Molecular Causality in the Age of Foundation Models

This manuscript (permalink) was automatically generated from slobentanzer/causality_perspective_2023@405bd49 on January 4, 2024.

Authors

- **Sebastian Lobentanzer** ✉
  0000-0003-3399-6695 · slobentanzer · slobentanzer
  Heidelberg University, Faculty of Medicine and Heidelberg University Hospital, Institute for Computational Biomedicine, Heidelberg, Germany

- **Pablo Rodriguez Mier**
  todo · tbd
  Heidelberg University, Faculty of Medicine and Heidelberg University Hospital, Institute for Computational Biomedicine, Heidelberg, Germany

- **Stefan Bauer**
  todo · tbd
  Helmholtz Zentrum München, -, Neuherberg, Germany

- **Julio Saez-Rodriguez** ✉
  0000-0002-8552-8976 · saezrodriguez · saezlab
  Heidelberg University, Faculty of Medicine and Heidelberg University Hospital, Institute for Computational Biomedicine, Heidelberg, Germany

✉ — Correspondence possible via GitHub Issues or email to Sebastian Lobentanzer <sebastian.lobentanzer@gmail.com>, Julio Saez-Rodriguez <pub.saez@uni-heidelberg.de>.

Introduction

Correlation is not causation. As simple as this widely agreed-upon statement may seem, scientifically defining causality and using it to drive our modern biomedical research is immensely challenging. Since being described by Aristotle approximately 2500 years ago [1], causal reasoning (CR) remained virtually unchanged until it experienced significant formal and mathematical advancements [2,3,4] and a resurgence in the field of machine learning (ML) [**doi?**/arxiv.2206.15475] only in recent times. In parallel, biomedicine has made major leaps in the past century, in particular in the development of high-throughput and large-scale methods.

In the field of systems biology, however, great hopes of causal insights from large-scale omics studies have largely been thwarted by the great complexity of molecular mechanisms and the inability of existing methods to distinguish between correlation and causation [5,6].

Randomised clinical trials show that, in a lower-dimensional context, we can reliably identify causal effects. By controlling "all" relevant covariates in a trial (via the principle of the gold-standard, randomised, double-blind, and placebo-controlled trial) we isolate the causal effect of the controlled variable, i.e., the treatment. In the vernacular of Pearl's Do-Calculus [7], we measure the outcome of, for instance, *do("Treat with Vemurafenib")* when conducting a clinical trial on V600E-positive melanoma [8]. However, translating this mode of reasoning into the high-dimensional space of modern omics is met with enormous challenges. The dramatically larger parameter space of models at the molecular level leads to problems in the performance of methods and the identifiability of results, as well as in model explainability.

With this perspective, we discuss the current connections between CR and molecular systems biology. We will elaborate on three main points:

- biases and what they mean for CR, particularly in the context of biomedical data
- the role of prior knowledge in CR and how to translate prior knowledge into suitable biases
- the role of foundation models in molecular systems biology and their relationship to CR

Background

Causal Discovery and Inference

The field of CR distinguishes between causal discovery - the process of building causal hypotheses from data - and causal inference - the process of predicting how a specific situation will turn out given data and the causal relationships known about the system.

Causal discovery is more expensive than inference both computationally and data-wise, because it needs to account for the variability in data generation while isolating generalisable relationships between single measured agents [9]. As a result, most inference mechanisms perform better when including prior knowledge at some point in the process, as has been observed in biomedical research [10]. For modern systems biology, this means that methods for causal discovery typically require large amounts of experiments. Highly parameterised models such as neural networks increase this requirement even further. As such, many regard causal discovery in molecular biomedicine as a scaling problem [11,12].

Causal inference, in contrast, focuses on quantifying the causal effects of one variable on another within the framework of already hypothesised causal relationships. This approach leverages prior knowledge about the assumed causal links, which in the causal field are often encoded using directed graphs. This allows researchers to represent both the causal connections between variables and the directionality, which is required to understand how changes in one variable might lead to changes in another. For instance, in the case of the RAF-MEK-ERK signalling pathway, a graph would depict RAF activation leading to MEK activation, which in turn leads to ERK activation. This clear representation of directionality is important for causal inference, as it ensures that analyses focus on the effect of upstream changes on downstream outcomes. For example, in analysing phosphoproteomic data to assess the impact of inhibiting RAF, a graph-based approach would guide researchers to correctly attribute subsequent changes in ERK to this specific intervention. Without this causal framework, one might mistakenly interpret correlations as bidirectional influences or overlook confounding factors, leading to incorrect conclusions. However, inference is also very sensitive to the completeness of the prior knowledge that is applied; most biomedical prior knowledge is far from complete [13]. For instance, the function of more than 95% of all the known phosphorylation events that occur in human cells is currently unknown [14,15]. In contrast to causal discovery, scaling therefore plays a smaller role in causal inference; here, the main problem is incompleteness and identifying the “right” biases to apply.

The Ladder of Causality

Orthogonally to the distinction between causal discovery and inference, we can also distinguish between different levels of causality. Pearl’s ladder of causality roughly distinguishes three types of CR in increasing order of power: observation, intervention, and counterfactuals [16]. While the inferences we wish to make in biomedical research are often of the counterfactual type (e.g., “would RAF inhibition lead to decrease in ERK activation if the media contained Epidermal Growth Factors?”), the data we have available is typically observational (e.g., “the levels of RAF and MEK activity are correlated”) and sometimes interventional (e.g., “targeting RAF with CRISPR leads to a decrease in ERK activity”). To generate interventional or even counterfactual inferences from observational data is a major challenge, if not impossible, depending on the characteristics of the system under study [17].

There are approaches to delineate interventional inference from observational data, such as the ‘natural experiments’ framework [3,4]. However, these approaches are by their nature even more data-hungry than when using interventional data, as they necessarily discard information that is not relevant to the intervention. Therefore, in biomedical research, there has been a push towards generating large-scale interventional data, for instance through the use of CRISPR/Cas9 screens with single-cell resolution [18]. Current developments of CR in the biomedical field therefore mostly focus on these types of data.

Deduction and Induction

Lastly, in CR, we can also distinguish between deductive and inductive reasoning. This is where certain biases are pivotal to the effectiveness of the CR method. Deductive reasoning is the process of deriving a conclusion from a set of fixed and known premises. “All men are mortal, Socrates is a man, therefore Socrates is mortal” is a classic example of deductive reasoning. In biomedical research, this is typically the process of deriving a conclusion from a set of prior knowledge. For instance, having prior knowledge of the linear activation cascade EGFR->RAS->RAF->MEK->ERK->Growth, and that Vemurafenib will inhibit RAF activity, allows us to deduce that giving Vemurafenib will inhibit growth of cancer cells [8].

Inductive reasoning, on the other hand, involves making generalisations from specific observations. Testing the hypothesis above, we apply Vemurafenib in a clinical trial of V600E-positive melanoma and find that it is clinically efficacious [8]. We could now infer (via induction) that Vemurafenib may be an effective remedy in other V600E-positive cancers as well, or, further, that inhibiting this cascade may be a general mechanism of action of anti-cancer agents in cancers that display MAPK pathway overactivation [19]. In the molecular realm, we could further infer that the inhibition of other components of the cascade, such as EGFR or MEK, may also be promising target leads [20].

The main difference between deduction and induction is that the former is logically complete - i.e., if the premises are true and the argument is valid, the conclusion must also be true. However, deduction is also more limited in scope than

induction. In biomedical research, we often have to rely on inductive reasoning because we cannot feasibly test all hypotheses in a deductive manner. In consequence, the inductive biases we introduce into our models (i.e., those mechanisms in the model that help with inductive reasoning) are a pivotal part of performing CR in biomedical research.

Bias

Meaning and examples of biases

Biases, generally, are systematic prejudices of a model towards certain outcomes. Humans make frequent use of biases to function in a complex world with limited cognitive resources. Our brain seems predisposed to doing causal inference, a skill which we learn and hone from a very early age [21]. In fact, we may be over-eager to deduce causality from observation (i.e., "jump to conclusions"), which is indicative of a strong inductive bias. A good *heuristic* is the application of a suitable bias to a problem, such that the solution can be considered acceptable despite limited resources.

In machine learning (ML), we can distinguish between useful and harmful biases. Harmful biases are common issues in the technical process of training models; they include, for instance, sampling bias, selection bias, confirmation bias, overfitting, and underfitting []. While addressing harmful biases is a crucial part of ML, we will not discuss them further in this perspective.

Useful biases, on the other hand, are biases that are introduced into a model to improve its performance. They can be relatively implicit, such as the choice of algorithm, architecture, or regularisation; or explicit, such as the choice of prior knowledge and how it is used. In the context of CR, useful biases are those that improve the performance of the model in terms of its ability to draw correct causal conclusions. Since most models developed in biomedical research and the broader ML community are inductive models, one of the most discussed useful biases is *inductive bias*.

Why do we need biases?

In biomedical research, we operate in a space that is very constrained in the amount and quality of data we can collect. This is due to the high cost of experiments, the limited availability of samples, and the high dimensionality of the data. These issues, in combination with the naturally high variability of biological measurements, lead to a relatively low signal-to-noise ratio of our observations. In addition, we are often trying to "climb" the ladder of causality with our CR approaches, which comes with additional data requirements. Lastly, we also lack a ground truth for most contexts in which we perform measurements. As a result, we need to introduce biases into our models to make the most of the data we have.

Some central questions then arise:

- How do we choose the right biases to introduce?
- How explicit should we be in introducing biases (i.e., should the model determine its own biases, or do we force them on the model)?
- How do we evaluate the biases we introduce?

Bias from prior knowledge

The first question alone is highly debated in the wider field of ML. The frequently quoted "Bitter Lesson" posits that we should refrain from inducing all but the most basic biases in our models, and that we should not view metrics as the ultimate measure of performance, but rather whether the model gets us closer to some truth [22]. However, it has been argued that many improvements that led to the models of today, such as convolution or attention, disprove this theory [23], and that the intrinsic complexity of real-world systems does not obviate, but rather necessitate, the integration of human insight into our learning frameworks [24,25].

In systems biology, specifically, there is much interest in finding models with suitable biases to deal with constraints specific to the field, such as data availability and the completeness of prior knowledge [6,6,26,27,28,29,29]. Considering these constraints, the question is not whether to include prior knowledge in our reasoning, but which knowledge, when, and how [25].

Prior knowledge

To be able to effectively use our knowledge in reasoning, we must be able to represent it robustly and in a way that is conducive to the reasoning task. Biomedical entities and relationships must be clearly defined and represented unambiguously. Additionally, the diversity in our tasks and knowledge sources requires a flexible representation that can be adapted to the task at hand. Knowledge representation frameworks can aid in walking the line between robust and transparent, reproducible knowledge representation on the one hand, and flexible, task-specific workflows on the other [30]. In addition, they can ground the biological entities that are subject to reasoning using domain-specific ontologies, which can be another useful source of bias. For instance, knowing the directionality of the central dogma of biology can be a useful bias in reasoning about gene expression.

Modelling on prior knowledge

Statistical, causal, and mechanistic models

Why modelling needs biases and how to introduce them

Benchmarks

Causality in foundation models

There has been an enormous spike of interest in attention-based neural network models, in large part due to the success of Large Language Models (LLMs). While the high performance of LLMs is based on myriad architectural improvements, the introduction of attention as an architectural bias has been a major contributor to their success [23]. This has led to the development of attention-based molecular models (most commonly for gene expression), which can also be considered "GPT" models: Generative Pre-trained Transformers [31,32,33]. Attention as a learning mechanism enables the integration of non-local information in a flexible manner. In a molecular model that reasons about gene expression, such as Geneformer, attention allows the integration of distant regulatory elements [32]. Notably, this mechanism comes with a computational cost that increases exponentially with respect to the length of the input sequence [34].

The generalist capabilities of LLMs have led to the designation of "foundation models" [35]. Foundation models are models that achieve high performance by training a generic architecture on extremely large amounts of data in an unsupervised manner. They can be fine-tuned for more specific tasks, because they are thought to derive generalisable representations and mechanisms by training on an amount of data large enough to learn the complexity of real-world systems. However, recent molecular foundation model benchmarks highlight clear discrepancies between the "foundational" aspirations of the pre-trained models and the real-world evaluation of their performance [36,37]. Briefly, the benchmarks found that, on single cell classification tasks, the proposed foundation models did not outperform simple baselines consistently. State-of-the-art methods such as scVI [38] and even the mere selection of highly variable genes was often statistically indistinguishable from the highly parameterised methods, and sometimes even yielded better classification outcomes. However, these are early models, and it could still be argued that, in line with the scaling hypothesis, models will improve via a combination of the right architecture with sufficient amounts of data.

Indeed, molecular foundation models lag behind in size: while current-generation LLMs have around 100 billion parameters or more and are trained on enormous text corpuses (hundreds of billions to trillions of tokens), molecular foundation models have tens of millions of parameters (scGPT: 53M, Geneformer: 10M) and are trained on corpuses of tens of millions of cells, which (optimistically) yields hundreds of billions of individual data points. Thus, LLMs are currently about 2000 times larger than molecular foundation models, while arguably also dealing with a less complicated system. The question whether scaling will lead to the emergence of "foundational behaviour" in molecular models is still a matter of much debate.

Attention - and large amounts of data - is all you need?

Given enough data to train on - and ample funds for compute - is attention "all you need" to induce reliable biases in your model? While there are doubts regarding the reasoning capabilities of LLMs, GPT arguably "understands" language very well already, to the point where it can flawlessly communicate and synthesise information [39]. This is what the term "foundation model" implies: the model has derived a generalisable representation of language, a tool that can be fine-tuned for a variety of language-related tasks. This behaviour is not possible without assuming some form of causality, even if it is not explicitly encoded in the model [11].

In this light, what are the reasons to be sceptical about the capacity of molecular foundation models to understand the "grammar" of the cell?

Explainability: For one, large transformer models are not explainable due to their large number of parameters and non-linearities. As such, there is no way to scrutinise their reasoning beyond the output they produce. What seems simple in

the case of language models - the famous Turing test can be performed by any human with a basic understanding of language - is exceedingly difficult in the molecular space, where many causal relationships are still unknown [39]. Yet the only way to scrutinise and subsequently improve the reasoning capabilities of a model is precisely this explicit validation of its predictions in an interpretable setting.

While the creation of explicit molecular models (e.g., logic, structural causal, or ODE-based models) and the self-supervised training of molecular foundation models are methodically very different, both can provide a hypothesis on causal structure that can be formulated as a network. Theodoris et al. explore the attention layers of their Geneformer foundation model to explain the model's reasoning [32]. While some layers show clear patterns of attention, such as attending to highly connected or highly expressed genes, other layers are not as readily interpretable, much less so than explicit molecular models.

Benchmarking: Whether these complex layers reflect the true complexity of the underlying biology or are rather evidence for overfitting to the training data is not clear. One argument in favour of overfitting is the poor generalisation of the model in independent benchmarks [36,37]. To determine whether molecular foundation models indeed capture generalisable causal representations of biology, dedicated benchmarks are needed.

Causal bias: The GPT-3 architecture that led to the recent breakthrough in LLM capabilities employs "causal self-attention," describing an implicit architectural bias that prevents the model from "looking into the future": for predicting the next token, only the previous tokens in the sentence can be used [34]. This leverages the implicit causality present in language, which incidentally is similar to one of the earliest formal descriptions of causality, that "the effect has regularly followed the cause in the past" [40]. Compared to language, the data that form the input of molecular foundation models do not have an implicit form of causality. The individual cells are in general not on a known temporal trajectory, and the genes that are masked as part of the training objective are masked at random, not because they are downstream (in some form) of the genes used for prediction. This fundamental difference between language and molecular models has so far not been explored theoretically or empirically.

Causal latent spaces

Latent encodings of explicit prior knowledge (GEARS)

Do 'causal latent spaces' exist, and how would we prove it?

How do we explore these latent spaces and use them for inference?

Stefan's comment: newer architectures (self-supervised) do not decode; how important is it for biological insights, particularly compared with scaling? Exploring and explaining the latent space...

- https://scholar.google.de/citations?view_op=view_citation&hl=de&user=soxv0s0AAAAJ&sortby=pubdate&citation_for_view=soxv0s0AAAAJ:2osOgNQ5qMEC (decoder important)
- https://proceedings.neurips.cc/paper_files/paper/2022/hash/87213955efbe48b46586e37bf2f1fe5b-Abstract-Conference.html (decoder not important)

References

1. **The Organon, Or Logical Treatises, Of Aristole, Vol. 1 Of 2**
   Aristotle
   Forgotten Books (2015)
   ISBN: 9781330267608

2. **Causality**
   Judea Pearl
   Cambridge University Press (2009-09-14) https://doi.org/ggd72q
   DOI: 10.1017/cbo9780511803161

3. **Identification of Causal Effects Using Instrumental Variables**
   Joshua D Angrist, Guido W Imbens, Donald B Rubin
   Journal of the American Statistical Association (1996-06) https://doi.org/gdz4f4
   DOI: 10.1080/01621459.1996.10476902

4. **Myth and measurement: the new economics of the minimum wage**
   David E Card, Alan B Krueger
   Princeton University Press (2016)
   ISBN: 9781400880874

5. **A map of human genome variation from population-scale sequencing** *Nature* (2010-10-27)
   https://doi.org/fmk7rw
   DOI: 10.1038/nature09534 · PMID: 20981092 · PMCID: PMC3042601

6. **The perpetual motion machine of AI-generated data and the distraction of ChatGPT-as-scientist**
   Jennifer Listgarten
   arXiv (2023) https://doi.org/gs8pnp
   DOI: 10.48550/arxiv.2312.00818

7. **The Do-Calculus Revisited**
   Judea Pearl
   arXiv (2012) https://doi.org/gtbf4r
   DOI: 10.48550/arxiv.1210.4852

8. **Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation**
   Paul B Chapman, Axel Hauschild, Caroline Robert, John B Haanen, Paolo Ascierto, James Larkin, Reinhard Dummer, Claus Garbe, Alessandro Testori, Michele Maio, ... Grant A McArthur
   New England Journal of Medicine (2011-06-30) https://doi.org/dsbxxt
   DOI: 10.1056/nejmoa1103782 · PMID: 21639808 · PMCID: PMC3549296

9. **Causal Structure Learning**
   Christina Heinze-Deml, Marloes H Maathuis, Nicolai Meinshausen
   Annual Review of Statistics and Its Application (2018-03-07) https://doi.org/ggh4pj
   DOI: 10.1146/annurev-statistics-031017-100630

10. **Inferring causal molecular networks: empirical assessment through a community-based effort**
    Steven M Hill, Laura M Heiser, Thomas Cokelaer, Michael Unger, Nicole K Nesser, Daniel E Carlin, Yang Zhang, Artem Sokolov, Evan O Paull, ... Sach Mukherjee
    Nature Methods (2016-02-22) https://doi.org/f3t7t4
    DOI: 10.1038/nmeth.3773 · PMID: 26901648 · PMCID: PMC4854847

11. **Can Foundation Models Talk Causality?**
    Moritz Willig, Matej Zečević, Devendra Singh Dhami, Kristian Kersting
    arXiv (2022) https://doi.org/gtb9wb
    DOI: 10.48550/arxiv.2206.10591

12. **The Scaling Hypothesis**
    Gwern Branwen
    (2020-05-28) https://gwern.net/scaling-hypothesis

13. **Integrating knowledge and omics to decipher mechanisms via large-scale models of signaling networks**
    Martin Garrido-Rodriguez, Katharina Zirngibl, Olga Ivanova, Sebastian Lobentanzer, Julio Saez-Rodriguez
    Molecular Systems Biology (2022-07) https://doi.org/gtb9v8

DOI: 10.15252/msb.202211036 · PMID: 35880747 · PMCID: PMC9316933

14. **Illuminating the dark phosphoproteome**
Elise J Needham, Benjamin L Parker, Timur Burykin, David E James, Sean J Humphrey
Science Signaling (2019-01-22) https://doi.org/gf8c3h
DOI: 10.1126/scisignal.aau8645 · PMID: 30670635

15. **The functional landscape of the human phosphoproteome**
David Ochoa, Andrew F Jarnuczak, Cristina Viéitez, Maja Gehre, Margaret Soucheray, André Mateus, Askar A Kleefeldt, Anthony Hill, Luz Garcia-Alonso, Frank Stein, ... Pedro Beltrao
Nature Biotechnology (2019-12-09) https://doi.org/ggd8n7
DOI: 10.1038/s41587-019-0344-3 · PMID: 31819260 · PMCID: PMC7100915

16. **The book of why: the new science of cause and effect**
Judea Pearl, Dana Mackenzie
Basic Books (2018)
ISBN: 9780465097616

17. **Causal inference in statistics: An overview**
Judea Pearl
Statistics Surveys (2009-01-01) https://doi.org/drt748
DOI: 10.1214/09-ss057

18. **Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens**
Atray Dixit, Oren Parnas, Biyu Li, Jenny Chen, Charles P Fulco, Livnat Jerby-Arnon, Nemanja D Marjanovic, Danielle Dionne, Tyler Burks, Raktima Raychowdhury, ... Aviv Regev
Cell (2016-12) https://doi.org/f9prjd
DOI: 10.1016/j.cell.2016.11.038 · PMID: 27984732 · PMCID: PMC5181115

19. **Vemurafenib: the first drug approved for BRAF-mutant cancer**
Gideon Bollag, James Tsai, Jiazhong Zhang, Chao Zhang, Prabha Ibrahim, Keith Nolop, Peter Hirth
Nature Reviews Drug Discovery (2012-10-12) https://doi.org/f4b975
DOI: 10.1038/nrd3847 · PMID: 23060265

20. **Targeting the ERK Signaling Pathway in Melanoma**
Paola Savoia, Paolo Fava, Filippo Casoni, Ottavio Cremona
International Journal of Molecular Sciences (2019-03-25) https://doi.org/gtb9v9
DOI: 10.3390/ijms20061483 · PMID: 30934534 · PMCID: PMC6472057

21. **A Theory of Causal Learning in Children: Causal Maps and Bayes Nets.**
Alison Gopnik, Clark Glymour, David M Sobel, Laura E Schulz, Tamar Kushnir, David Danks
Psychological Review (2004) https://doi.org/fkj59s
DOI: 10.1037/0033-295x.111.1.3 · PMID: 14756583

22. **The Bitter Lesson** http://www.incompleteideas.net/IncIdeas/BitterLesson.html

23. **Attention Is All You Need**
Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, Illia Polosukhin
arXiv (2017) https://doi.org/gpnmtv
DOI: 10.48550/arxiv.1706.03762

24. **A Better Lesson – Rodney Brooks** (2019-03-19) https://rodneybrooks.com/a-better-lesson/

25. **Thread by @shimon8282: "Rich Sutton has a new blog post entitled “The Bitter Lesson” (incompleteideas.net/IncIdeas/Bitte...) that I strongly disagree with. In it, he [...]"**
https://twitter.com/shimon8282
https://threadreaderapp.com/thread/1106534178676506624.html

26. **Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations**
Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, Olivier Bachem
arXiv (2018) https://doi.org/grx79c
DOI: 10.48550/arxiv.1811.12359

27. **Toward Causal Representation Learning**
Bernhard Scholkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, Yoshua Bengio

Proceedings of the IEEE (2021-05) https://doi.org/gjhqrh
DOI: 10.1109/jproc.2021.3058954

28. **Beyond Predictions in Neural ODEs: Identification and Interventions**
Hananeh Aliee, Fabian J Theis, Niki Kilbertus
arXiv (2021) https://doi.org/gszw4d
DOI: 10.48550/arxiv.2106.12430

29. **Inductive biases for deep learning of higher-level cognition**
Anirudh Goyal, Yoshua Bengio
Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences (2022-10)
https://doi.org/gs39f8
DOI: 10.1098/rspa.2021.0068

30. **Democratizing knowledge representation with BioCypher**
Sebastian Lobentanzer, Patrick Aloy, Jan Baumbach, Balazs Bohar, Vincent J Carey, Pornpimol Charoentong, Katharina Danhauser, Tunca Doğan, Johann Dreo, Ian Dunham, … Julio Saez-Rodriguez
Nature Biotechnology (2023-06-19) https://doi.org/gszqjr
DOI: 10.1038/s41587-023-01848-y · PMID: 37337100

31. **Effective gene expression prediction from sequence by integrating long-range interactions**
Žiga Avsec, Vikram Agarwal, Daniel Visentin, Joseph R Ledsam, Agnieszka Grabska-Barwinska, Kyle R Taylor, Yannis Assael, John Jumper, Pushmeet Kohli, David R Kelley
Nature Methods (2021-10) https://doi.org/gm2wv4
DOI: 10.1038/s41592-021-01252-x · PMID: 34608324 · PMCID: PMC8490152

32. **Transfer learning enables predictions in network biology**
Christina V Theodoris, Ling Xiao, Anant Chopra, Mark D Chaffin, Zeina R Al Sayed, Matthew C Hill, Helene Mantineo, Elizabeth M Brydon, Zexian Zeng, XShirley Liu, Patrick T Ellinor
Nature (2023-05-31) https://doi.org/gr9x63
DOI: 10.1038/s41586-023-06139-9 · PMID: 37258680

33. **scGPT: Towards Building a Foundation Model for Single-Cell Multi-omics Using Generative AI**
Haotian Cui, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Bo Wang
Cold Spring Harbor Laboratory (2023-05-01) https://doi.org/gshk6p
DOI: 10.1101/2023.04.30.538439

34. **HyperAttention: Long-context Attention in Near-Linear Time**
Insu Han, Rajesh Jayaram, Amin Karbasi, Vahab Mirrokni, David P Woodruff, Amir Zandieh
arXiv (2023) https://doi.org/gtb9wc
DOI: 10.48550/arxiv.2310.05869

35. **Stanford CRFM** https://crfm.stanford.edu/

36. **Assessing the limits of zero-shot foundation models in single-cell biology**
Kasia Z Kedzierska, Lorin Crawford, Ava P Amini, Alex X Lu
Cold Spring Harbor Laboratory (2023-10-17) https://doi.org/gszxk9
DOI: 10.1101/2023.10.16.561085

37. **A Deep Dive into Single-Cell RNA Sequencing Foundation Models**
Rebecca Boiarsky, Nalini Singh, Alejandro Buendia, Gad Getz, David Sontag
Cold Spring Harbor Laboratory (2023-10-23) https://doi.org/gszxmb
DOI: 10.1101/2023.10.19.563100

38. **Deep generative modeling for single-cell transcriptomics**
Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, Nir Yosef
Nature Methods (2018-11-30) https://doi.org/gfkw5z
DOI: 10.1038/s41592-018-0229-2 · PMID: 30504886 · PMCID: PMC6289068

39. **ChatGPT broke the Turing test — the race is on for new ways to assess AI**
Celeste Biever
Nature (2023-07-25) https://doi.org/gskd92
DOI: 10.1038/d41586-023-02361-7 · PMID: 37491395

40. **An enquiry concerning human understanding**
David Hume, PF Millican
Oxford University Press (2007)

ISBN: 9780199211586