Briefings in Bioinformatics, 2025, **26(1)**, bbae641
https://doi.org/10.1093/bib/bbae641
Review

Exploring the potential of large language model–based chatbots in challenges of ribosome profiling data analysis: a review

Zheyu Ding[1,2], Rong Wei[1,2], Jianing Xia[1,2], Yonghao Mu[1,2], Jiahuan Wang[1,2], Yingying Lin[1,2,*]

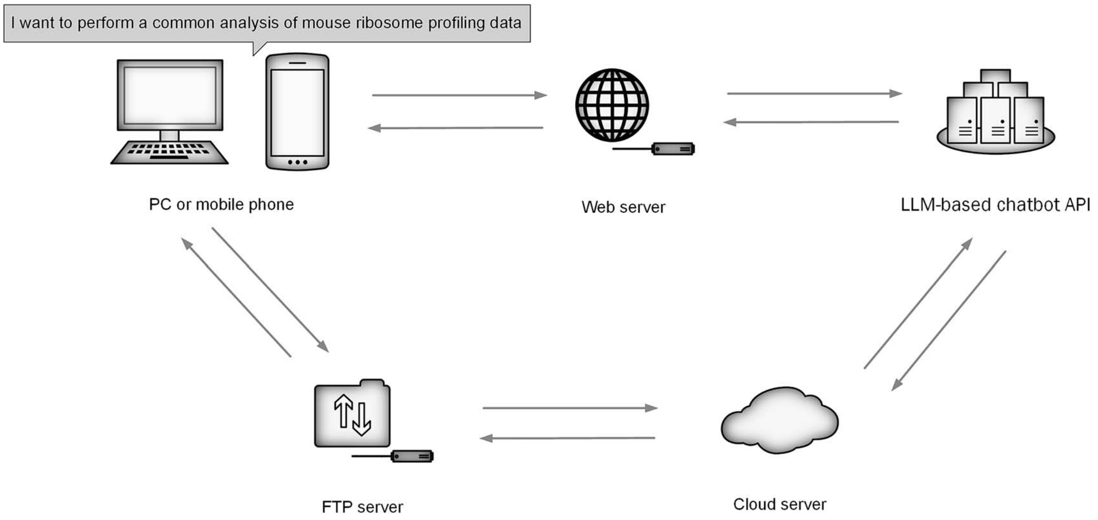
[1]School of Pharmacy, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China
[2]Key Laboratory of Elemene Class Anti-Cancer Chinese Medicines, Engineering Laboratory of Development and Application of Traditional Chinese Medicines, Collaborative Innovation Center of Traditional Chinese Medicines of Zhejiang Province, Hangzhou Normal University, Hangzhou, Zhejiang 311121, China

[*]Corresponding author. Yingying Lin, School of Pharmacy, Hangzhou Normal University, Hangzhou 311121, China. E-mail: YYLin@hznu.edu.cn

Abstract

Ribosome profiling (Ribo-seq) provides transcriptome-wide insights into protein synthesis dynamics, yet its analysis poses challenges, particularly for nonbioinformatics researchers. Large language model–based chatbots offer promising solutions by leveraging natural language processing. This review explores their convergence, highlighting opportunities for synergy. We discuss challenges in Ribo-seq analysis and how chatbots mitigate them, facilitating scientific discovery. Through case studies, we illustrate chatbots' potential contributions, including data analysis and result interpretation. Despite the absence of applied examples, existing software underscores the value of chatbots and the large language model. We anticipate their pivotal role in future Ribo-seq analysis, overcoming limitations. Challenges such as model bias and data privacy require attention, but emerging trends offer promise. The integration of large language models and Ribo-seq analysis holds immense potential for advancing translational regulation and gene expression understanding.

Graphical Abstract



Keywords: large language models; chatbots; next-generation sequencing; ribosome profiling; data analysis

Introduction

Ribosome profiling, also known as Ribo-seq, is based on high-throughput sequencing of ribosome-protected messenger RNAs (mRNAs) [1]. The method has facilitated the discovery of the regulation of gene expression underlying diverse and complex biological processes, the mechanism of protein synthesis, and even the identification of new proteins, by providing a systematic approach for experimental annotation of coding regions [2].

Received: July 9, 2024. **Revised:** November 2, 2024. **Accepted:** November 27, 2024

In parallel, the field of artificial intelligence has witnessed remarkable advancements, particularly in the realm of large language models (LLMs). These models, such as OpenAI's Generative Pretrained Transformer (GPT) series, are trained on vast amounts of text data and demonstrate exceptional capabilities in natural language understanding and generation. Beyond their initial applications in language processing tasks like text generation and summarization, LLMs have found utility across diverse domains including healthcare, finance, and education [3–7]. Their ability to comprehend and generate human-like text has spurred innovative applications, transforming how we interact with information and automate various tasks [8].

However, although there are so many wonderful toolkits for ribosome profiling data analysis designed for various aspects [9], there still exist challenges, especially for researchers lacking any coding experiences or limited bioinformatics expertise. For researchers without a background in bioinformatics, it demands a significant investment of time and effort to acquire knowledge of Linux, R, and Python which entails conducting ribosome profiling (Ribo-seq) data analysis. Depending solely on standardized analysis reports generated by sequencing companies may sometimes fail to meet researchers' personalized needs. Additionally, for novice bioinformatics researchers, the choice of different data analysis software for the same ribosome profiling raw data sometimes may yield disparate results [10]. Thus, optimizing data analysis workflows becomes an important consideration, but it is difficult for them. In summary, the immense volume and complexity of ribosome profiling datasets present formidable obstacles for data analysis, particularly for those without a background in bioinformatics and newcomers to the field.

It is within this context that the potential of LLM-based chatbots emerges as a compelling solution. By leveraging the natural language processing (NLP) capabilities of LLMs [11], chatbots may assist researchers in navigating the complexities of Ribo-seq data analysis. From data preprocessing and alignment to downstream analysis and interpretation, LLM-based chatbots have the potential to automate the appropriate workflows and empower researchers to extract meaningful insights from Ribo-seq datasets efficiently.

The intersection of LLMs and bioinformatics holds significant promise for addressing current limitations in Ribo-seq data analysis. LLM-based chatbots can democratize access to advanced bioinformatics tools and reduce the learning curve for new researchers. By integrating LLMs into the Ribo-seq data analysis workflow, we can overcome existing barriers and enhance the overall efficiency and accessibility of bioinformatics research.

In this review, we explore the convergence of Ribo-seq data analysis and LLM-based chatbots, highlighting the opportunities for synergy between these two domains. We examine the challenges inherent in Ribo-seq data analysis and discuss how LLM-based chatbots can mitigate these challenges, paving the way for accelerated scientific discovery in the field of bioinformatics and biomedical science [12].

Fundamentals of ribosome profiling

Ribosome profiling, also known as Ribo-seq, is a cutting-edge experimental technique developed to monitor translation *in vivo* at a genome-wide scale [13, 14]. The principle underlying Ribo-seq involves the selective isolation and sequencing of ribosome-protected mRNA fragments, known as ribosome-protected fragments (RPFs) [15]. These footprints represent the positions of actively translating ribosomes along the mRNA transcriptome [16].

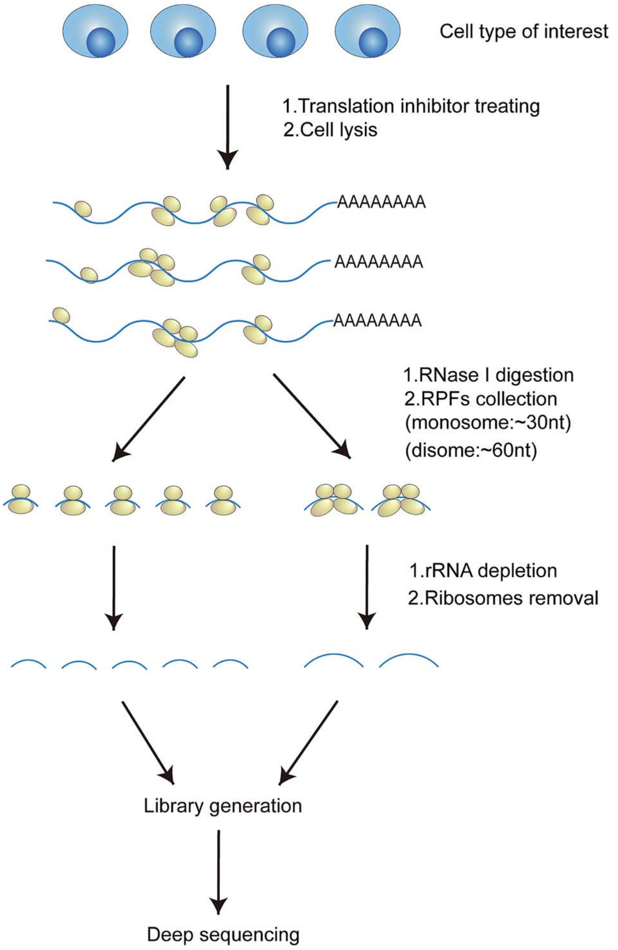


Figure 1. An overview of ribosome profiling library preparation.

Workflow of Ribo-seq technique

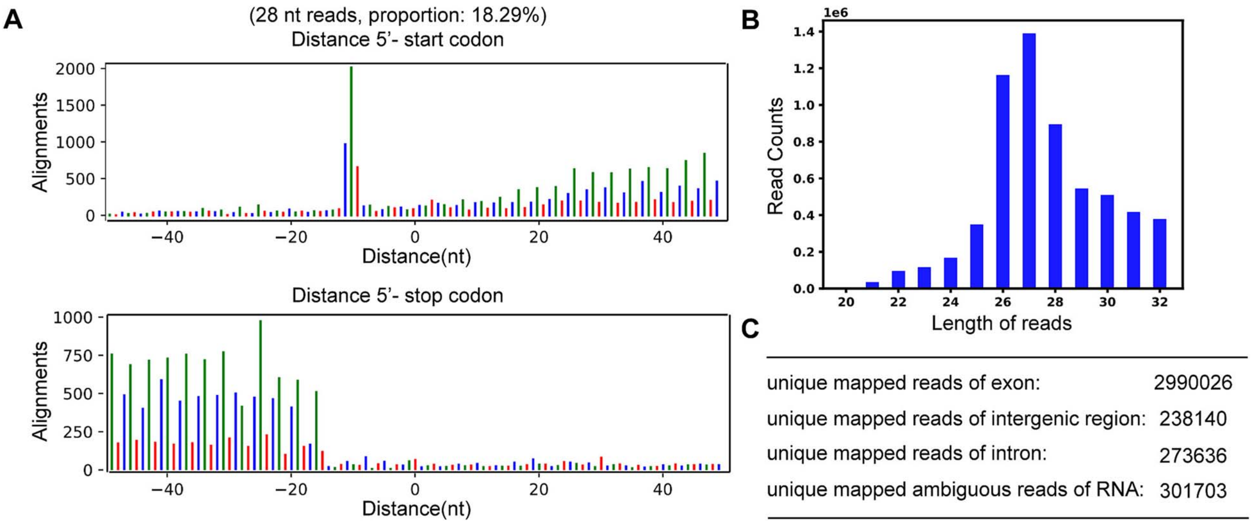
The workflow of Ribo-seq typically involves the following steps (Fig. 1):

The first step in ribosome profiling is the isolation of ribosomes and the preservation of their positions on mRNA molecules. Cells are typically treated with a translation inhibitor [17], such as cycloheximide, to freeze ribosomes in their current positions [18]. This step prevents ribosomes from translocating along the mRNA during the subsequent lysis and RNA isolation processes.

Next, the cells are lysed, the ribosome–mRNA complexes are captured, and the ribosome-protected footprints are isolated. This can be achieved through the addition of nucleases that degrade unprotected RNA molecules, leaving the ribosome-protected fragments (RPFs) intact. The resulting RPFs or ribosome footprints, represent the positions of ribosomes along the mRNA molecules at the time of lysis [19].

Following RPF generation, these fragments undergo several steps before being subjected to high-throughput sequencing. These steps include adenylation, adapter ligation, complementary DNA (cDNA) synthesis, and polymerase chain reaction (PCR) amplification. The sequencing process generates millions of short reads. Each clean read represents an individual ribosome footprint, monosome-protected footprints with typical lengths ∼30 nucleotides [20], and disome-protected footprints with typical lengths ∼60 nucleotides [21], corresponding to the size of ribosome footprints [22].

Once the sequencing raw data are obtained, the analysis phase begins. The raw sequencing reads are initially processed to remove



unique mapped reads of exon:	2990026
unique mapped reads of intergenic region:	238140
unique mapped reads of intron:	273636
unique mapped ambiguous reads of RNA:	301703

Figure 2. Post mapping quality control step of Ribo-seq data analysis. (A) Triplet periodicity plots, using 28 nt reads as an example. (B) Length distributions of all the RPF reads. (C) Numbers of reads mapped to different regions of the genome. The sample SRR9047194 in the dataset GSE131074 was used as an example here [37].

adapter sequences [23], remove reads with low quality, and remove ribosomal RNA (rRNA) contamination through quality control measures [24, 25]. These steps ensure the accuracy and reliability of downstream analyses.

After preprocessing, the ribosome profiling data are aligned to the reference genome or transcriptome using specialized alignment algorithms [26]. This step maps the ribosome footprints to their corresponding genomic locations, allowing for the identification of the genes and regions undergoing active translation [27].

The aligned data can then be further analyzed to extract various types of information, such as translation efficiency [28], ribosome density [29], cotranslation between protein [30], and translational dynamics [31–33]. Advanced computational tools and statistical methods are employed to infer translation-related parameters and identify differentially translated genes between different conditions or cell types.

Overall, the experimental workflow of ribosome profiling involves the isolation and preservation of ribosome–mRNA complexes, followed by the generation of ribosome footprints through nuclease digestion. The resulting footprints are then sequenced, and the data are processed, aligned, and analyzed to extract meaningful insights into translation dynamics and regulation [34]. By following this workflow, researchers can obtain genome-wide information on translation events and unravel the intricacies of protein synthesis at a transcriptome-wide scale [35].

Key points of Ribo-seq data analysis

Because of the unique characteristics of Ribo-seq data, one critical step of ribosome profiling data analysis is the evaluation of postmapping quality [36]. The key points are the assessment of triplet periodicity (Fig. 2A), read-length distribution (Fig. 2B), and DNA contamination (unique read count across genomic features such as exon, intergenic region, intron, and ambiguous reads of RNA) (Fig. 2C) [37].

Integrative analysis of Ribo-seq with other omics data

Ribo-seq is a quintessential example of sequencing-based omics data. The true potential of Ribo-seq is often realized when it is integrated with other omics datasets, such as transcriptomics, proteomics, and other translatomics. This integrative analysis enables a more comprehensive understanding of the intricate layers of gene regulation and protein synthesis.

One of the most common combinations is Ribo-seq with RNA sequencing (RNA-seq). This integration allows for the calculation of translation efficiency (TE) [28], which is a key metric that reflects the ratio of ribosome occupancy to mRNA abundance. By comparing Ribo-seq and RNA-seq data, researchers can distinguish between translational regulation and transcriptional changes, offering deeper insights into how gene expression is modulated at multiple levels. For example, differential TE across conditions can reveal mechanisms of selective translation that might not be apparent from RNA-seq data alone.

Polysome profiling, another powerful translatomics technique, when combined with Ribo-seq, provides a detailed view of the translational process. By integrating polysome profiling with ribosome profiling, it becomes possible to accurately quantify absolute ribosome density, which can then be used in models such as the totally asymmetric simple exclusion process to derive translation elongation efficiency and translation initiation efficiency [31]. This integration is particularly valuable for understanding the dynamics of ribosome movement along mRNAs.

The combination of Ribo-seq, RNA-seq, and proteomics data, particularly liquid chromatography–mass spectrometry (LC-MS), allows for a holistic view of gene expression from mRNA to protein. This integrative approach can uncover potential regulatory elements and networks at multiple levels of gene expression.

Challenges in Ribo-seq data analysis

When researchers get the Ribo-seq data, sometimes they are not satisfied with the standardized analysis reports generated by sequencing companies. It comes up with several challenges if the researcher requires customized data analysis due to the complexity of Ribo-seq data.

Especially, for researchers with no or limited bioinformatics expertise, although a lot of tools have been developed for Ribo-seq data analysis [9], it is difficult for them to get started with the software. There are two main reasons that shall be blamed for the difficulty. Firstly, the tools are often implemented in different programming languages, leading to an increased demand for learning different computer languages. Secondly, the software's installation often depends on language-centered running environments, leading to an increased difficulty of

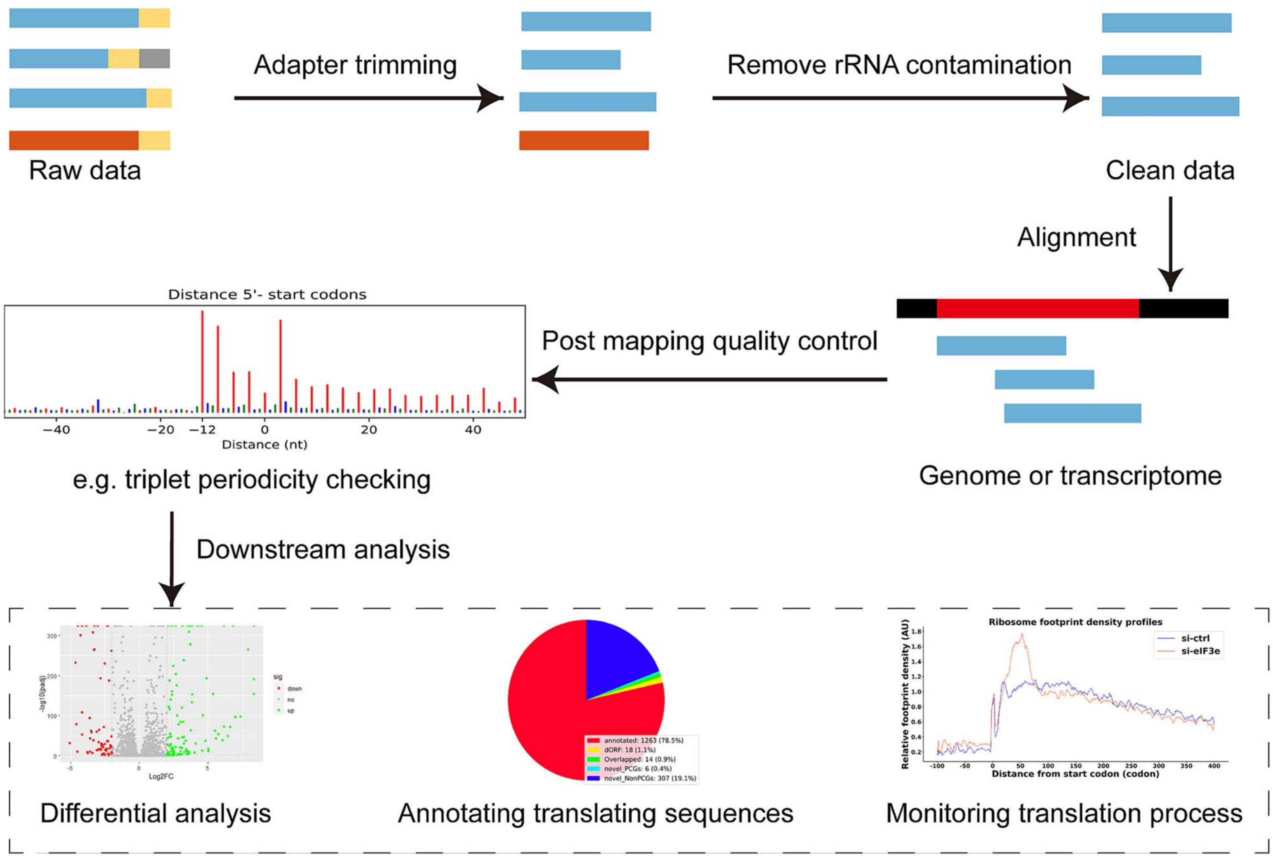


Figure 3. An overview of ribosome profiling data analysis.

software configuration management [36]. Hence, these users need to be acquainted with Linux, R, and Python, which demands a significant investment of time and effort. Steps of ribosome profiling data analysis including adapter trimming, removing rRNA contamination, alignment, and postmapping quality control often require execution on the Linux platform, necessitating proficiency in basic Linux commands. Processes including differential analysis typically involve learning the R programming language, while those that need statistics and computation may necessitate familiarity with Python (Fig. 3).

On the other hand, for researchers with a bioinformatics background, the choice of different data analysis software for the same ribosome profiling raw data sometimes may yield different results. As different raw data may require distinct analysis pipelines to achieve optimal outcomes, optimizing data analysis workflows needs to be strongly considered. Furthermore, due to variations in the library preparation methods for Ribo-seq, the raw data reads may contain different adapters, unique molecular identifiers (UMIs), and barcodes. Therefore, adjustments to the optimized pipeline should be based on the specific characteristics of each set of Ribo-seq raw data.

In summary, the complexity of ribosome profiling datasets presents formidable obstacles to data analysis, particularly for researchers without a background in bioinformatics and newcomers to the field. Therefore, there is a demand for tools to fulfill the following five features, as outlined in a paper: natural language understanding (NLU), artificial intelligence (AI), transparency, mobile and social media friendliness, and crowdsourcing [38]. In my opinion, integration of LLM-based chatbots in ribosome profiling data analysis may be a powerful weapon to achieving these requirements.

Role of chatbots in scientific discovery

Chatbots, also known as conversational agents or digital assistants [39], are software programs designed to simulate human conversation through NLP techniques [40]. Over the years, chatbot technology has undergone significant evolution, driven by advances in artificial intelligence, machine learning, and NLP algorithms [41]. Early chatbots relied on rule-based systems, where predefined rules determined their responses to user inputs. However, with the advent of machine learning and deep learning techniques [42], modern chatbots can learn from data and improve their conversational abilities over time [43]. Recently, there have been notable advancements in the development of LLMs such as GPT series, which have revolutionized chatbot capabilities by enabling more contextually relevant and human-like interactions.

Applications of chatbots in scientific research

Chatbots have shown diverse applications in scientific research, enhancing various aspects of the scientific discovery process.

- Literature mining: Chatbots can assist researchers in navigating the vast landscape of scientific literature by providing tailored recommendations, summarizing articles, and extracting key information from research papers [44]. This enables researchers to stay updated with the latest findings in their field and identify relevant literature for their studies.
- Data analysis: Chatbots equipped with analytical capabilities can fulfill data analysis tasks by automating data processing, visualization, and statistical analysis [45]. More importantly, chatbots possess features of being user-friendly and highly

interactive, enabling researchers to use them more easily [36]. In scientific disciplines like genomics, proteomics, and bioinformatics, chatbots can facilitate the interpretation of complex omics data and help researchers uncover meaningful patterns and insights [46].

- Hypothesis generation: Chatbots can aid researchers in hypothesis generation by synthesizing existing knowledge, identifying research gaps, and proposing novel hypotheses based on available data [47]. By leveraging machine learning algorithms and knowledge graphs, chatbots can assist researchers in formulating testable hypotheses and designing experiments to validate them.
- Collaboration facilitation: Chatbots serve as virtual collaborators, enabling seamless communication and collaboration among researchers across geographical locations and disciplinary boundaries [48, 49].

Advantages of using chatbots in scientific workflows

The integration of chatbots into scientific workflows offers several advantages.

- Automation: Chatbots automate routine tasks and workflows, freeing up researchers' time to focus on more creative and intellectually demanding activities [50]. By automating data analysis, literature search, and experiment design, chatbots enhance research productivity and efficiency [51].
- Scalability: Chatbots provide scalable solutions for handling large volumes of data and inquiries. Unlike human counterparts, chatbots can handle multiple inquiries simultaneously and operate the whole day, ensuring rapid response times and scalability to accommodate growing research demands [51–54].
- Accessibility: Chatbots build access to scientific knowledge and tools by providing intuitive interfaces and assistance to researchers with varying levels of expertise [55, 56]. Researchers with limited computational or technical skills can leverage chatbots to perform complex analyses and access advanced research tools with ease [57].

In summary, chatbots play a key role in accelerating scientific discovery by augmenting researchers' capabilities, automating routine tasks, and fostering collaboration and knowledge sharing within the scientific community [58, 59]. As chatbot technology continues to advance, its impact on scientific research is poised to grow, unlocking new opportunities for innovation and discovery across diverse domains [60].

Potential intersection of large language models in Ribo-seq data analysis

Recent years have witnessed significant advancements in LLMs, driven by breakthroughs in deep learning architectures and the availability of vast amounts of textual data for pretraining [61]. Notable examples include models like Llama (Large Language Model Meta AI), Claude, Mistral, NVLM, GPT-3.5 (Generative Pretrained Transformer 3.5) and its successors, such as GPT-4, GPT-4o, and GPT-4o mini, all these giving us unimaginable surprises [62–66]. These LLMs have achieved unprecedented performance in natural language understanding and generation tasks, surpassing earlier benchmarks and demonstrating capabilities akin to human-level language comprehension.

Utilization of large language models in natural language processing tasks relevant to bioinformatics

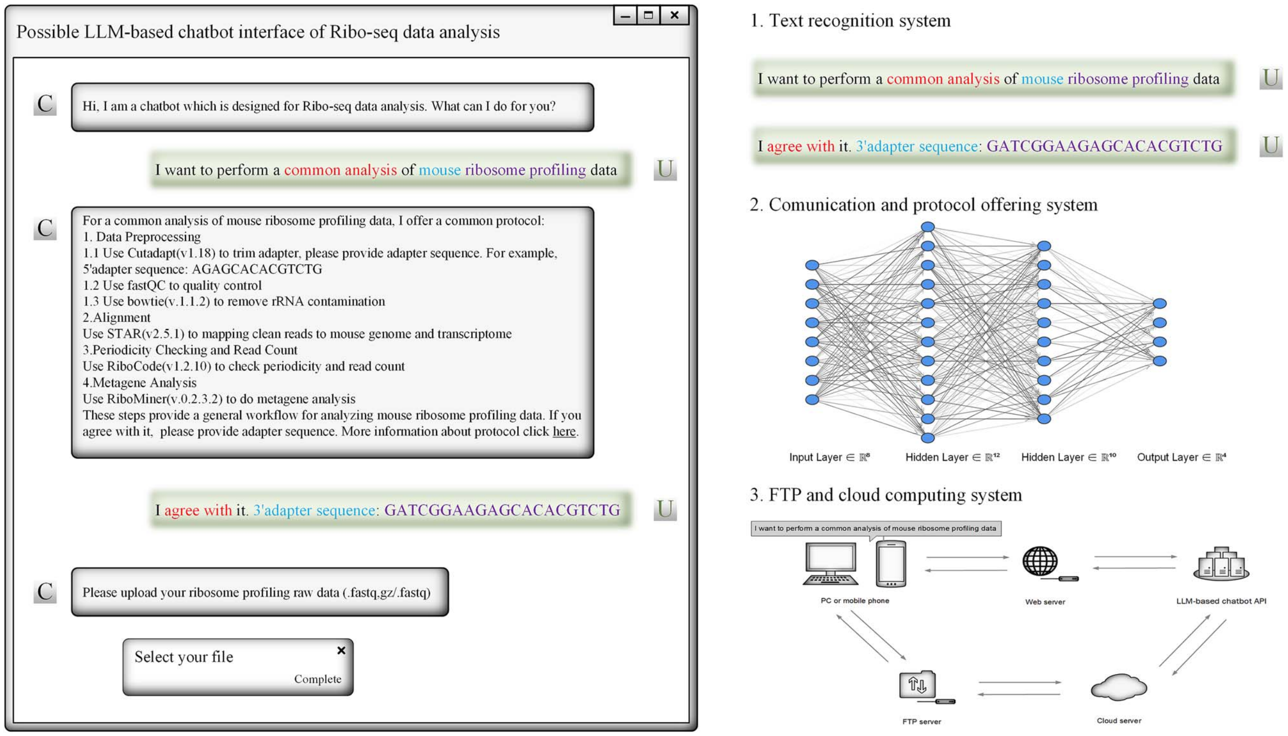
In the domain of bioinformatics, LLMs have been increasingly applied to various NLP tasks [67], leveraging their ability to extract insights from textual data, such as Med-BERT [68], BioBERT [69], and BioGPT [70].

- Literature mining: LLMs can analyze vast repositories of scientific literature to extract relevant information [71], identify key concepts, and summarize findings related to gene expression, protein synthesis, and translational regulation. This facilitates literature review and knowledge synthesis for researchers in the field.
- Textual data analysis: LLMs are capable of processing and analyzing textual data from diverse sources, including research articles, bioinformatics databases, and biological annotations [72]. They can extract structured information from unstructured text [73], enabling annotation for bioinformatics resources.
- Knowledge discovery: LLMs can uncover hidden patterns and relationships within bioinformatics datasets by analyzing textual descriptions of genes, proteins, and biological processes. By integrating information from multiple sources, LLMs facilitate knowledge discovery and hypothesis generation in biological research [3, 74].

Potential applications of large language model–based chatbots in Ribo-seq data analysis

LLM-based chatbots offer promising opportunities for enhancing various aspects of Ribo-seq data analysis. Figure 4 illustrates a potential framework where LLM-based chatbots could enhance Ribo-seq data analysis by integrating intuitive front-end interfaces for natural language queries with back-end automation of data processing, offering a hypothetical improvement in both user interaction and analytical accuracy.

- Data preprocessing: Fine-tuned LLM-based chatbots can assist researchers in preprocessing Ribo-seq data by performing quality control, adapter trimming, reads with low quality removal, UMI and barcode handling, and rRNA and transfer RNA (tRNA) removal. For example, as illustrated in Fig. 4, fine-tuned LLMs, through the communication and protocol offering system, can generate a protocol based on model calculations. Users can then review this protocol to determine if it aligns with their analytical needs. If suitable, they can provide the adapter sequences for subsequent analysis. However, if modifications are needed, such as setting constraints on minimum or maximum read lengths, users can request adjustments, and the fine-tuned LLM will promptly update the corresponding software parameters in the protocol. This capability, if fully realized, could significantly reduce the learning curve and time investment for biologists with limited programming skills.
- Feature extraction: Fine-tuned LLM-based chatbots can extract meaningful features from Ribo-seq data, such as ribosome occupancy profiles [75], translation initiation sites, and ribosomal pausing events [76, 77]. For example, as illustrated in Fig. 4, fine-tuned LLMs, through the file transfer protocol (FTP) and cloud computing system, can generate outputs that include metagene plots created using the RiboMiner software [29]. If the fine-tuned LLM, leveraging its multimodal learning capabilities, detects potential ribosomal



Possible LLM-based chatbot interface of Ribo-seq data analysis

C: Hi, I am a chatbot which is designed for Ribo-seq data analysis. What can I do for you?

U: I want to perform a common analysis of mouse ribosome profiling data

C: For a common analysis of mouse ribosome profiling data, I offer a common protocol:
1. Data Preprocessing
1.1 Use Cutadapt(v1.18) to trim adapter, please provide adapter sequence. For example, 5'adapter sequence: AGAGCACACGTCTG
1.2 Use fastQC to quality control
1.3 Use bowtie(v.1.1.2) to remove rRNA contamination
2.Alignment
Use STAR(v2.5.1) to mapping clean reads to mouse genome and transcriptome
3.Periodicity Checking and Read Count
Use RiboCode(v1.2.10) to check periodicity and read count
4.Metagene Analysis
Use RiboMiner(v.0.2.3.2) to do metagene analysis
These steps provide a general workflow for analyzing mouse ribosome profiling data. If you agree with it, please provide adapter sequence. More information about protocol click here.

U: I agree with it. 3'adapter sequence: GATCGGAAGAGCACACGTCTG

C: Please upload your ribosome profiling raw data (.fastq.gz/.fastq)

Select your file ✕
Complete

1. Text recognition system

I want to perform a common analysis of mouse ribosome profiling data

I agree with it. 3'adapter sequence: GATCGGAAGAGCACACGTCTG

2. Comunication and protocol offering system

3. FTP and cloud computing system

I want to perform a common analysis of mouse ribosome profiling data

Figure 4. An overview of possible front-end and back-end of LLM-based chatbot in ribosome profiling data analysis.

pausing events based on the metagene plot, it can alert the user to this finding. Such proactive insights could significantly contribute to scientific discovery by highlighting critical events that may otherwise go unnoticed.

- Pattern recognition: Fine-tuned LLMs can recognize patterns and regulatory motifs within Ribo-seq data, facilitating the identification of translationally active regions, regulatory elements [78], and sequence motifs associated with translation efficiency and ribosome dynamics [79]. In my opinion, the functionality of this pattern recognition could potentially draw inspiration from the design of an LLM for predicting antibiotic resistance genes (ARGs) prevalence data by Zhang *et al.* [80].
- Interpretation: LLM-based chatbots aid in the interpretation of Ribo-seq data by providing contextually relevant insights and annotations [81]. They can generate summaries, visualizations [82], and interactive reports to help researchers interpret complex patterns and biological implications from Ribo-seq datasets [83].

In summary, the integration of LLMs into Ribo-seq data analysis holds immense potential for advancing our understanding of translational regulation and gene expression dynamics [84, 85]. LLM-based chatbots offer scalable and intelligent solutions for data preprocessing, feature extraction, pattern recognition, and interpretation, thereby accelerating scientific discovery in bioinformatics and biomedical science [86].

Case studies and applications

Since the introduction of the ribosome profiling technique in 2009, its popularity has greatly increased [9]. In recent years, advancements in ribosome profiling techniques have led to the development of lower input methods [35, 87, 88]. Consequently, an increasing number of tools tailored for data analysis (Table 1) have emerged to accommodate these advancements.

Exploring potential solutions: case study of chatbot-assisted Ribo-seq data analysis

RiboChat, published in 2022, is the first online interactive platform for direct Ribo-seq data analysis implemented via a chat conversation [36]. Compared with existing tools, the primary added value of RiboChat is that the user interface with human language modalities enhances human–computer interaction, thereby greatly reducing the barrier to entry for Ribo-seq data analysis, particularly for users with no or limited programming or bioinformatics skills. While this software, as a chatbot, greatly facilitates researchers without programming skills by allowing them to perform top–down analyses through natural language, select from a predefined set of similar tools, and even modify certain key parameters of the software, the user experience can feel somewhat rigid and mechanical. This is primarily because RiboChat is not based on LLMs, which limits its ability to provide a more dynamic and flexible interaction.

Therefore, I explored using chatbots like GPT-4o to handle certain tasks in the ribosome profiling data analysis workflow. The results demonstrated that GPT-4o can provide a standard ribosome profiling data analysis pipeline and can adjust the software used in the protocol based on user needs (Supplementary Fig. 1). However, GPT-4o currently appears unable to directly process ribosome profiling fastq and bam files, lacking the ability to perform the specialized top–down data analysis provided by RiboChat. Additionally, I explored GPT-4o's ability to interpret line graphs from metagene analysis in published ribosome profiling data, and it demonstrated a fairly accurate understanding of the graphical results (Supplementary Fig. 2).

Fortunately, the development of certain software has revealed the potential intersection of LLM-based chatbots and ribosome profiling data analysis. First, Hou *et al.* demonstrate that the LLM GPT-4 can accurately annotate cell types using marker gene information in single-cell RNA sequencing analysis [105]. They developed an R software package GPTCelltype for GPT-4's

Table 1. Overview of tools tailored for ribosome profiling data analysis.

Name	Year	Introduction	URL
Ribodeblur [89]	2018	A Python package for estimating A-site locations of ribosomes from Ribo-seq	https://github.com/KingsfordGroup/ribodeblur
RiboCode [90]	2018	Detect translated open reading frames (ORFs) using ribosome-profiling data	https://github.com/xryanglab/RiboCode
RiboChat [36]	2022	A chat-style web interface for analysis and annotation of ribosome profiling data	https://db.cngb.org/ribobench/chat.html
RiboMiner [29]	2020	A Python toolset for mining multidimensional features of the translatome with ribosome profiling data	https://github.com/xryanglab/RiboMiner
RiboDiff [91]	2017	Tool to detect changes in translational efficiency based on ribosome footprinting data	https://github.com/ratschlab/RiboDiff
RiboDoc [92]	2021	A Docker-based package for ribosome profiling analysis	https://github.com/equipeGST/RiboDoc
Ribo-DT [93]	2022	A snakemake pipeline to infer single-codon and codon-pair dwell times	https://github.com/cgob/codonDT_snakemake
RiboGalaxy [94]	2023	A Galaxy-based Web Platform for Ribosome Profiling Data Processing	https://ribogalaxy.genomicsdatascience.ie/
RiboHMM [95]	2016	A mixture of hidden Markov models to infer translated sequences using ribosome footprint profiling	https://github.com/rajanil/riboHMM
Ribomap [75]	2016	A package that generates isoform-level ribosome profiles from ribosome profiling data	https://www.cs.cmu.edu/~ckingsf/software/ribomap/
RiboNT [96]	2021	A noise-tolerant predictor of open reading frames from ribosome-protected footprints	https://github.com/songbo446/RiboNT/
RibORF [97]	2018	Identifying genome-wide translated open reading frames using ribosome profiling	https://github.com/zhejilab/RibORF
RiboTaper [98]	2016	A new analysis pipeline for ribosome profiling experiments	https://ohlerlab.mdc-berlin.de/software/RiboTaper_126/
RiboToolkit [99]	2020	A convenient, freely available, web-based service to centralize Ribo-seq data analyses	http://rnabioinfor.tch.harvard.edu/RiboToolkit
RiboVIEW [100]	2020	Visualization, quality, and statistics for ribosome profiling	https://github.com/carinelegrand/RiboVIEW
riboviz [101]	2022	Analysis and visualization of ribosome profiling data	https://github.com/riboviz/RiboViz
riboWaltz [102]	2018	R package for calculation of optimal P-site offsets, diagnostic analysis, and visual inspection of ribosome profiling data	https://github.com/LabTranslationalArchitectomics/RiboWaltz
RPiso [103]	2021	A tool for analyzing and visualizing Ribo-seq data at the isoform level	http://cosbi7.ee.ncku.edu.tw/RPiso/
Xtail [28]	2016	Genome-wide assessment of differential translations with ribosome profiling data	https://github.com/xryanglab/xtail
RiboTools [104]	2015	A Galaxy toolbox for qualitative ribosome profiling analysis	https://testtoolshed.g2.bx.psu.edu/view/rlegendre/ribotools

automated cell type annotation, and it can considerably reduce the effort and expertise required for cell type annotation. This may suggest that implementing GPT for ribosome profiling data analysis might only require the development of a dedicated application programming interface (API), which was pretrained and fine-tuned specifically on ribosome profiling data. Second, Yang *et al*. developed a platform called ShennongAlpha for acquiring and translating knowledge about natural medicinal materials [106]. This platform includes an LLM-based chatbot named ShennongChat, which can provide information on natural medicinal materials (NMMs) through daily conversational exchanges in both Chinese and English. Users can access more detailed information via a provided ShennongSearch link, which includes the NMM ID, abstract, systematic nomenclature, and Chinese Pharmacopoeia associated with each natural medicinal material. Thanks to these advantages, users can obtain more accurate, detailed, and comprehensive information on target natural medicinal materials solely through natural language input. This approach successfully addresses the limitations of ChatGPT or Wikipedia in providing sufficient accuracy and completeness for explaining natural medicinal materials. The ShennongSearch link in ShennongChat suggests that linking mechanisms could be used to provide users with a more comprehensive explanation of related content, similar to the concept illustrated in Fig. 4. Third, a range of LLMs have been developed to address issues in specific research domains: EpiGePT by Gao *et al*. [107], scGPT by Cui *et al*. [108], GenePT by Chen *et al*. [109], CellPLM by Wen *et al*. [110], scMulan by Bian *et al*. [111], scBERT by Yang *et al*. [112], and DNABERT by Ji *et al*. [113], as well as further LLMs developed on this foundation by Zhang *et al*. [80] and NVLM by Dai *et al*. [66], which address challenges within the fields of epigenomics, scRNA-seq, genomics, and vision-language tasks, respectively.

In summary, with the concerted efforts of researchers in related fields, I hope that LLM-based tools for ribosome profiling data analysis will be developed in the future.

Potential contributions of chatbots in Ribo-seq data analysis

Through integrative analysis of Ribo-seq data with other omics datasets [38], the chatbot uncovered regulatory mechanisms controlling translational efficiency, mRNA stability, and post-translational modifications. By correlating translational activity with gene expression and protein abundance, the chatbot identified key regulatory factors and pathways involved in gene expression regulation [114].

In addition, the LLM-based chatbot can accurately predict translation initiation sites from Ribo-seq data by analyzing sequence features and ribosome occupancy profiles [115]. By identifying translation start codons and associated regulatory elements, the chatbot provided insights into the mechanisms of translation initiation and ribosome recruitment [116].

Moreover, the chatbot can be used to analyze temporal changes in ribosome occupancy and translation efficiency across different experimental conditions or cellular states [75]. By detecting shifts in translation dynamics and differential expression of translational regulators [35, 84], the chatbots elucidated the regulatory networks governing protein synthesis and cellular response to environmental stimuli.

With the development of LLMs, the chatbots show great potential contributions in Ribo-seq data analysis, mainly in these aspects including discovery of regulatory mechanisms, identification of translation initiation sites, and characterization of translation dynamics.

Comparison with traditional methods and existing tools

Compared with traditional methods and existing tools, chatbots exhibit several advantages.

- User-friendliness: Chatbots offered intuitive interfaces and user-friendly functionalities, making them accessible to researchers with diverse backgrounds and levels of expertise [117, 118]. Their interactive features, such as natural language interaction and visualizations, enhanced user experience and facilitated collaboration among interdisciplinary research teams.
- Efficiency: Chatbots enabled Ribo-seq data analysis workflows by automating repetitive tasks and leveraging parallel processing capabilities. Researchers experienced significant time savings and increased productivity [119, 120] allowing them to focus on higher-level analysis and interpretation tasks.

In summary, the mentioned software exemplifies the unique value of LLM-based chatbots applied in ribosome profiling data analysis. They demonstrate characteristics of user-friendliness and efficiency while maintaining the accuracy of data analysis. It is believed that in the future, LLM-based chatbots will be applied in ribosome profiling data analysis, leveraging their powerful NLP capabilities to address the current limitations of the software. They can be utilized for discovering regulatory mechanisms, identifying translation initiation sites, and characterizing translation dynamics.

Challenges and future directions

Challenges associated with large language model–based approaches in Ribo-seq data analysis

Though LLM-based chatbots show great advantages in potential Ribo-seq data analysis, limitations and challenges associated with LLM-based approaches are probably inevitable. Listed but not limited to model bias are data privacy problem and lower interpretability. By extension, LLMs may exhibit biases inherited from the training data [121] which can influence their performance and the accuracy of predictions. Biases in LLMs trained on general text corpora may affect their applicability to specific domains like bioinformatics and biomedical science [122–124], leading to suboptimal results in Ribo-seq data analysis.

Ribo-seq data often contain sensitive information about gene expression and translational regulation, raising concerns about data privacy and security [125]. Integrating LLMs into Ribo-seq data analysis workflows requires careful consideration of privacy regulations and ethical guidelines to protect sensitive unpublished ribosome profiling data [126, 127].

Besides, the black-box nature of LLMs poses challenges for interpreting their predictions and understanding the underlying mechanisms driving Ribo-seq data analysis [128–130]. Enhancing the interpretability of LLM-based models is essential for gaining insights into translational regulation and validating computational findings through experimental validation.

Strategies for mitigating challenges

To diminish the probable limitations and challenges of LLM-based chatbots, we can adopt suitable strategies to mitigate challenges. Topping the list is fine-tuned LLMs on domain-specific datasets [131], including the Ribo-seq dataset and omics literature, which can mitigate model bias and improve performance in ribosome profiling data analysis tasks [132]. Designing machine-understandable prompts like Fig. 4 and incorporating them into the software's help manuals is crucial for ensuring that LLMs accurately comprehend user intent and provide appropriate data analysis workflows. Additionally, to improve the accuracy of interpreting Ribo-seq result files and the ability to extract meaningful biological insights from them, the model's capacity to learn from translationally relevant biomedical literature should be reinforced during training, with specific methods such as those employed in BioGPT [70]. When misunderstandings of user intent or the provision of unsuitable solutions arise during training, it is essential to promptly relay this feedback to the LLM or its developers until the LLM consistently understands user intent and delivers suitable solutions.

Second, implementing privacy-preserving methods [133], such as federated learning and differential privacy [134, 135], can address concerns about data privacy while leveraging the collective intelligence of distributed datasets for model training. By decentralizing data storage and training, privacy-preserving approaches enable collaborative analysis of Ribo-seq data without compromising data confidentiality [136].

Last but not least, integrating interpretability techniques into LLM-based chatbots enables users to understand the rationale behind model predictions and identify influential features in Ribo-seq data analysis [137]. Methods such as attention mechanisms, feature importance ranking, and model visualization facilitate the interpretation of LLM-based predictions and enhance trust in computational findings [138].

Future directions

Combined with all the advantages and disadvantages of LLM-based chatbots in the applications of Ribo-seq data analysis, we put forward some emerging trends and future directions in the integration of LLMs and Ribo-seq data analysis. Multimodal learning has the potential to be an emerging trend. Integrating multimodal data sources, such as Ribo-seq data, transcriptomics, and epigenomics data, with textual information can enrich LLM-based models and enable comprehensive analysis of gene expression regulation [139]. Multimodal learning approaches enable synergistic analysis of complementary data types and facilitate a holistic understanding of biological processes [140].

Interactive chatbot interfaces are essential in the future of LLM-based chatbots. Developing interactive chatbot interfaces with natural language interaction capabilities and real-time feedback mechanisms enhances user engagement and collaboration in Ribo-seq data analysis [141]. Interactive chatbots enable researchers to interactively explore data [142], refine analysis parameters, and validate computational findings through iterative experimentation.

Given the power of LLM-based chatbots, ethical considerations should always address users' attention. Addressing ethical considerations and societal implications of LLM-based Ribo-seq data analysis is essential for responsible research conduct and equitable knowledge dissemination [143]. Engaging stakeholders, including researchers, policymakers, and community representatives, in ethical discussions and decision-making processes ensures that LLM-based technologies are deployed in a socially responsible manner.

In summary, addressing challenges associated with LLM-based approaches in Ribo-seq data analysis requires concerted efforts to improve model robustness, ensure data privacy, and enhance interpretability. Future directions in the integration of LLMs and Ribo-seq data analysis involve domain-specific training, privacy-preserving methods, and emerging trends such as multimodal learning and interactive chatbot interfaces, paving the way for innovative applications in bioinformatics and biomedical science.

Conclusion

In this review, we have explored the potential of leveraging LLM-based chatbots to address challenges in ribosome profiling data analysis. Ribosome profiling, a powerful technique for studying translation dynamics at the transcriptome-wide level, presents formidable obstacles in data analysis, particularly for researchers without a background in bioinformatics. The complexity of ribosome profiling datasets, coupled with the diverse needs of researchers, necessitates user-friendly and efficient tools for data analysis.

While there are currently no examples of LLM–based chatbots applied in ribosome profiling data analysis, existing software solutions exemplify the unique value of chatbots in this field. These tools demonstrate characteristics of user-friendliness and efficiency while maintaining the accuracy of data analysis. It is anticipated that in the future, LLM-based chatbots will be deployed in ribosome profiling data analysis, leveraging their powerful NLP capabilities to overcome current limitations of software. They have the potential to aid in discovering regulatory mechanisms, identifying translation initiation sites, and characterizing translation dynamics, thereby advancing our understanding of translational regulation and gene expression dynamics.

Despite the promising prospects, challenges remain in integrating large language models into ribosome profiling data analysis. Model bias, data privacy concerns, and interpretability issues pose significant hurdles that need to be addressed. Strategies such as domain-specific training, privacy-preserving methods, and interpretability techniques can enhance the performance and robustness of LLM-based chatbots in ribosome profiling data analysis.

Looking ahead, emerging trends such as multimodal learning and interactive chatbot interfaces offer exciting opportunities for innovation in bioinformatics and biomedical science [78, 144]. By embracing these advancements and addressing ethical considerations, we can harness the full potential of LLM-based chatbots to accelerate scientific discovery and unlock new insights into translation regulation and gene expression dynamics [145].

Potential collaborative research opportunities between bioinformaticians and AI researchers could greatly enhance the development and application of LLM-based tools in ribosome profiling data analysis. For instance, bioinformaticians can provide domain-specific knowledge and curate high-quality datasets, while AI researchers can focus on developing sophisticated models and algorithms tailored to bioinformatics challenges. Additionally, collaborative efforts could focus on creating benchmark datasets and standardized evaluation metrics, ensuring that LLM-based tools are reliable and effective across various biological contexts. By working together, bioinformaticians and AI researchers can push the boundaries of what is possible, driving forward innovations that benefit the entire scientific community.

Key Points

- Ribo-seq data analysis is complex, especially for researchers lacking bioinformatics skills, requiring proficiency in multiple programming languages and software environments, large language model (LLM)–based chatbots have great potential to overcome this difficulty.
- Beyond data analysis, chatbots have broad applications in literature mining, hypothesis generation, and collaboration facilitation.
- LLM-based methods have already been successfully applied to address challenges in single-cell transcriptomics, epigenetics, natural medicine, and genomics.
- The integration of LLMs and Ribo-seq analysis may advance translational regulation and gene expression research, but challenges like model bias, data privacy, and AI transparency need to be addressed.
- We hope to encourage scientists in translatomics and the LLM domain to collaborate in driving forward innovations that benefit the entire scientific community.

Funding

This study is supported by the National Natural Science Foundation of China (32100620).

Data availability

The public data of Fig. 2 are obtained from Gene Expression Omnibus (GEO) under the accession code GSE131074. The URLs of tools tailored for ribosome profiling data analysis are shown in Table 1.

Supplementary data

Supplementary data are available at *Briefings in Bioinformatics* online.

References

1. Kraus AJ, Cosentino RO. Ribosome profiling in Trypanosomatids. *Methods Mol Biol* 2019;**1971**:109–22.
2. Brar GA, Weissman JS. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat Rev Mol Cell Biol* 2015;**16**:651–64.
3. Xiao Z, Li W, Moon H. *et al.* Generative artificial intelligence GPT-4 accelerates knowledge mining and machine learning for synthetic biology. *ACS Synth Biol* 2023;**12**:2973–82.

4. Khan MS, Umer H. ChatGPT in finance: Applications, challenges, and solutions. *Heliyon* 2024;**10**:e24890. https://doi.org/10.1016/j.heliyon.2024.e24890.
5. Lee H. The rise of ChatGPT: Exploring its potential in medical education. *Anat Sci Educ* 2024;**17**:926–31. https://doi.org/10.1002/ase.2270.
6. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. *J Med Internet Res* 2023;**25**:e48568. https://doi.org/10.2196/48568.
7. Cascella M, Montomoli J, Bellini V. *et al.* Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios. *J Med Syst* 2023;**47**:33.
8. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: A conversation with ChatGPT and a call for papers. *JMIR Med Educ* 2023;**9**:e46885. https://doi.org/10.2196/46885.
9. Kiniry SJ, Michel AM, Baranov PV. Computational methods for ribosome profiling data analysis. *WIREs RNA* 2020;**11**:e1577. https://doi.org/10.1002/wrna.1577.
10. Blank HM, Perez R, He C. *et al.* Translational control of lipogenic enzymes in the cell cycle of synchronous, growing yeast cells. *EMBO J* 2017;**36**:487–502.
11. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: An overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell* 2023;**6**:1169595. https://doi.org/10.3389/frai.2023.1169595.
12. Zhao LP, Hu JH, Hu D. *et al.* Hyperprogression, a challenge of PD-1/PD-L1 inhibitors treatments: Potential mechanisms and coping strategies. *Biomed Pharmacother* 2022;**150**:112949. https://doi.org/10.1016/j.biopha.2022.112949.
13. Ingolia NT, Ghaemmaghami S, Newman JRS. *et al.* Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009;**324**:218–23.
14. Ingolia NT, Brar GA, Rouskin S. *et al.* The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 2012;**7**:1534–50.
15. Bartholomaus A, Del Campo C, Ignatova Z. Mapping the non-standardized biases of ribosome profiling. *Biol Chem* 2016;**397**:23–35.
16. Michel AM, Baranov PV. Ribosome profiling: A hi-Def monitor for protein synthesis at the genome-wide scale. *Wiley Interdiscip Rev RNA* 2013;**4**:473–90.
17. Fujita T, Kurihara Y, Iwasaki S. The plant Translatome surveyed by ribosome profiling. *Plant Cell Physiol* 2019;**60**:1917–26.
18. Gobet C, Naef F. Ribosome profiling and dynamic regulation of translation in mammals. *Curr Opin Genet Dev* 2017;**43**:120–7.
19. McGlincy NJ, Ingolia NT. Transcriptome-wide measurement of translation by ribosome profiling. *Methods* 2017;**126**:112–29.
20. Sawyer EB, Cortes T. Ribosome profiling enhances understanding of mycobacterial translation. *Front Microbiol* 2022;**13**:976550. https://doi.org/10.3389/fmicb.2022.976550.
21. Li F, Fang J, Yu Y. *et al.* Reanalysis of ribosome profiling datasets reveals a function of rocaglamide a in perturbing the dynamics of translation elongation via eIF4A. *Nat Commun* 2023;**14**:553.
22. Wang Y, Zhang H, Lu J. Recent advances in ribosome profiling for deciphering translational regulation. *Methods* 2020;**176**:46–54.
23. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 2011;**17**:3.
24. Langmead B, Trapnell C, Pop M. *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;**10**:R25.
25. Flanagan K, Li W, Greenblatt EJ., Dao Duc K End-to-end pipeline for differential analysis of pausing in ribosome profiling data. *STAR Protoc* 2022;**3**:1016055. https://doi.org/10.1016/j.xpro.2022.101605.
26. Dobin A, Davis CA, Schlesinger F. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21.
27. Bartholomaus A, Ignatova Z. Codon resolution analysis of ribosome profiling data. *Methods Mol Biol* 2021;**2252**:251–68.
28. Xiao Z, Zou Q, Liu Y. *et al.* Genome-wide assessment of differential translations with ribosome profiling data. *Nat Commun* 2016;**7**:11194. https://doi.org/10.1038/ncomms11194.
29. Li F, Xing X, Xiao Z. *et al.* RiboMiner: A toolset for mining multi-dimensional features of the translatome with ribosome profiling data. *BMC Bioinformatics* 2020;**21**:340.
30. Tian G, Hu C, Yun Y. *et al.* Dual roles of HSP70 chaperone HSPA1 in quality control of nascent and newly synthesized proteins. *EMBO J* 2021;**40**:e106183. https://doi.org/10.15252/embj.2020106183.
31. Szavits-Nossan J, Ciandrini L. Inferring efficiency of translation initiation and elongation from ribosome profiling. *Nucleic Acids Res* 2020;**48**:9478–90.
32. Erdmann-Pham DD, Son W, Dao Duc K. *et al.* EGGTART: A tool to visualize the dynamics of biophysical transport under the inhomogeneous l-TASEP. *Biophys J* 2021;**120**:1309–13.
33. Sabi R, Tuller T. Novel insights into gene expression regulation during meiosis revealed by translation elongation dynamics. *npj Syst Biol Appl* 2019;**5**:12. https://doi.org/10.1038/s41540-019-0089-0.
34. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 2011;**147**:789–802.
35. Xiong Z, Xu K, Lin Z. *et al.* Ultrasensitive Ribo-seq reveals translational landscapes during mammalian oocyte-to-embryo transition and pre-implantation development. *Nat Cell Biol* 2022;**24**:968–80.
36. Xie M, Yang L, Chen G. *et al.* RiboChat: A chat-style web interface for analysis and annotation of ribosome profiling data. *Brief Bioinform* 2022;**23**:bbab559. https://doi.org/10.1093/bib/bbab559.
37. Lin Y, Li F, Huang L. *et al.* eIF3 associates with 80S ribosomes to promote translation elongation, mitochondrial homeostasis, and muscle health. *Mol Cell* 2020;**79**:e577.
38. Li J, Chen H, Wang Y. *et al.* Next-generation analytics for omics data. *Cancer Cell* 2021;**39**:3–6.
39. Bibault J-E, Chaix B, Nectoux P. *et al.* Healthcare ex Machina: Are conversational agents ready for prime time in oncology? *Clin Transl Radiat Oncol* 2019;**16**:55–9.
40. Balderas A, García-Mena RF, Huerta M. *et al.* Chatbot for communicating with university students in emergency situation. *Heliyon* 2023;**9**:e19517.
41. Sin J. An AI chatbot for talking therapy referrals. *Nat Med* 2024;**30**:350–1.
42. Chen X, Wang X, Zhang K. *et al.* Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal* 2022;**79**:102444.
43. Castagna F, Garton A, McBurney P. *et al.* EQRbot: A chatbot delivering EQR argument-based explanations, Frontiers. *Artif Intell* 2023;**6**:1045614.
44. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus* 2023;**15**:e35179.
45. Anghelescu A, Firan FC, Onose G. *et al.* PRISMA systematic literature review, including with meta-analysis vs. Chatbot/GPT (AI) regarding current scientific data on the main effects of the

calf blood Deproteinized Hemoderivative medicine (Actovegin) in ischemic stroke. *Biomedicine* 2023;**11**:1623.
46. Shue E, Liu L, Li B. *et al.* Empowering beginners in bioinformatics with ChatGPT. *Quant Biol* 2023;**11**:105–8.
47. Naddaf M. ChatGPT generates fake data set to support scientific hypothesis. *Nature* 2023;**623**:895–6.
48. Temsah O, Khan SA, Chaiah Y. *et al.* Overview of early ChatGPT's presence in medical literature: Insights from a hybrid literature review by ChatGPT and human experts. *Cureus* 2023;**15**:e37281. https://doi.org/10.7759/cureus.37281.
49. Cheng X, Zhang X, Yang B. *et al.* An investigation on trust in AI-enabled collaboration: Application of AI-driven chatbot in accommodation-based sharing economy. *Electron Commer Res Appl* 2022;**54**:101164.
50. Anan T, Kajiki S, Oka H. *et al.* Effects of an artificial intelligence-assisted health program on workers with neck-/shoulder pain/stiffness and low Back pain: Randomized controlled trial. *JMIR Mhealth Uhealth* 2021;**9**:e27535. https://doi.org/10.2196/27535.
51. Roca S, Sancho J, García J. *et al.* Microservice chatbot architecture for chronic patient support. *J Biomed Inform* 2020;**102**:103305.
52. Nazareth S, Hayward L, Simmons E. *et al.* Hereditary cancer risk using a genetic Chatbot before routine care visits. *Obstet Gynecol* 2021;**138**:860–70.
53. Hernandez JPT. Network diffusion and technology acceptance of a nurse Chatbot for chronic disease self-management support: A theoretical perspective. *J Med Investig* 2019;**66**:24–30.
54. Dosovitsky G, Pineda BS, Jacobson NC. *et al.* Artificial intelligence Chatbot for depression: Descriptive study of usage. *JMIR Form Res* 2020;**4**:e17065.
55. Goodman RS, Patrinely JR, Stone CA. *et al.* Accuracy and reliability of Chatbot responses to physician questions. *JAMA Netw Open* 2023;**6**:e2336483.
56. Milne-Ives M, de Cock C, Lim E. *et al.* The effectiveness of artificial intelligence conversational agents in health care: Systematic review. *J Med Internet Res* 2020;**22**:e20346.
57. Rambaud K, van Woerden S, Palumbo L. *et al.* Building a Chatbot in a pandemic. *J Med Internet Res* 2023;**25.**:e42960.
58. Boscardin CK, Gin B, Golde PB. *et al.* ChatGPT and generative artificial intelligence for medical education: Potential impact and opportunity. *Acad Med* 2024;**99**:22–7.
59. King MR. The future of AI in medicine: A perspective from a Chatbot. *Ann Biomed Eng* 2022;**51**:291–5.
60. Lee P, Bubeck S, Petro J. *et al.* Benefits, limits, and risks of GPT-4 as an AI Chatbot for medicine. *N Engl J Med* 2023;**388**:1233–9.
61. Thirunavukarasu AJ, Ting DSJ, Elangovan K. *et al.* Large language models in medicine. *Nat Med* 2023;**29**:1930–40.
62. Li Y, Li Z, Zhang K. *et al.* ChatDoctor: A medical chat model fine-tuned on a large language model meta-AI (LLaMA) using medical domain knowledge. *Cureus* 2023;**15**:e40895. https://doi.org/10.7759/cureus.40895.
63. Wilhelm TI, Roos J, Kaczmarczyk R. Large language models for therapy recommendations across 3 clinical specialties: Comparative study. *J Med Internet Res* 2023;**25**:e49324. https://doi.org/10.2196/49324.
64. Hirosawa T, Harada Y, Yokose M. *et al.* Diagnostic accuracy of differential-diagnosis lists generated by generative Pretrained transformer 3 Chatbot for clinical vignettes with common chief complaints: A pilot study. *Int J Environ Res Public Health* 2023;**20**:3378.
65. Mondal H, Dash I, Mondal S. *et al.* ChatGPT in answering queries related to lifestyle-related diseases and disorders. *Cureus* 2023;**15**:e48296. https://doi.org/10.7759/cureus.48296.
66. Dai W, Lee N, Wang B. *et al.* NVLM: Open Frontier-Class Multimodal LLMs. arXiv e-prints 2024; arXiv:2409.11402.
67. Dergaa I, Chamari K, Zmijewski P. *et al.* From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biol Sport* 2023;**40**:615–22.
68. Zambrano Serrano CA, Carvajal OA. Diagnosis and hormonal treatment of male infertility. *Actas Urol Esp (Engl Ed)* 2020;**44**:321–7.
69. Lee J, Yoon W, Kim S. *et al.* BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020;**36**:1234–40.
70. Luo R, Sun L, Xia Y. *et al.* BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Brief Bioinform* 2022;**23**:bbac409.
71. Cheng SL, Tsai SJ, Bai YM. *et al.* Comparisons of quality, correctness, and similarity between ChatGPT-generated and human-written abstracts for basic research: Cross-sectional study. *J Med Internet Res* 2023;**25**:e51229. https://doi.org/10.2196/51229.
72. Garg RK, Urs VL, Agrawal AA. *et al.* Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review. *Health Promot Perspect* 2023;**13**:183–91.
73. Rezigalla AA. AI in medical education: Uses of AI in construction type a MCQs. *BMC Med Educ* 2024;**24**:247.
74. Ayub I, Hamann D, Hamann CR. *et al.* Exploring the potential and limitations of chat generative pre-trained transformer (ChatGPT) in generating board-style dermatology questions: A qualitative analysis. *Cureus* 2023;**15**:e43717. https://doi.org/10.7759/cureus.43717.
75. Wang H, McManus J, Kingsford C. Isoform-level ribosome occupancy estimation guided by transcript abundance with Ribomap. *Bioinformatics* 2016;**32**:1880–2.
76. Ngo DH, Koopman B. From free-text drug labels to structured medication terminology with BERT and GPT. *AMIA Annu Symp Proc* 2024;**2023**:540-49.
77. Wang X, Gao C, Han P. *et al.* PETrans: De novo drug design with protein-specific encoding based on transfer learning. *Int J Mol Sci* 2023;**24**:1146.
78. Zhang K, Zemke NR, Armand EJ. *et al.* A fast, scalable and versatile tool for analysis of single-cell omics data. *Nat Methods* 2024;**21**:217–27.
79. Ali S, Chourasia P, Patterson M. When protein structure embedding meets large language models. *Genes (Basel)* 2023;**15**:25.
80. Zhang J, Zhao L, Wang W. *et al.* Large language model for horizontal transfer of resistance gene: From resistance gene prevalence detection to plasmid conjugation rate evaluation. *Sci Total Environ* 2024;**931**:172466. https://doi.org/10.1016/j.scitotenv.2024.172466.
81. Benary M, Wang XD, Schmidt M. *et al.* Leveraging large language models for decision support in personalized oncology. *JAMA Netw Open* 2023;**6**:e2343689.
82. Sorin V, Glicksberg BS, Artsi Y. *et al.* Utilizing large language models in breast cancer management: Systematic review. *J Cancer Res Clin Oncol* 2024;**150**:140.
83. Shah A, Wahood S, Guermazi D. *et al.* Skin and syntax: Large language models in Dermatopathology. *Dermatopathology (Basel)* 2024;**11**:101–11.

84. Zou Z, Zhang C, Wang Q. *et al.* Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. *Science* 2022;**378**:abo7923. https://doi.org/10.1126/science.abo7923.
85. Legrand C, Duc KD, Tuorto F. Analysis of ribosome profiling data. *Methods Mol Biol* 2022;**2428**:133–56.
86. Hui Z, Wen H, Zhu J. *et al.* Discovery of plant-derived anti-tumor natural products: Potential leads for anti-tumor drug discovery. *Bioorg Chem* 2024;**142**:106957. https://doi.org/10.1016/j.bioorg.2023.106957.
87. Clamer M, Tebaldi T, Lauria F. *et al.* Active ribosome profiling with RiboLace. *Cell Rep* 2018;**25**:1097–1108.e1095.
88. Zhang C, Wang M, Li Y. *et al.* Profiling and functional characterization of maternal mRNA translation during mouse maternal-to-zygotic transition. *Sci Adv* 2022;**8**:eabj3967. https://doi.org/10.1126/sciadv.abj3967.
89. Wang H, Kingsford C, McManus CJ. Using the Ribodeblur pipeline to recover A-sites from yeast ribosome profiling data. *Methods* 2018;**137**:67–70.
90. Xiao Z, Huang R, Xing X. *et al.* De novo annotation and characterization of the translatome with ribosome profiling data. *Nucleic Acids Res* 2018;**46**:e61.
91. Zhong Y, Karaletsos T, Drewe P. *et al.* RiboDiff: Detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics* 2017;**33**:139–41.
92. François P, Arbes H, Demais S. *et al.* RiboDoc: A Docker-based package for ribosome profiling analysis. *Comput Struct Biotechnol J* 2021;**19**:2851–60.
93. Gobet C, Naef F. Ribo-DT: An automated pipeline for inferring codon dwell times from ribosome profiling data. *Methods* 2022;**203**:10–6.
94. Fedorova AD, Tierney JAS, Michel AM. *et al.* RiboGalaxy: A galaxy-based web platform for ribosome profiling data processing - 2023 update. *J Mol Biol* 2023;**435**:168043. https://doi.org/10.1016/j.jmb.2023.168043.
95. Raj A, Wang SH, Shim H. *et al.* Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling. *elife* 2016;**5**:e13328.
96. Song B, Jiang M, Gao L. RiboNT: A noise-tolerant predictor of open reading frames from ribosome-protected footprints. *Life (Basel)* 2021;**11**:701.
97. Ji Z. RibORF: Identifying genome-wide translated open reading frames using ribosome profiling. *Curr Protoc Mol Biol* 2018;**124**:e67. https://doi.org/10.1002/cpmb.67.
98. Calviello L, Mukherjee N, Wyler E. *et al.* Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods* 2016;**13**:165–70.
99. Liu Q, Shvarts T, Sliz P. *et al.* RiboToolkit: An integrated platform for analysis and annotation of ribosome profiling data to decode mRNA translation at codon resolution. *Nucleic Acids Res* 2020;**48**:W218–29.
100. Legrand C, Tuorto F. RiboVIEW: A computational framework for visualization, quality control and statistical analysis of ribosome profiling data. *Nucleic Acids Res* 2020;**48**:e7.
101. Cope AL, Anderson F, Favate J. *et al.* Riboviz 2: A flexible and robust ribosome profiling data analysis and visualization workflow. *Bioinformatics* 2022;**38**:2358–60.
102. Lauria F, Tebaldi T, Bernabò P. *et al.* riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput Biol* 2018;**14**:e1006169. https://doi.org/10.1371/journal.pcbi.1006169.
103. Wu WS, Tsao YH, Shiue SC. *et al.* A tool for analyzing and visualizing ribo-seq data at the isoform level. *BMC Bioinformatics* 2021;**22**:271.
104. Legendre R, Baudin-Baillieu A, Hatin I. *et al.* RiboTools: A galaxy toolbox for qualitative ribosome profiling analysis. *Bioinformatics* 2015;**31**:2586–8.
105. Hou W, Ji Z. Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. *Nat Methods* 2024;**21**:1462–5.
106. Yang Z, Yin Y, Kong C. *et al.* ShennongAlpha: An AI-driven sharing and collaboration platform for intelligent curation, acquisition, and translation of natural medicinal material knowledge. *arXiv e-prints* 2023;arXiv:2401.00020.
107. Gao Z, Liu Q, Zeng W. *et al.* EpiGePT: A Pretrained transformer model for epigenomics. bioRxiv [Preprint]. 2024 Feb 3:2023.07.15.549134.
108. Cui H, Wang C, Maan H. *et al.* scGPT: Toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods* 2024;**21**:1470–80.
109. Chen Y, Zou J. GenePT: A simple but effective foundation model for genes and cells built from ChatGPT. bioRxiv [Preprint]. 2024 Mar 5:2023.10.16.562533. https://doi.org/2023.2010.2016.562533.
110. Wen H, Tang W, Dai X. *et al.* CellPLM: Pre-training of cell language model beyond single cells. bioRxiv [Preprint]. 2023 :2023.10.03.560734. https://doi.org/2003.560734.
111. Bian H, Chen Y, Dong X. *et al.* scMulan: A multitask generative pre-trained language model for single-cell analysis. In: *Research in Computational Molecular Biology*. Cham, 2024, p. 479–82. https://doi.org/10.1007/978-1-0716-3989-4_57.
112. Yang F, Wang W, Wang F. *et al.* scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat Mach Intell* 2022;**4**:852–66.
113. Ji Y, Zhou Z, Liu H. *et al.* DNABERT: Pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. *Bioinformatics* 2021;**37**:2112–20.
114. Wang J, Cheng Z, Yao Q. *et al.* Bioinformatics and biomedical informatics with ChatGPT: Year one review. *Quant Biol* 2024;**12**:345–59.
115. Gleason AC, Ghadge G, Chen J. *et al.* Machine learning predicts translation initiation sites in neurologic diseases with nucleotide repeat expansions. *PLoS One* 2022;**17**:e0256411. https://doi.org/10.1371/journal.pone.0256411.
116. Andreev DE, Loughran G, Fedorova AD. *et al.* Non-AUG translation initiation in mammals. *Genome Biol* 2022;**23**:111.
117. Görtz M, Baumgärtner K, Schmid T. *et al.* An artificial intelligence-based chatbot for prostate cancer education: Design and patient evaluation study, digital. *Health* 2023;**9**:20552076231173304.
118. Garcia, Valencia OA, Suppadungsuk S, Thongprayoon C. *et al.* Ethical implications of Chatbot utilization in nephrology. *J Pers Med* 2023;**13**:1363.
119. Ahmed A, Hassan A, Aziz S. *et al.* Chatbot features for anxiety and depression: A scoping review. *Health Informatics J* 2023;**29**:14604582221146719. https://doi.org/10.1177/14604582221146719.
120. Kim S, Lee CK, Kim SS. Large language models: A guide for radiologists. *Korean J Radiol* 2024;**25**:126–33.
121. Nashwan AJ, Jaradat JH. Streamlining systematic reviews: Harnessing large language models for quality assessment and risk-of-bias evaluation. *Cureus* 2023;**15**:e43023. https://doi.org/10.7759/cureus.43023.

122. Fang X, Che S, Mao M. *et al.* Bias of AI-generated content: An examination of news produced by large language models. *Sci Rep* 2024;**14**:5224.
123. Goh E, Bunning B, Khoong E. *et al.* ChatGPT influence on medical decision-making, bias, and equity: A randomized study of clinicians evaluating clinical vignettes. medRxiv [Preprint]. 2023 Nov 27:2023.11.24.23298844.
124. Liu S, Li Q, Chen K. *et al.* The emerging molecular mechanism of m(6)a modulators in tumorigenesis and cancer progression. *Biomed Pharmacother* 2020;**127**:110098. https://doi.org/10.1016/j.biopha.2020.110098.
125. Liu Z, Zhang L, Wu Z. *et al.* Surviving ChatGPT in healthcare. *Front Radiol* 2023;**3**:1224682. https://doi.org/10.3389/fradi.2023.1224682.
126. Temsah MH, Altamimi I, Jamal A. *et al.* ChatGPT surpasses 1000 publications on PubMed: Envisioning the road ahead. *Cureus* 2023;**15**:e44769. https://doi.org/10.7759/cureus.44769.
127. Mukherjee P, Hou B, Lanfredi RB. *et al.* Feasibility of using the privacy-preserving large language model vicuna for Labeling radiology reports. *Radiology* 2023;**309**:e231147. https://doi.org/10.1148/radiol.231147.
128. Savage T, Nayak A, Gallo R. *et al.* Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. *NPJ Digit Med* 2024;**7**:20.
129. Altara R, Basson CJ, Biondi-Zoccai G. *et al.* Exploring the promise and challenges of artificial intelligence in biomedical research and clinical practice. *J Cardiovasc Pharmacol* **9900**:10. https://doi.org/1097/FJC.0000000000001546.
130. Cox LA Jr. Causal reasoning about epidemiological associations in conversational AI. *Glob Epidemiol* 2023;**5**:100102. https://doi.org/10.1016/j.gloepi.2023.100102.
131. Keloth VK, Hu Y, Xie Q. *et al.* Advancing entity recognition in biomedicine via instruction tuning of large language models. *Bioinformatics* 2024;**40**:btae163.
132. He Z, Bhasuran B, Jin Q. *et al.* Quality of answers of generative large language models vs peer patients for interpreting lab test results for lay patients: Evaluation study. ArXiv [Preprint]. 2024 Jan 23:arXiv:2402.01693v1.
133. Xu L, Sanders L, Li K. *et al.* Chatbot for health care and oncology applications using artificial intelligence and machine learning: Systematic review, JMIR. *Cancer* 2021;**7**:e27850.
134. Truhn D, Tayebi Arasteh S, Saldanha OL. *et al.* Encrypted federated learning for secure decentralized collaboration in cancer image analysis. *Med Image Anal* 2024;**92**:103059.
135. Wei K, Li J, Ding M. *et al.* Federated learning with differential privacy: Algorithms and performance analysis. *IEEE Trans Inf Forensics Secur* 2020;**15**:3454–69.
136. Ray PP, Majumder P. The double-edged sword of AI in biomedical engineering: ChatGPT's controversial impact on research and collaboration paradigms. *Ann Biomed Eng* 2023;**51**:1904–5.
137. Chen X, Zhang W, Zhao Z. *et al.* ICGA-GPT: Report generation and question answering for indocyanine green angiography images. *Br J Ophthalmol* 2024;**108**:1450–56.
138. Singh C, Askari A, Caruana R. *et al.* Augmenting interpretable models with large language models during training. *Nat Commun* 2023;**14**:7913.
139. Mesko B. The impact of multimodal large language models on health Care's future. *J Med Internet Res* 2023;**25**:e52865. https://doi.org/10.2196/52865.
140. Huang H, Zheng O, Wang D. *et al.* ChatGPT for shaping the future of dentistry: The potential of multi-modal large language model. *Int J Oral Sci* 2023;**15**:29.
141. Chou YH, Lin C, Lee SH. *et al.* User-friendly Chatbot to mitigate the psychological stress of older adults during the COVID-19 pandemic: Development and usability study. *JMIR Form Res* 2024;**8**:e49462. https://doi.org/10.2196/49462.
142. Booth F, Potts C, Bond R. *et al.* A mental health and well-being Chatbot: User event log analysis. *JMIR Mhealth Uhealth* 2023;**11**:e43052. https://doi.org/10.2196/43052.
143. Cohen IG. What should ChatGPT mean for bioethics? *Am J Bioeth* 2023;**23**:8–16.
144. Dong H, Wang M, Chang C. *et al.* Erianin inhibits the oncogenic properties of hepatocellular carcinoma via inducing DNA damage and aberrant mitosis. *Biochem Pharmacol* 2020;**182**:114266. https://doi.org/10.1016/j.bcp.2020.114266.
145. Back S, Aspuru-Guzik A, Ceriotti M. *et al.* Accelerated chemical science with AI. *Dig Dis* 2024;**3**:23–33.

Briefings in Bioinformatics, 2025, **26(1)**, bbae641
https://doi.org/10.1093/bib/bbae641
Review