characterize ribosome-associated lncRNAs compared with ribosome-associated messenger RNAs (mRNAs). We found that ribosome-associated lncRNAs had lower expression levels than the protein-coding genes (Mann–Whitney U test, $p < 2.2 \times e^{-16}$) (Figure 2A), which corroborates previous findings [51]. In addition, lncRNAs displayed shorter transcript lengths compared to protein-coding transcripts (Mann–Whitney U test, $p < 3.77 \times e^{-16}$), lower GC-content (Mann–Whitney U test, $p < 2.2 \times e^{-16}$) and fewer exons (Mann–Whitney U test, $p < 2.2 \times e^{-16}$) (Figure 2B–D). Besides these features, we analyzed the minimum folding energy (MFE) and determined that lncRNAs had a greater MFE in comparison with protein-coding transcripts (Kolmogorov–Smirnov test, $p < 2.2 \times e^{-16}$) (Figure 2E), which means that mRNAs fold more strongly than lncRNAs [52]. However, when we compare them with 5'UTR and 3'UTR sequences, lncRNAs have significantly lower MFE (Kolmogorov–Smirnov test, $p < 2.2 \times e^{-16}$) (Figure 2E). Regarding the chromosomal distribution, we noticed an enrichment of the expression of lncRNAs of chromosomes 1 and 17 (Figure 2F). Taken together, these results demonstrated that, despite ribosome occupancy, lncRNAs have fewer structural features and lower expression when compared with ribosome-associated protein-coding genes.
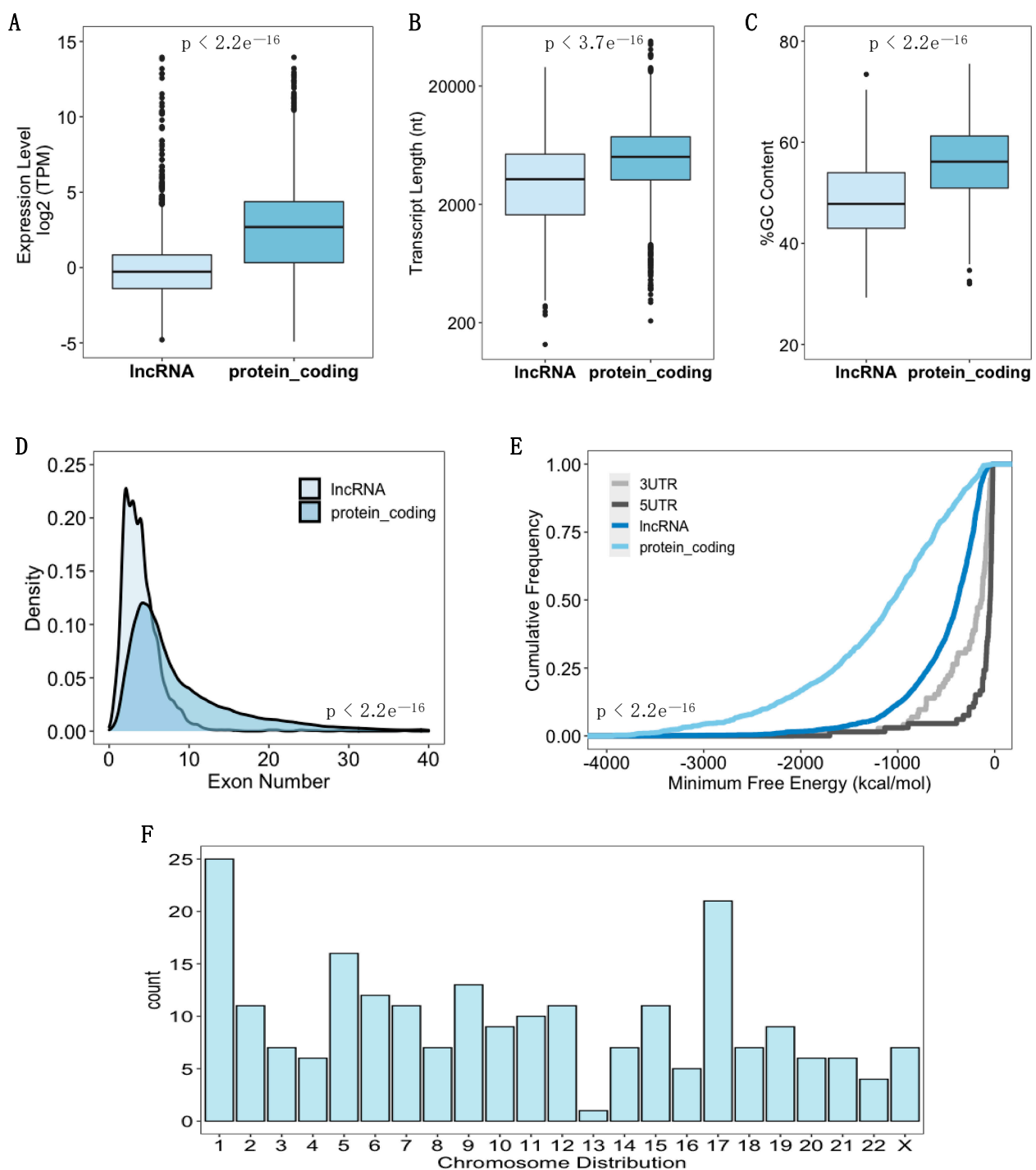
Figure 2. General characteristics of the lncRNAs identified in the Ribo-seq of hASC. (**A**) Boxplot comparing the expression levels of all lncRNAs and protein-coding genes identified in Ribo-seq. (**B**) Comparison of transcript lengths between lncRNAs and mRNAs. (**C**) Comparison of percentage of GC contents between lncRNAs and mRNAs. (**D**) Density plots show the number of exons in lncRNAs and mRNAs. (**E**) Cumulative distribution of minimum free energy (MFE) of lncRNAs, protein-coding genes, 5-untranslated-regions (5′UTR), and 3-untranslated-regions (3′UTR). (**F**) Histogram showing the chromosomal distribution of the lncRNAs identified in the Ribo-seq data.

To deepen our analysis of lncRNA translation, we used the RNA-seq data of Total and Polysomal fractions of undifferentiated hASC to compare with the generated Ribo-seq data. First, just to have a broader view, we performed a Gene Ontology (GO) analysis of the highest expressed protein-coding genes in these datasets, being identified as the most

relevant pathways related to the developmental process in the Ribo-seq, cotranslational protein targeting to membrane in the total RNA-seq, and translation initiation in the polysomal RNA-seq (Figure S2). As stated earlier, using our cut-off criterion, we demonstrated 222 lncRNAs with ribosome occupancy (Table S1). Using the same criterion, we identified 517 lncRNAs in the total RNA-seq and 352 lncRNAs in the polysomal RNA-seq (Figure S3, Tables S2 and S3). We proceeded by comparing the lncRNAs of Ribo-seq and RNA-seq (Total and Poly), finding 99 common lncRNAs among these datasets (Figure S3D, Table S4). Among them are the well-characterized lncRNAs such as H19, MALAT1, and NEAT1 (Figure S3E,F).

3.3. LncRNA-Encoded Microproteins in hASC

Considering only the 99 common lncRNAs identified in our datasets (Table S3), we performed an analysis of putative translated smORFs. Thus, we selected the smORFs with a length of ≥50 nt and with a canonical start codon (AUG) and found 11.048 smORFs with only 3.689 (33%) in-frame position. Using a stringent criterion of more than 10 reads in the ribosomal footprint within these smORFs, we identified 35 translated smORFs (within 15 lncRNAs). These smORF-derived microproteins have an average size of 52-aa (Table 2).

Four smORFs of identified lncRNAs have previously been confirmed as coding for microproteins [53–56]. For example, we show the smORF of LINC01420 that encodes the 68-aa microprotein NBDY (Figure 3A). NBDY (non-annotated P-body dissociating polypeptide) was initially characterized as a human microprotein component of the mRNA decapping complex, directly interacting with EDC4 and DCP1A being localized in P-bodies' granules [53,57]. It has recently been demonstrated that DCP1A is dispersed in granular structures in the hASC cytoplasm, partially co-located with DDX6 (another interaction partner of NBDY) [53,58]. Another example is the smORF of LINC00116 that encodes the 56-aa microprotein MTLN (Figure 3B). MTLN (Mitoregulin) is a single-pass transmembrane microprotein (located in the inner mitochondrial membrane) initially documented in murine myoblasts and cardiomyocytes; it plays a role in mitochondrial respiration, β-oxidation, reactive oxygen species (ROS) production, and Ca^{2+} retention capacity [54,59]. Moreover, MTLN regulates triglyceride clearance by regulating lipolysis and mitochondrial β-oxidation in human and murine adipocytes [60]. We previously found the LINC00116/MTLN differentially expressed (upregulated) in hASC submitted to 24 h of adipogenic differentiation [61], which indicates that this lncRNA/microprotein may play an important role in the cell metabolism during hASC differentiation, similar to the one demonstrated in the adipogenic differentiation of human embryonic stem cells (hESC) [60]. Furthermore, the smORF of TERC that encodes a 121-aa peptide (Figure 3C) [55] and the smORF located in the 5'UTR of TUG1 that encodes a 154-aa peptide (Figure 3D) are shown. Interestingly, TUG1 lncRNA possesses a triple molecular function, having a cis and a trans-DNA regulatory activity and also encoding this small protein [56]. To further confirm the translational region of the aforementioned lncRNAs, we analyzed publicly available ribosome profiling data with different cell types using GWIPs-viz [62] (Figure S4).

Table 2. Identified smORFs within lncRNAs with ribosome occupancy and their predicted microproteins.

| lncRNA	GeneID (Ensembl)	ORF Length (nt)	MP* Length (aa)	DeepLoc1.0 Pred.	DLoc[†] Score	Microprotein Sequence
CYTOR	ENSG00000222041.11	156	52	Nucleus	0.37	MTDTENHDSAPSSTSTCCPPITAGMQLKDSLG PGSNRPLWTLRPLHLRVVCL
EBLN3P	ENSG00000281649.2	141	47	Nucleus	0.77	MEEPMDTSEPLSALPFTGQQSFEPSGKFGQYPSMQMN HIQALGKWRT
EBLN3P	ENSG00000281649.2	78	26	Extracellular	0.52	MYVTDPESPAAWDPCLPSVSPAELWN
GAS5	ENSG00000234741.8	150	50	Extracellular	0.43	MVLGADAVWLWIAPYGQLCPQGRMRIATEVLKSKPN SSHWHTGIRQKAGS
GAS5	ENSG00000234741.8	81	27	Extracellular	1	MTCLGKDMKTVPVIPFKGTCFIDVNVN
LINC00968	ENSG00000246430.7	132	44	Extracellular	0.53	MFLQKLKSCLVKAFHKMVCVWDQEDRRLLKKRTGTL THFRLLHV
LINC01116	ENSG00000163364.10	249	83	Nucleus	0.71	MGPRFLADARGRGRVPGSRFSQAPIPAHARGPRPTHE APTPIVEAPPGKEVRLPLQAAPRGMGNRQEMTRTASL RLCSRPSLC
MEG3	ENSG00000214548.18	168	56	Nucleus	0.50	MPFERLEAKSIKHSWENTTGGTTRFSYTLGSHGEDRR EKKEVEREERAGETGEENN
MEG3	ENSG00000214548.18	444	148	Nucleus/ Cytoplasm	0.418/ 0.417	MRRLSIVMKNPWHSPHPQTHGSHSHTGPKATVSA AVAPVDIGKPGEGVEEISWPPAGSLGFCAQGSWSPK NFQKLTPHVPILLGFLDFSEAPAEGSRCSLECRGSPLTW LLESLLFLLLLPSSSSSLSISPSLCPSPVPDLAIPGCP
MEG3	ENSG00000214548.18	252	84	Cytoplasm	0.37	MEAAEEALMGPTIPDPSLLPGGPLVSFLVWAEAI TWMPTWEGTSNVGPQPLSSSKSLHSHGDTLHLFPRD RLDPETLDPGPPLE
MIR22HG	ENSG00000186594.14	279	93	Mitochondria	0.60	MGWEGPNSRVDDTFWASWRAFAQIGPARSGFRLE TLAGLRSRRLKQPKRLQEAVSVRFGG
MIR22HG	ENSG00000186594.14	66	22	Mitochondria	0.50	MIRFGQVGEPLPRLAQQGAVLD
MSC-AS1	ENSG00000235531.10	192	64	Nucleus	0.42	MSLETTGPQERQALSVLLLPWKKPAPTMPSATS KSSLRPPQKQMLSCFLYSCRTTSNHPNTREH
SNHG1	ENSG00000255717.7	87	29	Extracellular	0.47	MSYWAPVCRIYAHVGTEESSVVAPTRAYW
SNHG1	ENSG00000255717.7	153	51	Extracellular	0.73	MFSPQELTGEGMGQDPSLCKASVTVMFQVGVHGL CSYRGDLVDNHSMMNTK
SNHG16	ENSG00000163597.15	99	33	Nucleus	0.78	MATPVGVEHGEQSQAFSDDGAVSLSFQSRKRIL
SNHG16	ENSG00000163597.15	108	36	Nucleus	0.58	MATPVGVEHGEQSQAFSDDGWLGGLKVLDEKMLSKR
SNHG29	ENSG00000175061.18	405	135	Mitochondria	0.69	MFPGSLSRGRRAAVEMAWLPGSCARVAFAAGA AARYWTAWQGSAGPNPAAVAEAHGSLFCGRATSAR AWSLRRPGPGSPAHSGGVQTRENWVSWGRLAVWGTPR AVYVGKIVTVLLEDLFDCPDDTCNRKCRQKR
SNHG29	ENSG00000175061.18	285	95	Mitochondria	0.66	MFPGSLSRGRRAAVEMAWLPGSCARVAFAAGA AARYWTAWQGSAGPNPAAVAEAHGSLFCGRATSAR AWSLRRPGPGSPAHSGGVQTRENWVANS
SNHG29	ENSG00000175061.18	111	37	Extracellular	0.68	MDHSFVVGPHLPEPGVCEGRDPVPRPTVGVCKPERTG
SNHG29	ENSG00000175061.18	237	79	Nucleus/ Extracellular	0.243/ 0.230	MDHSFVVGPHLPEPGVCEGRDPVPRPTVGVCKPE RTGLQIREESASCLAAEYWSQEPAMRLYSQRMSVPRTS SCHQFGF
SNHG29	ENSG00000175061.18	210	70	Endoplasmic Reticulum	0.49	MLALCIRGHAQQIQEIYLATFSRKGTLGIIHYILEFF WVFFFFFETVLLYCPGWSVVAQSQLIASSITQA
SNHG29	ENSG00000175061.18	54	18	-	-	MYQRTCSADPRDIFGNFF
SNHG29	ENSG00000175061.18	237	79	Golgi Apparatus	0.42	MLSRSKRYIWQLFLEKAHWVSFITFLSFFGFFFFFLR QSCCIAQAGVWWHNHSSLHPQSPRPKQSSHLVAGTT AHSTPG
SNHG29	ENSG00000175061.18	78	26	Mitochondria	0.97	MLPRLVSGSWAQMVLLPQLPKAQAKL
SNHG5	ENSG00000203875.12	105	35	Mitochondria	0.26	MALSSVAQWSSSEDAKIHEKTSRTSGRIFNGKSLG
SNHG5	ENSG00000203875.12	72	24	Mitochondria	0.57	MQRYTKKLPEHLGEYLMENRLVKT
SNHG6	ENSG00000245910.8	75	25	Mitochondria	0.86	MPVWWRRRRLRARSWALRGARKPLR
SNHG8	ENSG00000269893.8	156	52	Mitochondria	0.68	MIIGPKLTALPKRQRSQDIGRSGAALETLKFTSMR GLECSLGRRASTCSPGP
SNHG8	ENSG00000269893.8	108	36	Mitochondria/ Nucleus	0.309/ 0.303	MDDGNIRLSRNPSGNGRSLFSIRQWTYRSWGNGCSE
ZFAS1	ENSG00000177410.13	75	25	Mitochondria	0.34	MDFGRGSHHWTSKEATCRHLQPSIS
ZFAS1	ENSG00000177410.13	60	20	-	-	MRVLEVEYIYTYKIETGDGI
ZFAS1	ENSG00000177410.13	99	33	Extracellular	0.44	MRVLEVEYIYTYKIGWEPRVPVCVDLGLIQSAL
ZFAS1	ENSG00000177410.13	153	51	Nucleus	0.57	MEYERSPLERKGQTLCFHESEDLAEPVPQGYCIH SLSLKGCAHFKNVIVRL
ZFAS1	ENSG00000177410.13	99	33	Extracellular	0.51	MRGALWKEKDRPCAFMKVKIWLNQFHKVTVYIA

MP*: Microprotein, DLoc[†]: DeepLoc1.0.
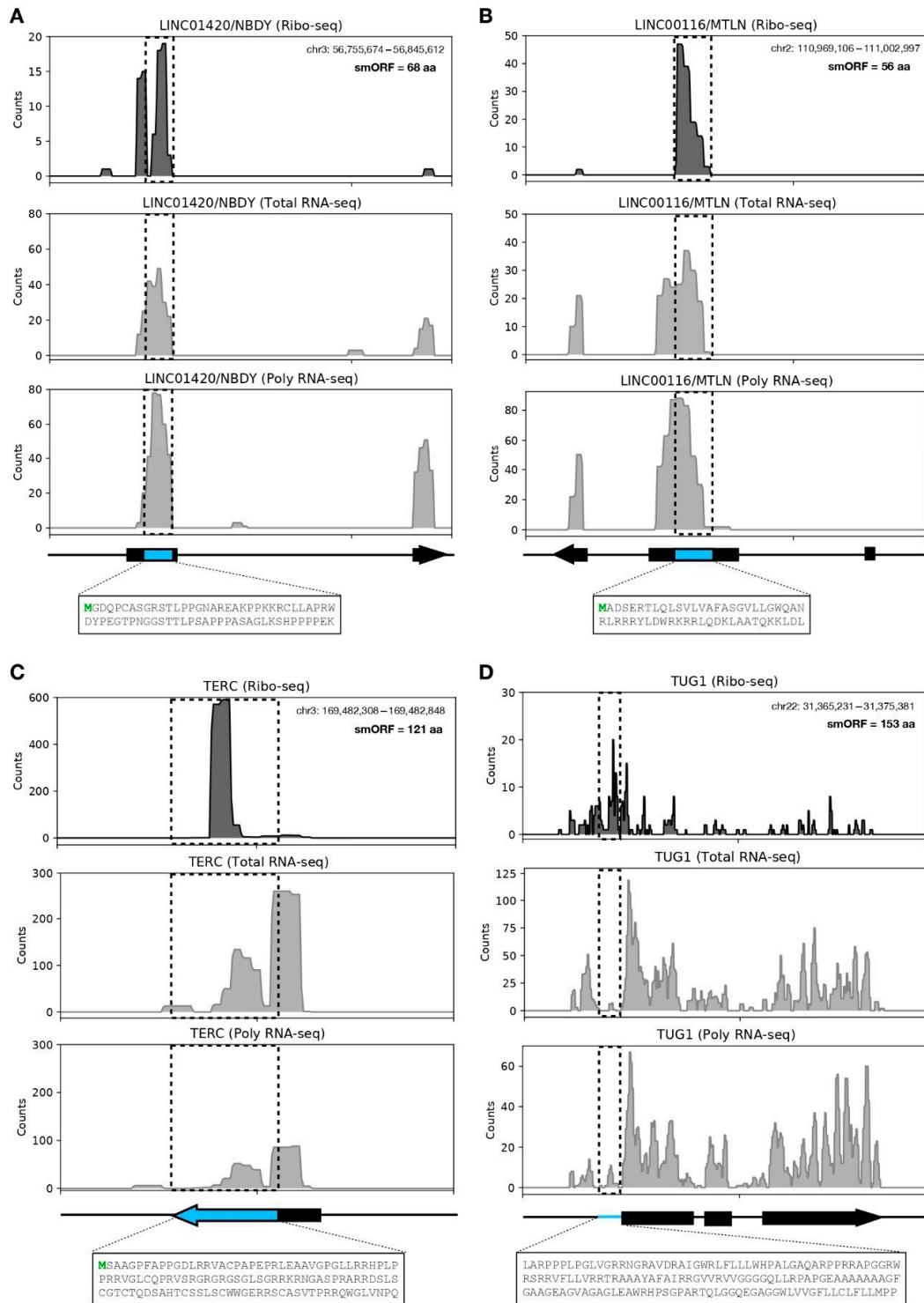
Figure 3. LncRNAs-encoded microproteins present in human adipose-derived stem cells. (**A**) Expression profile of microprotein NOBODY from LINC01420 with 68 aa. (**B**) Expression profile of microprotein MTLN from LINC0116 with 56 aa. (**C**) Expression profile of TERC with smORFs-encoded microprotein with 121 aa. (**D**) Expression profile of TUG1 with 5′UTR smORFs-encoded microprotein with 154 aa. Light blue boxes indicate the smORF location in transcripts, and the dashed boxes represent the smORF location in the read coverage plot. The upper panel represents Ribo-seq coverage, the middle panel represents Total fraction RNA-seq (Total RNA-seq) coverage, and the lower panel represents Polyso-mal fraction RNA-seq (Poly RNA-seq) coverage; the y-axis represents transcript raw counts.

As clear ribosomal coverage in smORFs-encoding microproteins was previously documented by us (Figure 3) and other groups, we proceeded with the analysis of the 35 identified smORF-encoded microproteins (Table 2). Therefore, for a qualitative analysis, we selected the lncRNAs EBLN3P, SNHG8, MIR22HG, and SNHG16. In Figure 4A, we highlighted the ribosome occupancy of the lncRNA EBLN3P, which has two translated smORFs with canonical start codons, namely the 78 nt (26-aa) smORFs within exon 1 (Figure 4C) and the 141 nt (47-aa) smORFs within exon 2 (Figure 4D). We found a high expression level in polysomal fraction compared to Ribo-seq from liposuction-derived hASC, Ribo-seq from dermolipectomy-derived hASC, and total fraction, despite the smORFs in exon 1 having a greater ribosome footprint (Figure 4C,D). Using DeepLoc 1.0 for predicted protein localization, we found that the microprotein of smORFs in exon 1 is extracellular (secreted), and the microprotein of smORFs in exon 2 is probably located in the nucleus (Figure 4C,D). Likewise, we performed a qualitative analysis of lncRNA SNHG8 demonstrating two translated smORFs, which may encode microproteins of 52-aa and 36-aa, respectively (Figure S5). The ribosome occupancies of lncRNAs MIR22HG and SNHG16 were also analyzed, revealing two smORFs with coding potential in MIR22HG and SNHG16 (Figure S6). These results shed new light on the dual function of lncRNAs, being possible regulatory and coding RNAs.

With this in mind, we compared the present Ribo-seq generated in this study with our previously performed ribosome profiling of hASC. However, unlike the ribosome profiling shown here, the previous hASC were isolated from a solid white adipose tissue (WAT) after a dermolipectomy procedure, followed by sequencing in a distinct high-throughput plataform [38]. Using the same cut-off criterion used for our analysis, we could identify only 81 lncRNAs with ribosome occupancy. However, we were able to verify a distinct ribosomal footprint in the smORFs of the lncRNAs identified here (Figure 4C,D). The reproducibility of our results in the datasets of others indicates the great possibility of using Ribo-seq to identify translating smORFs. In addition, we can confirm the translation and possible coding capacity of these smORFs in hASC.
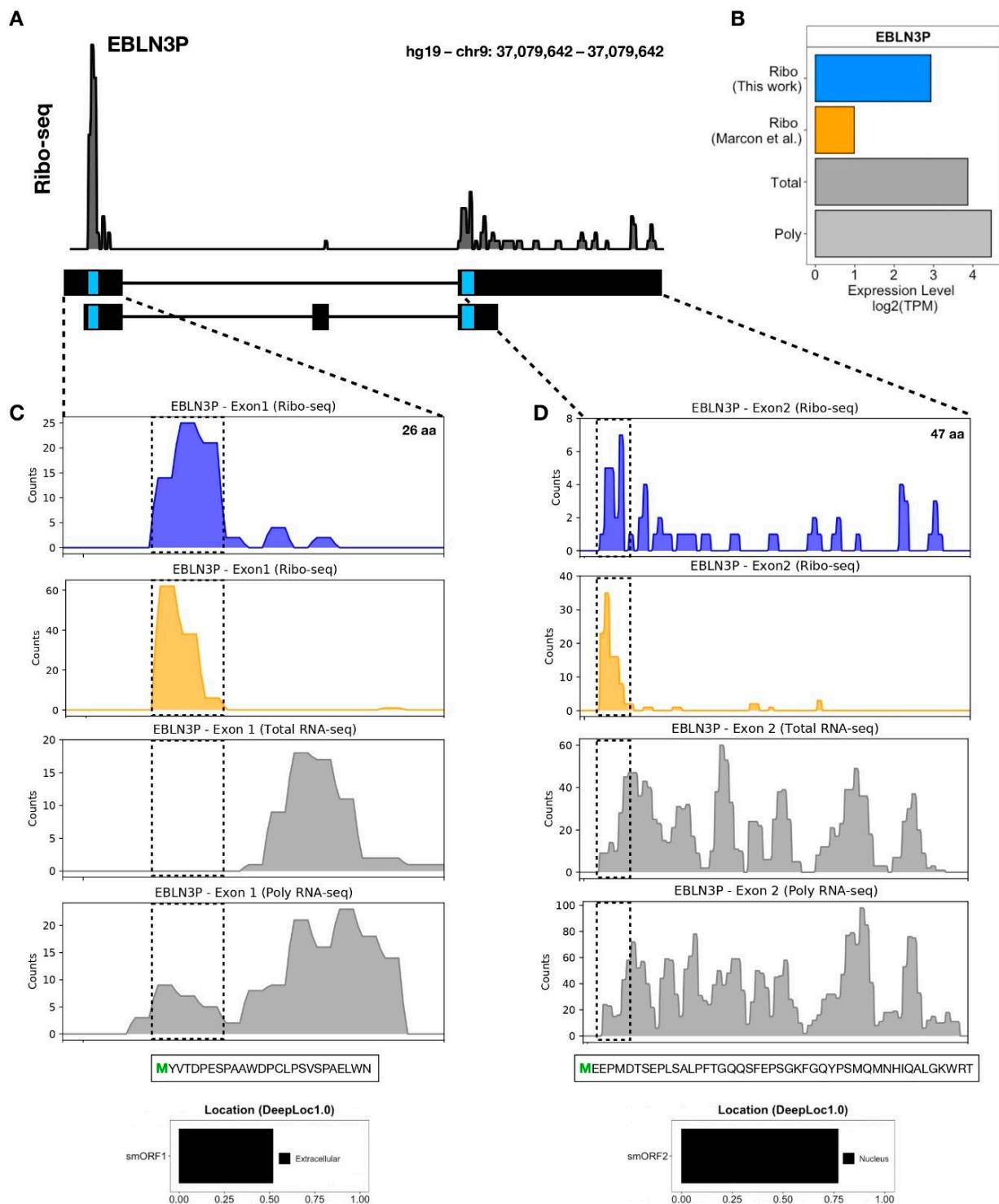
Figure 4. Putative smORFs-derived microproteins within EBLN3P lncRNA in hASC. (**A**,**B**) Ribosome coverage and the expression level of EBLN3P in different datasets, respectively. (**C**) Ribosome occupancy of smORFs located in exon 1 of EBLN3P, showing a putative 26-aa microprotein with extracellular localization prediction. (**D**) Ribosome occupancy of smORFs located in exon 2 or 3 of EBLN3P, showing a putative 47-aa microprotein with nuclear localization prediction. The upper panel represents Ribo-seq coverage from this study, the second panel represents Ribo-seq coverage from Marcon et al., the third panel represents Total RNA-seq coverage, and the lower panel represents Polysomal RNA-seq coverage; the y-axis represents transcript raw counts. Light blue boxes indicate the smORF location in transcripts, and the dashed boxes represent the smORF location in the read coverage plot.

4. Discussion

Human adipose-derived stem cells hold great promise in regenerative medicine due to their simple isolation procedure, yield, and proliferative capacity, and, above all, due to their paracrine effects on the injured site. This paracrine effect is directly related to the release of specific proteins and extracellular vesicles containing different biomolecules [19,63–65]. Therefore, discovering new microproteins derived from smORFs of previously annotated non-coding RNAs seems to be a promising step towards a better understanding of the regenerative capacity of hASC, its paracrine effects, and developing new treatments based on synthetic microproteins.

Ribosome profiling successfully identifies ribosome occupancy in lncRNAs, shedding light on the possible translation of smORFs. However, the following techniques must be combined to guarantee the existence of these microproteins: ribosome profiling, total and polysomal RNA-seq, mass spectrometry (MS)-based proteomics/peptidomics, in vitro and in vivo translation, immunofluorescence, and functional biochemistry, as well as molecular assays [25,66–69]. Technical limitations constitute a significant challenge in microprotein identification. Microprotein detection by MS-based proteomics/peptidomics (the gold standard for protein detection) is difficult due to low concentration/expression, low peptide recovery from the isolation protocols, signal interference of peptides derived from abundant proteins, protease activity, turnover rates, and weak signal from these microproteins [70,71]. For example, Slavoff et al. combined transcriptomic and peptidomic analyses to discover new microproteins and found only eight coding smORFs of lncRNAs [72]. However, a proteogenomic approach (RNA-seq/Ribo-seq and MS spectra) is strong evidence of microprotein existence.

Over the past years, distinct groups have used experiments and proteogenomics analysis to discover microproteins quantitatively, demonstrating microproteins derived from 5′ untranslated regions (5′UTR) and 3′ untranslated regions (3′UTR) of mRNAs and also from smORFs within lncRNAs [66,72–74]. Several studies have demonstrated that these lncRNAs-derived microproteins are functional and play an important role in certain cells. However, these transcripts, in addition to encoding microproteins, have a regulatory function. This represents a dual role of lncRNAs, increasing the plethora of molecular mechanisms of action from these RNAs [34,54,56,75,76].

Here, we were able to identify several translated smORFs within lncRNAs in hASC. As shown in Figure 4, the lncRNA EBLN3P presents two translated smORFs with a canonical start codon. It has recently been demonstrated that EBLN3P may act as a ceRNA, regulating the expression of DOCK4 through miR-144-3p sequestration, promoting liver cancer cell proliferation, migration, and invasion [77]. Similarly, EBLN3P promotes the recovery of impaired spiral ganglion neurons by competitively binding miR-204-5p, regulating TMPRSS3 expression [78]. Despite showing a regulatory function in different cancer types [79,80], here we demonstrated a coding capacity of EBLN3P in hASC, which represents a possible dual-function of this transcript in stem cell biology, including a putative secreted smORF-derived microprotein.

Similarly, we demonstrated the presence of translated smORFs in lncRNAs SNHG8, SNHG16, and MIR22HG. Recently, mitochondrial microproteins derived from smORFs in lncRNAs SNHG8 and SNHG16 were demonstrated by immunofluorescence of FLAG-tagged microproteins, while their function and mechanism of action remain unknown [73]. Regarding MIR22HG, the Uniprot database [81] records only one putative uncharacterized 57-aa protein encoded by MIR22HG (Q0VDD5) with protein uncertain classification. Meanwhile, an 86-aa microprotein from MIR22HG was demonstrated by Western blot in control cells and upon viral infection [82]. Nevertheless, both SNHG16 and MIR22HG are shown to play an important role in bone metabolism and promoting osteogenic differentiation of human bone marrow mesenchymal stem cells (hBMSC) [83,84].

As mentioned previously, the discovery of microproteins has a significant impact on molecular and cellular biology as well as medicine. Huang et al. demonstrated a microprotein (named MP31) encoded from 5′UTR of PTEN mRNA and localized in mi-

tochondria, which limits lactate–pyruvate conversion. Recombinant MP31 administered intraperitoneally penetrated the blood–brain barrier and inhibited glioblastoma xenografts, establishing its tumor-suppressing activity and, in turn, clinical use [85]. Similarly, a microprotein encoded by lncRNA MIR155HG (named P155) was shown to regulate antigen presentation and to function as a likely suppressor of inflammatory diseases. Synthetic P155 administered intravenously displayed therapeutic effects on autoinflammatory conditions in mice [86].

Here, we begin this journey of discovering new microproteins in hASC, showing the ribosome occupancy in lncRNAs, with future perspectives to characterize these possible small proteins. Microproteins may play important roles in adult stem cell self-renewal, differentiation, and paracrine effects.

5. Conclusions

In summary, using the state-of-the-art ribosome profiling technique followed by next-generation sequencing (Ribo-seq) of hASC isolated from liposuction (liquid WAT) or dermolipectomy (solid WAT) and bioinformatics analysis of different sequenced fractions (Total and Polysomal RNA-seq), we were able to identify a subset of lncRNAs with ribosome occupancy in smORFs, which may indicate a translational capacity and coding potential of functional microproteins.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Oswaldo Cruz Foundation (CAAE: 48374715.8.0000.5248).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The Ribo-seq raw data obtained in this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB47140. The SOLiD Ribo-seq raw data are available in the Short Read Archive (SRA) of NCBI under the accession number PRJNA328260. The RNA-seq raw data are available in the ArrayExpress repository under the accession number E-MTAB-6298.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rinn, J.L.; Chang, H.Y. Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* **2012**, *81*, 145–166. [CrossRef] [PubMed]
2. Ulitsky, I.; Bartel, D.P. lincRNAs: Genomics, Evolution, and Mechanisms. *Cell* **2013**, *154*, 26–46. [CrossRef] [PubMed]
3. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Merkel, A.; Gonzalez, D.; Lagarde, J.; et al. The GENCODE v7 Catalogue of Human Long Non-Coding RNAs: Analysis of Their Structure, Evolution and Expression. *Genome Res.* **2012**, *22*, 1775–1789. [CrossRef]
4. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, 1915–1927. [CrossRef]
5. Brannan, C.I.; Dees, E.C.; Ingram, R.S.; Tilghman, S.M. The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.* **1990**, *10*, 28–36. [CrossRef]
6. Brockdorff, N.; Ashworth, A.; Kay, G.F.; McCabe, V.M.; Norris, D.P.; Cooper, P.J.; Swift, S.; Rastan, S. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* **1992**, *71*, 515–526. [CrossRef]
7. Brown, C.; Hendrich, B.; Rupert, J.L.; Lafrenière, R.G.; Xing, Y.; Lawrence, J.; Willard, H.F. The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **1992**, *71*, 527–542. [CrossRef]
8. Fatica, A.; Bozzoni, I. Long non-coding RNAs: New players in cell differentiation and development. *Nat. Rev. Genet.* **2014**, *15*, 7–21. [CrossRef]
9. Guttman, M.; Donaghey, J.; Carey, B.W.; Garber, M.; Grenier, J.K.; Munson, G.; Young, G.; Lucas, A.B.; Ach, R.; Bruhn, L.; et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nat. Cell Biol.* **2011**, *477*, 295–300. [CrossRef]
10. Tripathi, V.; Ellis, J.D.; Shen, Z.; Song, D.Y.; Pan, Q.; Watt, A.T.; Freier, S.M.; Bennett, C.F.; Sharma, A.; Bubulya, P.A.; et al. The Nuclear-Retained Noncoding RNA MALAT1 Regulates Alternative Splicing by Modulating SR Splicing Factor Phosphorylation. *Mol. Cell* **2010**, *39*, 925–938. [CrossRef]
11. Sharma, V.; Khurana, S.; Kubben, N.; Abdelmohsen, K.; Oberdoerffer, P.; Gorospe, M.; Misteli, T. A BRCA 1-interacting lnc RNA regulates homologous recombination. *EMBO Rep.* **2015**, *16*, 1520–1534. [CrossRef] [PubMed]
12. Graf, M.; Bonetti, D.; Lockhart, A.; Serhal, K.; Kellner, V.; Maicher, A.; Jolivet, P.; Teixeira, M.T.; Luke, B. Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. *Cell* **2017**, *170*, 72–85.e14. [CrossRef]
13. Li, J.; Meng, H.; Bai, Y.; Wang, K. Regulation of lncRNA and Its Role in Cancer Metastasis. *Oncol. Res. Featur. Preclin. Clin. Cancer Ther.* **2016**, *23*, 205–217. [CrossRef] [PubMed]
14. Bonilauri, B.; Dallagiovanna, B. Long Non-coding RNAs Are Differentially Expressed after Different Exercise Training Programs. *Front. Physiol.* **2020**, *11*, 567614. [CrossRef]
15. Noh, J.H.; Kim, K.M.; McClusky, W.G.; Abdelmohsen, K.; Gorospe, M. Cytoplasmic functions of long noncoding RNAs. *Wiley Interdiscip. Rev. RNA* **2018**, *9*, e1471. [CrossRef] [PubMed]
16. Wang, L.; Cho, K.B.; Li, Y.; Tao, G.; Xie, Z.; Guo, B. Long Noncoding RNA (lncRNA)-Mediated Competing Endogenous RNA Networks Provide Novel Potential Biomarkers and Therapeutic Targets for Colorectal Cancer. *Int. J. Mol. Sci.* **2019**, *20*, 5758. [CrossRef]
17. Robert, A.; Angulski, A.B.B.; Spangenberg, L.; Shigunov, P.; Pereira, I.T.; Bettes, P.S.L.; Naya, H.; Correa, A.; Dallagiovanna, B.; Stimamiglio, M.A. Gene expression analysis of human adipose tissue-derived stem cells during the initial steps of in vitro osteogenesis. *Sci. Rep.* **2018**, *8*, 4739. [CrossRef]
18. Robert, A.W.; Marcon, B.H.; Dallagiovanna, B.; Shigunov, P. Adipogenesis, Osteogenesis, and Chondrogenesis of Human Mesenchymal Stem/Stromal Cells: A Comparative Transcriptome Approach. *Front. Cell Dev. Biol.* **2020**, *8*, 561. [CrossRef]
19. Si, Z.; Wang, X.; Sun, C.; Kang, Y.; Xu, J.; Wang, X.; Hui, Y. Adipose-derived stem cells: Sources, potency, and implications for regenerative therapies. *Biomed. Pharmacother.* **2019**, *114*, 108765. [CrossRef]
20. Tye, C.; Gordon, J.A.; Martin-Buley, L.A.; Stein, J.L.; Lian, J.B.; Stein, G.S. Could lncRNAs be the Missing Links in Control of Mesenchymal Stem Cell Differentiation? *J. Cell. Physiol.* **2015**, *230*, 526–534. [CrossRef]
21. Xie, Z.-Y.; Wang, P.; Wu, Y.-F.; Shen, H.-Y. Long non-coding RNA: The functional regulator of mesenchymal stem cells. *World J. Stem Cells* **2019**, *11*, 167–179. [CrossRef] [PubMed]
22. Ingolia, N.T.; Brar, G.A.; Stern-Ginossar, N.; Harris, M.S.; Talhouarne, G.J.; Jackson, S.E.; Wills, M.; Weissman, J.S. Ribosome Profiling Reveals Pervasive Translation Outside of Annotated Protein-Coding Genes. *Cell Rep.* **2014**, *8*, 1365–1379. [CrossRef] [PubMed]
23. Ruiz-Orera, J.; Messeguer, X.; Subirana, J.; Alba, M.M. Long non-coding RNAs as a source of new peptides. *eLife* **2014**, *3*, e03523. [CrossRef]
24. Ruiz-Orera, J.; Villanueva-Cañas, J.L.; Albà, M.M. Evolution of new proteins from translated sORFs in long non-coding RNAs. *Exp. Cell Res.* **2020**, *391*, 111940. [CrossRef]
25. Van Heesch, S.; Van Iterson, M.; Jacobi, J.; Boymans, S.; Essers, P.B.; De Bruijn, E.; Hao, W.; MacInnes, A.W.; Cuppen, E.; Simonis, M. Extensive localization of long noncoding RNAs to the cytosol and mono- and polyribosomal complexes. *Genome Biol.* **2014**, *15*, R6. [CrossRef]

26. Dallagiovanna, B.; Pereira, I.T.; Origa-Alves, A.C.; Shigunov, P.; Naya, H.; Spangenberg, L. lncRNAs are associated with polysomes during adipose-derived stem cell differentiation. *Gene* **2017**, *610*, 103–111. [CrossRef]

27. Guttman, M.; Russell, P.; Ingolia, N.T.; Weissman, J.S.; Lander, E.S. Ribosome Profiling Provides Evidence that Large Noncoding RNAs Do Not Encode Proteins. *Cell* **2013**, *154*, 240–251. [CrossRef]

28. Niazi, F.; Valadkhan, S. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3′ UTRs. *RNA* **2012**, *18*, 825–843. [CrossRef]

29. Zeng, C.; Fukunaga, T.; Hamada, M. Identification and analysis of ribosome-associated lncRNAs using ribosome profiling data. *BMC Genom.* **2018**, *19*, 1–14. [CrossRef] [PubMed]

30. Ji, Z.; Song, R.; Regev, A.; Struhl, K. Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **2015**, *4*, e08890. [CrossRef] [PubMed]

31. Yeasmin, F.; Yada, T.; Akimitsu, N. Micropeptides Encoded in Transcripts Previously Identified as Long Noncoding RNAs: A New Chapter in Transcriptomics and Proteomics. *Front. Genet.* **2018**, *9*, 144. [CrossRef]

32. Huang, J.; Chen, M.; Chen, D.; Gao, X.-C.; Zhu, S.; Huang, H.; Hu, M.; Zhu, H.; Yan, G.-R. A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. *Mol. Cell* **2017**, *68*, 171–184.e6. [CrossRef]

33. 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.; et al. A Micropeptide Encoded by a Putative Long Noncoding RNA Regulates Muscle Performance. *Cell* **2015**, *160*, 595–606. [CrossRef] [PubMed]

34. 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. A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. *Science* **2016**, *351*, 271–275. [CrossRef]

35. Horinouchi, C.D.; Barisón, M.J.; Robert, A.W.; Kuligovski, C.; Aguiar, A.M.; Dallagiovanna, B. Influence of donor age on the differentiation and division capacity of human adipose-derived stem cells. *World J. Stem Cells* **2020**, *12*, 1640–1651. [CrossRef] [PubMed]

36. Marcon, B.H.; Spangenberg, L.; Bonilauri, B.; Robert, A.W.; Angulski, A.B.B.; Cabo, G.C.; Cofré, A.R.; Bettes, P.S.L.; Dallagiovanna, B.; Shigunov, P. Data describing the experimental design and quality control of RNA-Seq of human adipose-derived stem cells undergoing early adipogenesis and osteogenesis. *Data Brief* **2020**, *28*, 105053. [CrossRef] [PubMed]

37. Dominici, M.; Le Blanc, K.; Mueller, I.; Slaper-Cortenbach, I.; Marini, F.; Krause, D.; Deans, R.; Keating, A.; Prockop, D.; Horwitz, E. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* **2006**, *8*, 315–317. [CrossRef]

38. Marcon, B.H.; Holetz, F.B.; Eastman, G.; Origa-Alves, A.C.; Amorós, M.A.; de Aguiar, A.M.; Rebelatto, C.K.; Brofman, P.R.; Sotelo-Silveira, J.; Dallagiovanna, B. Downregulation of the protein synthesis machinery is a major regulatory event during early adipogenic differentiation of human adipose-derived stromal cells. *Stem Cell Res.* **2017**, *25*, 191–201. [CrossRef]

39. Hsu, P.; Calviello, L.; Wu, H.-Y.L.; Li, F.-W.; Rothfels, C.J.; Ohler, U.; Benfey, P.N. Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E7126–E7135. [CrossRef] [PubMed]

40. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 10–12. [CrossRef]

41. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]

42. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]

43. Liao, Y.; Smyth, G.; Shi, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **2013**, *41*, e108. [CrossRef] [PubMed]

44. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]

45. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef]

46. Ramírez, F.; Ryan, D.P.; Grüning, B.; Bhardwaj, V.; Kilpert, F.; Richter, A.S.; Heyne, S.; Dündar, F.; Manke, T. deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **2016**, *44*, W160–W165. [CrossRef] [PubMed]

47. Clote, P.; Ferre', F.; Kranakis, E.; Krizanc, D. Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA* **2005**, *11*, 578–591. [CrossRef]

48. Armenteros, J.J.A.; Sønderby, C.K.; Sønderby, S.K.; Nielsen, H.; Winther, O. DeepLoc: Prediction of protein subcellular localization using deep learning. *Bioinformatics* **2017**, *33*, 3387–3395. [CrossRef] [PubMed]

49. Calviello, L.; Ohler, U. Beyond Read-Counts: Ribo-seq Data Analysis to Understand the Functions of the Transcriptome. *Trends Genet.* **2017**, *33*, 728–744. [CrossRef]

50. Ingolia, N.T.; Hussmann, J.; Weissman, J.S. Ribosome Profiling: Global Views of Translation. *Cold Spring Harb. Perspect. Biol.* **2019**, *11*, a032698. [CrossRef]

51. Wang, H.; Wang, Y.; Xie, S.; Liu, Y.; Xie, Z. Global and cell-type specific properties of lincRNAs with ribosome occupancy. *Nucleic Acids Res.* **2016**, *45*, 2786–2796. [CrossRef] [PubMed]

52. Yang, J.-R.; Zhang, J. Human Long Noncoding RNAs Are Substantially Less Folded than Messenger RNAs. *Mol. Biol. Evol.* **2015**, *32*, 970–977. [CrossRef] [PubMed]

53. D'Lima, N.G.; Ma, J.; Winkler, L.; Chu, Q.; Loh, K.H.; Corpuz, E.O.; Budnik, B.A.; Lykke-Andersen, J.; Saghatelian, J.M.Q.C.A.; A Slavoff, N.G.D.L.W.S. A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol.* **2017**, *13*, 174–180. [CrossRef]

54. Stein, C.S.; Jadiya, P.; Zhang, X.; McLendon, J.; Abouassaly, G.M.; Witmer, N.; Anderson, E.J.; Elrod, J.; Boudreau, R.L. Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. *Cell Rep.* **2018**, *23*, 3710–3720.e8. [CrossRef] [PubMed]

55. Rubtsova, M.; Naraykina, Y.; Vasilkova, D.; Meerson, M.; Zvereva, M.; Prassolov, V.; Lazarev, V.; Manuvera, V.; Kovalchuk, S.; Anikanov, N.; et al. Protein encoded in human telomerase RNA is involved in cell protective pathways. *Nucleic Acids Res.* **2018**, *46*, 8966–8977. [CrossRef]

56. Lewandowski, J.P.; Dumbović, G.; Watson, A.R.; Hwang, T.; Jacobs-Palmer, E.; Chang, N.; Much, C.; Turner, K.M.; Kirby, C.; Rubinstein, N.D.; et al. The Tug1 lncRNA locus is essential for male fertility. *Genome Biol.* **2020**, *21*, 1–35. [CrossRef]

57. Na, Z.; Luo, Y.; Schofield, J.A.; Smelyansky, S.; Khitun, A.; Muthukumar, S.; Valkov, E.; Simon, M.D.; Slavoff, S.A. The NBDY Microprotein Regulates Cellular RNA Decapping. *Biochemistry* **2020**, *59*, 4131–4142. [CrossRef]

58. Marcon, B.H.; Rebelatto, C.K.; Cofré, A.R.; Dallagiovanna, B.; Correa, A. DDX6 Helicase Behavior and Protein Partners in Human Adipose Tissue-Derived Stem Cells during Early Adipogenesis and Osteogenesis. *Int. J. Mol. Sci.* **2020**, *21*, 2607. [CrossRef] [PubMed]

59. Makarewich, C.A.; Baskin, K.K.; Munir, A.Z.; Bezprozvannaya, S.; Sharma, G.; Khemtong, C.; Shah, A.M.; McAnally, J.R.; Malloy, C.; Szweda, L.I.; et al. MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid β-Oxidation. *Cell Rep.* **2018**, *23*, 3701–3709. [CrossRef]

60. Friesen, M.; Warren, C.R.; Yu, H.; Toyohara, T.; Ding, Q.; Florido, M.H.; Sayre, C.; Pope, B.; Goff, L.A.; Rinn, J.L.; et al. Mitoregulin Controls β-Oxidation in Human and Mouse Adipocytes. *Stem Cell Rep.* **2020**, *14*, 590–602. [CrossRef] [PubMed]

61. Marcon, B.H.; Shigunov, P.; Spangenberg, L.; Pereira, I.; De Aguiar, A.M.; Amorín, R.; Rebelatto, C.K.; Correa, A.; Dallagiovanna, B. Cell cycle genes are downregulated after adipogenic triggering in human adipose tissue-derived stem cells by regulation of mRNA abundance. *Sci. Rep.* **2019**, *9*, 1–10. [CrossRef]

62. Michel, A.M.; Fox, G.; Kiran, A.M.; De Bo, C.; O'Connor, P.B.F.; Heaphy, S.M.; Mullan, J.P.A.; Donohue, C.A.; Higgins, D.G.; Baranov, P.V. GWIPS-viz: Development of a ribo-seq genome browser. *Nucleic Acids Res.* **2014**, *42*, D859–D864. [CrossRef] [PubMed]

63. Xu, P.; Xin, Y.; Zhang, Z.; Zou, X.; Xue, K.; Zhang, H.; Zhang, W.; Liu, K. Extracellular vesicles from adipose-derived stem cells ameliorate ultraviolet B-induced skin photoaging by attenuating reactive oxygen species production and inflammation. *Stem Cell Res. Ther.* **2020**, *11*, 1–14. [CrossRef] [PubMed]

64. Li, X.; Ma, T.; Sun, J.; Shen, M.; Xue, X.; Chen, Y.; Zhang, Z. Harnessing the secretome of adipose-derived stem cells in the treatment of ischemic heart diseases. *Stem Cell Res. Ther.* **2019**, *10*, 1–13. [CrossRef]

65. Mitchell, R.; Mellows, B.; Sheard, J.; Antonioli, M.; Kretz, O.; Chambers, D.; Zeuner, M.-T.; Tomkins, J.; Denecke, B.; Musante, L.; et al. Secretome of adipose-derived mesenchymal stem cells promotes skeletal muscle regeneration through synergistic action of extracellular vesicle cargo and soluble proteins. *Stem Cell Res. Ther.* **2019**, *10*, 1–19. [CrossRef] [PubMed]

66. 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.; et al. Pervasive functional translation of noncanonical human open reading frames. *Science* **2020**, *367*, 1140–1146. [CrossRef]

67. Schlesinger, D.; Elsässer, S.J. Revisiting sORFs: Overcoming challenges to identify and characterize functional microproteins. *FEBS J.* **2021**, 15769. [CrossRef]

68. Martinez, T.F.; Chu, Q.; Donaldson, C.; Tan, D.; Shokhirev, M.N.; Saghatelian, A. Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* **2020**, *16*, 458–468. [CrossRef]

69. 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. Noncanonical open reading frames encode functional proteins essential for cancer cell survival. *Nat. Biotechnol.* **2021**, *39*, 697–704. [CrossRef]

70. Choi, S.-W.; Kim, H.-W.; Nam, J.-W. The small peptide world in long noncoding RNAs. *Briefings Bioinform.* **2019**, *20*, 1853–1864. [CrossRef]

71. Hartford, C.C.R.; Lal, A. When Long Noncoding Becomes Protein Coding. *Mol. Cell. Biol.* **2020**, *40*, e00528-19. [CrossRef]

72. Slavoff, S.; Mitchell, A.J.; Schwaid, A.G.; Cabili, M.N.; Ma, J.; Levin, J.; Karger, A.; Budnik, B.A.; Rinn, J.; Saghatelian, A. Peptidomic discovery of short open reading frame–encoded peptides in human cells. *Nat. Chem. Biol.* **2012**, *9*, 59–64. [CrossRef] [PubMed]

73. 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. The Translational Landscape of the Human Heart. *Cell* **2019**, *178*, 242–260.e29. [CrossRef] [PubMed]

74. Chong, C.; Müller, M.; Pak, H.; Harnett, D.; Huber, F.; Grun, D.; Leleu, M.; Auger, A.; Arnaud, M.; Stevenson, B.J.; et al. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat. Commun.* **2020**, *11*, 1–21. [CrossRef] [PubMed]

75. Gaertner, B.; van Heesch, S.; Schneider-Lunitz, V.; Schulz, J.F.; Witte, F.; Blachut, S.; Nguyen, S.; Wong, R.; Matta, I.; Hübner, N.; et al. A human ESC-based screen identifies a role for the translated lncRNA LINC00261 in pancreatic endocrine differentiation. *eLife* **2020**, *9*, 58659. [CrossRef]

76. Spencer, H.L.; Sanders, R.; Boulberdaa, M.; Meloni, M.; Cochrane, A.; Spiroski, A.-M.; Mountford, J.; Emanueli, C.; Caporali, A.; Brittan, M.; et al. The LINC00961 transcript and its encoded micropeptide, small regulatory polypeptide of amino acid response, regulate endothelial cell function. *Cardiovasc. Res.* **2020**, *116*, 1981–1994. [CrossRef]

77. Li, H.; Wang, M.; Zhou, H.; Lu, S.; Zhang, B. Long Noncoding RNA *EBLN3P* Promotes the Progression of Liver Cancer via Alteration of microRNA-144-3p/DOCK4 Signal. *Cancer Manag. Res.* **2020**, *12*, 9339–9349. [CrossRef]

78. Jiang, W.; Peng, A.; Chen, Y.; Pang, B.; Zhang, Z. Long non-coding RNA EBLN3P promotes the recovery of the function of impaired spiral ganglion neurons by competitively binding to miR-204-5p and regulating TMPRSS3 expression. *Int. J. Mol. Med.* **2020**, *45*, 1851–1863. [CrossRef] [PubMed]

79. Mathias, C.; Muzzi, J.C.D.; Antunes, B.B.; Gradia, D.F.; Castro, M.A.A.; de Oliveira, J.C. Unraveling Immune-Related lncRNAs in Breast Cancer Molecular Subtypes. *Front. Oncol.* **2021**, *11*, 692170. [CrossRef]

80. Dai, S.; Li, N.; Zhou, M.; Yuan, Y.; Yue, D.; Li, T.; Zhang, X. LncRNA EBLN3P promotes the progression of osteosarcoma through modifying the miR-224-5p/Rab10 signaling axis. *Sci. Rep.* **2021**, *11*, 1–12. [CrossRef] [PubMed]

81. The UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef]

82. Razooky, B.S.; Obermayer, B.; O'May, J.B.; Tarakhovsky, A. Viral Infection Identifies Micropeptides Differentially Regulated in smORF-Containing lncRNAs. *Genes* **2017**, *8*, 206. [CrossRef]

83. Jin, C.; Jia, L.; Tang, Z.; Zheng, Y. Long non-coding RNA MIR22HG promotes osteogenic differentiation of bone marrow mesenchymal stem cells via PTEN/ AKT pathway. *Cell Death Dis.* **2020**, *11*, 1–13. [CrossRef]

84. Asila, A.; Yang, X.; Kaisaer, Y.; Ma, L. SNHG16/miR-485-5p/BMP7 axis modulates osteogenic differentiation of human bone marrow-derived mesenchymal stem cells. *J. Gene Med.* **2021**, *23*, e3296. [CrossRef]

85. Huang, N.; Li, F.; Zhang, M.; Zhou, H.; Chen, Z.; Ma, X.; Yang, L.; Wu, X.; Zhong, J.; Xiao, F.; et al. An Upstream Open Reading Frame in Phosphatase and Tensin Homolog Encodes a Circuit Breaker of Lactate Metabolism. *Cell Metab.* **2021**, *33*, 128–144.e9. [CrossRef]

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