

Foundation models for bioinformatics

Ziyu Chen^{1,2} | Lin Wei^{1,2} | Ge Gao^{1,2}

¹State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Biomedical Pioneering Innovative Center (BOPIC) & Beijing Advanced Innovation Center for Genomics (ICG), Center for Bioinformatics (CBI), Peking University, Beijing, China

²Changping Laboratory, Beijing, China

Correspondence

Ge Gao.

Email: gaog@mail.cbi.pku.edu.cn

Funding information

National Key Research and Development Program of China, Grant/Award Number: 2022ZD0115004

Abstract

Transformer-based foundation models such as ChatGPTs have revolutionized our daily life and affected many fields including bioinformatics. In this perspective, we first discuss about the direct application of textual foundation models on bioinformatics tasks, focusing on how to make the most out of canonical large language models and mitigate their inherent flaws. Meanwhile, we go through the transformer-based, bioinformatics-tailored foundation models for both sequence and non-sequence data. In particular, we envision the further development directions as well as challenges for bioinformatics foundation models.

KEYWORDS

ChatGPT, foundation models, large language models, transformer

1 | INTRODUCTION

Deep learning is undergoing a paradigm shift with the boom of large-scale foundation models pre-trained on a large corpus of data and adapted to multiple downstream tasks [1, 2]. The foundation nature of these models not only highlights their broad applications but also implies their incompleteness. Here, we will focus on two topics: canonical textual large language models (LLMs) for text-based bioinformatics data mining and foundation models adapted for biological data (Figure 1).

Most, if not all, foundation models are based on transformer architecture [3]. The key concept, “attention”, behind transformer, which emphasizes inter-token relationship, first came into public sight in an “additive attention” setup over a bi-LSTM network [5]. Later in 2017, attention was further switched to a “dot-product/multiplicative attention” setup, which was named transformer [3], for high scalability up to billions of parameters, enabling parallel pre-training on a massive corpus of unlabeled data in a self-supervised manner. The pre-training strategy mainly goes into two categories: encoder-only, BERT-like [6, 7]

architecture which takes an auto-encoding mask language model training objective and decoder-only, GPT-like [8, 9] architecture which takes an autoregressive causal language model loss.

After chat-oriented fine-tuning and alignment [10, 11], these pre-trained models become canonical textual LLMs we now come across in daily life, including ChatGPTs, Claude, Gemini, as well as LLaMA and BLOOM family [12, 13]. The biological and biomedical information mining has been widely adopted as a direct downstream application of these LLMs. Meanwhile, transformer architecture has been adapted to biological sequences such as DNA, RNA, proteins, and serializable data such as single-cell omics and small molecules, benefiting from the scalability and capacity of transformer.

2 | TEXTUAL LLMs FOR BIOINFORMATICS TEXT MINING

LLMs’ inherent logical reasoning capabilities [14], rooted from their abilities to extract syntactic and semantic structure of input text [15], turn LLMs into

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Quantitative Biology* published by John Wiley & Sons Australia, Ltd on behalf of Higher Education Press.

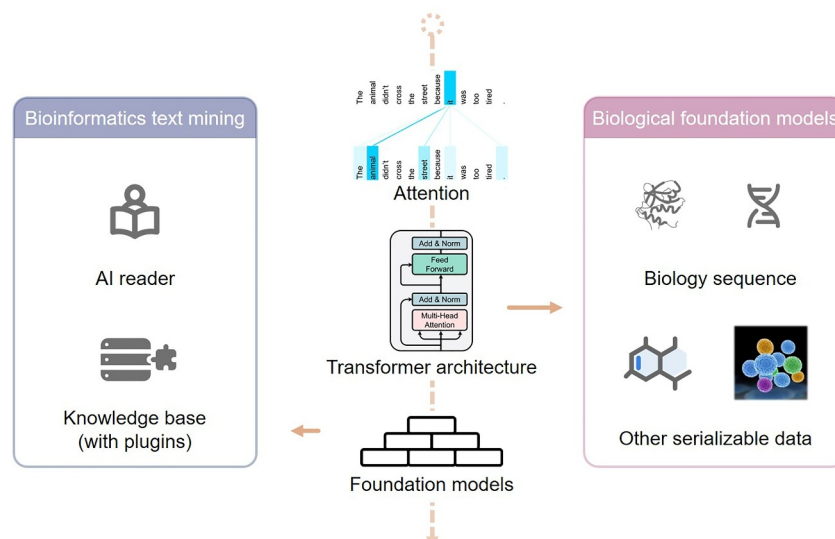


FIGURE 1 Applications of foundation models for bioinformatics. Textual foundation models, represented by large language models, can help with bioinformatics text mining, functioning as an AI reader or knowledge base with plugins. Transformer architecture can also be adapted for biological sequences and other serializable data to build large-scale biological foundation models. Part of the illustration was adapted from Refs [3, 4].

an effective “AI reader” for context-sensitive summarizing across a large corpus of literature [16]. These abilities enable LLMs to participate in the traditional manual curation process of databases. Furthermore, LLMs with further contrastive fine-tuning can generate semantic feature vector representations of the given sentences [17, 18] (also see OpenAI’s embedding application programming interfaces (APIs) [19]), extending their abilities from key-word-based searching to semantics-based searching. Meanwhile, LLMs can be regarded as a “knowledge base” about biological concepts [20]. Several works test the performance of LLMs on biological question answering, including gene interactions, biological pathways [21], genomic functions [22], and marker-gene-based cell type annotation [23]. In particular, it has been well demonstrated that biomedical-tailored LLMs trained over medical literature are able to better understand and answer domain-specific questions [24–26].

Of note, the fact that current LLMs suffer from hallucination (which is defined as “generating syntactically and semantically correct but unfaithful or nonsensical text” [27]) makes them good copilots, other than reliable advisers, for bioinformatics workflows. For instance, ChatGPT is queried for additional participants of the Circadian Clock pathway in a Reactome curation trial run [28]. ChatGPT proposed 13 candidates, among which 7 have literature supports but are overlooked in traditional manual curation, 5 cannot be confirmed, and 1 is inaccurate.

Several strategies have been shown to improve LLMs’ real-world usability.

1. Prompt tuning is always your first resort when coming into problems with LLMs. Previous work [29] demonstrated the importance of role-prompting, chain-of-thought prompting, and in-context learning in gene relationship mining and promoted an iterative prompt refinement strategy to boost performance.
2. It is well known that LLMs are prone to hallucination and cannot tell what is unknown, especially when asked for detailed questions or questions out of the knowledge base of the training corpus [27]. Though the popular LLMs’ pre-training corpus includes several biological databases, such as PubMed abstracts and PubMed Central full texts, incorporating more biological texts into the pre-training process may still be needed to mitigate hallucination in biology-related tasks.
3. Retrieval augmented generation (RAG) [30, 31] has been suggested for suppressing hallucination. The key idea behind RAG is to provide extra information, which is retrieved by text-embedding-based database search, directly into the chat context of LLMs. The reasoning process of LLMs can leverage this information for better handling of downstream tasks.
4. More generally, as modern chatbots such as ChatGPTs have been pre-trained to understand APIs and return in JavaScript Object Notation mode, they know when and how to call functions or tools provided as plugins to strengthen themselves. Additional abilities can be injected into LLMs by these plugin functions to support vector database searching for RAG, web browsing, and PubMed

searching, etc. [32]. These plugins and growing model scale may further increase the abilities of LLMs to handle biological tasks.

3 | FOUNDATION MODELS FOR BIOLOGICAL DATA

Next, we move on to discuss about foundation models pre-trained on biological data, including biological sequence data (DNA, RNA, and proteins) and data that is serializable (small molecules and single cell omics).

Transformer architecture is naturally suitable for biological sequences. There are several key choices when designing a foundation model for sequence data. (1) Whether we should choose a BERT-like structure, which excels at extracting meaningful embeddings, or a GPT-like structure that harnesses the generation abilities. Briefly, for BERT-like models, representations learned from large-scale pre-training data are contextualized representation for each token (i.e., amino acid or nucleotide) and are to some extent a replacement for multi-sequence alignment (MSA) [33, 34]. The contextualized representation can be further employed for multiple downstream tasks including structure prediction, mutation effects inference, and functional properties prediction (see more details for protein [34–37], DNA [38–42], and RNA [43, 44]). Oppositely, GPT-like models can be adapted for control tag-based *de novo* protein generation or protein engineering tasks with high diversity and success rate [45–49]. (2) Whether to pre-train at a single sequence or the MSA level [50]. (3) The scale we tokenize and model sequences: DNABERT-2 [41] utilizes Byte-Pair Encoding algorithms to tokenize DNA sequences and treats multiple residues as one token to improve compute efficiency, while ESM All-Atom [51] chooses to expand the residue representation to the atom level during pre-training for finer grain modeling and incorporates the small molecule modality.

Just like transformers' application in computer vision [52, 53], it can also be adapted to other serializable biological data. Small molecules can be easily serialized as SMILES strings and fed into transformer to learn representations to predict molecular properties [54, 55], drug-target interaction [56], and other functional tasks, thus promoting drug design. Serialization and application of transformer architecture are more challenging for single-cell omics [57]. Geneformer [58], GeneCompass [59], and Nicheformer [60] serialize the single-cell count matrix by only taking the relative ranking of the normalized expression of individual genes into consideration. Meanwhile, scBERT [61], scGPT [62], scFoundation [63], and xTrimoGene [64] replace the position embedding in classic transformer models into gene embedding and regard binned or transformed expression as tokens. Contextualized gene-level embeddings, cell-level embeddings, and attention patterns from the pre-trained

models can be tailored for downstream tasks, including cell type annotation [58, 61, 62], perturbation analysis [62, 64], regulatory network inference [58, 62] etc.

Another line of work introduces textual LLMs as external knowledge for single-cell analysis. GenePT [65] and scELMo [66] utilize textual LLMs-based embeddings as an orthogonal approach to aforementioned expression-based embeddings for downstream tasks. As it has been demonstrated that LLMs can generate semantic embeddings of sentences, these studies provide text summaries of genes and cells to textual LLMs to generate textual gene/cell embeddings and aggregate them based on expression profiles to generate cell embeddings for downstream tasks. These studies pose a new direction of integrating textual and biology foundation models to improve performance, usability, and interpretability.

4 | DISCUSSION AND PERSPECTIVE

Up till now, foundation models based on transformer still suffer from several limitations. Interpretability was initially assumed to be an advantage of transformer architecture, given that “attention” should highlight inter-relationship inside sequences. However, later works find that there are still gaps between interpretation and “attention” [67], and additional steps are needed [68]. Though the transformer operation itself can be applied on the sequence of any length if not taking position embedding into account, its space and time complexity grow quadratically with the length of sequence and consequently limits the input context length. This field is calling for a new generation of drop-in replacement for transformer operators without hurting its scalability and $O(1)$ information path length between long-range dependencies in the network [3]. This object may be achieved by (1) better implementation handling memory bounding [69], (2) sparse or low rank approximation [70–72], and (3) utilizing state space models or some other forms [42, 73–76].

Further development of foundation models for bioinformatics points out several directions.

1. The capability boundary of textual LLMs on more biological information retrieval tasks needs further testing and can be augmented with plugins to LLMs.
2. In particular, as the field has learned “the bitter lesson” [77] that scaling law [78, 79] is the secret of success of LLMs nowadays, foundation models extending to billions of parameters for biological data remain to be tested. And more research is needed to answer what emergent properties [80] are in this field. Growing model size in turn presses the need for growing volume of well-cleaned open-source biological data.

3. Taking inspirations from the success of multimodal models across texts, images, and voices, multimodal models across different kinds of biological sequences, different single cell omics, biomedical images and omics, are also interesting topics and attract growing attention.

AUTHOR CONTRIBUTIONS

Ziyu Chen: Investigation; visualization; writing – original draft. **Lin Wei:** Writing – review & editing. **Ge Gao:** Project administration; supervision.

ACKNOWLEDGMENTS

The authors would like to thank lab colleagues Yu Wang and Cheng Li for their helpful discussions and comments during the writing. This work was supported partly by the National Key Research and Development Program of China (2022ZD0115004).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

ETHICS STATEMENT

This article is a perspective article and does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

- [1] Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, et al. On the opportunities and risks of foundation models. 2021. Preprint at arXiv: 2108.07258.
- [2] Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. 2023. Preprint at arXiv: 2303.18223.
- [3] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. 2017. Preprint at arXiv: 1706.03762.
- [4] Uszkoreit J. Transformer: a novel neural network architecture for language understanding. Google Research Blog. 2017.
- [5] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. 2014. Preprint at arXiv: 1409.0473.
- [6] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. 2018. Preprint at arXiv: 1810.04805.
- [7] Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized bert pretraining approach. 2019. Preprint at arXiv: 1907.11692.
- [8] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. 2020. Preprint at arXiv: 2005.14165.
- [9] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI blog. 2019;1:9.
- [10] Wei J, Bosma M, Zhao VY, Guu K, Yu AW, Lester B, et al. Finetuned language models are zero-shot learners. 2021. Preprint at arXiv: 2109.01652.
- [11] Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, Mishkin P, et al. Training language models to follow instructions with human feedback. 2022. Preprint at arXiv: 2203.02155.
- [12] Touvron H, Lavril T, Izacard G, Martinet X, Lachaux M-A, Lacroix T, et al. LLaMA: open and efficient foundation language models. 2023. Preprint at arXiv: 2302.13971.
- [13] Workshop B, Le Scao T, Fan A, Akiki C, Pavlick E, Ilić S, et al. BLOOM: a 176b-parameter open-access multilingual language model. 2022. Preprint at arXiv: 2211.05100.
- [14] Liu H, Ning R, Teng Z, Liu J, Zhou Q, Zhang Y. Evaluating the logical reasoning ability of ChatGPT and GPT-4. 2023. Preprint at arXiv: 2304.03439.
- [15] Rogers A, Kovaleva O, Rumshisky A. A primer in BERTology: what we know about how BERT works. 2020. Preprint at arXiv: 2002.12327.
- [16] Elicit. Elicit: the AI research assistant. 2023.
- [17] Xiao S, Liu Z, Shao Y, Cao Z. RetroMAE: pre-training retrieval-oriented language models via masked auto-encoder. 2022. Preprint at arXiv: 2205.12035.
- [18] Xiao S, Liu Z, Zhang P, Muennighoff N, Lian D, Nie JY. C-pack: packaged resources to advance general Chinese embedding. 2023. Preprint at arXiv: 2309.07597.
- [19] OpenAI. OpenAI embeddings guides. 2024.
- [20] Wang J, Cheng Z, Yao Q, Liu L, Xu D, Hu G. Bioinformatics and biomedical informatics with ChatGPT: year one review. Quantitative Biology. 2024;1–15.
- [21] Azam M, Chen Y, Arowolo MO, Liu H, Popescu M, Xu D. A comprehensive evaluation of large language models in mining gene relations and pathway knowledge. Quantitative Biology. 2024;1–15.
- [22] Hou W, Ji Z. Geneturing tests GPT models in genomics. 2023. Preprint at bioRxiv: 2023.03.11.532238.
- [23] Hou W, Ji Z. Assessing GPT-4 for cell type annotation in single-cell RNA-seq analysis. Nat Methods. 2024.
- [24] Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. 2019. Preprint at arXiv: 1901.08746.
- [25] Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. 2022. Preprint at arXiv: 2210.10341.
- [26] Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172–80.
- [27] Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. 2022. Preprint at arXiv: 2202.03629.
- [28] Tiwari K, Matthews L, May B, Shamovsky V, Orlic-Milacic M, Rothfels K, et al. ChatGPT usage in the reactome curation process. 2023. Preprint at bioRxiv: 2023.11.08.566195.
- [29] Chen Y, Gao J, Petruc M, Hammer RD, Popescu M, Xu D. Iterative prompt refinement for mining gene relationships from ChatGPT. 2023. Preprint at bioRxiv: 2023.12.23.573201.
- [30] Borgeaud S, Mensch A, Hoffmann J, Cai T, Rutherford E, Millican K, et al. Improving language models by retrieving from trillions of tokens. 2021. Preprint at arXiv: 2112.04426.
- [31] Gao L, Ma X, Lin J, Callan J. Precise zero-shot dense retrieval without relevance labels. 2022. Preprint at arXiv: 2212.10496.
- [32] Chase H. Langchain. 2022.
- [33] Chowdhury R, Bouatta N, Biswas S, Floristean C, Kharkar A, Roy K, et al. Single-sequence protein structure prediction using a language model and deep learning. Nat Biotechnol. 2022; 40(11):1617–23.
- [34] Meier J, Rao R, Verkuil R, Liu J, Sercu T, Rives A. Language models enable zero-shot prediction of the effects of mutations on protein function. 2021. Preprint at bioRxiv: 2021.07.09.450648.
- [35] Rives A, Meier J, Sercu T, Goyal S, Lin Z, Liu J, et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. Proc Natl Acad Sci USA. 2021;118(15):e2016239118.

- [36] Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. 2023;379(6637):1123–30.
- [37] Hsu C, Verkuil R, Liu J, Lin Z, Hie B, Sercu T, et al. Learning inverse folding from millions of predicted structures. 2022. Preprint at bioRxiv: 2022.04.10.487779.
- [38] Ji Y, Zhou Z, Liu H, Davuluri RV. DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. *Bioinformatics*. 2021;37(15):2112–20.
- [39] Benegas G, Batra SS, Song YS. DNA language models are powerful predictors of genome-wide variant effects. *Proc Natl Acad Sci USA*. 2023;120(44):e2311219120.
- [40] Dalla-Torre H, Gonzalez L, Mendoza-Revilla J, Carranza NL, Grzywaczewski AH, Oteri F, et al. The nucleotide transformer: building and evaluating robust foundation models for human genomics. 2023. Preprint at bioRxiv: 2023.01.11.523679.
- [41] Zhou Z, Ji Y, Li W, Dutta P, Davuluri R, Liu H. DNABERT-2: efficient foundation model and benchmark for multi-species genome. 2023. Preprint at arXiv: 2306.15006.
- [42] Nguyen E, Poli M, Durrant MG, Thomas AW, Kang B, Sullivan J, et al. Sequence modeling and design from molecular to genome scale with Evo. 2024. Preprint at bioRxiv: 2024.02.27.582234.
- [43] Chen J, Hu Z, Sun S, Tan Q, Wang Y, Yu Q, et al. Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. 2022. Preprint at arXiv: 2204.00300.
- [44] Wang X, Gu R, Chen Z, Li Y, Ji X, Ke G, et al. UNI-RNA: universal pre-trained models revolutionize RNA research. 2023. Preprint at bioRxiv: 2023.07.20.11.548588.
- [45] Madani A, Krause B, Greene ER, Subramanian S, Mohr BP, Holton JM, et al. Large language models generate functional protein sequences across diverse families. *Nat Biotechnol*. 2023;41(8):1099–106.
- [46] Nijkamp E, Ruffolo JA, Weinstein EN, Naik N, Madani A. ProGen2: exploring the boundaries of protein language models. *Cell Syst*. 2023;14(11):968–78.e963.
- [47] Ferruz N, Schmidt S, Höcker B. ProtGPT2 is a deep unsupervised language model for protein design. *Nat Commun*. 2022;13(1):4348.
- [48] Ferruz N, Höcker B. Controllable protein design with language models. *Nat Mach Intell*. 2022;4(6):521–32.
- [49] Winnifrid A, Outeiral C, Hie BL. Generative artificial intelligence for de novo protein design. *Curr Opin Struct Biol*. 2024;86:102794.
- [50] Rao R, Liu J, Verkuil R, Meier J, Canny JF, Abbeel P, et al. MSA transformer. 2021. Preprint at bioRxiv: 2021.02.12.430858.
- [51] Zheng K, Long S, Lu T, Yang J, Dai X, Zhang M, et al. ESM all-atom: multi-scale protein language model for unified molecular modeling. 2024. Preprint at arXiv: 2403.12995.
- [52] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. 2020. Preprint at arXiv: 2010.11929.
- [53] Chen M, Radford A, Child R, Wu J, Jun H, Luan D, et al. Generative pretraining from pixels. In: *Proceedings of the 37th international conference on machine learning*. 2020;119:1691–703.
- [54] Wang S, Guo Y, Wang Y, Sun H, Huang J. Smiles-bert: large scale unsupervised pre-training for molecular property prediction. In: *Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics*; 2019. p. 429–36.
- [55] Rong Y, Bian Y, Xu T, Xie W, Wei Y, Huang W, et al. Self-supervised graph transformer on large-scale molecular data. In: *Proceedings of the 34th international conference on neural information processing systems*; 2020. Article 1053.
- [56] Kalakoti Y, Yadav S, Sundar D. TransDTI: transformer-based language models for estimating DTIs and building a drug recommendation workflow. *ACS Omega*. 2022;7(3):2706–17.
- [57] Hao M, Wei L, Yang F, Yao J, Theodoris CV, Wang B, et al. Current opinions on large cellular models. *Quantitative Biology*. 2024;1–11.
- [58] Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, et al. Transfer learning enables predictions in network biology. *Nature*. 2023;618(7965):616–24.
- [59] Yang X, Liu G, Feng G, Bu D, Wang P, Jiang J, et al. GeneCompass: deciphering universal gene regulatory mechanisms with knowledge-informed cross-species foundation model. 2023. Preprint at bioRxiv: 2023.09.26.559542.
- [60] Schaar AC, Tejada-Lapueta A, Palla G, Gutgesell R, Halle L, Minaeva M, et al. Nicheformer: a foundation model for single-cell and spatial omics. 2024. Preprint at bioRxiv: 2024.04.15.589472.
- [61] Yang F, Wang W, Wang F, Fang Y, Tang D, Huang J, 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(10):852–66.
- [62] Cui H, Wang C, Maan H, Pang K, Luo F, Duan N, et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods*. 2024;21:1470–80.
- [63] Hao M, Gong J, Zeng X, Liu C, Guo Y, Cheng X, et al. Large scale foundation model on single-cell transcriptomics. 2023. Preprint at bioRxiv: 2023.05.29.542705.
- [64] Gong J, Hao M, Zeng X, Liu C, Ma J, Cheng X, et al. xTrimoGene: an efficient and scalable representation learner for single-cell RNA-seq data. 2023. Preprint at bioRxiv: 2023.03.24.534055.
- [65] Chen Y, Zou J. GenePT: a simple but effective foundation model for genes and cells built from ChatGPT. 2024. Preprint at bioRxiv: 2023.10.16.562533.
- [66] Liu T, Chen T, Zheng W, Luo X, Zhao H. scELMo: embeddings from language models are good learners for single-cell data analysis. 2023. Preprint at bioRxiv: 2023.12.07.569910.
- [67] Jain S, Wallace BC. Attention is not explanation. 2019. Preprint at arXiv: 1902.10186.
- [68] Abnar S, Zuidema W. Quantifying attention flow in transformers. 2020. Preprint at arXiv: 2005.00928.
- [69] Dao T, Fu DY, Ermon S, Rudra A, Ré C. Flashattention: fast and memory-efficient exact attention with IO-awareness. 2022. Preprint at arXiv: 2205.14135.
- [70] Child R, Gray S, Radford A, Sutskever I. Generating long sequences with sparse transformers. 2019. Preprint at arXiv: 1904.10509.
- [71] Zaheer M, Guruganesh G, Dubey A, Ainslie J, Alberti C, Ontanon S, et al. Big bird: transformers for longer sequences. 2020. Preprint at arXiv: 2007.14062.
- [72] Choromanski K, Likhoshesterov V, Dohan D, Song X, Gane A, Sarlos T, et al. Rethinking attention with performers. 2020. Preprint at arXiv: 2009.14794.
- [73] Peng B, Alcaide E, Anthony Q, Albalak A, Arcadinho S, Biderman S, et al. Rvk: reinventing RNNs for the transformer era. 2023. Preprint at arXiv: 2305.13048.
- [74] Poli M, Massaroli S, Nguyen E, Fu DY, Dao T, Baccus S, et al. Hyena hierarchy: towards larger convolutional language models. 2023. Preprint at arXiv: 2302.10866.
- [75] Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. 2023. Preprint at arXiv: 2312.00752.
- [76] Nguyen E, Poli M, Faizi M, Thomas A, Birch-Sykes C, Wornow M, et al. HyenaDNA: long-range genomic sequence modeling at single nucleotide resolution. 2023. Preprint at arXiv: 2306.15794.
- [77] Sutton R. The bitter lesson. 2019.
- [78] Kaplan J, McCandlish S, Henighan T, Brown TB, Chess B, Child R, et al. Scaling laws for neural language models. 2020. Preprint at arXiv: 2001.08361.

- [79] Hoffmann J, Borgeaud S, Mensch A, Buchatskaya E, Cai T, Rutherford E, et al. Training compute-optimal large language models. 2022. Preprint at arXiv: 2203.15556.
- [80] Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, et al. Emergent abilities of large language models. 2022. Preprint at arXiv: 2206.07682.

How to cite this article: Chen Z, Wei L, Gao G. Foundation models for bioinformatics. *Quantitative Biology*. 2024;12(4):339–44. <https://doi.org/10.1002/qub2.69>