LINC00493/SMIM26 and Its Dual Functioning

In the present study, we investigate the widely expressed lncRNA LINC00493. We determine the structure of the LINC00493 transcript, its cell localization and influence on cell physiology. Our data demonstrate that LINC00493 has an influence on cell viability in a cell-type-specific manner. Furthermore, it was recently shown that LINC00493 has a sORF that is translated into small protein SMIM26. The results of our knockdown and overexpression experiments suggest that both LINC00493/SMIM26 transcript and protein affect cell viability, but in the opposite manner.

1. Introduction

Advances in sequencing techniques revealed the transcription of non-coding regions of the genome which correspond to different groups, such as long noncoding RNA genes, small noncoding RNA genes, pseudogenes and immunoglobulin/T-cell receptor gene segments. Long noncoding RNAs (lncRNAs) are transcripts with lengths more than 200 nucleotides that are not translated into functional proteins. According to the GENCODE project (release 38) [1], the human genome contains 17,944 lncRNA genes. Data from the FANTOM CAT project revealed 27,919 human lncRNA loci [2]. The number of lncRNA genes is comparable to the number of protein-coding genes (19,954). However, lncRNAs make up only 0.03–0.20% of total RNA mass in the cell, whereas mRNAs make up 3–7% of it [3].

Initially thought to be transcriptional noise, several lncRNAs were discovered to be involved in gene expression regulation processes and affect cellular functions [4]. LncRNAs realize their functions through different intermolecular interactions: formation of a DNA–lncRNA triplex, formation of an lncRNA–RNA duplex and formation of an lncRNA–protein or lncRNA–chromatin complex [5]. These complexes may affect gene expression at the transcriptional [6][7][8][9] or post-transcriptional levels [10][11][12] and thus affect cellular phenotype.

Many lncRNAs are expressed in a tissue-specific manner, and their effect can vary in different cell types [13]. Moreover, lncRNAs' expression can be altered in different pathological conditions, and their dysregulation may play an important role in disease progression [14]. To date, the lncRNADisease2.0 database contains entries about 19,166 lncRNAs associated with 529 diseases [15], including heart failure, cerebral injury, hypertension, acute kidney injury and cancer [16][17][18]. Genome-wide association studies (GWAS) indicate that lncRNA genes are enriched for trait- or disease-linked polymorphisms. Over 90% of all GWAS hits lie outside of known coding genes [19][20].

LncRNAs have an average length of about 3 kb and could contain up to 120 small open reading frames (sORFs) with a median of six sORFs per lncRNA. Recent studies proved that about 10,000 lncRNA genes in the mammalian genome contain sORFs less than 300 nt in length [21][22][23]. These sORFs could be translated into short peptides with key biological functions [24]. The presence of small peptides encoded by lncRNAs suggests that in some cases lncRNAs may have a dual function, or that the observed biological effect is contributed by the small protein, which means that this class of genes should be reclassified as protein coding [25][26][27].

In the present study, we investigate the widely expressed lncRNA *LINC00493*. We determine the structure of the *LINC00493* transcript, its localization, protein-coding potential and its influence on cell physiology. Our data demonstrate the cell-type-specific role of *LINC00493*. During our work, *LINC00493* was predicted to contain a sORF that could translate a small protein—SMIM26. We collected and described all the existing data on this protein. Using knockdown and overexpression experiments, we obtain data suggesting that both *LINC00493/SMIM26* transcript and protein affect cell viability, but in the opposite manner.

2. Investigation of LINC00493/SMIM26 Gene Suggests Its Dual Functioning

2.1. LINC00493 Transcript Structure

The *LINC00493* gene was predicted through the ENCODE project. The UCSC Genome Browser [28] shows that *LINC00493* is located on human chromosome 20p11.23 and consists of two or three exons, according to mRNA and EST sequence data from RefSeq and Ensembl databases. In contrast to the protein-coding genes, the lncRNA gene annotations tend to have poorly defined boundaries, because of weak conservation, low and tissue-specific expression and lack of characteristic hallmarks of transcription initiation and termination [29]. Therefore, to define the exact structure of the *LINC00493* transcript we performed reverse transcription (RT)-PCR and rapid amplification of cDNA ends (RACE) analysis on total RNA from HEK293T, HeLa cell lines and human primary skin fibroblasts. RT-PCR analysis revealed that the *LINC00493* transcript consists of two exons and RACE showed the exact 5′ and 3′ cDNA ends (**Figure 1**A).
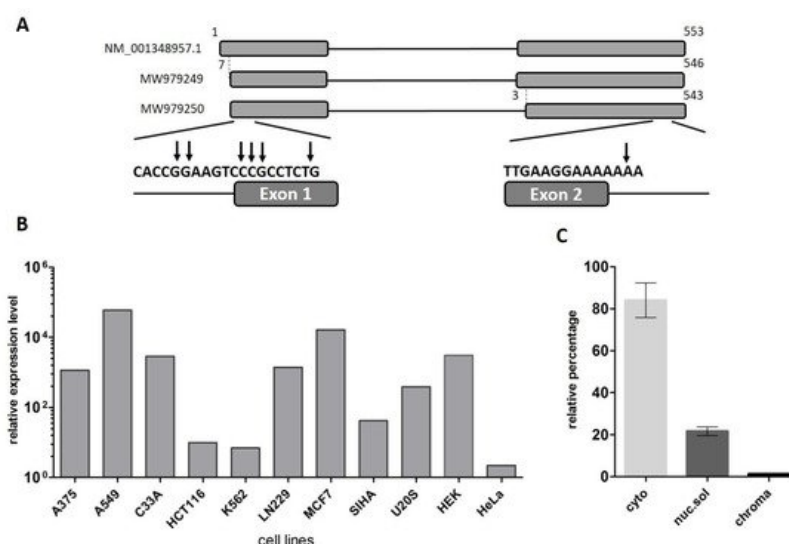
Figure 1. Analysis of the structure and expression of the *LINC00493* transcript. (**A**) Scheme of the *LINC00493* gene (according to RefSeq NM_001348957.1 sequence). Results of 5′ and 3′ RACE analysis are presented under the scheme. The vertical arrows represent the genomic position of exact 5′ and 3′ ends. Nucleotide numbering was based on reference sequence NM_001348957.1 (**B**) Relative expression of *LINC00493* transcript in 11 human cell lines was detected by RT-qPCR. (**C**) Subcellular localization of *LINC00493* was detected by qPCR of RNA isolated from cytoplasmic (cyto), nuclear-soluble (nuc.sol) and chromatin-bound (chroma) fractions of HEK293T cells. The error bars represent SEM (standard error mean).

According to the Ensembl database, there are two *LINC00493* isoforms, ENST00000411646.1 and ENST00000435844.3, which differ by three nucleotides at the beginning of the second exon. Our RT-PCR analysis confirmed that *LINC00493* has two isoforms, expressed in all analyzed cell lines. The total length of the short and long isoforms was 500 and 497 bp, respectively. The nucleotide sequences of short and long isoforms were deposited into GenBank under accession numbers MW979249 and MW979250. The difference in the sequencing signal suggests that the major long isoform is expressed at an approximately 3-fold higher level than the minor one.

2.2. LINC00493 Is Widely Expressed in Human Tissues and Cell Lines

To identify *LINC00493* expression profile, we provided analysis of the FANTOM5 and GTEx expression data. We observed that *LINC00493* is highly expressed in most human cell lines and tissues. An expression profile of *LINC00493* in 889 human samples from FANTOM5 classified this gene as a housekeeping gene [30]. We confirmed the high widespread expression level of this transcript using RT-qPCR analysis of 11 human cell lines, as well as human primary skin fibroblasts (**Figure 1**B). The highest expression level was observed in A549, MCF7 and HEK293T cell lines.

## 2.3. Cytoplasmic Localization of LINC00493

LncRNA subcellular localization is closely related to its biological function. Some lncRNAs play a role in a transcriptional regulation through their interaction with chromatin, while others are found in the cytoplasm and affect the post-transcriptional control of gene expression or could be translated into small peptides. We investigated the subcellular localization of the *LINC00493* transcript using the soft lysis method. RNA was isolated from cytoplasmic, nuclear and chromatin-bound fractions of cells. To determine the level of the investigated transcript in each fraction we performed RT-qPCR. Our analysis revealed that the *LINC00493* transcript is localized predominantly in cytoplasm (**Figure 1**C). This

result suggests that the function of the *LINC00493* transcript is not related to the transcription regulation and chromatin binding. The observed results highlight similarities between the investigated lncRNA and mRNAs, such as sequence length, high expression level and accumulation in the cytoplasm.

2.4. Knockdown of LINC00493 Affects Cell Growth in a Cell-Type-Specific Manner

To determine the function of *LINC00493*, we analyzed previously published CRISPRi-based data for functional long noncoding RNA loci in human cells [13] and discovered that *LINC00493* modified cell growth in a cell-type-specific manner. To confirm the cell-type-specific role of *LINC00493*, we performed knockdown experiments using RNA interference in three human cell lines: HEK293T, A375 and MDA-MB-231.

LINC00493 knockdown efficiency was about 60–70% (**Figure 2**A). After knockdown, cell proliferation was measured by MTT assay, and cell migration was examined using wound-healing assay. We revealed that *LINC00493* knockdown reduced cell viability in HEK293T and A375 cell lines, while the opposite effect was observed in MDA-MB-231 (**Figure 2**C). Thus, knockdown experiments confirmed that downregulation of *LINC00493* affected cell proliferation activity in a cell-type-specific manner. On the other hand, wound-healing assay revealed that *LINC00493* knockdown did not affect cell migration (**Figure 2**B,D).
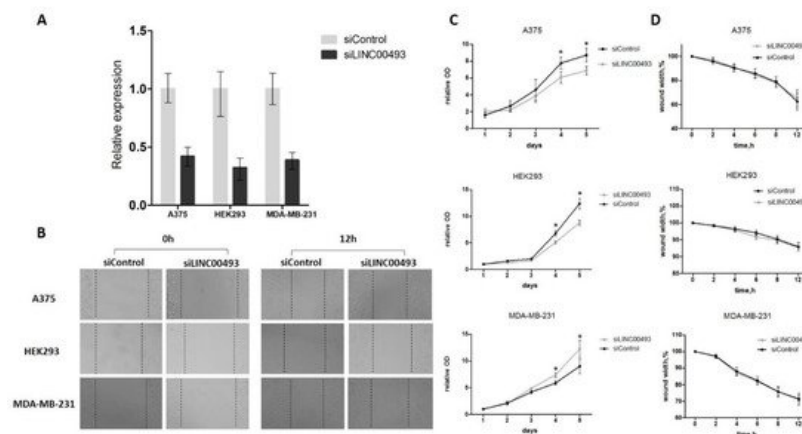
Figure 2. LINC00493 knockdown has a cell-type-specific effect. (**A**) *LINC00493* knockdown efficiency in A375, HEK293T and MDA-MB-231 cell lines. (**B**) The migration ability of A375, HEK293T and MDA-MB-231 cells after transfection of siLINC00493 and siControl was measured by wound-healing assay. Representative images of wound-healing experiments are shown. (**C**) MTT assay reveals the effect of *LINC00493* knockdown in 3 cell lines. Relative optical density value of A375, HEK293 and MDA-MB-231 cells with *LINC00493* knockdown and control cells in the MTT assay. (**D**) Summary graph showing typical wound-healing rates by A375, HEK293 and MDA-MB-231 cells after *LINC00493* knockdown. Error bars represent the mean ± SEM (standard error mean) of three independent experiments. * $p < 0.01$, vs. control (according to Mann–Whitney U test).

2.5. Small Protein Is Translated from LINC00493 RNA

According to the Human Protein Atlas [31], *LINC00493* contains a small open reading frame that could be translated to a 94/95-amino-acid protein—SMIM26. Using GWIPS-viz [32] and Trips-Viz [33], which provide ribosome profiling data, we confirmed that *LINC00493* has a strong ribosome association in the predicted sORF region (**Figure 3**C). The predicted secondary and tertiary structures of SMIM26, constructed using an improved predictor of protein structure [34], showed that the protein contains two alpha helixes; one of them is supposed to be a transmembrane domain (**Figure 3**B,D). A number of tools (Phobius, SPOCTOPUS, MEMPACK) predict that SMIM26 is localized in the membrane. Despite this fact, the Human Protein Atlas immunocytochemistry analysis revealed that this protein is localized mainly in the nucleoplasm and partially in the cytosol [31]. However, the function of the protein remains unknown. Comparative analysis of amino-acid and nucleic sequences of *LINC00493* reveals that the gene is evolutionarily conserved among mammals both at the RNA and protein level (**Figure 3**A). This fact supports a potential functional role for SMIM26, while the influence of this protein on cell migration and viability was not shown previously.
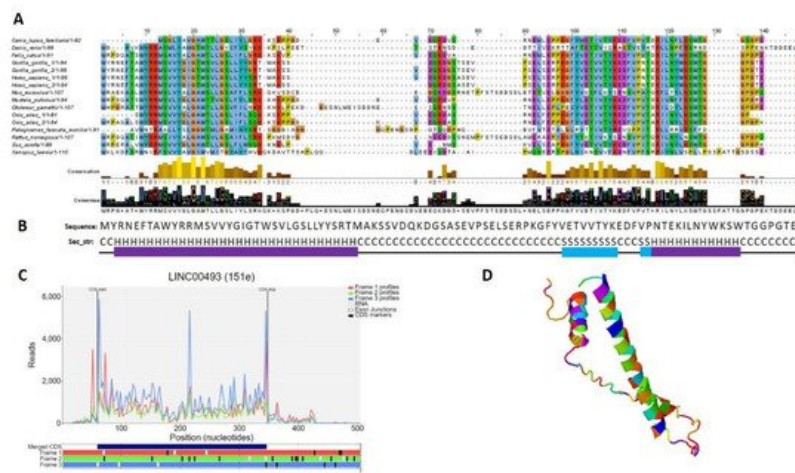
Figure 3. Analysis of the structure of the SMIM 26. (**A**) Multiple-species alignment of the amino acid sequences of the *LINC00493*-encoded small protein SMIM26. (**B**) Predicted secondary structure (H: helix; S: strand; C: coil) of SMIM26. (**C**) Distribution of the RPF reads in the *LINC00493*-ORF. (**D**) Predicted 3D structure of SMIM26.

2.6. SMIM26 Protein Affects Cell Viability

Data from knockdown experiments revealed that the *LINC00493* gene is important for cell viability. However, it remains unclear whether the observed effect is associated with small protein or with the RNA itself. To investigate separately the effect of SMIM26 protein and the *LINC00493* transcript on cell viability, we cloned a full-length *LINC00493* cDNA into pcDNA3.1-GFP vector (pcDNA3.1) and obtained LINC00493_WT plasmid. Next, we mutated the SMIM26 start-codon by site-directed mutagenesis to prevent the translation of small protein and obtained the LINC00493_MUT plasmid. Wild-type and mutant constructs, along with an empty pcDNA3.1, were transfected into A375 and HEK293T cells and cell viability was measured using the MTT test. The expression of *LINC00493* was increased about a thousand-fold times by qPCR (**Figure 4**A).
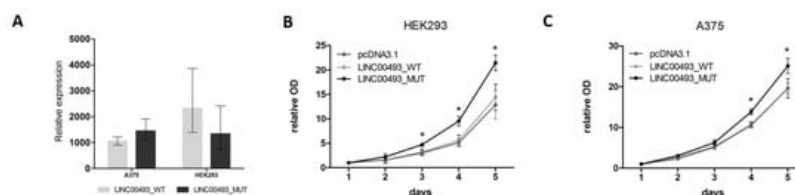
Figure 4. Influence of *LINC00493/SMIM26* overexpression on cell viability. (**A**) LINC00493_WT and LINC00493_MUT overexpression efficiency in A375 and HEK293 cell lines. (**B**,**C**) Relative optical density value of HEK293 and A375 cells treated by LINC00493_WT, LINC00493_MUT and control cells in the MTT assay. Cells with overexpression of mutated *LINC00493* showed higher viability than those transfected with WT *LINC00493* or empty pcDNA3.1 vectors. Error bars represent the mean ± SEM (standard error mean) of three independent experiments. * $p < 0.01$, vs. control (according to Mann–Whitney U test).

Interestingly, overexpression of WT *LINC00493* did not influence the number of cells, while overexpression of *LINC00493* with a mutated start-codon demonstrated increased cell viability (**Figure 4**B,C). This observation suggests that, even in the absence of SMIM26 protein translation, the *LINC00493* transcript itself has an effect on cell viability. However, the difference between WT and MUT *LINC00493* overexpression indicates that SMIM26 protein could also affect cell viability, but in the opposite manner.

References

1. Frankish, A.; Diekhans, M.; Ferreira, A.M.; Johnson, R.; Jungreis, I.; Loveland, J.; Mudge, J.M.; Sisu, C.; Wright, J.; Armstrong, J.; et al. Gencode reference annotation for the human and mouse genomes. Nucleic Acids Res. 2019, 47, D766–D773.

2. Hon, C.C.; Ramilowski, J.A.; Harshbarger, J.; Bertin, N.; Rackham, O.J.; Gough, J.; Denisenko, E.; Schmeier, S.; Poulsen, T.M.; Severin, J.; et al. An atlas of human long non-coding RNAs with accurate 5' ends. Nature 2017, 543, 199–204.

3. Dykes, I.M.; Emanueli, C. Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. Genom. Proteom. Bioinform. 2017, 15, 177–186.

4. St Laurent, G.; Wahlestedt, C.; Kapranov, P. The Landscape of long noncoding RNA classification. Trends Genet. TIG 2015, 31, 239–251.

5. Panni, S.; Lovering, R.C.; Porras, P.; Orchard, S. Non-coding RNA regulatory networks. Biochim. Biophys. Acta Gene Regul. Mech. 2020, 1863, 194417.

6. Cai, X.; Cullen, B.R. The imprinted H19 noncoding RNA is a primary microRNA precursor. RNA 2007, 13, 313–316.

7. Pandey, R.R.; Mondal, T.; Mohammad, F.; Enroth, S.; Redrup, L.; Komorowski, J.; Nagano, T.; Mancini-Dinardo, D.; Kanduri, C. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell 2008, 32, 232–246.

8. Orom, U.A.; Derrien, T.; Beringer, M.; Gumireddy, K.; Gardini, A.; Bussotti, G.; Lai, F.; Zytnicki, M.; Notredame, C.; Huang, Q.; et al. Long noncoding RNAs with enhancer-like function in human cells. Cell 2010, 143, 46–58.

9. Orom, U.A.; Derrien, T.; Guigo, R.; Shiekhattar, R. Long noncoding RNAs as enhancers of gene expression. Cold Spring Harb. Symp. Quant. Biol. 2010, 75, 325–331.

10. Gong, C.; Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature 2011, 470, 284–288.

11. Wang, Y.; Xu, Z.; Jiang, J.; Xu, C.; Kang, J.; Xiao, L.; Wu, M.; Xiong, J.; Guo, X.; Liu, H. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev. Cell 2013, 25, 69–80.

12. Keniry, A.; Oxley, D.; Monnier, P.; Kyba, M.; Dandolo, L.; Smits, G.; Reik, W. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat. Cell Biol. 2012, 14, 659–665.

13. Liu, S.J.; Horlbeck, M.A.; Cho, S.W.; Birk, H.S.; Malatesta, M.; He, D.; Attenello, F.J.; Villalta, J.E.; Cho, M.Y.; Chen, Y.; et al. CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science 2017, 355.

14. Sparber, P.; Filatova, A.; Khantemirova, M.; Skoblov, M. The role of long non-coding RNAs in the pathogenesis of hereditary diseases. BMC Med. Genom. 2019, 12, 42.

15. Bao, Z.; Yang, Z.; Huang, Z.; Zhou, Y.; Cui, Q.; Dong, D. LncRNADisease 2.0: An updated database of long non-coding RNA-associated diseases. Nucleic Acids Res. 2019, 47, D1034–D1037.

16. Lorenzen, J.M.; Thum, T. Long noncoding RNAs in kidney and cardiovascular diseases. Nat. Rev. Nephrol. 2016, 12, 360–373.

17. Chen, Y.; Zhou, J. LncRNAs: Macromolecules with big roles in neurobiology and neurological diseases. Metab. Brain Dis. 2017, 32, 281–291.

18. Chi, Y.; Wang, D.; Wang, J.; Yu, W.; Yang, J. Long Non-Coding RNA in the Pathogenesis of Cancers. Cells 2019, 8, 1015.

19. St Laurent, G.; Vyatkin, Y.; Kapranov, P. Dark matter RNA illuminates the puzzle of genome-wide association studies. BMC Med. 2014, 12, 97.

20. Hindorff, L.A.; Sethupathy, P.; Junkins, H.A.; Ramos, E.M.; Mehta, J.P.; Collins, F.S.; Manolio, T.A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA 2009, 106, 9362–9367.

21. Chew, G.L.; Pauli, A.; Rinn, J.L.; Regev, A.; Schier, A.F.; Valen, E. Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. Development 2013, 140, 2828–2834.

22. Aspden, J.L.; Eyre-Walker, Y.C.; Phillips, R.J.; Amin, U.; Mumtaz, M.A.; Brocard, M.; Couso, J.P. Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. eLife 2014, 3, e03528.

23. Bazzini, A.A.; Johnstone, T.G.; Christiano, R.; Mackowiak, S.D.; Obermayer, B.; Fleming, E.S.; Vejnar, C.E.; Lee, M.T.; Rajewsky, N.; Walther, T.C.; et al. Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J. 2014, 33, 981–993.

24. Choi, S.W.; Kim, H.W.; Nam, J.W. The small peptide world in long noncoding RNAs. Brief. Bioinform. 2019, 20, 1853–1864.

25. Stein, C.S.; Jadiya, P.; Zhang, X.; McLendon, J.M.; Abouassaly, G.M.; Witmer, N.H.; Anderson, E.J.; Elrod, J.W.; Boudreau, R.L. Mitoregulin: A lncRNA-Encoded Microprotein that Supports Mitochondrial Supercomplexes and Respiratory Efficiency. Cell Rep. 2018, 23, 3710–3720.e3718.

26. Chugunova, A.; Loseva, E.; Mazin, P.; Mitina, A.; Navalayeu, T.; Bilan, D.; Vishnyakova, P.; Marey, M.; Golovina, A.; Serebryakova, M.; et al. LINC00116 codes for a mitochondrial peptide linking respiration and lipid metabolism. Proc. Natl. Acad. Sci. USA 2019, 116, 4940–4945.

27. Konina, D.O.; Filatova, A.Y.; Skoblov, M.Y. LINC01420 RNA structure and influence on cell physiology. BMC Genom. 2019, 20, 298.

28. Haeussler, M.; Zweig, A.S.; Tyner, C.; Speir, M.L.; Rosenbloom, K.R.; Raney, B.J.; Lee, C.M.; Lee, B.T.; Hinrichs, A.S.; Gonzalez, J.N.; et al. The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. 2019, 47, D853–D858.

29. Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D.G.; et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. 2012, 22, 1775–1789.

30. FANTOM Consortium; The RIKEN PMI; CLST (DGT); Forrest, A.R.; Kawaji, H.; Rehli, M.; Baillie, J.K.; de Hoon, M.J.; Haberle, V.; Lassmann, T.; et al. A promoter-level mammalian expression atlas. Nature 2014, 507, 462–470.

31. Uhlen, M.; Fagerberg, L.; Hallstrom, B.M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; et al. Proteomics. Tissue-based map of the human proteome. Science 2015, 347, 1260419.

32. Michel, A.M.; Kiniry, S.J.; O'Connor, P.B.F.; Mullan, J.P.; Baranov, P.V. GWIPS-viz: 2018 update. Nucleic Acids Res. 2018, 46, D823–D830.

33. Kiniry, S.J.; O'Connor, P.B.F.; Michel, A.M.; Baranov, P.V. Trips-Viz: A transcriptome browser for exploring Ribo-Seq data. Nucleic Acids Res. 2019, 47, D847–D852.

34. Yang, J.; Anishchenko, I.; Park, H.; Peng, Z.; Ovchinnikov, S.; Baker, D. Improved protein structure prediction using predicted interresidue orientations. Proc. Natl. Acad. Sci. USA 2020, 117, 1496–1503.