non-coding RNA

MDPI

*Article*

# FuncPEP v2.0: An Updated Database of Functional Short Peptides Translated from Non-Coding RNAs

**Swati Mohapatra [1,2], Anik Banerjee [2,3], Paola Rausseo [1,4], Mihnea P. Dragomir [5,6,7], Ganiraju C. Manyam [8], Bradley M. Broom [8] and George A. Calin [1,9,*]**

1 Department of Translational Molecular Pathology, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; smohapatra@mdanderson.org (S.M.); rausseopaola@gmail.com (P.R.)
2 The University of Texas MD Anderson Cancer Center UTHealth Houston Graduate School of Biomedical Sciences, Houston, TX 77030, USA; anik.banerjee@uth.tmc.edu
3 Department of Neurology, University of Texas McGovern Medical School, Houston, TX 77030, USA
4 Scripps College, Claremont, CA 91711, USA
5 Institute of Pathology, Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, 10117 Berlin, Germany; mihnea.p.dragomir@gmail.com
6 German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
7 Berlin Institute of Health at Charité, 10117 Berlin, Germany
8 Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; gcmanyam@mdanderson.org (G.C.M.)
9 Center for RNA Interference and Non-Coding RNAs, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
* Correspondence: gcalin@mdanderson.org

**Abstract:** Over the past decade, there have been reports of short novel functional peptides (less than 100 aa in length) translated from so-called non-coding RNAs (ncRNAs) that have been characterized using mass spectrometry (MS) and large-scale proteomics studies. Therefore, understanding the bivalent functions of some ncRNAs as transcripts that encode both functional RNAs and short peptides, which we named ncPEPs, will deepen our understanding of biology and disease. In 2020, we published the first database of functional peptides translated from non-coding RNAs—FuncPEP. Herein, we have performed an update including the newly published ncPEPs from the last 3 years along with the categorization of host ncRNAs. FuncPEP v2.0 contains 152 functional ncPEPs, out of which 40 are novel entries. A PubMed search from August 2020 to July 2023 incorporating specific keywords was performed and screened for publications reporting validated functional peptides derived from ncRNAs. We did not observe a significant increase in newly discovered functional ncPEPs, but a steady increase. The novel identified ncPEPs included in the database were characterized by a wide array of molecular and physiological parameters (i.e., types of host ncRNA, species distribution, chromosomal density, distribution of ncRNA length, identification methods, molecular weight, and functional distribution across humans and other species). We consider that, despite the fact that MS can now easily identify ncPEPs, there still are important limitations in proving their functionality.

**Keywords:** non-coding RNAs; ncRNA-encoded peptides; immunity; micro-peptides



**Citation:** Mohapatra, S.; Banerjee, A.; Rausseo, P.; Dragomir, M.P.; Manyam, G.C.; Broom, B.M.; Calin, G.A. FuncPEP v2.0: An Updated Database of Functional Short Peptides Translated from Non-Coding RNAs. *Non-Coding RNA* **2024**, *10*, 20. https://doi.org/10.3390/ncrna10020020

Academic Editor: Luca Agnelli

Received: 5 March 2024
Revised: 27 March 2024
Accepted: 28 March 2024
Published: 9 April 2024



## 1. Introduction

Conventional studies have defined non-coding RNAs (ncRNAs) as having no protein-coding potential [1]. However, several ncRNAs with small open reading frames (smORFs) have been reported to be translated into short functional peptides, which, by definition, are less than 100 aa in length [2]. By comparison, more than 95% of the proteins translated from coding regions are significantly longer than 100 aa [3–5]. Peptides encoded by ncRNAs are termed ncRNA-encoded peptides (ncPEPs). ncPEPs have been extensively characterized in

silico [6,7], but validation by wet lab experiments using high-throughput mass spectrometry (MS) and ribosome profiling is still lacking for many [8,9]. In addition, the functional characterization of ncPEPs has not been extensively studied compared to peptides derived from coding regions. smORFs in the transcripts of ncPEPs lack evolutionary conservation compared to the peptides derived from coding regions [9], leading to tissue-specific expression in certain species [1] and most likely in specific pathological states. ncPEPs can be used as biomarkers for tissue differentiation under physiological conditions and during disease progression. There was a spike in the number of discoveries and reports of FuncPEPs in 2020, the year the first version of a FuncPEP was published [10] and, therefore, we aim to keep the database updated to motivate researchers in the field to be on the lookout for novel functional peptides while they study the classical and nonclassical functions of ncRNAs [11,12].

Almost all classes of ncRNAs have been shown to encode ncPEPs. ncRNAs are broadly classified according to their length into long ncRNAs (lncRNAs), which are longer than 200 bp, and short non-coding RNAs (sncRNAs), which are shorter than 200 nucleotides (nts). Furthermore, different classes of lncRNAs are reported based on their genomic location, such as long intergenic ncRNAs (lincRNAs), or based on their conservation across species such as transcribed ultraconserved regions (T-UCRs), which are transcribed from ultraconserved genomic regions. Similarly, sncRNAs are classified into several categories, including microRNAs (miRNAs), PIWI-interacting RNAs (piRNAs), and transfer RNAs (tRNAs). Circular RNAs (circRNAs) or small nucleolar RNAs (snoRNAs) belong to both classes depending on their bp length. lncRNA and sncRNA have been extensively studied and reported to be involved in maintaining homeostasis [13] and contributing to pathological conditions [14,15]. The function of ncRNA transcripts is complex, and in recent years we have seen that most of these classes can also function by encoding ncPEPs, which broadens their mechanistic spectrum.

ncPEPs are mainly identified by computational prediction software for potential ORFs, and out of the many predicted ncPEPs, only a handful of them have been characterized by MS and techniques such as Western blotting, immunohistochemistry, and ribosome profiling. Moreover, the functional characterization of ncPEP is difficult and limited; thus, more effective techniques for functional validation are warranted to expand this field. Only a few experimentally discovered ncPEPs have been functionally characterized in the last decade with an aim to understand their role in ncRNA-associated pathology and in tissue homeostasis [16,17]. The biologically active short ncPEPs are functionally distinct from peptides translated from highly conserved messenger RNAs (mRNAs) [18] or short peptides generated through the breakdown of large proteins. For example, human smORFs are not always conserved beyond primates, while several human proteins are highly conserved across species [19,20]. Regarding the function of human ncPEPs, they are mostly involved in rapidly changing adaptive processes such as immunity and tumorigenesis [13,21].

Studies reporting ncPEPs using indirect (in silico analyses and peptide sequencing data without molecular validation) and predictive methods need more characterization and cannot be argued as being functional. Therefore, in 2020, we put together the first database [10] of **fu**nctional **nc**RNA-encoded **pep**tides (FuncPEPs) only including ncPEPs that had been validated by experimental methods and functionally characterized. As proteomics technologies have significantly developed, we hypothesized that, in recent years, the discovery of ncPEPs has escalated and, therefore, we performed an update of our initial database. This revised version of the database (FuncPEP v2.0) includes the new functional ncPEPs that have been published in the last 3 years along with the categorization of the host ncRNA from which they are derived.

## 2. Systematic Profiling, Data Collection, and Construction of Database

### *2.1. Collection of Data and Database Construction*

The selected peptides were translated from ncRNAs containing smORFs and were included in the database as per the same criteria as in the first version of FuncPEP: (1) val-

idated by molecular biology techniques (MS, Western blotting, immunohistochemistry, immunofluorescence) and/or indirect methods such as ribosome profiling and loss/gain of function studies by wet lab experiments; (2) functionally characterized by demonstrating their involvement in physiological homeostasis or disease; and (3) size less than 100 aa in length. Some of the ncPEPs are validated only by indirect methods and are marked with an asterisk in the database and require further validation. ncPEPs identified by computational methods only were not included in the database. We used the same inclusion criteria as in the first version of FuncPEP [10], performing searches of PubMed and Google Scholar to identify and characterize ncPEPs in the period from 2020 to present using keywords ("functional peptides", "antisense RNA", OR "lincRNA", OR "lncRNA", OR "miRNA", OR "circRNA", OR "rRNA", OR "tRNA", OR "ncPEP"). Furthermore, these terms were filtered by incorporatingterms regarding "translation" ("smORF", OR "Ribo-seq", OR "ribosome profiling", OR "mass spectrometry", OR "translation").

### *2.2. Selection Criteria for ncPEPs Included in FuncPEP*

A PubMed search incorporating the keywords was performed and screened for validated functional peptides derived from non-coding RNAs. The novel identified ncPEPs included in our database were characterized by a wide array of molecular and physiological parameters (i.e., types of host ncRNA, species distribution, chromosomal density, distribution of ncRNA length, identification methods, molecular weight and functional distribution across humans and other species). Exclusive to our database, we only included ncPEPs that were detected by direct means (MS, Western blotting, immunohistochemistry, immunofluorescence) or indirect means (ribosome profiling, loss/gain of function studies) of experimental confirmation. All ncPEPs from the database are also functionally characterized and are linked with physiological and pathological processes as well.

## 3. Database Construction and Results

### *3.1. Systemic Review and Database Interface*

A total of 15,805 articles were identified for the period August 2020 to July 2023. After accounting for duplicate articles, 10,502 articles were initially screened by title and abstract. Subsequently, 927 articles were selected for full-text evaluation to identify a wide range of biological and molecular features (i.e., species specificity, type of RNA source, amino acid length, etc.), which are easily accessible through FuncPEP's user-friendly interface. Of these 927 candidate studies, only 40 studies, with a total of 40 validated functional ncPEPs, were included in the FuncPEP v2.0 database, according to the stated inclusion criteria (Figure 1A). The updated version of the website, described in this manuscript, includes the most recently investigated and characterized FuncPEPs to date (Figure 1B). Moreover, it was noted that there has not been a large increase in the number of ncPEPs identified in the last 3 years (Figure 1B). Throughout the last decade, the number of identified functional peptides has increased due to advances in standardized techniques.

FuncPEP is designed to have a better, user-friendly interface, allowing researchers to access all the validated ncPEPs discovered so far through the website. The website can be accessed via the following link: https://bioinformatics.mdanderson.org/Supplements/FuncPEP/ (accessed on 27 March 2024). The major sections of the website are divided into the (1) Home section, which includes a description of FuncPEP; (2) Database section, which includes a dynamic hyperlinked table browser providing detailed information about all the ncPEPs; (3) Method section, which includes a description of the workflow that we used to collect and curate the ncPEP characteristics; and (4) Help section, which includes details and aids the user in website navigation (Figure 2A–D).

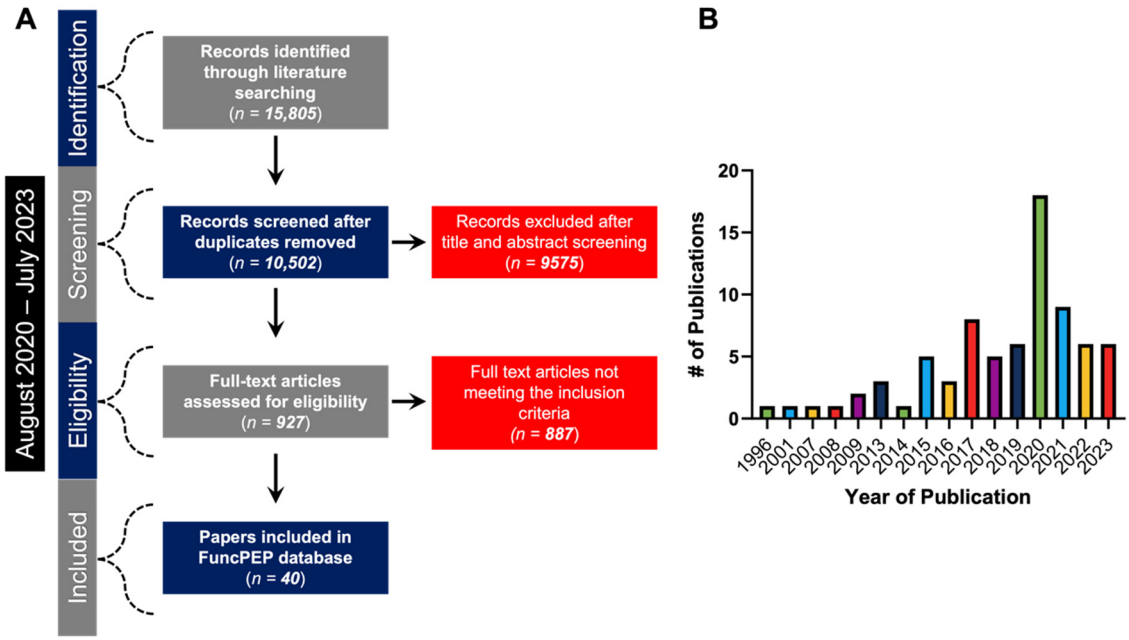


**Figure 1.** (**A**) Workflow demonstrating the selection criteria (i.e., identification, screening, eligibility testing, and inclusion) for FuncPEPs included in the database. (**B**) Analysis indicating the number of publications pertaining to functional ncPEPs. # means "Numbers".


**Figure 2.** Database interface, as previously described. (**A**–**D**) The wide array of tabs the user can access to find pertinent information regarding the functional peptides included in our database: about, index, info, and help, respectively. The red boxes indicate the pertinent tab being observed for the corresponding panel.

### *3.2. Characteristics of ncPEPs from the FuncPEP v2.0 Database*

FuncPEP v2.0 has been designed to provide a dynamic resource for accumulating ncPEP data from peer-reviewed scientific literature. Our database also provides information on the molecular characteristics of the included ncPEPs. The overall function of the database website contains information and characteristics of ncRNAs, as follows: the chromosomal position, ncRNA length, amino acid sequence, the method of identification, and the physiological function across all species. The analysis of the ncPEPs' lengths in aa showed that the average length range of the included ncPEPs is 51 to 60 aa (Figure 3A). FuncPEP v2.0 also provides the molecular weight of the selected ncPEPs across all species,

with an average range of 2 to 5.99 kDa (Figure 3B). Next, we sought to analyze the types of ncRNA that encode for ncPEPs. As expected, the majority of the ncPEPs are encoded by lncRNAs (69.4%, 107), followed by miRNAs (9.71%, 15), and circRNAs (3.89%, 6) (Figure 3C). Previous reports have shown that the majority of circRNAs and pri-miRNAs are also derived from lncRNAs [22]. In addition, FuncPEP v2.0 provides the host ncRNA length of the selected ncPEPs, across all species, with a range of 1000 to 10,000 bp (Figure 3D), indicating that lncRNAs have regions that may become spliced before the initiation of translation. Further investigation of the altered splicing frequencies in ncRNAs correlating with their biological function and structural characteristics are warranted.

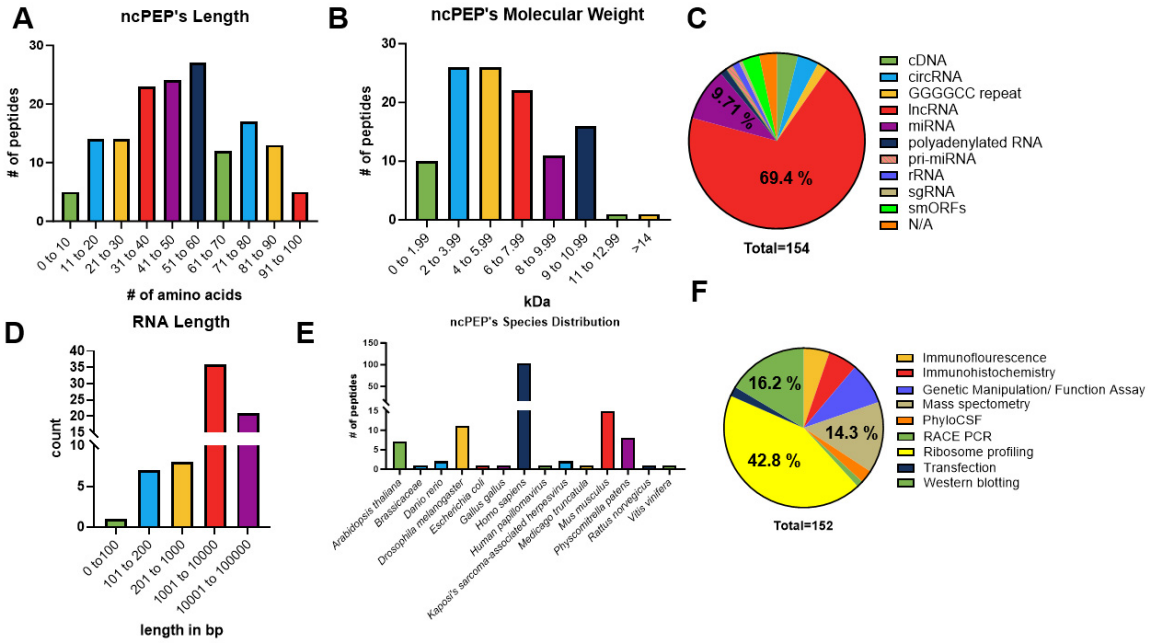


**Figure 3.** The characterization of ncPEPs included in the FuncPEP database. (**A**) ncPEPs' length displayed in amino acids. (**B**) ncPEPs' molecular weight represented as (kDa). (**C**) The classification of all ncPEPs' host ncRNAs across all species through a part-to-a-whole analysis. (**D**) Distribution of host ncRNAs' lengths of the included FuncPEPs in the database. (**E**) The distribution of species for the identification of ncPEPs. (**F**) Identification methods incorporated to detect ncPEPs as displayed by a part-to-a-whole analysis. %: ncPEPs. # means "Number".

We then analyzed in which species the selected ncPEPs were discovered, of which, the majority were found in *Homo sapiens* (66.2%, 102), followed by *Mus musculus* (9.74%, 15), and *Drosophila melanogaster* (7.14%, 11) (Figure 3E). In addition, bioinformatically predicted smORFs located in ncRNA regions were experimentally confirmed to be translated. Advances in metagenomics and transcriptomic platforms have been shown to improve the accuracy of predicting the translation of smORFs to ncPEPs [23]. Therefore, bioinformatics tools for smORF prediction were included in the FuncPEP database. The majority of the experimental methods were ribosome profiling (42.8%, 66), followed by Western blotting (16.2%, 25), and MS (14.3%, 22) (Figure 3F).

The ncPEPs' chromosome distribution demonstrated that chromosomes 17, 19, and 20 had a higher number compared to the other chromosomes carrying the candidate ncPEPs (Figure 4). Interestingly, an analysis of the chromosomal distribution of the human ncPEPs demonstrated that chromosome 17 displayed the largest presence of ncPEPs. Furthermore, an excess of genes on chromosome 19 has also been reported for protein-coding genes [24]. This suggests that the higher the gene density a chromosome exhibits, the more likely that it has an encoding region for an ncPEP. Further supporting this assumption, we observed that the ncPEPs are found in all chromosomes except chromosome 13.

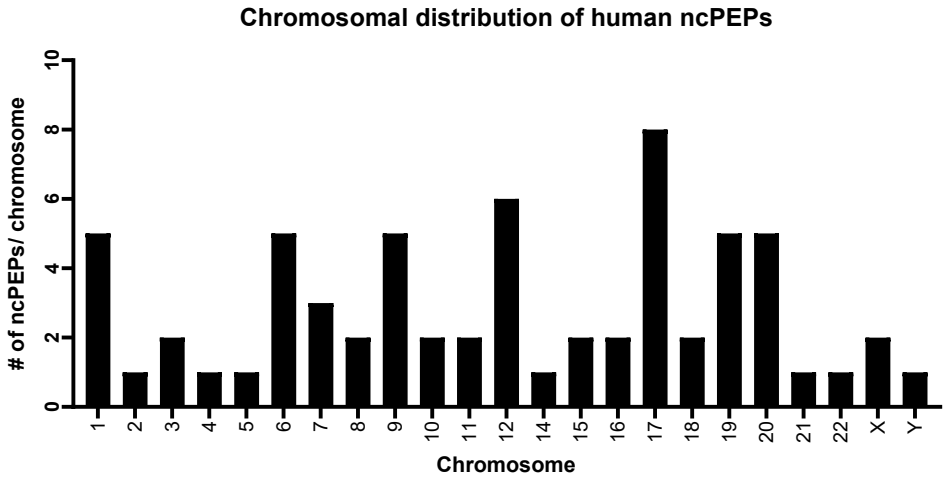


**Figure 4.** The classification of chromosomal dynamics of the ncPEPs included in the FuncPEP database version 2.0. The chromosomal distribution of human ncPEPs. # means "Number".

### *3.3. The Functions of ncPEPs*

The selected ncPEPs in the FuncPEP database are observed to play cell-specific roles in regulating multiple cellular processes, including metabolism and immunity. The function of ncPEPs in tumor-promoting cellular pathways has been well recognized [25], more specifically, in myeloid malignancies [26]. Furthermore, biomarker discovery platforms have associated the higher expression of these ncPEPs with poor prognosis and survival rates [27]. To decipher the role of the retrieved funcPEPs, we performed a functional summary and meta-analysis of the candidate ncPEPs included in the v2.0 database across all species and then *Homo sapiens* separately (Figure 5A,B). Interestingly, immunity was the most enriched function across all species, including *Homo sapiens* (Figure 5A,B). Several of the included ncPEPs in the FuncPEP database predominantly regulated host pro- or anti-tumor immunity. Other top functions across all species were development ($n$ = 27, 17.53%), tumor suppressor ($n$ = 13, 8.44%), tumor-promoting ($n$ = 12, 7.79%), metabolism ($n$ = 6, 3.89%), and muscle contractility ($n$ = 5, 3.24%) (Figure 5A). Interestingly, similar patterns of functionality were observed in the ncPEPs identified in *Homo sapiens*, the top five enriched functions being immunity ($n$ = 63, 61.16%), tumor-promotion ($n$ = 12, 11.65%), tumor suppression ($n$ = 12, 11.65%), neural regeneration ($n$ = 3, 2.91%), and metabolism ($n$ = 3, 2.91%) (Figure 5B). This functional analysis reveals that these classes of molecules have a diverse set of functions in a wide array of regulatory cellular processes, which warrant future study.

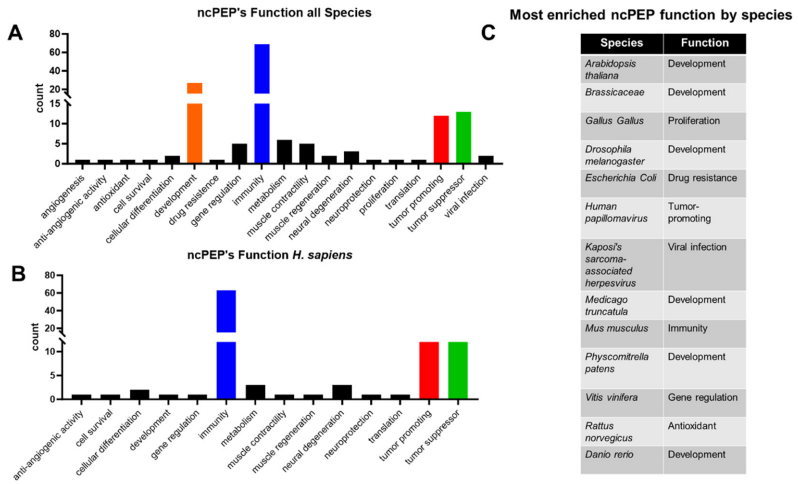


**Figure 5.** Comprehensive overview of the physiological functions of ncPEPs. (**A**) Distribution of the physiological functions of ncPEPs across all species included in the database. Top enriched

functionally relevant physiological pathways are indicated by colored bars. (**B**) Distribution of the physiological functions of human ncPEPs included in the database. Top enriched functionally relevant physiological pathways are indicated by colored bars. (**C**) Table demonstrated the most enriched ncPEP function stratified across each species.

Further, Figure 5C highlights the most enriched pathways individually stratified based on the species. A comprehensive overview of the wide array of functions driven by the ncPEPs in the database revealed that across all species, the majority of the ncPEPs are involved in immunity ($n$ = 69, 44.81%). Interestingly, the most common enriched function across each species separately was development, specifically for *Arabidopsis thaliana*, *Brassicacease*, *Drosophila melanogaster*, *Medicago truncatula*, *Physcomitrella patens*, and *Danio rerio*.

## 4. Discussion

The FuncPEP v2.0 database is designed to be a comprehensive resource using experimentally confirmed ncPEPs translated from ncRNAs and not only by bioinformatic tools. The database has been updated from the previous version based on functionally and experimentally validated ncPEPs. Ideally, the use of prediction tools for the initial screening step is robust, but it is important to experimentally characterize and validate their expression, localization, abundance, and other biochemical properties. Based on the criteria set in our previous version, we maintain our decision to not include ncPEPs discovered using ribosome profiling without independent experimental validation in version 2.0.

This observation could mean that (i) ncPEPs are not as abundant as originally predicted; or (ii) the current methodology for discovering, mapping, and validating ncPEPs is limited. The preliminary analysis of the candidate functional ncPEPs revealed that a wide range of species harbor a diverse and mechanistically versatile group of ncRNAs that are translated into functional small peptides. Although the functional relevance of these ncPEPs in biological settings has increasingly been recognized, the mechanistic underpinnings and cell-specific functions in different species are inconclusive due to the paucity of identification techniques. The clinical implications and discovery platforms for the identified ncPEPs are warranted for future investigation as a mode of patient stratification for disease severity and progression. More recently, multiple methods have been incorporated to comprehensively identify ncPEPs, primarily encompassing computational (i.e., intrinsic sequence features as a proxy of open reading frame (ORF) length, sequence homology to documented protein sequences, nucleotide composition, and substitution ratio used to putatively characterize the coding potential of ncRNA reading frames) [28–30] and experimental (i.e., ribosome profiling, mass spectrometry, and global translation initiation sequencing) [31,32] approaches for ncPEP classification. Understanding the biological roles of ncRNAs can provide a tool for examining the molecular mechanisms of understanding cell growth, proliferation, and development, as well as their responses to environmental stressors.

The downstream functional analysis of the peptides included in the FuncPEP v2.0 database revealed that the majority of the characterized ncPEPs play regulatory roles in immune function and development. For example, miPEP155 selectively binds to HSC70, rendering the modulation of MHC class II presentation in dendritic cells (DCs). Targeting miPEP155 in a murine model of autoinflammation exhibiting a dampening of DC-mediated autoimmunity [33] highlighted the importance of an ncPEP–immune axis, warranting future investigation. To gain translational insight, the authors also classified the potential of miPEP155 to interact with HSC70 in human DCs [33]. Fluorescent-tagged miPEP155 was observed to localize within the cytoplasmic and nuclear regions of human DC subsets. A large number of small peptides identified in our database regulate the function of antigen presentation [33], which could potentially provide an additional layer of refined immune regulation as well as potential peptide drug candidates for therapeutic intervention in a wide range of immunogenic diseases.

Studies investigating the functional roles of ncPEPs could help us understand their roles in cell-to-cell communication and delineate the complexity of ncRNAs with transla-

tional capabilities. The addition of several novel ncRNAs along with multiple regulatory roles has inspired researchers to investigate their multifaceted roles. However, with the discovery of novel functional peptides being translated from non-coding RNA, a new horizon has opened for researchers to investigate the function of non-coding regions at the genomic, transcriptomic, and proteomic levels.

We plan to continue to regularly screen for newly discovered ncPEPs with functional roles and include them in the FuncPEP database. In addition, we encourage researchers to inform us about new ncPEPs for immediate inclusion in our database. This field has the potential to expand the world of ncRNA biology and translate ncRNAs into clinical practice as biomarkers, with uses such as the identification of circulating ncPEP plasma markers. New avenues could be opened, enabling a deeper understanding of multi-level peptide–peptide, peptide–DNA, and peptide–ncRNA interactions, which could lead to the discovery of new therapies.

**Author Contributions:** Conceptualization, S.M. and G.A.C.; methodology, S.M. and P.R.; software, G.C.M. and B.M.B.; validation, S.M. and A.B.; formal Analysis, A.B.; investigation, G.A.C.; resources, G.A.C.; data curation, S.M. and A.B.; writing—original draft preparation, S.M. and A.B.; writing—review and editing, S.M., A.B. and M.P.D.; visualization, S.M. and A.B.; supervision, G.A.C.; project administration, G.A.C.; funding acquisition, G.A.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** Calin is Felix L. Haas's endowed professor in Basic Science. Work in G.A.C.'s laboratory is supported by the NCI grants 1R01 CA182905-01 and 1R01CA222007-01A1, NIGMS grant 1R01GM122775-01, DoD Idea Award W81XWH-21-1-0030, a Team DOD grant in Gastric Cancer W81XWH-21-1-0715, a Chronic Lymphocytic Leukemia Moonshot Flagship project, a CLL Global Research Foundation 2019 grant, a CLL Global Research Foundation 2020 grant, a CLL Global Research Foundation 2022 grant, The G. Harold & Leila Y. Mathers Foundation, two grants from Torrey Coast Foundation, an Institutional Research Grant and Development Grant associated with the Brain SPORE 2P50CA127001. S.M. was supported by the Cancer Prevention and Research Institute of Texas (CPRIT) Research Training Grant No. RP210028, the Schissler Foundation Fellowship, and Andrew Sowell-Wade Huggins Scholarship in Cancer Research, Steve Lasher and Janiece Longoria Graduate Student Research Award, and 2023-24 SIC Academic Achievement Award. A.B.'s work was supported by the American Heart Association (AHA) National Predoctoral Fellowship (Grant No. 23PRE1027421), the John J. Kopchick Research Award, and the Jesse B. Heath, Jr. Family Legacy Award. P.R received a summer fellowship from Partnership for Careers in Cancer Science and Medicine. M.P.D. was supported by the Berlin Institute of Health (Clinician Scientist Program) and DKTK Berlin (Young Investigator Grant 2022). B.M.B was supported by NIH/NCI grant CA016672 (the MD Anderson Cancer Center Support Grant for the Bioinformatics Shared Resource).

**Data Availability Statement:** FuncPEP v2.0 is freely accessible at https://bioinformatics.mdanderson.org/Supplements/FuncPEP/ (accessed on 27 March 2024).

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

## References

1. 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. [CrossRef] [PubMed]
2. Ruiz-Orera, J.; Messeguer, X.; Subirana, J.A.; Alba, M.M. Long non-coding RNAs as a source of new peptides. *eLife* **2014**, *3*, e03523. [CrossRef] [PubMed]
3. Patraquim, P.; Magny, E.G.; Pueyo, J.I.; Platero, A.I.; Couso, J.P. Translation and natural selection of micropeptides from long non-canonical RNAs. *Nat. Commun.* **2022**, *13*, 6515. [CrossRef] [PubMed]
4. Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M.C.; Maeda, N.; Oyama, R.; Ravasi, T.; Lenhard, B.; Wells, C.; et al. The transcriptional landscape of the mammalian genome. *Science* **2005**, *309*, 1559–1563. [CrossRef] [PubMed]
5. 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. [CrossRef] [PubMed]
6. Makarewich, C.A.; Olson, E.N. Mining for Micropeptides. *Trends Cell Biol.* **2017**, *27*, 685–696. [CrossRef] [PubMed]

7. Lin, M.F.; Jungreis, I.; Kellis, M. PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **2011**, *27*, i275–i282. [CrossRef] [PubMed]
8. Slavoff, S.A.; Mitchell, A.J.; Schwaid, A.G.; Cabili, M.N.; Ma, J.; Levin, J.Z.; Karger, A.D.; Budnik, B.A.; Rinn, J.L.; Saghatelian, A. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.* **2013**, *9*, 59–64. [CrossRef] [PubMed]
9. 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. [CrossRef]
10. Dragomir, M.P.; Manyam, G.C.; Ott, L.F.; Berland, L.; Knutsen, E.; Ivan, C.; Lipovich, L.; Broom, B.M.; Calin, G.A. FuncPEP: A Database of Functional Peptides Encoded by Non-Coding RNAs. *Noncoding RNA* **2020**, *6*, 41. [CrossRef]
11. Banerjee, A.; Chokkalla, A.K.; Shi, J.J.; Lee, J.; Venna, V.R.; Vemuganti, R.; McCullough, L.D. Microarray Profiling Reveals Distinct Circulating miRNAs in Aged Male and Female Mice Subjected to Post-stroke Social Isolation. *Neuromolecular Med.* **2021**, *23*, 305–314. [CrossRef]
12. Sloane, R.A.S.; White, M.G.; Witt, R.G.; Banerjee, A.; Davies, M.A.; Han, G.; Burton, E.; Ajami, N.; Simon, J.M.; Bernatchez, C.; et al. Identification of MicroRNA-mRNA Networks in Melanoma and Their Association with PD-1 Checkpoint Blockade Outcomes. *Cancers* **2021**, *13*, 5301. [CrossRef] [PubMed]
13. Mohapatra, S.; Pioppini, C.; Ozpolat, B.; Calin, G.A. Non-coding RNAs regulation of macrophage polarization in cancer. *Mol. Cancer* **2021**, *20*, 24. [CrossRef] [PubMed]
14. Nemeth, K.; Bayraktar, R.; Ferracin, M.; Calin, G.A. Non-coding RNAs in disease: From mechanisms to therapeutics. *Nat. Rev. Genet.* **2024**, *25*, 211–232. [CrossRef] [PubMed]
15. Geisler, S.; Coller, J. RNA in unexpected places: Long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 699–712. [CrossRef] [PubMed]
16. Wagner, V.; Kern, F.; Hahn, O.; Schaum, N.; Ludwig, N.; Fehlmann, T.; Engel, A.; Henn, D.; Rishik, S.; Isakova, A.; et al. Characterizing expression changes in noncoding RNAs during aging and heterochronic parabiosis across mouse tissues. *Nat. Biotechnol.* **2024**, *42*, 109–118. [CrossRef] [PubMed]
17. 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] [PubMed]
18. Li, L.J.; Leng, R.X.; Fan, Y.G.; Pan, H.F.; Ye, D.Q. Translation of noncoding RNAs: Focus on lncRNAs, pri-miRNAs, and circRNAs. *Exp. Cell Res.* **2017**, *361*, 1–8. [CrossRef] [PubMed]
19. Ramakrishnan, A.; Janga, S.C. Human protein-RNA interaction network is highly stable across mammals. *BMC Genom.* **2019**, *20*, 1004. [CrossRef]
20. Lopez-Bigas, N.; De, S.; Teichmann, S.A. Functional protein divergence in the evolution of Homo sapiens. *Genome Biol.* **2008**, *9*, R33. [CrossRef]
21. Barczak, W.; Carr, S.M.; Liu, G.; Munro, S.; Nicastri, A.; Lee, L.N.; Hutchings, C.; Ternette, N.; Klenerman, P.; Kanapin, A.; et al. Long non-coding RNA-derived peptides are immunogenic and drive a potent anti-tumour response. *Nat. Commun.* **2023**, *14*, 1078. [CrossRef] [PubMed]
22. Jeck, W.R.; Sorrentino, J.A.; Wang, K.; Slevin, M.K.; Burd, C.E.; Liu, J.; Marzluff, W.F.; Sharpless, N.E. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **2013**, *19*, 141–157. [CrossRef] [PubMed]
23. Kim, C.; Pongpanich, M.; Porntaveetus, T. Unraveling metagenomics through long-read sequencing: A comprehensive review. *J. Transl. Med.* **2024**, *22*, 111. [CrossRef] [PubMed]
24. Grimwood, J.; Gordon, L.A.; Olsen, A.; Terry, A.; Schmutz, J.; Lamerdin, J.; Hellsten, U.; Goodstein, D.; Couronne, O.; Tran-Gyamfi, M.; et al. The DNA sequence and biology of human chromosome 19. *Nature* **2004**, *428*, 529–535. [CrossRef] [PubMed]
25. Polycarpou-Schwarz, M.; Gross, M.; Mestdagh, P.; Schott, J.; Grund, S.E.; Hildenbrand, C.; Rom, J.; Aulmann, S.; Sinn, H.P.; Vandesompele, J.; et al. The cancer-associated microprotein CASIMO1 controls cell proliferation and interacts with squalene epoxidase modulating lipid droplet formation. *Oncogene* **2018**, *37*, 4750–4768. [CrossRef] [PubMed]
26. Papaioannou, D.; Petri, A.; Dovey, O.M.; Terreri, S.; Wang, E.; Collins, F.A.; Woodward, L.A.; Walker, A.E.; Nicolet, D.; Pepe, F.; et al. The long non-coding RNA HOXB-AS3 regulates ribosomal RNA transcription in NPM1-mutated acute myeloid leukemia. *Nat. Commun.* **2019**, *10*, 5351. [CrossRef] [PubMed]
27. Li, Z.; Jin, J.; He, W.; Long, W.; Yu, H.; Gao, X.; Nakai, K.; Zou, Q.; Wei, L. CoraL: Interpretable contrastive meta-learning for the prediction of cancer-associated ncRNA-encoded small peptides. *Brief. Bioinform.* **2023**, *24*, bbad352. [CrossRef]
28. Sun, L.; Luo, H.; Bu, D.; Zhao, G.; Yu, K.; Zhang, C.; Liu, Y.; Chen, R.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **2013**, *41*, e166. [CrossRef]
29. Wang, Y.; Li, Y.; Wang, Q.; Lv, Y.; Wang, S.; Chen, X.; Yu, X.; Jiang, W.; Li, X. Computational identification of human long intergenic non-coding RNAs using a GA-SVM algorithm. *Gene* **2014**, *533*, 94–99. [CrossRef]
30. Fan, X.N.; Zhang, S.W. lncRNA-MFDL: Identification of human long non-coding RNAs by fusing multiple features and using deep learning. *Mol. Biosyst.* **2015**, *11*, 892–897. [CrossRef]
31. Weaver, J.; Mohammad, F.; Buskirk, A.R.; Storz, G. Identifying Small Proteins by Ribosome Profiling with Stalled Initiation Complexes. *mBio* **2019**, *10*. [CrossRef] [PubMed]

32. Su, D.; Ding, C.; Qiu, J.; Yang, G.; Wang, R.; Liu, Y.; Tao, J.; Luo, W.; Weng, G.; Zhang, T. Ribosome profiling: A powerful tool in oncological research. *Biomark. Res.* **2024**, *12*, 11. [CrossRef] [PubMed]
33. 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] [PubMed]