doi 10.18699/bgrs2024-12.1-09

The virtual cell

Kolpakov F.A.*, Akberdin I.R., Kutumova E.O.
Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia
** kolpakov.fa@talantiuspeh.ru*

Key words: virtual cell; mathematical modeling; systems biology; GTRD, BioUML

Motivation and Aim: With the beginning of the 21st century, systems biology, an interdisciplinary science that studies complex interactions in living systems, has received active development. Systems biology involves a specific cycle of research, consisting of theory, analytical or computer modeling to formulate hypotheses about the system, experimental testing of the hypotheses, and then using the obtained data to describe cellular processes, allowing to improve the computer model or refine the theory. Since the ultimate goal is to build a portrait model of interactions in a complex system, experimental techniques used in systems biology must be as detailed as possible. For this purpose, experimental approaches such as transcriptomics, metabolomics, proteomics and other high-throughput omics technologies are used to obtain numerical data. Thus, success in the development of high-throughput automated biomedical technologies has led to a renaissance in modeling biological systems at the molecular-cellular level.

The Virtual Cell project suggests a new approach that can be represented as a pyramid (Fig. 1) where each layer contains less information (10 to 10 000 times) than previous one.

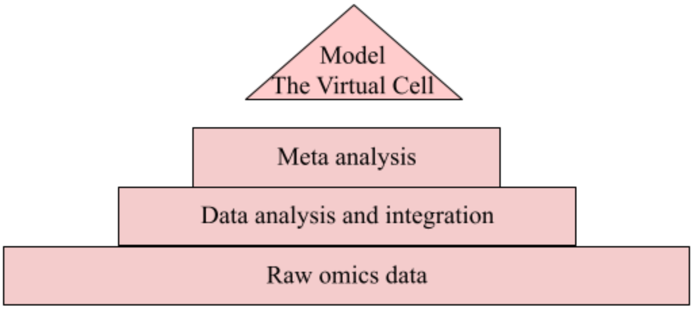

Fig. 1. The virtual cell as a model to aggregate and replace raw omics data

Finally, as the project goal, the constructed virtual cell model should accurately reproduce (replace) the experimental omics data obtained for a given human cell type (cell line):

- gene expression level (transcriptomics);
- protein quantities (proteomics);
- concentration of metabolites (metabolomics).

The developed virtual cell model verified on these omics data will serve as a basis for *in silico* experiments aimed at creating a state-of the-art rational strategy to fundamentally understand and improve the target properties of cells.

Methods and Algorithms: Architecture of The Virtual Cell shown in Figure 2. The key ideas are:

- agent based approach – main cellular processes (transcription, translation, metabolism, etc.) are presented by agents;
- different agents use different approaches to simulate their dynamics. For example, a neutral network is used to predict gene expression level using information from the GTRD [1], KnockTF [2] and FANTOM5 [3]. Atlas of human metabolism provides a genome-scale model of metabolite fluxes [4]. Ordinary differential equations are used to simulate cellular signal transduction pathways and energy metabolism [5]. Corresponding new plugin for BioUML platform [6] is developed;
- there are 3 main shared molecule pools: RNAs, proteins and their complexes and other small molecules and metabolites. They serve as input and output for all agents. Each agent specifies which molecules it accepts as inputs or outputs.
- iterative development of the model allows incrementally refine the model and add new agents;
- there are a number of the virtual cell builds that correspond different cell types (lines) as well as different iterations.

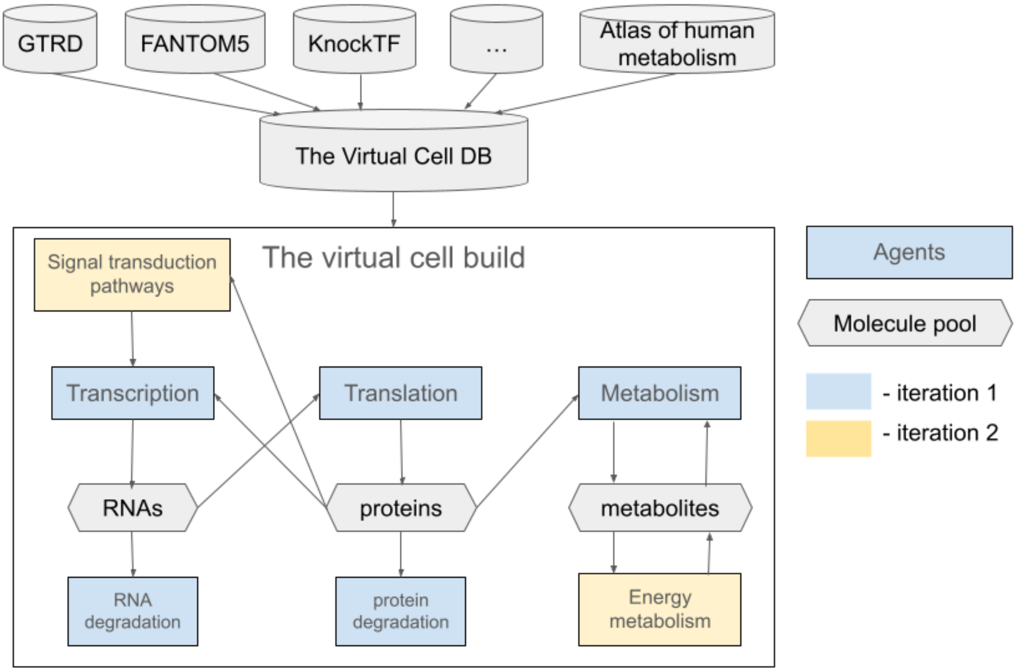

Fig. 2. The Virtual Cell architecture. First two iterations are shown

Conclusion: Here we are suggesting a new iterative approach to build a complicated computational model that should accurately reproduce (replace) the experimental whole genome omics data obtained for a given cell type. The Virtual Cell is an open source project available at https://gitlab.sirius-web.org/virtual-cell.

We look forward to working with other teams to join forces to tackle this challenging task.

Funding: The study was financially supported by the Russian Science Foundation (project No. 24-14-20031, https://www.rscf.ru/en/project/24-14-20031/).

References

1. Kolmykov S., Yevshin I., Kulyashov M. et al. GTRD: an integrated view of transcription regulation. *Nucleic Acids Res.* 2021;49(D1):D104-D111. doi 10.1093/nar/gkaa1057
2. Feng C., Song C., Liu Y. et al. KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. *Nucleic Acids Res.* 2020;48(D1):D93-D100. doi 10.1093/nar/gkz881
3. FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al. A promoter-level mammalian expression atlas. *Nature.* 2014;507:462-470. doi 10.1038/nature13182
4. Wang H., Robinson J.L., Kocabas P. et al. Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *PNAS.* 2021;118,e2102344118. doi 10.1073/pnas.2102344118
5. Akberdin I.R., Kiselev I.N., Pintus S.S. et al. A Modular Mathematical Model of Exercise-Induced Changes in Metabolism, Signaling, and Gene Expression in Human Skeletal Muscle. *Int J Mol Sci.* 2021;22(19):10353. doi 10.3390/ijms221910353
6. Kolpakov F., Akberdin I., Kiselev I. et al. BioUML-towards a universal research platform. *Nucleic Acids Res.* 2022;50(W1):W124-W131. doi:10.1093/nar/gkac286