
Article:
Świrski, M.I., Tierney, J.A.S., Albà, M.M. et al. (2025) Translon: a single term for translated regions. Nature Methods. ISSN: 1548-7091

https://doi.org/10.1038/s41592-025-02810-3

Translon: a single term for translated regions.

by

Michał I. Świrski[1,*], Jack A. S. Tierney[2,£,*], M. Mar Albà[3,4], Dmitry E. Andreev[5,6], Julie L. Aspden[7,8,9], John F. Atkins[2,10], Michal Bassani-Sternberg[11,12], Marla J. Berry[13], Stefano Biffo[14,15], Kathleen Boris-Lawrie[16], Mark Borodovsky[17,18], Ian Brierley[19], Matthew Brook[20], Marie A. Brunet[21,22], Janusz M. Bujnicki[23], Neva Caliskan[24,25], Lorenzo Calviello[26], Anne-Ruxandra Carvunis[27,28], Jamie H. D. Cate[29], Can Cenik[30], Kung Yao Chang[31], Yiwen Chen[32], Sonia Chothani[33], Jyoti S. Choudhary[34], Patricia L. Clark[35], Jim Clauwaert[36], Lynn Cooley[37], Erik Dassi[38], Kellie Dean[2], Jean-Jacques Diaz[39], Christoph Dieterich[40,41], Rivka Dikstein[42], Jonathan D. Dinman[43,44], Sergey E. Dmitriev[6,45], Olga A. Dontsova[5,46,47], Christine M. Dunham[48], Sandeep M. Eswarappa[49], Philip J. Farabaugh[50], Pouya Faridi[51,52], Ivo Fierro-Monti[53], Andrew E. Firth[19], David Gatfield[54], Fátima Gebauer[55,56], Mikhail S. Gelfand[47], Nicola K. Gray[57], Rachel Green[58], Chris H. Hill[59,60,61], Ya-Ming Hou[62], Norbert Hübner[63,64,65,66], Zoya Ignatova[67], Pavel Ivanov[68,69], Shintaro Iwasaki[70,71], Rory Johnson[72], Ahmad Jomaa[73,74], Marko Jovanovic[75], Irwin Jungreis[76,77], Manolis Kellis[76,77], Jeffrey S. Kieft[78], Alex V. Kochetov[79], Eugene V. Koonin[80], Andrei A. Korostelev[81], Joanna Kufel[1], Ivan V. Kulakovskiy[82,83], Leo Kurian[84,85,86], Denis L.J. Lafontaine[87,88], Ola Larsson[89], Gary Loughran[90], Julius Lukeš[91,92], Marco Mariotti[93], Elena S. Martens-Uzunova[94], Thomas F. Martinez[95,96,97], Akinobu Matsumoto[98], Joel McManus[99,100], Jan Medenbach[101], Sergey V. Melnikov[102], Gerben Menschaert[103], Catharina Merchante[104], Martin Mikl[105], W. Allen Miller[106], Oliver Mühlemann[107], Olivier Namy[108], Danny D. Nedialkova[109,110], Jozef Nosek[111], Sandra Orchard[53], Petar Ozretić[112], Mihaela Pertea[113], Dmitri D. Pervouchine[47], Luísa Romão[114,115], David Ron[116], Xavier Roucou[117,118], Maria P. Rubtsova[5,46], Jorge Ruiz-Orera[63], Alan Saghatelian[119,120], Steven L. Salzberg[121], Lucia A. Seale[13], Cathal Seoighe[122], Petr V. Sergiev[6], Premal Shah[123,124], Nikolay Shirokikh[125], Sarah A. Slavoff[126], Nahum Sonenberg[127,128], Timothy J. Stasevich[129], Roman J. Szczesny[130], Tiina Tamm[131], Marek Tchorzewski[132], Ivan Topisirovic[127,133], Michel L. Tremblay[127,128], Tamir Tuller[134], Igor Ulitsky[135], Leoš Shivaya Valášek[136], Petra Van Damme[137], Gabriella Viero[138], Juan Antonio Vizcaino[53], Christine Vogel[139], Edward W. J. Wallace[140], Jonathan S. Weissman[141,142,143], Eric Westhof[144,145], Nicola Whiffin[146,147], Daniel N. Wilson[67], Zhi Xie[148], Jonathan W. Yewdell[149], Martina M. Yordanova[2], Chien-Hung Yu[150,151], Vyacheslav Yurchenko[152], Bojan Zagrovic[153], TRANSLACORE, Eivind Valen[154,155,*], Pavel V. Baranov[2,*]

[1]University of Warsaw, Faculty of Biology, Institute of Genetics and Biotechnology, Warsaw, Poland

[2]School of Biochemistry and Cell Biology, University College Cork, Cork, T12 CY82, Ireland

[3]Hospital del Mar Research Institute, Dr. Aiguader 88, Barcelona 08003, Spain

[4]Catalan Institution for Research and Advanced Studies, Passeig Lluís Companys 23, Barcelona 08010, Spain

[5]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, 117997, Russia

[6]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, 119234, Russia.

[7]School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, UK

[8]Leeds Omics, University of Leeds, Leeds, LS2 9JT, UK

[9]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK,

[10]Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA

[11]University Hospital of Lausanne, Lausanne, 1005, Switzerland

[12]Ludwig Institute for Cancer Research, Lausanne, 1005, Switzerland

[13]Pacific Biosciences Research Center, University of Hawaii at Manoa, Honolulu, HI, 96822, USA

[14]National Institute of Molecular Genetics, Milan, 20132, Italy

[15]University of Milan, 20126 Milan, Italy

[16]Department of Veterinary and Biomedical Sciences, University of Minnesota, Saint Paul, MN 55108, USA

[17]Wallace H. Coulter Department of Biomedical Engineering, Georgia Tech, Atlanta, Georgia, 30332, USA

[18]School of Computational Science and Engineering, Georgia Tech, Atlanta, Georgia, 30332, USA

[19]Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QP, UK

[20]Centre for Cardiovascular Sciences, Institute for Neuroscience and Cardiovascular Research, University of Edinburgh, Edinburgh, EH16 4TJ, UK

[21]University of Sherbrooke, Sherbrooke, J1E 4K8, Canada

[22]Cancer Research Institute of the University of Sherbrooke (IRCUS), Sherbrooke, J1E 4K8, Canada

[23]International Institute of Molecular and Cell Biology in Warsaw, Trojdena 4, 02-109 Warsaw, Poland

[24]Regensburg Center for Biochemistry (RCB), University of Regensburg, 93053 Regensburg, Germany

[25]Helmholtz Institute for RNA-Based Infection Research, Helmholtz Centre for Infection Research (HIRI-HZI), Josef-Schneider-Strasse 2, 97080 Würzburg, Germany

[26]Human Technopole, 20157 Milan, Italy

[27]Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA

[28]Pittsburgh Center for Evolutionary Biology and Medicine (CEBaM), School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA

[29]University of California Berkeley; Berkeley, CA 94720, USA

[30]Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, 78712, USA

[31]Graduate Institute of Biochemistry, NCHU, Taiwan

[32]The Department of Bioinformatics & Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA

[33]Centre for Computational Biology, Duke-NUS medical school, Singapore

[34]Functional Proteomics team, Chester Beatty Laboratories, The Institute of Cancer Research, London SW3 6JB, UK

[35]Department of Chemistry & Biochemistry, University of Notre Dame, Notre Dame, IN, 46556, USA

[36]University of Michigan, Ann Arbor, 48109, USA

[37]Department of Genetics, Yale University School of Medicine, New Haven, CT, 06510, USA

[38]Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, 38123, Trento, Italy

[39]Centre de Recherche en Cancérologie de Lyon, Inserm U1052, CNRS UMR5286, Université de Lyon, Université Claude Bernard Lyon, Centre Léon Bérard, LabEx Dev2CAN, CEDEX 08, Lyon, France

[40]Klaus Tschira Institute for Integrative Computational Cardiology, Heidelberg, 69120, Germany

[41]German Center for Cardiovascular Research (DZHK) - Partner site Heidelberg/Mannheim, Germany

[42]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, 76100, Israel

[43]Department of Cell Biology and Molecular Genetics, College Park MD, 20742, USA

[44]Institute for Bioscience and Biotechnology Research, Rockville MD, 20850, USA

[45]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, 119234, Russia

[46] Faculty of Chemistry, Lomonosov Moscow State University, Moscow, 119991, Russia

[47]Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, 143025, Russia

[48]Department of Chemistry, Emory University, Atlanta, Georgia, USA

[49]Department of Biochemistry, Indian Institute of Science, Bangalore, 560012, India

[50]University of Maryland Baltimore County, Baltimore, Maryland 21250, USA

[51]Centre for Cancer Research, Hudson Institute of Medical Research, Melbourne, Australia, 3168

[52]Monash Proteomics and Metabolomics Platform, Department of Medicine, School of Clinical Sciences, Monash University, Melbourne, 3168, Australia

[53]European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge. CB10 1SD. UK

[54]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland

[55]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, 08003, Spain

[56]Universitat Pompeu Fabra (UPF), Barcelona, 08003, Spain

[57]Centre for Reproductive Health, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh, EH16 4UU, UK

[58]HHMI, Johns Hopkins University School of Medicine, Baltimore, Maryland, 21210 USA

[59]Department of Biology, University of York, York, YO10 5DD, UK

[60]York Structural Biology Laboratory, University of York, York, YO10 5DD, UK

[61]York Biomedical Research Institute, University of York, York, YO10 5DD, UK

[62]Department of Biochemistry and Molecular Biology, Thomas Jefferson University, Philadelphia, PA, USA

[63]Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), 13125, Berlin, Germany

[64]Charite - Iniversitätsmedizin Berlin, Germany

[65]Helmholtz-Institute for Translational AngioCardioScience (HI-TAC) of the Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC) at Heidelberg University, Heidelberg 69117, Germany

[66]German Center for Cardiovascular Research (DZHK) - Partner site Berlin, Germany

[67]Institute for Biochemistry and Molecular Biology, University of Hamburg, 20146, Hamburg

[68]Division of Rheumatology, Inflammation and Immunity, Brigham and Women's Hospital, Boston, MA, 02115, USA

[69]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

[70]RNA Systems Biochemistry Laboratory, Pioneering Research Institute, Wako, Saitama 351-0198, Japan

[71]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8561, Japan

[72]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland.

[73]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA, 22903, USA

[74]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22903, USA

[75]Columbia University, New York, NY 10025, USA

[76]MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, 02139 USA

[77]Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA

[78]New York Structural Biology Center, New York, NY, 10027, USA

[79]Institute of Cytology & Genetics, Novosibirsk, Russia

[80]Computational Biology Branch, DIvision of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

[81]RNA Therapeutics Institute, University of Massachusetts Chan Medical School. Worcester, MA 01605, USA

[82]Institute of Protein Research, Russian Academy of Sciences, Pushchino, 142290, Russia

[83]Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, 119991, Russia

[84]Institute for Cardiovascular Physiology, Goethe University Frankfurt, Germany

[85]German Center for Cardiovascular Research (DZHK), Partner Site Rhein-Main, D-60590 Frankfurt, Germany.

[86]Cardio-Pulmonary Institute (CPI), Goethe University Frankfurt, Frankfurt, Germany.

[87]RNA Molecular Biology, Université libre de Bruxelles, Biopark campus, B-6041 Gosselies, Belgium

[88]Fonds de la Recherche Scientifique (F.R.S./FNRS), B-1000 Brussels, Belgium

[89]Department of Oncology-Pathology, Science for Life Laboratory, Karolinska Institutet, 17165 Stockholm, Sweden

[90]EIRNA Bio, Cork, Ireland

[91]Institute of Parasitology, Biology Centre, 37005 České Budějovice, Czech Republic

[92]Faculty of Science, University of South Bohemia, 37005 České Budějovice, Czech Republic

[93]Department of Genetics, Microbiology and Statistics, Universitat de Barcelona, Barcelona, Catalonia, 08024, Spain

[94]Erasmus MC Cancer Institute, University Medical Center Rotterdam, Department of Urology, 3000CA Rotterdam, The Netherlands

[95]Department of Pharmaceutical Sciences, University of California, Irvine, Irvine, CA 92617, USA

[96]Department of Biological Chemistry, University of California, Irvine, Irvine, CA 92617, USA

[97]Chao Family Comprehensive Cancer Center, University of California, Irvine, Irvine, CA 92617, USA

[98]Division of Biological Science, Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan

[99]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

[100]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

[101]University Hospital RWTH Aachen, 52074 Aachen, Germany

[102]Biosciences Institute, Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

[103]Faculty of Bioscience Engineering, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, 9000, Belgium

[104]Instituto de Hortofruticultura Subtropical y Mediterránea La Mayora (IHSM-UMA-CSIC). Boulevard Louis Pasteur, 49. 29010 Málaga, Spain

[105]Department of Human Biology, University of Haifa, Haifa, 3498838, Israel

[106]Plant Pathology, Entomology & Microbiology Department, Iowa State University, Ames, Iowa, 50011, USA

[107]Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, 3012 Bern, Switzerland

[108]CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Université Paris-Saclay, Gif-sur-Yvette, 91198, France

[109]Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

[110]Department of Bioscience, TUM School of Natural Sciences, 85748 Garching, Germany

[111]Comenius University Bratislava, Faculty of Natural Sciences, 842 15 Bratislava, Slovakia

[112]Laboratory for Hereditary Cancer, Division of Molecular Medicine, Ruđer Bošković Institute, 10000 Zagreb, Croatia

[113]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21211, USA

[114]Instituto Nacional de Saúde Doutor Ricardo Jorge, 1649-016 Lisboa, Portugal

[115]BioISI-Biosystems Integrative Sciences Institute, Faculty of Sciences, University of Lisboa, 1749-016 Lisboa, Portugal

[116]Cambridge Institute for Medical Research, Department of Clinical Biochemistry, University of Cambridge, Cambridge CB2 0XY, UK

[117]Department of biochemistry and functional genomics, Université de Sherbrooke, 3201 Jean Mignault, Sherbrooke, QC J1E 4K8, Canada

[118]Centre de Recherche du Centre Hospitalier Universitaire de Sherbrooke (CRCHUS), Sherbrooke, QC J1H 5N4, Canada

[119]Salk Institute, La Jolla, CA, 92037, USA

[120]UCSD, La Jolla, CA, 92037, USA

[121]Center for Computational Biology, Johns Hopkins University, Baltimore, Maryland 21211, USA

[122]School of Mathematical and Statistical Sciences, University of Galway, Galway, Ireland

[123]Convergence Bio, MA USA 02139

[124]Rutgers University, NJ USA 08807

[125]School of Human Sciences, University of Western Australia, Crawley, M309, Perth, 6009, Western Australia, Australia

[126]Department of Chemistry, Yale University, New Haven, CT 06520, USA

[127]Department of Biochemistry, McGill University, Montréal, QC, H3G 1Y6, Canada

[128]Rosalind and Morris Goodman Cancer Research Centre, McGill University, Montréal, QC, H3A 1A3, Canada

[129]Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, CO, 80525, USA

[130]Laboratory of RNA Biology, Institute of Biochemistry and Biophysics Polish Academy of Sciences, Warsaw, 02-106, Poland

[131]Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia

[132]Department of Molecular Biology, Institute of Biological Sciences, Maria Curie-Skłodowska University, Lublin, Poland

[133]Lady Davis Institute, Gerald Bronfman Department of Oncology, Division of Experimental Medicine, McGill University, Montréal H3A 1A3, Canada

[134]Department of Biomedical Engineering, Tel-Aviv University, Tel Aviv, Israel

[135]Department of Immunology and Regenerative Biology, Weizmann Institute of Science, Rehovot 76100, Israel

[136]Laboratory of Regulation of Gene Expression Institute of Microbiology of the Czech Academy of Sciences Videnska 1083, 142 20, Prague 4, Czech Republic

[137]iRIP Unit, Laboratory of Microbiology, Department of Biochemistry and Microbiology, Ghent University, 9000 Ghent, Belgium

[138]Institute of Biophysics CNR, Via Sommarive 18 38123 Trento, Italy

[139]Department of Biology, New York University, New York, 10003, USA

[140]School of Biological Sciences, University of Edinburgh, Edinburgh, EH9 3FF, UK

[141]Whitehead Institute, Cambridge, MA 02139, USA

[142]Department of Biology, MIT, Cambridge, MA 02139 USA

[143]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

[144]Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire, Architecture et Réactivité de l'ARN, CNRS UPR9002, 2 allée Konrad Roentgen, F-67084 Strasbourg, France

[145]Engineering Research Center of Clinical Functional Materials and Diagnosis & Treatment Devices of Zhejiang Province, Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, China

[146]Big Data Institute, University of Oxford, Oxford, OX3 7LF, UK

[147]Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK

[148]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China

[149]Cellular Biology Section, Laboratory of Viral Diseases, Division of Intramural Research, NIAID, Bethesda, 20892, USA

[150]Department of Biochemistry and Molecular Biology, National Cheng Kung University, Tainan, 701, Taiwan

[151]Institute of Basic Medical Sciences, College of Medicine, National Cheng Kung University, Tainan, 701, Taiwan

[152]Life Science Research Centre, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czech Republic

[153]Department of Structural and Computational Biology, Max Perutz Labs, University of Vienna, Campus Vienna Biocenter 5, 1030, Vienna, Austria

[154]Department of Biosciences, University of Oslo, Kristine Bonnevies hus, Blindernveien 31, 0731 Oslo, Norway

[155]Computational Biology Unit, Department of Informatics, University of Bergen, Thormøhlensgate 55, 5008 Bergen, Norway.

[£]Current affiliation: European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge. CB10 1SD, UK

*To whom correspondence should be addressed: michal.swirski@uw.edu.pl, jackt@ebi.ac.uk, eivind.valen@ibv.uio.no, p.baranov@ucc.ie

During translation, ribosomes synthesise polypeptides using RNA molecules as templates. All cellular proteins are products of translation, and the identification of protein-coding regions is the primary goal of genome annotation. Beyond protein synthesis, translation has long been known to have regulatory functions independent of its products[1,2]. However, only with the advent of ribosome profiling, was the broad scale and complexity of translated regions fully appreciated.

Recent interest in pervasive translation has exposed a lack of general terminology for translated regions that does not depend on the properties of their products or their sequence. In the absence of such terminology, a range of inconsistently defined terms are currently used to describe them. These terms are typically variations on Open Reading Frame (ORF), for example, ***n***on-***c***anonical, ***RiboSeq***, ***alt***ernative, ***t***ranslated, ***sm***all, ***s***hort, and other ***x***ORFs-based labels. Such terms largely overlap and redundantly describe the same core concept, that the region in question is translated.

Denoting translated regions with ORF-based terms is problematic for two main reasons. First, in sequence analysis, ORFs are defined purely by the nucleotide sequence and the relevant genetic code (starts/stops; Fig. 1a). Thus, ORFs are found everywhere in the genome, including regions that are not transcribed. This ORF definition is useful for the prediction of potential protein-coding regions. However, using the same term to specify regions that are known or confidently predicted to be translated leads to confusion. Indeed, for annotated protein-coding regions, a different technical term, CDS (historically from 'Coding DNA Sequence'), is commonly used.

The second problem is that, although most known protein-coding regions in the majority of model organisms conform to an ORF, starting at an AUG and terminating at the first in-frame stop, there are numerous alternatives. In addition to AUG, other triplets can also initiate translation[3] (Fig. 1a), making it challenging to determine start codons purely from sequence. The incorporation of non-standard proteinogenic amino acids (selenocysteine/pyrrolysine) usually occurs at stop codons within the CDS, with over a hundred in CDSs encoding selenoproteins in some species[4] (Fig. 1a). Furthermore, in some organisms, termination codons are defined not only by their sequence, but also by their position within mRNA[5].

Ribosomal frameshifting (Fig. 1a), which involves translation of two different reading frames to produce a single protein, is common in viruses and also occurs in nuclear-encoded genes of most organisms, in some species, in 5-20% of the genes[6]. Because the products of translation of frameshifted CDSs cannot be derived automatically by converting nucleotide triplets to amino acids, current annotation practice is to introduce a "pseudo-intron" between two partial ORFs in the annotation of these genes (Fig. 1b). Most concerning is the practice of automatically modifying RNA sequences inferred from the genome to obtain full-length protein product via triplet translation (Fig. 1c). Such artificial fitting of well-established translation mechanisms to a simplified model reinforces a widespread fallacy that all translated regions can be represented as single ORFs. This leads to an inaccurate and oversimplified representation of genetic information and molecular composition of the cells, with potentially serious consequences for those unaware of the underlying makeshift solutions. Similar terminological confusion, such as the routine conflation of "exons" and "protein-coding regions"[7], highlights the need for precise vocabulary relating to gene expression.

The above problems can be solved or at least alleviated with the introduction of a specific term for a translated region that would be defined without reference to the product of translation or to the sequence of the region. For this purpose, we suggest using the term *translon* (short for **transl**ated regi**on**), which has previously been introduced but failed to gain traction[8]. It aligns well

with other terms describing gene structures: intron (**intr**agenic regi**on**) and exon (**ex**pressed regi**on**)[9].

The term translon denotes any region that is decoded by the ribosome. This ranges from minimal sequences with detectable translation (AUG followed by a stop) to sequences encoding long proteins in multiple reading frames, and even those disrupted by non-coding sequences as in translational bypassing[10] (Fig. 1a). It also facilitates the efforts to characterise unannotated translation: newly identified translated regions can be described as novel translons. Their biological roles, if any, can remain enigmatic until specific information is obtained. These roles can be purely regulatory, independent of the translation products, or involving those products, such as short peptides modulating functions of other proteins, signalling peptides or antigens; some translons could encode neutral or even harmful "protein junk".

Translon would fill the gap in the vocabulary used to describe units of genetic information (Fig. 1d). Unlike the existing terminology, translon is directly defined by the process it aims to capture instead of doing so indirectly, through sequence or function. We expect that this term will reduce confusion when discussing translated regions and facilitate development of more biologically realistic annotations.

ACKNOWLEDGMENTS

We wish to thank the participants of 2024 EMBO Workshop "Recoding and the diversity of genetic decoding" and of 2024 Dagstuhl Seminar "Deep Learning for RNA Regulation and Multidimensional Transcriptomics" as well as the members of TransCODE consortium and TRANSLACORE network (supported by COST action CA21154) for invaluable discussions that led to the development of this proposal.

AUTHOR CONTRIBUTIONS

M.I.S., J.A.S.T, E.V., and P.V.B developed the concept and drafted the correspondence. M.M.A., D. E.A., J.L.A., J.F.A., M.B.S., M.J.B., S.B., K.B.-L., M.Borodovsky, I.B., M.Brook, M.A.B., J.M.B., N.C., L.Calviello, A.-R.C., J.H.D.C., C.C., K.Y.C., Y.C., S.C., J.S.C., P.L.C, J.C., L.Cooley, E.D., K.D., J.J. D., C.D., R.D., J.D.D., S. E.D., O.A.D., C.M.D., S.M.E., P.J.F., P.F., I.F. -M., A.E.F., D.G., F.G., M.S.G., N. K.G., R.G., C.H.H., Y.-M.H., N.H., Z.I., P.I., S.I., R.J., A.J., M.J., I.J., M.K., J.S.K., A.V.K., E.V.K., A.A.K., J.K., I.V.K., L.K., D.L.J.L., O.L., G.L., J.L, M.Mariotti, E.S.M.-U., T.F.M., A.M., J.McManus, J.Medenbach, S.V.M., G.M., C.M., M.Mikl, W.A.M., O.M., O.N., D.D.N., J.N., S.O., P.O., M.P., D.D.P., L.R, D.R., X.R., M.P.R., J.R.-O., A.S., S.L.S., L.A.S., C.S., P.V.S., P.S., N.Shirokikh, S.A.S., N.Sonenberg, T.J.S., R.J.S., T. Tamm, M.T., I.T., M.L.T., T.Tuller, I.U., L.S.V., P.V.D., G.V., J.A.V., C.V., E.W.J.W., J.S.W, E.W., N.W., D.N.W, Z.X., J.W.Y., M.M.Y., C.-H.Y., V.Y., and B.Z. critically discussed the concept and/or edited the manuscript at different stages of its development.

ETHICS DECLARATION

Competing interests

A.-R.C. is on the scientific advisory board of ProFound Therapeutics. G.L. & P.V.B. are co-founders and shareholders of Eirnabio Ltd. The remaining authors declare no competing interests.

REFERENCES

1. Miller, P. F. & Hinnebusch, A. G. cis-acting sequences involved in the translational control of GCN4 expression. Biochim Biophys Acta 1050, 151–154 (1990).

2. Orbach, M. J., Sachs, M. S. & Yanofsky, C. The Neurospora crassa arg-2 locus. Structure and expression of the gene encoding the small subunit of arginine-specific carbamoyl phosphate synthetase. J Biol Chem 265, 10981–10987 (1990).

3. Andreev, D. E. et al. Non-AUG translation initiation in mammals. Genome Biol 23, 111 (2022).

4. Baclaocos, J. et al. Processive Recoding and Metazoan Evolution of Selenoprotein P: Up to 132 UGAs in Molluscs. J Mol Biol 431, 4381–4407 (2019).

5. Zinshteyn, B. & Green, R. When stop makes sense. Science 354, 1106 (2016).

6. Lobanov, A. V. et al. Position-dependent termination and widespread obligatory frameshifting in Euplotes translation. Nat Struct Mol Biol 24, 61–68 (2017).

7. Aspden, J. L., Wallace, E. W. J. & Whiffin, N. Not all exons are protein coding: Addressing a common misconception. Cell Genom 3, 100296 (2023).

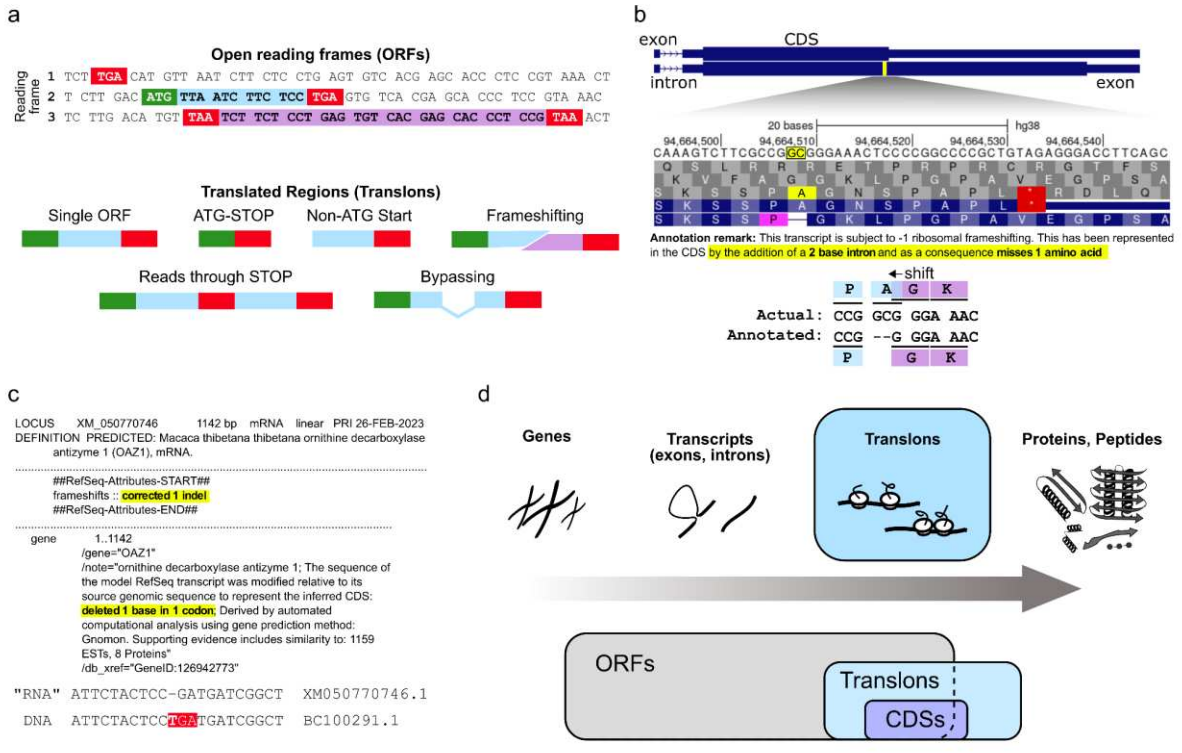
8. Goel, S. C. Transcription Unit. Nature 245, 397–397 (1973).

9. Gilbert, W. Why genes in pieces? Nature 271, 501 (1978).

10. Nosek, J., Tomaska, L., Burger, G. & Lang, B. F. Programmed translational bypassing elements in mitochondria: structure, mobility, and evolutionary origin. Trends Genet 31, 187–194 (2015).

FIGURE LEGEND

Figure 1| Open Reading Frames (ORFs) and translated regions. **a.** Any nucleotide sequence can be represented as three sequences of nucleotide triplets (codons), i.e. triplets read in three different reading frames. An ORF is defined as a sequence of sense codons bounded by a stop codon (purple bar) or from a start (most commonly ATG) to a stop (light blue bar). Below are examples of how different translated regions (translons) relate to ORF. **b.** Graphical representation of human PEG10 annotation zoomed in the area of ribosomal frameshifting site. Two different transcripts are used to represent two different proteins synthesised during translation of PEG10 mRNA. The one representing the full-length protein is artificially modified with a two-nucleotide "pseudo-intron". As a result of this modification, the product of conceptual translation differs from the full-length product of actual translation by one amino acid **c.** A fragment of macaca *OAZ1* gene transcript (XM_050770746.1) annotation illustrating automatic "correction" of mRNA sequences relative to the genomic sequence. The alignment of predicted RNA and genomic source is shown below, with the conserved stop codon at the *OAZ1* frameshifting site shown in red and deleted T in bold. **d.** Relationship of the translon to other terms used when interpreting genome information. Translon fills a missing step in the description of genetic information flow in the cells. Below is a relationship of translon to ORFs and CDSs: most ORFs are not transcribed. Of those that are, most are not translated; all CDSs are translons, but not all translons and CDSs are ORFs.


a
Open reading frames (ORFs)
Reading frame
1 TCT TGA CAT GTT AAT CTT CTC CTG AGT GTC ACG AGC ACC CTC CGT AAA CT
2 T CTT GAC ATG TTA ATC TTC TCC TGA GTG TCA CGA GCA CCC TCC GTA AAC
3 TC TTG ACA TGT TAA TCT TCT CCT GAG TGT CAC GAG CAC CCT CCG TAA ACT
Translated Regions (Translons)
Single ORF
ATG-STOP
Non-ATG Start
Frameshifting
Reads through STOP
Bypassing
b
exon
CDS
intron
exon
20 bases
hg38
94,664,500
94,664,510
94,664,520
94,664,530
94,664,540
CAAAGTCTTCGCCGGCGGGAAACTCCCCGGCCCCGCTGTAGAGGGACCTTCAGC
Annotation remark: This transcript is subject to -1 ribosomal frameshifting. This has been represented in the CDS by the addition of a 2 base intron and as a consequence misses 1 amino acid
shift
P A G K
Actual: CCG GCG GGA AAC
Annotated: CCG --G GGA AAC
P G K
c
LOCUS XM_050770746 1142 bp mRNA linear PRI 26-FEB-2023
DEFINITION PREDICTED: Macaca thibetana thibetana ornithine decarboxylase antizyme 1 (OAZ1), mRNA.
##RefSeq-Attributes-START##
frameshifts :: corrected 1 indel
##RefSeq-Attributes-END##
gene 1..1142
/gene="OAZ1"
/note="ornithine decarboxylase antizyme 1; The sequence of the model RefSeq transcript was modified relative to its source genomic sequence to represent the inferred CDS: deleted 1 base in 1 codon; Derived by automated computational analysis using gene prediction method: Gnomon. Supporting evidence includes similarity to: 1159 ESTs, 8 Proteins"
/db_xref="GeneID:126942773"
"RNA" ATTCTACTCC-GATGATCGGCT XM050770746.1
DNA ATTCTACTCCTGATGATCGGCT BC100291.1
d
Genes
Transcripts (exons, introns)
Translons
Proteins, Peptides
ORFs
Translons
CDSs