ISSCR

A conversation with ChatGPT on the role of computational systems biology in stem cell research

Welcome to this special issue of *Stem Cell Reports* on computational systems biology. As a computationalists at heart, we are going to adhere to the advice that Larry Wall (the creator of the Perl programming language) gave to budding programmers: " […] develop the three great virtues of a programmer: laziness, impatience, and hubris." In the normal world, these traits are typically considered vices. However, in the world of programming, they are considered virtues because they encourage writing code instead of performing repetitive tasks manually (laziness), documenting the code and program so that people do not harass you with incessant and perhaps inane questions (impatience), and fixing and updating the program rather than letting your reputation be tarnished by buggy or outdated code (hubris).

In this editorial, we will rely on our hard-earned skill of laziness by using a new a deep-learning program to perform the somewhat repetitive task of generating bits of content, as indicated below, to introduce this special issue. The tool that we are using, ChatGPT, was just released by OpenAI, and it has created quite an uproar. Some corners of social media are afraid that ChatGPT and friends will have a host of negative consequences on society, including the loss of knowledge-centric careers. Other corners of social media are excited about potential applications, for example one-on-one tutoring. No matter your position on this topic, just as machine learning is already impacting our day-to-day lives, it is inexorably making inroads in our field (see Ouyang et al. in this issue). Here, we are using it to illustrate how advances in computation (maybe not computational systems biology, per se, but the underlying theory and methods are shared) can help practitioners across the broader stem cell research enterprise: by saving time and thus freeing us to do more research! (Our choice to use it has nothing to do with the fact that the deadline for the editorial is today.)

Without further ado, below is a record of our correspondence with ChatGPT. The text following "prompt:>" is our input, and the text following "ChatGPT:" is the output from the model. We have made some minor edits, and we interrupt the conversation at points to call out relevant papers in this issue.

> prompt:> Help us to write an editorial about how computational biology and systems biology has impacted stem cell research.

ChatGPT: Computational biology and systems biology have had a significant impact on stem cell research in

Barbara Treutlein (left) and Patrick Cahan (right), guest editors of *Stem Cell Reports*' **special edition on computational systems biology**

recent years. These fields use computational and mathematical methods to analyze and model biological systems, including stem cells, allowing researchers to better understand their behavior and potential applications. One of the key benefits of using computational biology and systems biology in stem cell research is the ability to process and analyze large amounts of data. This is particularly important in stem cell research, where the behavior of individual cells can vary greatly and understanding the underlying mechanisms can be complex. By using computational methods to analyze data, researchers can identify patterns and trends that would be difficult to detect using other approaches.

While the application of computation to detect patterns in data might seem obvious, it bears considering that many trends and patterns are often not readily discernible with standard computational approaches. This is especially true for noisy single-cell omics data, and so specialized methods are need. In this issue, Radley et al. introduce a method called "entropy sorting" that finds cell identity genes, which the authors use to uncover, for the first time, inner cell mass cells of the human pre-implantation embryo. Similarly, some patterns or trends are only discernible after proper integration with other datasets to increase power (see Moya-Jódar et al.) or after integration across data modalities (for example, see Alavattam et al. and Ganguly et al.). Estévez-Priego et al. showcase how advanced analysis of long-term imaging data can be used to quantitatively compare neuronal activity in disparate models, ranging from primary rodent neurons to human induced pluripotent stem cell (hiPSCs)-neurons. Rounding out this theme of using computational tools to discover novel patterns in big data is the application by Schmidt et al., who find a minimal DNA methylation signature that tracks with hPSC

differentiation. Notably, the authors distill this discovery down to a practical assay that monitors reprogramming status and differentiation capacity of iPSCs.

> ChatGPT: In addition, computational biology and systems biology can also be used to optimize stem cell culture conditions, allowing researchers to grow stem cells more efficiently and with greater control. This can improve the quality of stem cells and make them more suitable for use in research and potential clinical applications.

Indeed, Nishimura et al. harvest and analyze single-cell RNA sequencing (scRNA-seq) data to create hypotheses to improve human DA neural differentiation.

> ChatGPT: Another advantage of computational biology and systems biology in stem cell research is the ability to create detailed models of stem cell behavior. These models can be used to simulate the behavior of stem cells under different conditions, providing valuable insights into how they respond to various stimuli and how they can be manipulated for different purposes. For example, researchers can use computational models to understand how stem cells differentiate into different cell types and how they can be directed to form specific tissues or organs.

In this issue, we have several articles that generate and use detailed models of cell regulation to predict their behavior. For example, Kamimoto et al. use a combination of gene regulatory network (GRN) inference, single-cell lineage tracing, and simulation to uncover how GRN topology of the source cell population influences cell-fate engineering and to identify novel factors to improve the reprogramming process. Similarly, Zheng et al. use prior regulatory networks and pre-defined signatures of responses to small molecules to identify compounds that modulate the activity of transcription factors to enact fate changes. Cockerell et al. review techniques and applications of biophysical modeling and simulations of mammalian embryogenesis, and Kaul et al. perform simulations on cell "agents" that explicitly account for intra-cellular factors (i.e., GRN topology) and extra-cellular factors (i.e., localization and morphogen concentration) to explore testable mechanisms of pattern formation in hPSC differentiation.

> Prompt:> What are the major challenges in using computational techniques to address questions in stem cell biology?

> ChatGPT:> The need for large amounts of data: To accurately model stem cell behavior, researchers need large amounts of data about the behavior of individual cells and the interactions between different cell types. This can be challenging to collect and process, especially when dealing with large numbers of cells.

One of the downsides to the modern era of genomics data is that the volume of data generated vastly exceeds the capacity of a single journal article to describe it in any depth. However, the wider adherence to policies for standardization and accessibility of data has enabled the re-use of previously published data. For example, Li et al. perform data mining on scRNA-seq data to explore the prevalence and role of alternative polyadenylation in early embryonic development in human and mouse. The accumulation of data also allows for retrospective meta-analyses, as performed by Becker et al. where they survey existing scRNA-seq data to establish the relevance of long non-coding RNAs (lncRNAs) in the ventricular-subventricular zone. An outstanding example of the power of combining massive datasets is the case of Kim et al., who harmonize multiple retina atlases to then compare protocols for generating retina organoids.

Here, we will end our conversation with ChatGPT. It was helpful in articulating how computational systems biology will aid in stem cell research and in identifying some obstacles. However, we excluded several threads of our conversation because they were repetitive, and overall, many of the responses lacked depth and insight. In fact, to be able to extract insight, especially causal insight, from sophisticated machine-learning computational tools, MacArthur argues that their use must be coupled to an explicit and sound theory. Haghverdi and Ludwig move in this direction in their review, in which they not only cover current mathematical models of cell fate from single-cell multiomics data but also propose the exploration of chaos theory as a starting point to handle the inherent complexity of development. Similarly, Okano et al. describe a mathematical theory of cell types in the context of scRNA-seq data.

In conclusion, we hope that you enjoy the exciting examples of computational systems biology as applied to stem cells in this special issue.

Patrick Cahan[1,2,*] and Barbara Treutlein[3,4,5,*]
[1]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205, USA
[2]Institute for Cell Engineering, Johns Hopkins University, Baltimore, MD 21205, USA
[3]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland
[4]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[5]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany
*Correspondence: patrick.cahan@jhmi.edu (P.C.), barbara.treutlein@bsse.ethz.ch (B.T.)
https://doi.org/10.1016/j.stemcr.2022.12.009