

# UC Santa Cruz

## UC Santa Cruz Previously Published Works

### Title

Short open reading frame genes in innate immunity: from discovery to characterization.

### Permalink

<https://escholarship.org/uc/item/9zh9g4hk>

### Journal

Trends in Immunology, 43(9)

### Authors

Malekos, Eric  
Carpenter, Susan

### Publication Date

2022-09-01

### DOI

10.1016/j.it.2022.07.005

Peer reviewed



Published in final edited form as:

*Trends Immunol.* 2022 September ; 43(9): 741–756. doi:10.1016/j.it.2022.07.005.

## Short open reading frame genes in innate immunity: from discovery to characterization

Eric Malekos<sup>1,2</sup>, Susan Carpenter<sup>2,3,\*</sup>

<sup>1</sup>Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA, USA

<sup>2</sup>Genomics Institute, University of California Santa Cruz, Santa Cruz, CA, USA

<sup>3</sup>Department of Molecular Cell and Developmental Biology, University of California Santa Cruz, Santa Cruz, CA, USA

### Abstract

Next-generation sequencing (NGS) technologies have greatly expanded the size of the known transcriptome. Many newly discovered transcripts are classified as long noncoding RNAs (lncRNAs) which are assumed to affect phenotype through sequence and structure and not via translated protein products despite the vast majority of them harboring short open reading frames (sORFs). Recent advances have demonstrated that the noncoding designation is incorrect in many cases and that sORF-encoded peptides (SEPs) translated from these transcripts are important contributors to diverse biological processes. Interest in SEPs is at an early stage and there is evidence for the existence of thousands of SEPs that are yet unstudied. We hope to pique interest in investigating this unexplored proteome by providing a discussion of SEP characterization generally and describing specific discoveries in innate immunity.

### sORFs in innate immunity

#### Gene annotation and sORFs

Beginning with the sequencing of the yeast *Saccharomyces cerevisiae* genome in the 1990s, the scientific community has shown considerable interest in comprehensive identification and annotation of protein-coding genes in eukaryotes. Early efforts focused on ATG-initiated ORFs capable of encoding a polypeptide of at least 100 amino acids [1,2]. ORFs that did not meet this length cut-off, **short open reading frames (sORFs)** (see Glossary), required additional evidence to merit a protein-coding designation. In subsequent assemblies of eukaryotic genomes, similar heuristics – often complemented by homology searches against earlier annotations – have propagated this length bias [1]. These, and other conservative assumptions, have reduced spurious annotations, but imply the existence of a ‘**dark proteome**’ that is currently unannotated and understudied [3]. The ORFs and sORFs that give rise to this proteome have been found in many contexts, including 5’ **untranslated**

\*Correspondence: [sucarpen@ucsc.edu](mailto:sucarpen@ucsc.edu) (S. Carpenter).

Declaration of interests

S.C. is a paid consultant to NextRNA Therapeutics. No interests are declared by E.M.

**regions (UTRs)** and 3' UTRs, as extensions of canonical coding sequences or nested within them. Moreover, translation from sORFs has been documented in 'noncoding' transcripts, including long noncoding RNAs (lncRNAs), circular RNAs, and ribosomal RNAs [3-5].

## **lncRNAs**

lncRNAs are an increasingly appreciated class of genes that are defined by a length of at least 200 nucleotides (nts) and their lack of coding potential. Their abundance in the genome is similar to that of protein-coding genes, and many have been shown to contribute to various phenotypes through interactions with proteins, RNA, DNA, or combinations thereof [6,7]. Although only a few per cent of catalogued lncRNAs have been characterized, those that have are implicated in key biological processes including immunity [8,9], cancer [10], viability, and differentiation [6,7]. lncRNAs are often processed similarly to mRNAs – spliced, capped, and adenylated – and those that are exported to the cytoplasm are frequently found bound to ribosomes. Although association with ribosomes does not necessarily imply active translation [11], many genes classified as lncRNAs harbor sORFs that are translated to functional SEPs (also called microproteins) [12]. Indeed, an analysis of human, mouse, and fruit fly indicates that 98% of lncRNAs, across all three species, contain an AUG-initiated ORF, with a median of six such ORFs per lncRNA [4]. Furthermore, improved computational, proteomic, and sequencing-based approaches have resulted in the discovery of coding potential in some lncRNAs and hinted at that possibility in many more [12].

In this review, we focus on translated sORFs originating from lncRNAs in the innate immune systems of mice and humans. Reviews of sORFs in more general contexts are available [3,4,13,14]. We discuss the merits and limitations of both high-throughput and focused approaches to sORF discovery and characterization. We catalog the small number of newly discovered SEPs in innate immunity and suggest the most likely areas of near-future discovery. For each catalogued SEP, we summarize the steps taken to establish sORF translation and SEP function. Furthermore, we highlight the dual nature of these genes which often exhibit both coding and noncoding functionality. Finally, we discuss the particular merits of studying innate immune SEPs as a pursuit of scientific discovery and for the possibility of uncovering novel therapeutics.

## **Discovery of translated sORFs**

### **Sequence analysis**

Discovering new SEPs begins with sequence analysis. The standard scanning model of translation initiation involves a preinitiation complex binding the RNA 5'-cap and traversing the transcript until reaching a Kozak sequence centered at the start codon AUG. There, the remaining translational machinery engages, and elongation of the polypeptide occurs. Once a stop codon is encountered, translation is terminated, and the ribosome dissociates from the transcript [15]. This model implies a simple sequence analysis approach to identifying putative coding sORFs: to identify the 5'-most, AUG-initiated ORF of a cytosol-localized transcript. Indeed, this approach has been adequate for discovering some translated sORFs [16,17].

However, despite the prominence of the scanning model, it is not uncommon for ribosomes to exhibit alternative modes of translation, including leaky scanning, shunting, readthrough, and initiation at internal ribosome entry sites (IRESs) [18]. Additionally, it has long been appreciated that the majority of eukaryotic protein-coding genes lack an optimal Kozak context, and that translation can initiate at near-cognate start codons (CUG/GUG/ACG/etc.) [19-22]. Add to this the hundreds of thousands of uncharacterized ORFs in the human transcriptome and it becomes clear that more sophisticated approaches are required to identify functionally translated sORFs. For example, in a statistical analysis in which the human NCBI Reference Sequence (RefSeq) transcriptome sequences were shuffled 100 times, it was concluded that close to 90% of AUG-initiated ORFs in the human transcriptome would be expected by random chance. Nonetheless, this leaves tens of thousands of sORFs in excess of what would be expected randomly [23]. Therefore, it is probable that many uncharacterized sORFs are functional, but determining which cannot be fully accomplished by single-species sequence analysis (Figure 1, Key figure).

### Conservation and coding potential

Because functional coding sequences are expected to exhibit significant cross-species codon conservation, it is common to use sequence similarity to support further investigation of sORF translation [24]. Popular conservation scoring tools rely on multispecies alignments which, in turn, rely on high-quality genome assemblies [24-27]. Incomplete assemblies and heuristics employed by genome-wide alignment tools [28] can obfuscate homologous sORFs. Furthermore, it is characteristic of many sORFs to exhibit lower conservation scores than known protein-coding sequences and to be limited to smaller clades [29]. The functional sORF within HOXB-AS3, for instance, was found to be conserved in primates but not in other species (see later) [30]. Examples such as this indicate that conservation scores calculated from large sets of diverse species – including, for example, the 100-vertebrate PhyloP base-wise conservation track and the 58-mammal PhyloCSF hub on the University of California Santa Cruz (UCSC) Genome Browser – are at risk of ‘washing out’ true homologs that are restricted to specific clades [24,26,31,32]. Thus, relying on sequence conservation runs the risk of rejecting truly functional sORFs.

A complementary approach to nucleotide sequence analysis is an assessment of the protein domains of the putative SEP. Of the SEPs that are characterized, many are secreted as signaling molecules, found in membranes, or are known to contain intrinsically disordered domains [13]. The European Molecular Biology Laboratory (EMBL)-hosted tool, InterProScan, integrates many protein domain databases [33]. For further lists of bioinformatics tools that can aid in predicting coding potential, see [13,14].

### Ribosome sequencing

As already mentioned, translation requires physical association between ribosomes and transcripts. Researchers have used this association, along with NGS, to gain a better understanding of the translome. There are multiple approaches for isolating polysome-bound transcripts for sequencing; however, they are broadly similar and involve pausing translation, isolating ribosomes, and sequencing the ribosome-bound transcripts [34-36]. For the purposes of discovering translated sORFs, ribosome profiling (or ribosome footprinting

or Ribo-Seq) is notable for combining NGS with high-resolution mapping, allowing well-informed predictions of translational potential and efficiency. In contrast to other methods which extract full RNA transcripts from polysome complexes, Ribo-Seq includes a nuclease treatment step that degrades unprotected RNA, leaving only ribosome-protected fragments (~30 nucleotides in length) which are blocked from nuclease activity by virtue of being within the ribosome translation pocket [35,37,38]. Sequencing and mapping provide information on the locations of ribosomal occupancy on the transcript and can be used to predict translated sORFs. In standard Ribo-Seq, the elongation inhibitor cycloheximide can be used. In experiments aimed at identifying translation initiation sites (TI-Seq), lactimidomycin or harringtonine are used in place of elongation inhibitors and may be followed by puromycin to deplete elongating ribosomes and enrich initiating monosomes [39,40]. See Table 1 for a compilation of Ribo-Seq studies in cells and cell lines of the innate immune system.

Following sequencing, mapped fragments are passed to ORF prediction programs (see <http://rnabioinform.tch.harvard.edu/RiboToolkit/links.php> for a comprehensive list of software [41]). Alternatively, the investigator may skip this step altogether by making use of the many Ribo-Seq ORF prediction databases already available (Table 2).

### Candidate validation, approaches, and drawbacks

Translated sORF predictions based on ribosomal association likely misestimate the coding potential of transcripts and say nothing about the function of predicted SEPs (including whether they are functional at all). To confirm novel peptide production, candidate sORFs are typically validated via peptide tagging and microscopy or immunoprecipitation (IP). SEPs frequently influence phenotype by complexing with larger protein partners [16,17,70-72]; therefore, determining these partners with coimmunoprecipitation mass spectrometry (MS) [73] can inform SEP function. A UV-induced crosslinking tag has been reported to be particularly well suited to identifying SEP binding partners [74].

There are important considerations when conducting peptide-tag experiments. First, there is evidence that N or C terminus tagging can bias protein travel to different compartments, confusing claims about localization [75]. Additionally, it can affect the ability to detect an SEP, perhaps due to the destabilization of an important domain [76]. Alternatively, it is possible that the tag will stabilize an otherwise unstable and uninteresting SEP, resulting in a false discovery. Finally, the use of bulkier fluorescent tags such as green fluorescent protein (GFP) (~240 amino acids) has been observed to impair the colocalization of SEPs with binding partners [16,77]. An alternative to full fluorescent protein tagging is the use of a **split fluorescent system** which may reduce steric hindrance and allow the SEP to act in a wild-type manner while still providing a fluorescent readout [78]. For a thorough discussion of peptide tags and a list of available tags see [79].

Each of these techniques can be implemented in ectopic expression systems in the cell type of interest; however, conclusions drawn from these artificial expression systems should be used primarily for further hypothesis generation. Ultimately, validation should include the detection of endogenous SEP production either by splicing a molecular tag into the genome

or with a validated antibody (Ab) against the SEP. Here, Ab validation is preferred because it leaves the SEP and cell line in their wild-type states, but it is likely to be more costly, and may be unfeasible depending on the characteristics of the SEP in question.

## High-throughput validation

Individual, high-resolution characterization of SEPs will be an important part of correctly annotating the genome and characterizing SEP–protein interaction networks. However, the large numbers of putatively coding sORFs detected from Ribo-Seq and sequence analysis argues for the application of high-throughput methods to validate translation *en masse*. Broadly, there are two approaches: peptidomics via MS (Box 1) and genome editing with CRISPR-Cas (Box 2). In both cases the challenge is for the investigator to formulate a comprehensive **ORFeome** that encompasses translated sORFs in a given experiment while minimizing the total size of that space to maintain statistical power. For example, a combined six-frame genome-wide and three-frame transcriptome-wide proteome may capture all potentially translated ORFs, but in the case of MS, the large percentage of null results and multi-mappings to ORFs that are not even transcribed will reduce the statistical significance of truly translated ORFs [80]. Ribo-Seq-guided ORF prediction is well suited to this problem. If such data are not available, *de novo* transcriptome assembly and three-frame translation has proved capable of discovering novel translated ORFs [81]. Polysome profiling or RiboTag pulldown could provide additional evidence of translation from ‘noncoding’ transcripts [82-84].

## Characterized SEPs in innate immunity

### An emerging class: bifunctional genes

Here we describe SEPs that were recently discovered and characterized in innate immune (and innate immune-derived) cell lines. We also note instances where the RNA itself is known to contribute to a phenotype distinctly from the SEP; this is the case in four of the five examples shown in Figure 2. Although the sample size is too small to make strong inferences, the high representation of these ‘coding-noncoding’ or ‘bifunctional’ [85,86] genes suggests that future studies of SEPs would do well to consider the impact of the transcript itself.

### MIR155HG

The lncRNA MIR155HG has been the subject of intense study for the contribution of its miRNA product (miR-155) in inflammation in both innate and adaptive immune responses in human and mouse [87-89]. This includes increased MHC II presentation in human primary monocytes and mouse bone-marrow-derived dendritic cells (BMDCs) [90,91]. Recently it was demonstrated that this gene also produces a 17-amino-acid SEP (miPEP155) in humans but not in mice [16]. This sORF was considered because it began with the 5′-most AUG codon. Endogenous peptide production was confirmed by CRISPR-Cas9-mediated GFP insertion inframe with the genomic sORF in human HEK293T cells [16]. Additionally, an Ab against miPEP155 was used to confirm peptide production in human monocyte-derived dendritic cells (MDCs). Immunoprecipitation with anti-miPEP155 Ab

showed that miPEP155 bound a protein partner, and liquid chromatography MS (LC-MS) was used to identify the chaperone protein HSC70.

HSC70 is required for antigen presentation in DCs, and further study in miPEP155-treated mouse BMDCs showed reduced MHCII expression compared with untreated control cells [16]. As mentioned, the sORF in question is not found in mice, but there is a high degree of homology between mouse and human HSC70 proteins, raising the hypothesis that miPEP155 could function across species. Subsequent experiments in mice demonstrated that intravenous injection of miPEP155 improved outcomes in mouse models of psoriasis and experimental autoimmune encephalomyelitis (EAE) (a mouse model for multiple sclerosis) as evidenced by reduced IL-17A production and skewed T-cell polarization as a result of altered antigen presentation from DCs [16]. In this case, the RNA and SEP functions appear to be antagonistic, enhancing or reducing MHCII presentation, respectively. Although it would take a targeted study to make a strong claim about what purpose this serves, we can speculate that the importance of miR-155 in modulating many inflammatory pathways necessitates its expression for reasons other than MHCII expression; then, miPEP155 might act as a targeted ‘reducing valve’ on this singular miR-155-controlled pathway, resulting in a well calibrated MHCII presentation phenotype.

## NMES1

*NMES1* (*Nmes1* in mouse) is a second example of a bifunctional gene with miRNA and SEP products. Both the miRNA miR-147b, and the sORF encoding an 83-amino acid SEP, C15ORF48, are highly conserved [92]. Furthermore, both the *NMES1* transcript and miR-147b were found to be expressed in primary human monocyte-derived macrophages (MDMs) and mouse bone-marrow-derived macrophages (BMDMs) following **lipopolysaccharide (LPS)** stimulation [92]. A second group observed the same result in THP-1-derived M1 macrophages (T-M1s) treated with interleukin-1 $\beta$  (IL-1 $\beta$ ) [93]. MiR-147b was known to target the *NDUFA4* transcript [92], and this was confirmed by reduced *NDUFA4* transcript amounts following transfection of miR-147b mimic in MDMs [92].

An 82-amino-acid peptide encoded by *NDUFA4* showed structural similarity to C15ORF48 [92,93], and both SEPs shared protein–protein interaction partners, in particular the subunits of the cytochrome *c* oxidase complex (C $c$ O) [94]. C15ORF48 was detected by Western blot in LPS-stimulated BMDMs and MDMs [92] and IL-1 $\beta$ -stimulated human aortic endothelial cells (HAECs) and A549 cells, but not in IL-1 $\beta$  stimulated T-M1s [93]. Aside from the obvious difference in stimulus (LPS vs. IL-1 $\beta$ ) and cell type, we cannot provide an adequate explanation for discrepant C15ORF48 detection between the two human monocyte-derived macrophage lines (MDMs and T-M1s).

Regardless, this observation – along with the replacement of *NDUFA4* with C15ORF48 – supports a model of competitive binding at C $c$ O in response to either LPS [92] or IL-1 $\beta$  [93]. Here, the SEP and RNA exhibit complementary functions: miR-147b targets *NDUFA4* mRNA and C15ORF48 promotes *NDUFA4* protein degradation by excluding it from its binding pocket in MDMs and HAECs when stimulated with LPS or IL-1 $\beta$ , respectively [92].

**1810058I24Rik**

The expression of the uncharacterized lncRNA *1810058I24Rik* was significantly decreased in mouse BMDMs stimulated with five separate **pathogen-associated molecular patterns (PAMPs)** *in vitro* [76]. The same was seen in human monocytes and MDCs stimulated with LPS [76]. Moreover, the transcript localized to the cytosol and harbored a highly conserved sORF that produced a 47-amino-acid SEP, Mm47. A cDNA expression vector containing the sORF and FLAG-tags at either the N terminus or C terminus was expressed, and translation occurred in two of the three constructs, as evidenced by an anti-FLAG immunoblot. Although N terminus tagging alone resulted in no detectable peptide – perhaps due to disruption of a predicted signaling domain – simultaneous tagging of the N and C termini resulted in detectable protein [76].

Protein product formation was demonstrated endogenously via anti-Mm47 Ab pulldown. SignalP [95] suggested mitochondrial localization for the protein, and this was confirmed via immuno-staining against FLAG-tagged Mm47 [76]. In immortalized BMDMs, genomic knockout via CRISPR-Cas9 and rescue with an ectopic Mm47 expression vector were combined with LPS and **nigericin** or ATP to link Mm47 to activation of the **Nlrp3 inflammasome** as measured by IL-1 $\beta$  release. This was further supported via *1810058I24Rik* targeting siRNA knockdown in primary BMDMs which also showed reduced IL-1 $\beta$  release in response to nigericin [76].

This case highlights the importance of epitope tag placement and used immunofluorescence imaging to show that both FLAG-tagged and endogenous Mm47 localize to the mitochondria. Ultimately the experimental results led to a slightly counterintuitive conclusion: Mm47 must be present for Nlrp3 inflammasome activation, but both *1810058I24Rik* transcript and Mm47 are downregulated by LPS stimulation. This might indicate that degradation of Mm47 from baseline levels following inflammatory stimulation can act as a molecular timer, allowing immediate Nlrp3 activation, but also attenuating that inflammatory pathway when Mm47 is lost, thereby functioning as a built-in safeguard against chronic Nlrp3 inflammasome activation [76].

**Aw112010**

The mouse-specific lncRNA *Aw112010* has been observed to be highly expressed in microglia, astrocytes, CD4<sup>+</sup> T cells, and macrophages under various inflammatory conditions [8,82,96]. In a study of CD4<sup>+</sup> T cells polarized into inflammatory Th1 cells, chromatin isolation by RNA purification (CHIRP) was used to show that *Aw112010* was enriched in proximity to the anti-inflammatory *Il10* gene, and that *Aw112010* depletion by siRNA led to increased *Il10* expression [8]. Additionally, *Aw112010* coprecipitated with KDM5A, a demethylase that removes the activity-promoting H3K4me3 marker; under wild-type conditions, H3K4me3 at the *Il10* locus was lost. By contrast, when *Aw112010* RNA was knocked down with siRNA, H3K4me3 was maintained. An 84-nucleotide deletion was introduced in the *Aw112010* gene (*Aw112010*<sup>430-514</sup>) downstream of the primary sORF (see next paragraph) in RAW264.7 macrophages and resulted in decreased proinflammatory *Il6* and increased *Il10* expression [8]. Taken together, these findings supported a regulatory



role for *Aw112010* RNA in KDM5A-directed demethylation of the *Il10* gene in Th1 cells and an uncharacterized regulatory role of *Il10* and *Il6* transcription in RAW264.7 cells [8].

In another study, which utilized RiboTag and Ribo-Seq approaches in BMDMs, this lncRNA coded an 84-amino-acid SEP from a CUG start codon [82]. In comparative experiments between WT *Aw112010* mice and a mutant strain with an early stop codon (*Aw112010<sup>stop</sup>*), decreased *Il6* and *Il12p40* release were observed in response to intraperitoneal LPS injection, as well as the increased proliferative capacity of orally administered *Salmonella Typhimurium* serovar in *Aw112010<sup>stop</sup>* mice [82]. However, the introduction of a premature stop codon triggered **nonsense-mediated decay** which reduced both RNA and SEP concentrations, and made it impossible to discern which was contributing to the change in phenotype. To overcome this, transcripts were constructed so that the sORF was composed of synonymous codon substitutions; this substantially altered the predicted RNA structure and sequence but encoded the same peptide. This altered transcript rescued IL-12p40 cytokine production as measured by ELISA in BMDMs and supported the claim that the SEP itself was responsible for regulating expression of *Il12p40* [82].

These experiments demonstrated RNA-driven reduction of anti-inflammatory IL-10 production in CD4<sup>+</sup> T cells [8], and SEP-driven increase in IL-12p40 in BMDMs [82]. The observation that – following LPS treatment and relative to wild-type, *Aw112010<sup>430–514</sup>* RAW264.7 cells exhibited decreased *Il6* transcription [8] and *Aw112010<sup>stop</sup>* BMDMs exhibited decreased IL-6 cytokine release [82] – could be explained by *Aw112010* RNA function (but not SEP function) in both cases. In total, the evidence suggests that *Aw112010* RNA and SEP work synergistically to promote a robust immune response in mouse macrophages, but the full contribution of either component is uncertain.

### HOXB-AS3

In a study of NPM1 mutations in the context of acute myeloid leukemia (AML), NPM1<sup>mut</sup>-AML cells and K562 cells were used to thoroughly characterize human *HOXB-AS3* function [10]. Knockdown with anti-*HOXB-AS3* gapmers implicated the *HOXB-AS3* transcript in proliferation [10]. RNA immunoprecipitation coupled with tandem MS revealed the *HOXB-AS3* protein-binding partners, with the protein EBP1 being the most prominent; 100-nucleotide deletions of the transcript showed that loss of nucleotides 95–195 decreased association with EBP1. RNA antisense purification (RAP-DNA) confirmed enrichment of the EBP1-*HOXB-AS3* at the rDNA promoter and indicated that a different 100-nucleotide section was responsible for that interaction. After recording insubstantial association of *HOXB-AS3* with polysomes, it was concluded that translation was not relevant for *HOXB-AS3* function in the NPM1<sup>mut</sup>-AML cell line [10].

A second study of an alternate *HOXB-AS3* transcript in colon cancer cells found a 53-amino-acid SEP, but not the lncRNA itself, to be an important mediator in colon cancer proliferation [30]. Ribo-Seq results indicated translational potential which was confirmed with GFP and FLAG tagging. Coimmunoprecipitation using an anti-SEP Ab followed by MS analysis was used to determine that the SEP interacted with a number of proteins involved in RNA splicing. A focused interrogation of a specific binding partner, hnRNPA1, was undertaken. After mutating hnRNPA1's functional domains, the *HOXB-AS3* SEP

interacted with the RNA-binding RGG box and influenced hnRNPA1 activity. Despite contributing to a pronounced phenotype, this sORF was found only in primates [30]. In the context of colon cancer, HOXB-AS3 SEP reduced proliferation in contrast to the proliferative phenotype induced by *HOXB-AS3* transcript in the AML cell lines. The marked contrast in both the mechanisms of action and the phenotypic results from these two *HOXB-AS3* transcripts highlights the importance of making cell-type- and context-specific claims regarding the action of these complex genes.

### In summary

Taken together these examples offer a window into the complex nature of bifunctional genes. Though few in number, the fact that four of the five genes have been shown to exert phenotypic effects through both RNA and SEP emphasizes the inadequacy of the current ‘protein-coding/noncoding’ dichotomy. Furthermore, it causes us to wonder whether the understandable emphasis on studying proteins has resulted in the scientific community failing to consider RNA function from mRNAs.

Alternatively, or additionally, it may be the case that RNAs with coding sORFs are particularly apt to exert both coding and noncoding functions. It has been previously hypothesized that sORF transcripts represent **proto-genes**, lowly translated coding sequences which, through evolutionary time and selection, stabilize into canonical protein-coding genes [36,97]. One criticism of this hypothesis is that maintaining a latent pool of translated proto-genes would cost cellular resources and provide a negligible-or-negative survival advantage. Bifunctional genes offer an answer to this criticism: if the sORF-bearing RNA is itself contributing to evolutionary fitness, then the drawbacks of aberrant sORF translation could be overcome by the RNA’s provided benefit. Furthermore, RNA function would be robust against indel mutations that modulate the sORF; thus, these transcripts might provide the material for *de novo* protein-coding gene creation. Although this is speculative, it suggests that many proto-genes may turn out to be functional RNAs, akin to a recapitulation of the RNA world, observable in the present day.

### Concluding remarks

Nuanced biomolecular interrogation has allowed researchers to differentiate between SEP and RNA activity, and there are many databases containing thousands of sORFs and lncRNAs that are yet to be investigated (Table 2). Immunologists might find the study of this expanded proteome particularly fruitful. It is well established that the transcriptome is drastically changed under conditions of inflammatory stimuli. Furthermore, it is reported that multiple components of translation initiation machinery become altered under conditions of cellular stress, including in innate immune cells treated with LPS and other PAMPs [98,99]. These changes shift the cells from the standard scanning model of translation towards an atypical mode that can enhance translation from IRES, initiate with Leu-tRNAs, and involve other aberrant initiation mechanisms [100-102]. These changes in the transcriptome and translatoome imply the existence of a unique and context-specific proteome for immunologists to investigate and understand.

A particular emerging area of interest is in the field of immunometabolism. Described SEPs show a positive charge bias and enrichment for transmembrane  $\alpha$ -helices [22]. Amphipathic positively charged peptides can cross the outer mitochondrial membrane, and many characterized SEPs are components of the mitochondrial proteome [23,74,76,93,103-105]. Metabolic reprogramming is important in innate immune responses, and SEPs are likely to be further implicated in this process [106,107].

Another emerging area of interest is immuno-oncology. Peptides originating from noncoding regions are greatly enriched in MHCI presentation complexes in many cancers [108-110]. As the MHC peptidomes of cancers are further resolved, therapeutics with high specificity against cancer types might be discovered. There are already promising studies making use of DCs and peptide vaccines in this regard (see Clinician's corner).

Although previously overlooked, both SEPs and lncRNAs are increasingly recognized as important contributors to various phenotypes. NGS and improved computational tools have improved our ability to identify sORFs with translation potential. High-throughput functional characterization methods are required to rapidly determine which sORFs are biologically relevant. Advancements in peptidomic technologies, including the possibility of direct protein sequencing, may eventually allow high-throughput characterization of the peptidome with high sensitivity (Box 1). For now, CRISPR-Cas-based approaches appear to be the most effective method for mass characterization. Given the proliferation of predicted SEPs in databases and the growing appreciation of SEP functionality, we expect a large number of important SEP discoveries in coming years. Furthermore, although many questions remain (see Outstanding questions), and as argued earlier, we think that immunologists may be particularly well positioned to characterize functional SEPs and utilize this information in the development of novel candidate therapeutics and vaccines.

## Acknowledgments

S.C. is supported by R01AI148413 from National Institute of Allergy and Infectious Diseases and R35GM137801 from the National Institute of General Medical Sciences. E.M. is supported by T32HG012344 and in part by R35GM137801.

## Glossary

### **Dark proteome**

understudied and under-characterized proteins and peptides, including those that arise from UTRs and noncoding RNAs.

### **FASTQs**

the standard short-read sequencing format for bioinformatic sequencing analysis.

### **Homology-directed repair (HDR)**

repair of DNA breaks using a homologous template, allowing the insertion of genetic material.

### **Lipopolysaccharide (LPS)**

a PAMP component of Gram-negative bacterial cell walls. In its purified form, it is commonly used as an inflammation-inducing ligand in studies of innate immunity.

**Nigericin**

a PAMP isolated from the Gram-positive bacterium *Streptomyces hygroscopicus*, it is capable of activating the Nlrp3 inflammasome.

**Nlrp3 inflammasome**

a large multiprotein complex required for the production of IL-1 $\beta$  and IL-18 and conserved in human and mouse. Nigericin and ATP are PAMPs known to activate this inflammasome.

**Nonhomologous end-joining (NHEJ)**

cellular repair mechanism for double-stranded DNA breaks such as those caused by CRISPR-Cas9. It typically results in small nucleotide insertions or deletions at the repair site.

**Nonsense-mediated decay**

a eukaryotic RNA surveillance pathway that prevents the production of aberrant proteins containing premature termination codons (PTCs).

**ORFeome**

set of all ORFs, including known coding sequences, sORFs, and ORFs that are not translated.

**Pathogen-associated molecular pattern (PAMP)**

conserved features specific to microbes, essential for their survival; a PAMP is recognized as foreign and initiates an innate immune response during an infection.

**Perturb-Seq**

a high-throughput technique that combines pooled screening (perturbations) with single cell sequencing, allowing for accurate identification of gene targets or phenotypes following single perturbations.

**Proto-gene**

a genomic element which undergoes transient translation and may act as the 'raw material' for *de novo* gene generation.

**Short open reading frame (sORF)**

a nucleotide sequence of 300 codons or less and giving rise to a sORF-encoded peptide (SEP). Also called small open reading frame (smORF).

**Split fluorescent system**

experimental approach in which a fluorescent tag is split into two nonfluorescent protein subunits. When the subunits are coexpressed and joined together, the unit fluoresces. Typically, one subunit (the tag) is much smaller than the other.

**Untranslated regions (UTRs)**

regulatory regions at the 5' end and 3' end of protein-coding genes that are typically not thought of as having coding potential.

## References

1. Harrison PM et al. (2002) A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.* 30, 1083–1090 [PubMed: 11861898]
2. Goffeau A et al. (1996) Life with 6000 genes. *Science* 274, 563–567
3. Wright BW et al. (2022) The dark proteome: translation from noncanonical open reading frames. *Trends Cell Biol.* 32, 243–258 [PubMed: 34844857]
4. Couso J-P and Patraquim P (2017) Classification and function of small open reading frames. *Nat. Rev. Mol. Cell Biol* 18, 575–589 [PubMed: 28698598]
5. Li X et al. (2018) The biogenesis, functions, and challenges of circular RNAs. *Mol. Cell* 71, 428–442 [PubMed: 30057200]
6. Rinn JL and Chang HY (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem* 81, 145–166 [PubMed: 22663078]
7. Robinson EK et al. (2020) The how and why of lncRNA function: an innate immune perspective. *Biochim. Biophys. Acta Gene Regul. Mech* 1863, 194419 [PubMed: 31487549]
8. Yang X et al. (2020) Long NONCODING RNA AW112010 promotes the differentiation of inflammatory T cells by suppressing IL-10 expression through histone demethylation. *J. Immunol* 205, 987–993 [PubMed: 32690657]
9. Robinson EK et al. (2022) lincRNA-Cox2 functions to regulate inflammation in alveolar macrophages during acute lung injury. *J. Immunol* 208, 1886–1900 [PubMed: 35365562]
10. Papaioannou D et al. (2019) The long non-coding RNA HOXB-AS3 regulates ribosomal RNA transcription in NPM1-mutated acute myeloid leukemia. *Nat. Commun* 10, 5351 [PubMed: 31767858]
11. Guttman M et al. (2013) Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240–251 [PubMed: 23810193]
12. Ji Z et al. (2015) Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* 4, e08890 [PubMed: 26687005]
13. Schlesinger D and Elsässer SJ (2022) Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins. *FEBS J.* 289, 53–74 [PubMed: 33595896]
14. Leong AZ-X et al. (2022) Short open reading frames (sORFs) and microproteins: an update on their identification and validation measures. *J. Biomed. Sci* 29, 19 [PubMed: 35300685]
15. Hinnebusch AG (2014) The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem* 83, 779–812 [PubMed: 24499181]
16. Niu L et al. (2020) A micropeptide encoded by lncRNA MIR155HG suppresses autoimmune inflammation via modulating antigen presentation. *Sci. Adv* 6, eaaz2059 [PubMed: 32671205]
17. Matsumoto A et al. (2017) mTORC1 and muscle regeneration are regulated by the LINC00961-encoded SPAR polypeptide. *Nature* 541, 228–232 [PubMed: 28024296]
18. Kwan T and Thompson SR (2019) Noncanonical translation initiation in eukaryotes. *Cold Spring Harb. Perspect Biol* 11, a032672 [PubMed: 29959190]
19. Hann SR et al. (1988) A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* 52, 185–195 [PubMed: 3277717]
20. Smith E et al. (2005) Leaky ribosomal scanning in mammalian genomes: significance of histone H4 alternative translation in vivo. *Nucleic Acids Res.* 33, 1298–1308 [PubMed: 15741183]
21. Peabody DS (1989) Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem* 264, 5031–5035 [PubMed: 2538469]
22. Acevedo JM et al. (2018) Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence. *Sci. Rep* 8, 4018 [PubMed: 29507361]

23. Samandi S et al. (2017) Deep transcriptome annotation enables the discovery and functional characterization of cryptic small proteins. *eLife* 6, e27860 [PubMed: 29083303]
24. Mudge JM et al. (2019) Discovery of high-confidence human protein-coding genes and exons by whole-genome PhyloCSF helps elucidate 118 GWAS loci. *Genome Res.* 29, 2073–2087 [PubMed: 31537640]
25. Armstrong J et al. (2019) Whole-genome alignment and comparative annotation. *Annu. Rev. Anim. Biosci* 7, 41–64 [PubMed: 30379572]
26. Lin MF et al. (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* 27, i275–i282 [PubMed: 21685081]
27. Siepel A et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050 [PubMed: 16024819]
28. Prakash A and Tompa M (2007) Measuring the accuracy of genome-size multiple alignments. *Genome Biol.* 8, R124 [PubMed: 17594489]
29. Fields AP et al. (2015) A regression-based analysis of ribosome-profiling data reveals a conserved complexity to mammalian translation. *Mol. Cell* 60, 816–827 [PubMed: 26638175]
30. Huang J-Z et al. (2017) A peptide encoded by a putative lncRNA HOXB-AS3 suppresses colon cancer growth. *Mol. Cell* 68, 171–184.e6 [PubMed: 28985503]
31. Pollard KS et al. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20, 110–121 [PubMed: 19858363]
32. Kent WJ et al. (2002) The human genome browser at UCSC. *Genome Res.* 12, 996–1006 [PubMed: 12045153]
33. Blum M et al. (2021) The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–D354 [PubMed: 33156333]
34. Yoshikawa H et al. (2018) Efficient analysis of mammalian polysomes in cells and tissues using Ribo Mega-SEC. *eLife* 7, e36530 [PubMed: 30095066]
35. Ingolia NT et al. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223 [PubMed: 19213877]
36. Brar GA and Weissman JS (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol* 16, 651–664 [PubMed: 26465719]
37. Ingolia NT et al. (2012) The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc* 7, 1534–1550 [PubMed: 22836135]
38. McGlincy NJ and Ingolia NT (2017) Transcriptome-wide measurement of translation by ribosome profiling. *Methods* 126, 112–129 [PubMed: 28579404]
39. Lee S et al. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A* 109, E2424–E2432 [PubMed: 22927429]
40. Gao X et al. (2015) Quantitative profiling of initiating ribosomes in vivo. *Nat. Methods* 12, 147–153 [PubMed: 25486063]
41. Liu Q et al. (2020) RiboToolkit: an integrated platform for analysis and annotation of ribosome profiling data to decode mRNA translation at codon resolution. *Nucleic Acids Res.* 48, W218–W229 [PubMed: 32427338]
42. Jovanovic M et al. (2015) Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. *Science* 347, 1259038 [PubMed: 25745177]
43. Wang H et al. (2019) Mettl3-mediated mRNA m(6)A methylation promotes dendritic cell activation. *Nat. Commun* 10, 1898 [PubMed: 31015515]
44. Zhang X et al. (2017) Translation repression via modulation of the cytoplasmic poly(A)-binding protein in the inflammatory response. *eLife* 6, e27786 [PubMed: 28635594]
45. Barry KC et al. (2017) Global analysis of gene expression reveals mRNA superinduction is required for the inducible immune response to a bacterial pathogen. *eLife* 6, e22707 [PubMed: 28383283]
46. Guo H et al. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840 [PubMed: 20703300]

47. Hornstein N et al. (2016) Ligation-free ribosome profiling of cell type-specific translation in the brain. *Genome Biol.* 17, 149 [PubMed: 27380875]
48. Scheckel C et al. (2020) Ribosomal profiling during prion disease uncovers progressive translational derangement in glia but not in neurons. *eLife* 9, e62911 [PubMed: 32960170]
49. Schott J et al. (2021) Nascent Ribo-Seq measures ribosomal loading time and reveals kinetic impact on ribosome density. *Nat Methods* 18, 1068–1074 [PubMed: 34480152]
50. Liu T-Y et al. (2017) Time-resolved proteomics extends ribosome profiling-based measurements of protein synthesis dynamics. *Cell Syst* 4, 636–644.e9 [PubMed: 28578850]
51. Wiita AP et al. (2013) Global cellular response to chemotherapy-induced apoptosis. *eLife* 2, e01236 [PubMed: 24171104]
52. Fritsch C et al. (2012) Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res.* 22, 2208–2218 [PubMed: 22879431]
53. Mills EW et al. (2016) Dynamic regulation of a ribosome rescue pathway in erythroid cells and platelets. *Cell Rep.* 17, 1–10 [PubMed: 27681415]
54. Martinez TF et al. (2020) Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol* 16, 458–468 [PubMed: 31819274]
55. Calviello L et al. (2020) Quantification of translation uncovers the functions of the alternative transcriptome. *Nat. Struct. Mol. Biol* 27, 717–725 [PubMed: 32601440]
56. Zinshteyn B et al. (2020) Nuclease-mediated depletion biases in ribosome footprint profiling libraries. *RNA* 26, 1481–1488 [PubMed: 32503920]
57. Su X et al. (2015) Interferon- $\gamma$  regulates cellular metabolism and mRNA translation to potentiate macrophage activation. *Nat. Immunol* 16, 838–849 [PubMed: 26147685]
58. Li Q et al. (2022) Low-input Rnase footprinting for simultaneous quantification of cytosolic and mitochondrial translation. *Genome Res.* 32, 545–557 [PubMed: 35193938]
59. Michel AM et al. (2014) GWIPS-viz: development of a ribo-seq genome browser. *Nucleic Acids Res.* 42, D859–D864 [PubMed: 24185699]
60. Kiniry SJ et al. (2019) Trips-Viz: a transcriptome browser for exploring Ribo-Seq data. *Nucleic Acids Res.* 47, D847–D852 [PubMed: 30239879]
61. Wang H et al. (2019) RPFdb v2.0: an updated database for genome-wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res.* 47, D230–D234 [PubMed: 30335166]
62. Olexiouk V et al. (2018) An update on [sORFs.org](https://sORFs.org): a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* 46, D497–D502 [PubMed: 29140531]
63. Choteau SA et al. (2021) MetamORF: a repository of unique short open reading frames identified by both experimental and computational approaches for gene and metagenome analyses. *Database J. Biol. Databases Curation* 2021, baab032
64. Hao Y et al. (2018) SmProt: a database of small proteins encoded by annotated coding and non-coding RNA loci. *Brief. Bioinform* 19, 636–643 [PubMed: 28137767]
65. Brunet MA et al. (2021) OpenProt 2021: deeper functional annotation of the coding potential of eukaryotic genomes. *Nucleic Acids Res.* 49, D380–D388 [PubMed: 33179748]
66. Neville MDC et al. (2021) A platform for curated products from novel open reading frames prompts reinterpretation of disease variants. *Genome Res.* 31, 327–336 [PubMed: 33468550]
67. Liu T et al. (2022) LncPep: a resource of translational evidences for lncRNAs. *Front. Cell Dev. Biol* 10, 795084 [PubMed: 35141219]
68. Luo X et al. (2022) SPENCER: a comprehensive database for small peptides encoded by noncoding RNAs in cancer patients. *Nucleic Acids Res.* 50, D1373–D1381 [PubMed: 34570216]
69. Huang Y et al. (2021) cncRNAdb: a manually curated resource of experimentally supported RNAs with both protein-coding and noncoding function. *Nucleic Acids Res.* 49, D65–D70 [PubMed: 33010163]
70. D’Lima NG et al. (2017) A human microprotein that interacts with the mRNA decapping complex. *Nat. Chem. Biol* 13, 174–180 [PubMed: 27918561]
71. Cloutier P et al. (2020) Upstream ORF-encoded ASDURF is a novel prefoldin-like subunit of the PAQosome. *J. Proteome Res* 19, 18–27 [PubMed: 31738558]

72. Fisher ME et al. (2021) Dwarf open reading frame (DWORF) is a direct activator of the sarcoplasmic reticulum calcium pump SERCA. *eLife* 10, e65545 [PubMed: 34075877]
73. Low TY et al. (2021) Recent progress in mass spectrometry-based strategies for elucidating protein-protein interactions. *Cell. Mol. Life Sci. CMLS* 78, 5325–5339 [PubMed: 34046695]
74. Koh M et al. (2021) A short ORF-encoded transcriptional regulator. *Proc. Natl. Acad. Sci* 118, e2021943118 [PubMed: 33468658]
75. Weill U et al. (2019) Assessment of GFP tag position on protein localization and growth fitness in yeast. *J. Mol. Biol* 431, 636–641 [PubMed: 30550779]
76. Bhatta A et al. (2020) A mitochondrial micropeptide is required for activation of the Nlrp3 inflammasome. *J. Immunol* 204, 428–437 [PubMed: 31836654]
77. Shibata T et al. (2018) Addition of an EGFP-tag to the N-terminal of influenza virus M1 protein impairs its ability to accumulate in ND10. *J. Virol. Methods* 252, 75–79 [PubMed: 29174083]
78. Chen J et al. (2020) Pervasive functional translation of noncanonical human open reading frames. *Science* 367, 1140–1146 [PubMed: 32139545]
79. Vandemoortele G et al. (2019) Pick a tag and explore the functions of your pet protein. *Trends Biotechnol.* 37, 1078–1090 [PubMed: 31036349]
80. Blakeley P et al. (2012) Addressing statistical biases in nucleotide-derived protein databases for proteogenomic search strategies. *J. Proteome Res* 11, 5221–5234 [PubMed: 23025403]
81. Ma J et al. (2018) The influence of transcript assembly on the proteogenomics discovery of microproteins. *PLoS One* 13, e0194518 [PubMed: 29584760]
82. Jackson R et al. (2018) The translation of non-canonical open reading frames controls mucosal immunity. *Nature* 564, 434–438 [PubMed: 30542152]
83. Sanz E et al. (2019) RiboTag: ribosomal tagging strategy to analyze cell-type-specific mRNA expression in vivo. *Curr. Protoc. Neurosci* 88, e77 [PubMed: 31216392]
84. Chasse H et al. (2017) Analysis of translation using polysome profiling. *Nucleic Acids Res.* 45, e15 [PubMed: 28180329]
85. Nam J-W et al. (2016) Incredible RNA: dual functions of coding and noncoding. *Mol. Cells* 39, 367–374 [PubMed: 27137091]
86. Kumari P and Sampath K (2015) cncRNAs: bi-functional RNAs with protein coding and non-coding functions. *Semin. Cell Dev. Biol* 47-48, 40–51 [PubMed: 26498036]
87. Jablonski KA et al. (2016) Control of the inflammatory macrophage transcriptional signature by miR-155. *PLoS One* 11, e0159724 [PubMed: 27447824]
88. Rothchild AC et al. (2016) MiR-155-regulated molecular network orchestrates cell fate in the innate and adaptive immune response to *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U. S. A* 113, E6172–E6181 [PubMed: 27681624]
89. Li X et al. (2016) miR-155 acts as an anti-inflammatory factor in atherosclerosis-associated foam cell formation by repressing calcium-regulated heat stable protein 1. *Sci. Rep* 6, 21789 [PubMed: 26899994]
90. Hodge J et al. (2020) Overexpression of microRNA-155 enhances the efficacy of dendritic cell vaccine against breast cancer. *OncolImmunology* 9, 1724761 [PubMed: 32117588]
91. Olsson AM et al. (2022) miR-155-overexpressing monocytes resemble HLAhighISG15+ synovial tissue macrophages from patients with rheumatoid arthritis and induce polyfunctional CD4+ T-cell activation. *Clin. Exp. Immunol* 207, 188–198 [PubMed: 35020867]
92. Clayton SA et al. (2021) Inflammation causes remodeling of mitochondrial cytochrome c oxidase mediated by the bifunctional gene C15orf48. *Sci. Adv* 7, eabl5182 [PubMed: 34878835]
93. Lee CQE et al. (2021) Coding and non-coding roles of MOCCI (C15ORF48) coordinate to regulate host inflammation and immunity. *Nat. Commun* 12, 2130 [PubMed: 33837217]
94. Floyd BJ et al. (2016) Mitochondrial protein interaction mapping identifies regulators of respiratory chain function. *Mol. Cell* 63, 621–632 [PubMed: 27499296]
95. Almagro Armenteros JJ et al. (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol* 37, 420–423 [PubMed: 30778233]



96. Madeddu S et al. (2015) Identification of glial activation markers by comparison of transcriptome changes between astrocytes and microglia following innate immune stimulation. *PLoS One* 10, e0127336 [PubMed: 26214311]
97. Carvunis A-R et al. (2012) Proto-genes and de novo gene birth. *Nature* 487, 370–374 [PubMed: 22722833]
98. Potter MW et al. (2001) Endotoxin (LPS) stimulates 4E-BP1/PHAS-I phosphorylation in macrophages. *J. Surg. Res* 97, 54–59 [PubMed: 11319880]
99. Starck SR et al. (2016) Translation from the 5' untranslated region shapes the integrated stress response. *Science* 351, aad3867 [PubMed: 26823435]
100. Min K-W et al. (2017) eIF4E phosphorylation by MST1 reduces translation of a subset of mRNAs, but increases lncRNA translation. *Biochim. Biophys. Acta Gene Regul. Mech* 1860, 761–772 [PubMed: 28487214]
101. Starck SR et al. (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* 336, 1719–1723 [PubMed: 22745432]
102. Komar AA and Merrick WC (2020) A retrospective on eIF2A – and not the alpha subunit of eIF2. *Int J. Mol. Sci* 21, E2054
103. van Heesch S et al. (2019) The translational landscape of the human heart. *Cell* 178, 242–260.e29 [PubMed: 31155234]
104. Stein CS et al. (2018) Mitoregulin: a lncRNA-encoded microprotein that supports mitochondrial supercomplexes and respiratory efficiency. *Cell Rep.* 23, 3710–3720.e8 [PubMed: 29949756]
105. Zhang S et al. (2020) Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly. *Nat. Commun* 11, 1312 [PubMed: 32161263]
106. O'Neill LAJ and Pearce EJ (2016) Immunometabolism governs dendritic cell and macrophage function. *J. Exp. Med* 213, 15–23 [PubMed: 26694970]
107. van den Bossche J et al. (2017) Macrophage immunometabolism: where are we (going)? *Trends Immunol.* 38, 395–406 [PubMed: 28396078]
108. Ruiz Cuevas MV et al. (2021) Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Rep.* 34, 108815 [PubMed: 33691108]
109. Laumont CM et al. (2018) Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med* 10, eaau5516 [PubMed: 30518613]
110. Chong C et al. (2020) Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat. Commun* 11, 1293 [PubMed: 32157095]
111. Ahrens CH et al. (2022) A practical guide to small protein discovery and characterization using mass spectrometry. *J. Bacteriol* 204, e0035321 [PubMed: 34748388]
112. Leblanc S and Brunet MA (2020) Modelling of pathogen-host systems using deeper ORF annotations and transcriptomics to inform proteomics analyses. *Comput. Struct. Biotechnol. J* 18, 2836–2850 [PubMed: 33133425]
113. Prensner JR et al. (2021) Noncanonical open reading frames encode functional proteins essential for cancer cell survival. *Nat. Biotechnol* 39, 697–704 [PubMed: 33510483]
114. Erhard F et al. (2018) Improved Ribo-seq enables identification of cryptic translation events. *Nat. Methods* 15, 363–366 [PubMed: 29529017]
115. Omenn GS et al. (2017) Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project. *J. Proteome Res* 16, 4281–4287 [PubMed: 28853897]
116. Bartel J et al. (2020) Optimized proteomics workflow for the detection of small proteins. *J. Proteome Res* 19, 4004–4018 [PubMed: 32812434]
117. Petruschke H et al. (2020) Enrichment and identification of small proteins in a simplified human gut microbiome. *J. Proteome* 213, 103604
118. Catherman AD et al. (2014) Top down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun* 445, 683–693 [PubMed: 24556311]
119. Caron E et al. (2015) An open-source computational and data resource to analyze digital maps of immunopeptidomes. *eLife* 4, e07661 [PubMed: 26154972]

120. Purcell AW et al. (2019) Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. *Nat. Protoc* 14, 1687–1707 [PubMed: 31092913]
121. Wilhelm M et al. (2021) Deep learning boosts sensitivity of mass spectrometry-based immunopeptidomics. *Nat. Commun* 12, 3346 [PubMed: 34099720]
122. Wang Y et al. (2021) Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol* 39, 1348–1365 [PubMed: 34750572]
123. Zhang S et al. (2021) Bottom-up fabrication of a proteasome-nanopore that unravels and processes single proteins. *Nat. Chem* 13, 1192–1199 [PubMed: 34795436]
124. Brinkerhoff H et al. (2021) Multiple rereads of single proteins at single-amino acid resolution using nanopores. *Science* 374, 1509–1513 [PubMed: 34735217]
125. Bock C et al. (2022) High-content CRISPR screening. *Nat. Rev. Methods Primer* 2, 1–23
126. Wang T et al. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80–84 [PubMed: 24336569]
127. Gilbert LA et al. (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159, 647–661 [PubMed: 25307932]
128. Liu SJ et al. (2017) CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 355, eaah7111
129. Covarrubias S et al. (2020) High-throughput CRISPR screening identifies genes involved in macrophage viability and inflammatory pathways. *Cell Rep.* 33, 108541 [PubMed: 33378675]
130. Dixit A et al. (2016) Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167, 1853–1866.e17 [PubMed: 27984732]
131. Schwinn MK et al. (2020) A simple and scalable strategy for analysis of endogenous protein dynamics. *Sci. Rep* 10, 8953 [PubMed: 32488146]
132. Leonetti MD et al. (2016) A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc. Natl. Acad. Sci. U. S. A* 113, E3501–E3508 [PubMed: 27274053]
133. Cho NH et al. (2022) OpenCell: endogenous tagging for the cartography of human cellular organization. *Science* 375, eabi6983 [PubMed: 35271311]
134. Wang L et al. (2022) Therapeutic peptides: current applications and future directions. *Signal Transduct. Target. Ther* 7, 48 [PubMed: 35165272]
135. Carreno BM et al. (2015) Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* 348, 803–808 [PubMed: 25837513]
136. He X et al. (2021) Immunization with short peptide particles reveals a functional CD8(+) T-cell neoepitope in a murine renal carcinoma model. *J. Immunother. Cancer* 9, e003101 [PubMed: 34862254]
137. He Q et al. (2019) Targeting cancers through TCR-peptide/MHC interactions. *J. Hematol. Oncol* 12, 139 [PubMed: 31852498]

**Box 1.****Proteomics for SEP discovery**

Standard MS workflows are optimized for large, abundant proteins and may fail to detect SEPs for several reasons, including low sORF translation, high SEP turnover rate, loss of small peptides in sample preparation, and a lack of trypsin digest sites [111]. For these reasons there has been only moderate success in reanalyzing MS data with updated peptide predictions [112,113], and deep peptidomic analyses aimed at detecting SEPs commonly result in intolerable false discovery rates [103,114,115]. Nonetheless, direct detection of SEP production is possible and desirable; here are some approaches to peptide detection, current and future.

**Bottom-up or shotgun proteomics**

This is the most common approach to proteomic MS. Following sample collection, a protease (typically trypsin) is used to fragment the protein. Protein fragments are subject to LC, and are then treated to MS. In the case of tandem MS, measured peptides are further fragmented and undergo an additional round of MS (see Figure 1 in main text). Typically, this is optimized for peptides in the range of 8–25 amino acids, and SEPs may have too few digestion sites to produce fragments in that range [111]. Fortunately, recent excitement in the field of SEP discovery is leading to improved workflows that enhance the discovery of small peptides [14,111,116,117].

**Top-down proteomics**

The distinguishing feature from the bottom-up technique is the lack of a protease digestion step, enabling direct MS on the intact peptide; this avoids the problem of ambiguous peptide fragment spectra. The main drawbacks are technical and practical: this approach is less well suited to complex protein lysates, and fewer core MS facilities have expertise in it [111,118]. However, in the near future it could become the preferred method for SEP detection [111].

**Immunopeptidomics**

This involves isolating MHC–peptide complexes and eluting out the peptides, which can be analyzed by LC-MS/MS [119] (see Figure 1B). Conveniently, the size ranges of both MHCI and MHCII complex peptides are appropriate for standard analysis without a digestion step [120]. This approach may be particularly attractive as SEPs are overrepresented in MHC complexes, a phenomenon partly explained by SEPs entering the ‘defective ribosomal product’ pathway at a high rate [114]. This approach has proved highly successful in identifying SEPs associated with cancer [108–110,120,121].

**Nanopore protein sequencing**

Nanopore long-read sequencing of DNA and RNA has had a lasting impact on DNA and RNA sequencing [122]. The technology works by measuring ion flow through a membrane-spanning nanopore. The pore is equipped with a combination motor protein/helicase which unwinds nucleotide sequences and pulls them through the pore. The nucleotide sequence in the pore causes a characteristic change in the current from which

the sequence can be inferred [122]. The same idea applied to protein sequences faces added challenges, including the greater variety of amino acids compared to nucleotides, and the stability of protein tertiary structures. Nonetheless, recent advances show that this approach is feasible in principle [123,124]. If brought to fruition, this technology could revolutionize peptide discovery by reducing dependence on sophisticated MS devices and allowing protein sequence determination without prior predictions.

**Box 2.****CRISPR-Cas for SEP discovery and characterization****CRISPR-Cas screens**

Genome-wide CRISPR-Cas screens have proved to be valuable tools in protein-coding gene and lncRNA characterization [125-128]. Here, they offer the same ability to quickly narrow a large number of putative coding ORFs to those that are involved in a given phenotype. In a recent CRISPR-Cas9 screen against 553 novel ORFs in four human cancer cell lines, 57 were implicated in growth and viability; roughly half of these induced a consistent phenotype across all four cell lines [113]. Another CRISPR-Cas9 growth screen against 2353 noncanonical ORFs in induced pluripotent stem cells and K562 cells linked growth defects to hundreds of targeted ORFs. This included 229 lncRNA ORFs in either cell type and 51 lncRNA ORFs in both cell types [78].

**Perturb-Seq**

Screens, such as mentioned in the previous section, demonstrate that a large number of putatively functional ORFs can be reduced, in a high-throughput manner, to a set of high-confidence ORFs in a given context. However, they are limited to a single phenotype: typically viability, growth, differentiation, or a pathway reporter [125,128,129]. An alternative approach that provides a more comprehensive readout of cell state is offered by **Perturb-Seq** (or Crop-Seq or CRISPR-Seq) [125,130]. Here, a pool of perturbed cells undergoes single-cell RNA sequencing and transcriptomic readout is used to infer function (see Figure 1 in main text). In the work referenced previously [78], a comprehensive CRISPR-Cas9 screen was performed, and selected hits were followed up with Perturb-Seq. Gene set enrichment analysis allowed direct mapping of particular sORF disruption to biological processes [78]. Such information could be instrumental in informing further experiments aimed at elucidating sORF contribution to specific pathways.

**Homology-directed repair**

In the studies described in the previous sections, CRISPR-Cas was employed in its disruptive **nonhomologous end-joining (NHEJ)** capacity. However, the system can also be used constructively to insert peptide tags in-frame with ORFs via **homology-directed repair (HDR)**. Several recent studies have demonstrated successful use of split-fluorescent systems for visualization of dozens of proteins in single experiments with N- or C-terminal tagging [131,132]. This approach has not yet been applied to the study of sORFs, but it holds significant promise for high-throughput validation of coding potential while simultaneously providing localization information and a handle for MS pulldown to determine SEP-protein interactions [133]. Caveats to consider are the low efficiency of HDR, which frequently results in insertion at only one allele, and a dearth of usable guide RNA sites in close proximity to 5' or 3' termini of the sORF [132]. The OpenCell study [133] successfully tagged over 1000 protein-coding genes using the split-GFP system by inserting the tag directly into portions of the coding sequence that were predicted to avoid disrupting protein function, thereby overcoming the restriction

to N- or C-terminal adjacent sites [133]. Although such an inframe internal insertion could work in sORFs, the small size of sORFs seems likely to make these insertions deleterious.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Significance**

Coding from noncoding: in this review we explore the emerging roles of bifunctional genes in contributing to the innate immune response.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

### Highlights

New approaches have implicated hundreds of long noncoding RNAs as potential protein coding genes through overlooked short open reading frames (sORFs).

There are many thousands of sORFs, with multiple lines of evidence supporting production of sORF-encoded peptides (SEPs), compiled in databases but uninvestigated.

Confirmation of production and functional characterization can be nuanced, but thoughtful interrogation has already expanded the known proteome and our understanding of important biological pathways. This includes contributions to innate immune function in mouse and human.

Although the expanding proteome is likely to interest investigators from all fields, there is reason to believe that immunologists are particularly well positioned to make impactful discoveries.



### Clinician's corner

Beginning with the commercialization of the 51-amino-acid insulin in 1921, small peptides have had important roles as therapeutics. Today there are 80 peptide therapeutics and hundreds more under development. Progress is being made in addressing drawbacks such as short half-lives and low oral bioavailability [134].

Although sORFs show low sequence conservation, the protein complexes that their encoded peptides interact with are likely to be conserved, allowing for cross-species therapeutics. To wit, the sORF encoding miPEP155 is reportedly conserved only in primates, but its 17-amino-acid SEP effectively resolves psoriasis and autoimmune encephalomyelitis in mice [16].

An orthogonal clinical application is related to the prevention, detection, and treatment of cancer. A recent study demonstrated a pipeline for discovering cancer-specific MHCI-presented peptides and showed that mice injected with peptide-pulsed DCs could effectively resolve subsequent cancer challenge [135]. A second approach avoided vaccination with DCs, using instead peptide–adjuvant complexes [136]. In both cases, endogenous T cells were primed against future detection of these cancer-associated antigens.

An alternative to relying on endogenous T cells is to select or engineer T cells with affinity for specific peptide–MHC complexes. In clinical studies of such systems, one major problem is on-target/off-tumor effects, that is, a failure to discriminate between tumors and healthy tissue [137]. In studies undertaken thus far, the MHCI peptidome has been heavily populated by SEP-derived peptides [108-110], and we expect these to be a source of tumor-specific antigens in the future.

### Outstanding questions

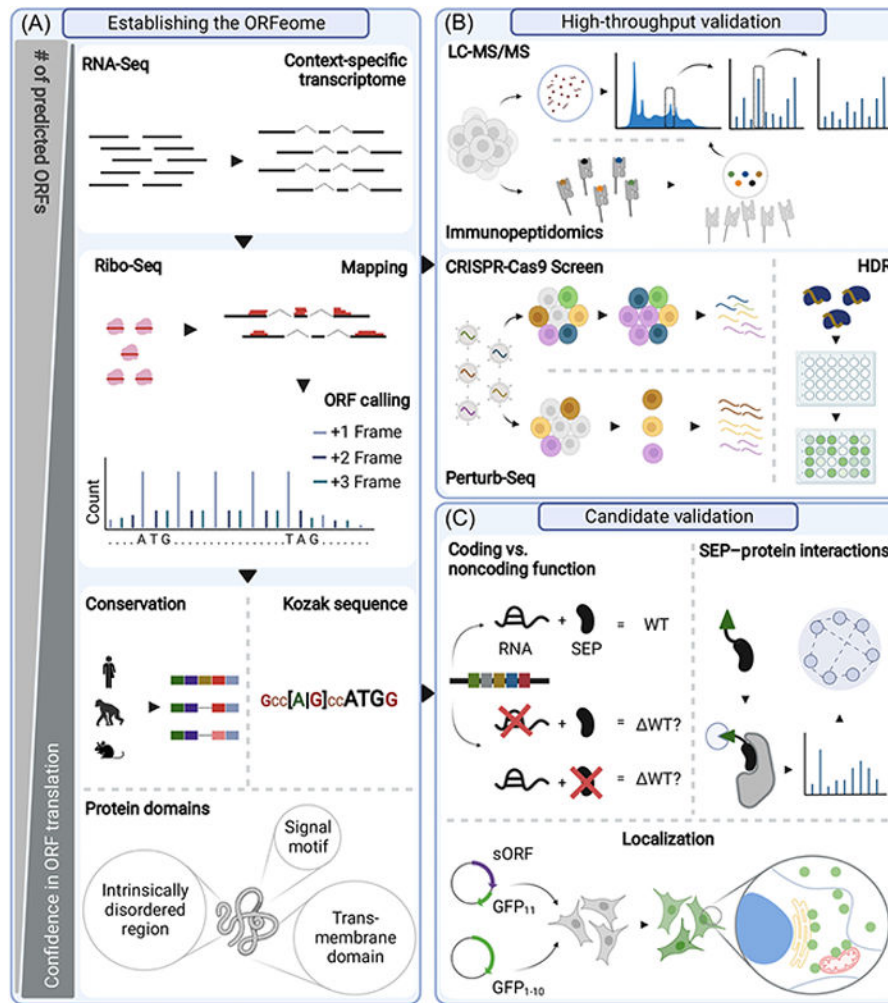
From the perspective of broad scientific inquiry, at least two fundamental questions remain unanswered: Which sORFs are translated? Of the resulting SEPs, which are biologically functional?

There are many databases with coding predictions for uncharacterized ORFs based on hundreds of Ribo-Seq experiments. Would time and energy be better spent parsing these datasets for candidate sORFs, or is sORF translation so context-specific that more Ribo-Seq experiments are required?

There are many coding prediction tools, including dozens based on Ribo-Seq alone. How well do predictions from these tools coincide? As additional functional characterizations of SEPs are generated, can these data be incorporated to improve prediction tools?

How often do sORF coding genes operate in a bifunctional capacity? To what extent do functional lncRNAs and proto-genes overlap? And to what extent is the community's built-in bias preventing us from understanding RNA contribution to phenotype in coding genes more broadly?

How important is the role of sORFs in regulating adaptive immunity through their antigenic presentation on MHC I? Is this a regulated process? Is there a link between dysregulated SEP production and subsequent MHC presentation and autoimmune diseases?



**Figure 1. Key figure. Discovering and characterizing novel peptides**

(A) Steps in discovering translated open read frames (ORFs). Top: determining the transcriptome of the cell type of interest. Middle: Ribo-Seq specifies ribosome-associated ORFs. Additional confidence gained from three-nucleotide periodicity. Bottom: computational analyses including conservation score, identification of Kozak sequences, and predicted protein domains increase confidence in ORF translation. (B) Approaches to high-throughput validation of functional translation. Proteomics: proteins are measured from whole-cell lysate or MHC-peptide complexes. CRISPR-Cas9 Screen: GuideRNAs (gRNAs) against putative short ORFs (sORFs) are sequenced in bulk, and changes in distribution of guides imply sORF-encoded peptide (SEP) function (i.e., disappearance of guides over time, indicating a SEP critical for viability). Perturb-Seq: single-cell RNA sequencing links sORF disruption to phenotype. Homology directed repair (HDR): epitope tag insertion inframe with the sORF via nucleofection with Cas9-complexed RNAs (Box 2). (C) Approaches to validating sORF coding potential and determining SEP function. Coding versus noncoding function: ablating the start codon of the sORF leaves RNA function intact. Alternatively, re-engineering sORF with synonymous mutations alters RNA sequence and structure but leaves SEP intact. Conclusions about the relative importance of SEP

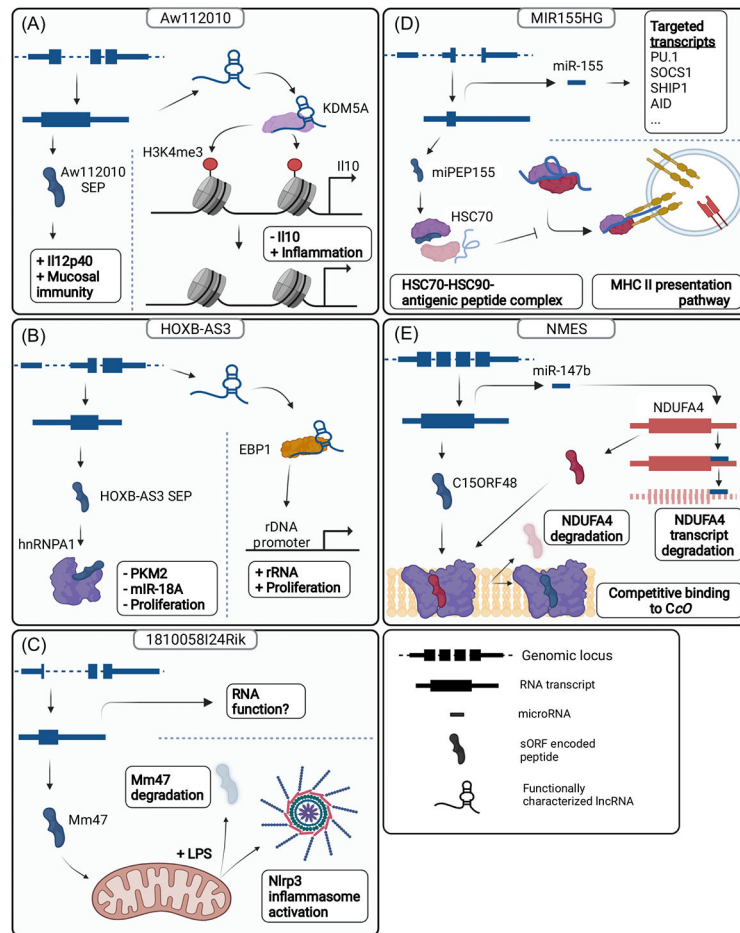
versus RNA contribution to phenotype may be context-specific. SEP-protein interactions: a tagged SEP, or an anti-SEP antibody, can be used to pull down the SEP and interaction partners which can be identified by mass spectrometry (MS). Localization: determined by epitope tagging and immunostaining. Split fluorescent systems may reduce disruption of wild-type (WT) SEP activity while providing direct fluorescent confirmation of translation. Abbreviations: LC, liquid chromatography; GFP, green fluorescent protein. Figure generated using [Biorender.com](https://biorender.com).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 2. Newly characterized short open reading frame (sORF)-encoded peptide (SEP) in innate immunity.**

(A) In mouse bone-marrow-derived macrophages (BMDMs), *Aw112010* SEP is essential to robust mucosal immunity. In mouse CD4<sup>+</sup> T cells *Aw112010* RNA guides KDM5A demethylase to histones at the *Ii10* locus, reducing expression [8,82]. (B) In K562s and an acute myeloid leukemia (AML) cell line (human), a *HOXB-AS3* transcript produces a SEP that antagonistically binds at the hnRNPA1 RNA-binding domain, reducing the amounts of cancer-associated transcripts and decreasing proliferation. In a colon cancer line, an alternative *HOXB-AS3* transcript guides DNA methylase EBP1 to the ribosomal DNA locus, increasing transcription and contributing to a proliferative phenotype [10,30]. (C) In BMDMs, *1810058I24Rik* produces Mm47, which localizes to the mitochondria and contributes to Nlrp3 inflammasome generation in response to lipopolysaccharide (LPS). LPS also triggers degradation of Mm47, perhaps as a timer on the resolution of inflammation [76]. (D) in BMDMs, *MIR155HG* produces miPEP155 which antagonistically binds HSC70 and reduces MHCII display [16]. This transcript also serves as a precursor to miRNA miR-155 which regulates many immune-related RNAs. (E) In human monocyte-derived macrophages, *NMES* produces SEP C15ORF48 which competes with NDUFA4 in binding cytochrome *c* oxidase (*CcO*). Additionally, miRNA miR-147b downregulates the *NDUFA4* transcript [92]. Figure generated using [Biorender.com](https://www.biorender.com).

**Table 1.**Ribo-Seq in innate immune cells<sup>a</sup>

Species	Cell type	Experimental conditions	Translation inhibitors	GEO accession	Refs
Mouse	BMDC	NT	CHX	GSE59793	[42]
	BMDC	LPS	CHX, HARR, LTM	GSE74139	[29]
	BMDC	Mettl3 KO	CHX	GSE108331	[43]
	BMDM	LPS	CHX	GSE99787	[44]
	BMDM	LPS	CHX	GSE120762	[82]
	BMDM	<i>Legionella pneumophila</i> strains	CHX, HARR	GSE89184	[45]
	BMDN	Mir-223 KO	CHX	GSE22004	[46]
	Microglia from whole brain tissue <sup>b</sup>	AZD-8055	CHX	GSE78163	[47]
	Microglia	PrP <sup>Sc</sup>	CHX	GSE149805	[48]
	RAW264.7	Nascent Ribo-Seq <sup>c</sup>	CHX	GSE155236	[49]
Human	MM1.S	Bortezomib	CHX	GSE69047	[50]
	MM1.S	Bortezomib	CHX	GSE48785	[51]
	THP-1	NT	PURO, CHX	GSE39561	[52]
	K562, erythrocytes, reticulocytes, platelets	PELO overexpression, shABCE1	CHX	GSE85864	[53]
	K562	NT	CHX, HARR	GSE125218	[54]
	K562	NT	CHX	GSE129061	[55]
	K562	Survey of rRNA depletion techniques	CHX	GSE147324	[56]
	Primary CD14 <sup>+</sup> monocytes	IFN- $\gamma$ , Pam3CSK4	CHX	GSE66810	[57]
	K562, Spleen tissue, primary PBMCs, primary monocytes, primary B and T cells	RNAse footprinting <sup>d</sup>	CHX	GSE151989 GSE151986 GSE151987 GSE151988 GSE153411	[58]

<sup>a</sup> Abbreviations: BMDC, bone-marrow-derived dendritic cell; BMDM, bone-marrow-derived macrophage; BMDN, bone-marrow-derived neutrophil; CHX, cycloheximide; HARR, harringtonine; IFN, interferon; KO, knockout; LPS, lipopolysaccharide; LTM, lactimidomycin; NT, no treatment; PBMC, peripheral blood mononuclear cell; PURO, puromycin.

<sup>b</sup> Uses cell-type-specific gene expression to computationally deconvolute individual cell types from whole tissue.

<sup>c</sup> Introduces a modified technique, nascent Ribo-Seq, for investigating nascent-mRNA-ribosome loading kinetics.

<sup>d</sup> Introduces a low-input technique, RNAse footprinting, that is distinct from Ribo-Seq but ultimately results in similar output.

**Table 2.**Select sORF and SEP databases and resources<sup>a</sup>

Primary use	Resource	ORF/SEP supporting evidence	Notable features	URL	Refs
Visualization	GWIPS-Viz	Ribo-Seq, Ti-Seq and RNA-Seq visualization from individual experiments and global aggregates	Online genome browser for visualizing processed Ribo-Seq data	<a href="https://gwips.ucc.ie/">https://gwips.ucc.ie/</a>	[59]
	Trips-Viz	Ribo-Seq and MS data from individual experiments	Online browser for visualizing processed Ribo-Seq data. Can plot single nucleotide read intensity and predict translated ORFs	<a href="https://trips.ucc.ie/">https://trips.ucc.ie/</a>	[60]
Transcriptome-wide ORF predictions	RPFdb v2.0	Ribo-Seq	Comprehensive collection of Ribo-Seq studies with ORF predictions in a searchable browser	<a href="http://sysbio.gzzoc.com/rpfdb/index.html">http://sysbio.gzzoc.com/rpfdb/index.html</a>	[61]
	sORFs.org	Conservation Ribo-Seq MS	Incorporates multiple sORF scoring metrics and annotates sORFs with dozens of attributes	<a href="http://www.sorfs.org/">http://www.sorfs.org/</a>	[62]
	Metamorf	Conservation Coding potential Kozak context Ribo-Seq MS	Includes UCSC genome browser session for visualization and introduces an ORF nomenclature	<a href="https://metamorf.hb.univ-amu.fr/">https://metamorf.hb.univ-amu.fr/</a>	[63]
	smProt	Literature mining Database mining Ribo-Seq MS	Includes variant and disease specific annotations and maintains a high-confidence set	<a href="http://bigdata.ibp.ac.cn/SmProt/">http://bigdata.ibp.ac.cn/SmProt/</a>	[64]
	OpenProt	Ribo-Seq MS	Generates ORFeome by 3-frame transcriptome translation, then checks for evidence of translation	<a href="https://www.openprot.org/">https://www.openprot.org/</a>	[65]
	nORFs.org	sORFs.org openProt.org	Aggregates information from other databases and presents a fast, user-friendly interface, including a built-in genome browser	<a href="https://norfs.org/home">https://norfs.org/home</a>	[66]
ORFs in noncoding RNA	LncPEP	Coding potential Conserved protein domains Ti-Seq Ribo-Seq m6A RNA modification	Focuses on lncRNAs. Determines a coding score based on a normalized sum of six input variables	<a href="http://www.shenglilabs.com/LncPep/">http://www.shenglilabs.com/LncPep/</a>	[67]
	SPENCER	MS	SEPs from ncRNA in cancer including predictions of MHC I affinity, stability, and TCR recognition probability	<a href="http://spencer.renlab.org/#/home">http://spencer.renlab.org/#/home</a>	[68]
	Coding and noncoding RNA database	Literature review	Curated RNAs that have coding and noncoding functions with links to literature and supporting data	<a href="http://www.rna-society.org/cncrnadb/">http://www.rna-society.org/cncrnadb/</a>	[69]
Ribo-Seq web application	RiboToolkit	Predictions from uploaded <b>FASTQs</b>	Webserver that accepts Ribo-Seq FASTQs for analysis and implements a full prediction pipeline	<a href="http://rnabioinfor.tch.harvard.edu/RiboToolkit/">http://rnabioinfor.tch.harvard.edu/RiboToolkit/</a>	[41]

<sup>a</sup>These resources enable investigators to circumvent the technical and computational challenges of sORF prediction and move directly to the hypothesis-driven characterization phase. Indeed, given the thousands of predicted but unstudied sORFs, the greater benefit to the scientific community collectively might come from investigations of current predictions rather than new exploratory experiments.