

## PERSPECTIVE

# Building digital life systems for future biology and medicine

Xuegong Zhang<sup>1,2,3,\*</sup>, Lei Wei<sup>1</sup>, Rui Jiang<sup>1</sup>, Xiaowo Wang<sup>1,2</sup>, Jin Gu<sup>1</sup>, Zhen Xie<sup>1,2</sup>, Hairong Lv<sup>1</sup>

<sup>1</sup> MOE Key Lab of Bioinformatics and Bioinformatics Division of BNRIST, Department of Automation, Tsinghua University, Beijing 100084, China

<sup>2</sup> Center of Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China

<sup>3</sup> School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China

\* Correspondence: zhangxg@tsinghua.edu.cn

Received May 19, 2023; Accepted May 22, 2023

The rapid development of biological technology (BT) and information technology (IT) especially of genomics and artificial intelligence (AI) is bringing great potential for revolutionizing future medicine. We propose the concept and framework of Digital Life Systems or dLife as a new paradigm to unleash this potential. It includes the multi-scale and multi-granule measure and representation of life in the digital space, the mathematical and/or computational modeling of the biology behind physiological and pathological processes, and ultimately cyber twins of healthy or diseased human body in the virtual space that can be used to simulate complex biological processes and deduce effects of medical treatments. We advocate that dLife is the route toward future AI precision medicine and should be the new paradigm for future biological and medical research.

**Keywords:** digital life systems; digital twin; artificial intelligence; precision medicine

## INTRODUCTION

The understanding of life and the pursuit of health have always been a common goal of human beings' unrelenting exploration since ancient times. The exploration of human anatomy during the ancient Egyptian period opened the prelude to the decomposed research on the human body [1]. The approach of "looking, listening, inquiring, and pulse-taking" developed in ancient China around the 5th century BC started the holistic phenotypic probing of human body as a system. During the long journey, most of our ancestors' efforts for understanding human body and combating diseases had been based on naked eye observations and philosophical reasoning. The invention of microscopes and its use for medical observations in 1674 [2], and the invention of stethoscope in 1816 [3] marked the start of human's use of scientific equipment to help inspecting human body. The trend was accelerated with the revolution of modern science and technology in the 20th

century. The X-ray discovered by Wilhelm Roentgen in 1895 soon became a major technology to look inside the human body without surgical operation [4]. Electrocardiogram (ECG) [5] and electroencephalogram (EEG) [6] was invented in 1903 and 1929, respectively, followed by the invention of medical ultrasound technology in 1953 [7]. In 1971, computer tomography (CT) and magnetic resonance imaging (MRI) technology came into clinical applications [8,9]. These technologies for clinical observation of human physiology and pathology, together with the many drugs and surgical treatment technologies in the 20th century, as well as the development of industrial and agricultural production and public health infrastructures, jointly promoted the human life expectancy from around 30–40 years at the beginning of the 20th century to around 70–80s in the 21st century [10].

At the same time, people's exploration on the biology of life was also quickly deepening. Mendel discovered the basic laws of inheritance in 1865 [11]. The discovery

of DNA, RNA and proteins in the 19th century seeded the understanding of the material basis behind inheritance [12–14]. This year celebrates the 70th anniversary of the discovery of DNA double-helix, and the 20th anniversary of the completion of Human Genome Project. The discovery of DNA double helix structure in 1953 confirmed that DNA is the basic carrier of the genetic information [15]. In 1968, Wu measured the first DNA sequence [16], opening up the journey of human beings to read our own genomic information. The Human Genome Project (HGP) completed from 1990 to 2003 [17] established a reference map of all human genetic information, led to the development of many genomics and other omics technologies, and made it possible to carry out precision medicine based on individual's omics characteristics. We can now not only to read, but also edit or write the genomic information. The Human Cell Atlas (HCA) project and several other similar projects, driven by the development of single-cell sequencing technology, aims to decipher the molecular characteristics of various cell types in all tissues and organs as the reference for all future biological and medical studies [18–20].

These great progresses in medical science and biology have significantly improved our ability to combat diseases. However, with the increase in people's life expectancy and the ever-upgrading pursuit of better life quality, there are new challenges for improving the effectiveness and efficiency of medicine. For example, the morbidity and mortality of cardiovascular diseases in many countries still remain high, and that of many malignant tumors are still rising [21,22]. The accessibility of high-quality medical services varies significantly among different regions in China and many other developing countries, and there is a huge gap in management and rehabilitation of many chronic diseases outside hospitals. These challenges can be summarized as the insufficient and unbalanced development of medical sciences and services. They are rooted in the lacking of sufficient understanding of the biology of diseases, and the lacking of effective and efficient applications of existing knowledges and technologies. New paradigms in the way we conduct research in human biology and medicine are needed to better tackle these challenges.

## **KEYS TO AI MEDICINE: INFORMATION ACQUISITION, SYSTEM MODELING AND KNOWLEDGE DELIVERY**

The rapid development of artificial intelligence (AI) in the past two decades, especially the explosive growth of large AI models like generative pretraining transformers (GPTs) in recent years, has aroused great expectation on

using AI to solve challenges in medicine [23]. Many advancements have been made toward this direction but there is still a long way to go before AI can really help to provide systematic solutions to the challenges. The nature of medicine and the biology behind medicine have many unique features comparing to other fields where AI has shown great success. The realization of AI medicine requires the solution of three key questions: information acquisition, system modeling, and knowledge delivery.

The human body is a super complex system composed of trillions of cells organized in multiple hierarchies and networks. Each cell is programed with tens of thousands of genes and many other genomic and epigenomic factors. Obtaining information of multiple dimensionalities, multiple angles and multiple granularities of the human body in health and diseases is the basis for understanding how the system works. Modern clinical inspection and imaging technologies provide rich phenotypic information of human organs, systems and the whole body. Current biological technologies such as high-throughput omics can provide genomic, transcriptomic, epigenomic, proteomic and metabolic information of tissues and organs at single-cell and single-molecular resolution. Various wearable and implantable devices being developed rapidly in recent years are expected to provide real-time physiological surveillance of the human body and measurement of life styles or environmental factors. The collection of all these information will eventually provide a holographic portrait of an individual at all levels. Currently many of these technologies are still imperfect, and the information we can get for a patient is limited. New developments in physics, chemistry, material science, and information sciences are constantly catalyzing technological advancements in ways to obtain information of life.

Capturing multiscale, multifaceted and multidomain information of a human body is the first step for understanding life, but the goal cannot be reached without the decoding and modeling of the underlying connections and relations in the information. Scientists have discovered tons of biological and medical knowledge, but most of them are scattered, local, and qualitative. They have not yet been developed to the level of being described by mathematical models or being implemented by computing models. A central step toward AI medicine is to achieve quantitative understanding of complex biological phenomena and underlying laws and to establish their mathematical and/or computational models, based on the ever-growing biological/medical data and knowledge. Such models should mirror real life by being able to reproduce or simulate major biological processes and mechanisms in the digital space.

Beyond information acquisition and system modeling, an equally important but often overlooked aspect in medicine is the effective presentation of knowledge and implementation of technologies in practice. With the advancement of modern biology and medicine, the information and knowledge required for today's and future medical decision-making are increasingly exceeding the level that an average human doctor can effectively handle. This problem will only get more and more prominent if biological and medical research continues to follow the current paradigm of depositing new discoveries into the literature for human doctors to digest and apply. Figure 1 illustrates the situation under the current paradigm: scientific research finds answers to part of the questions about life and health. Only a small portion of the findings can be transferred to applicable principles, drugs, technologies, or equipment with clinical potentials. Among them, that part can really pass clinical trials and go into practice. Due to technical, administrative or financial reasons, even a smaller part of them can finally benefit specific patients. The effect can be further discounted because of factors like differences in individuals' biological and socio-economic backgrounds and their compliance. This status quo indicates that the existing paradigm of centralizing all information and knowledge to clinicians for decision-making and technology-implementation has encountered a ceiling problem. More effective knowledge delivery and application paradigms need to be invented.

### DIGITAL LIFE SYSTEMS: A CYBER-LIFE MODEL AND THE BIOLOGICAL AND MEDICAL STUDY *in data, in silico* AND *in math*

We propose the concept and framework of “Digital Life Systems” or dLife to tackle the three key problems. The idea is to build and run models of life in the virtual

space defined by data, mathematics and computation. All biological discoveries will be deposited into dLife models. Clinicians can use the models to simulate and test alternative treatments to help make decisions and evaluate treatment effects. Individuals can also use the models to know themselves in health or diseases.

Real biological human beings are carbon-based life. We named dLife as “Shu-based Life” in Chinese, where “Shu” is the Chinese character “数” that possesses the meanings of *data*, *digital*, *quantitative*, *mathematics* and *computing* at the same time. There seems to be no single equivalent word in English that has all these meanings. dLife system is a novel model of life that we propose to build, and it is also a new paradigm of life science research that we initiate. As a model it is the reproduction of life phenomena or a cyber-life in the virtual space. As a research paradigm it is study of human biology and medicine *in data*, *in silico* and *in math* beyond current *in vivo* and *in vitro* approaches. The existing term “*in silico*” studies share a similar meaning with dLife partially. We coined the new term “*dLife*” because silicon-based computers may not be the only technology for performing computation. dLife emphasizes more on the systematic mirroring of carbon life systems in the data/math/computing space from molecular and cellular levels to the level of clinical phenotypes.

A complete dLife system will be composed of three major layers, corresponding to the three key questions for AI medicine.

The first and basic layer of dLife is the perception and digitization of biological/medical information. The hierarchical components that compose the human life will be represented as entities or tokens in the digital space, with holographic portraits describing their cellular, molecular and physicochemical properties at multiple scales, with multi-modal data and from angles of other related entities.

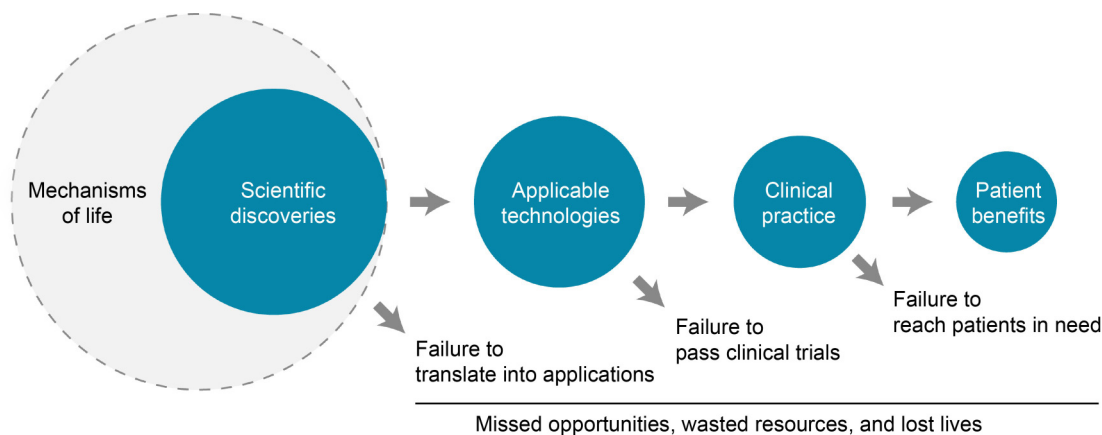


Figure 1. The current medical paradigm. (Figure designed inspired by a figure in [24].).

The second and core layer of dLife is system modeling, which makes the entities live and interact with each other to reproduce life phenomena in the virtual space. The modeling will be cross multiple levels, scales and granularities, and will include both mathematical models for well-deciphered relations and large neural-net-like models trained with massive data. The complete model of dLife will be a general cyber life that can represent major developmental, physiological or pathological phenomena and processes in the data/math/computing spaces.

The third layer of dLife is the intelligent implementation of dLife as a personalized cyber-twins or digital twins of individuals or groups of individuals sharing the same biological properties. Unlike digital twins in the field of manufacturing, dLife can simulate not only the geometric and physical properties of the individual, but more importantly the molecular and cellular properties of key tissues and organs that are responsible for the medical phenotypes of interest. It mirrors not only the body but also the life inside the body. This layer involves the individualization of the general dLife model with the biological, medical and life-style data of individuals, and the fine-tuning of general models pretrained with population data for specific downstream

tasks. Such digital cyber-twins will open new windows for both healthcare applications and for medical researches. Doctors can use the cyber-twins interact with generic knowledge and individual's information, to simulate, deduce and compare different treatment strategies for each individual patient, and researchers can use cyber-twins to do *in data* or *in-silico* experiments on cyber individuals to study certain biological mechanism, or on a cyber population to conduct virtual clinical trials of a drug candidate. Figure 2 diagramed a conceptual structure of dLife systems and examples of their potential applications.

Building complete dLife systems is a long-term goal. However, we do not have to wait for the completion of a full system to unfold its advantages in scientific research and medical practices. For example, we have recently succeeded in building the first version of human Ensemble Cell Atlas (hECA 1.0) as a prototypic example of a dLife system [20]. It is composed of ~1 million cells of 38 healthy human organs featured by their transcriptomic properties, organized as an abstract virtual human body of digital cells. Based on this virtual body, we invented the *in data* cell experiment technique that enables testing for possible side-effects for targeted cancer therapy [20]. This invention showcased the future

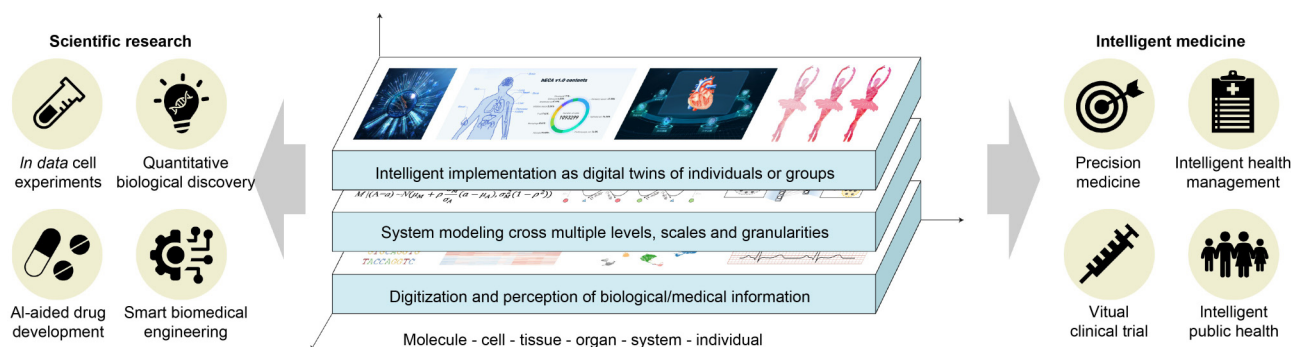


Figure 2. A conceptual structure of dLife systems.

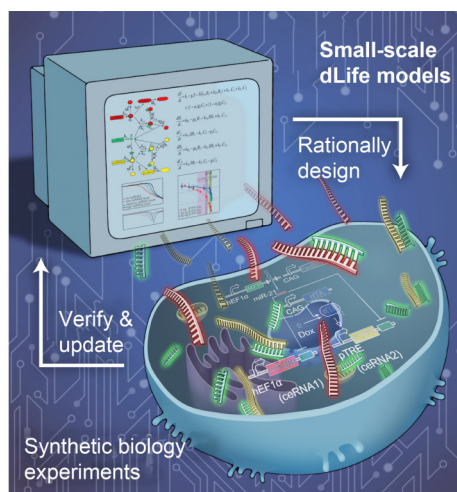


Figure 3. Showcases of *in data* drug experiment using dLife models.



of *in data* drug experiment and virtual clinical trial using dLife models (Fig. 3). In another example, following the conceptual idea of dLife, we were also able to define new quantitative index on CT images of non-small cell lung cancer (NSCLC) patients to reveal the internal heterogeneity of tumors that are crucial for making treatment decisions [25].

dLife also provides a new paradigm of future biological research by linking the carbon space of cells and genes with the digital space of numbers and vectors. Small-scale dLife models can be built for specific biological questions using available data and knowledge. New biological laws can be speculated from mathematical and/or computational simulations in the digital space, and then verified or updated through rationally designed synthetic biology experiments. We call this paradigm of biological study as “*digi-carbon experiments*”. In our recent works, we have applied this paradigm to discover gene regulation patterns and design synthetic gene circuits for targeted cancer therapy (Fig. 4), which is a significant breakthrough for future intelligent drug development [26–29].



**Figure 4.** A schematic diagram of *digi-carbon experiments* using dLife models.

## RELATED CONCEPTS AND ROADMAP TOWARD DIGITAL LIFE SYSTEMS

There have been several new terms such as medical digital twins, digital humans and digital biology in recent years. They have some similarity in the concepts and terminology with the digital life system or dLife we proposed, but they have different intension and extensions. “Digital humans” is a term widely used in the Virtual Reality (VR) community, usually to refer the realization of a character with a 3D human-looking body

and face in artificially generated videos that can perform some human-like motions or facial and language interactions with users [30,31]. “Digital biology” has been first used in a National Institutes of Health (NIH) conference in 2003 [32]. It was used to mean the broad field of computationally-oriented bio-sciences, with key areas of scientific data integration, multi-scale modeling and networked science [32]. In recent years, medical scientists have realized the future of using digital twins to help improve medical decisions. For example, Corral-Acero *et al.* proposed the idea of digital twin for precision cardiology [33]. Subramanian proposed the idea of developing a virtual liver as a digital twin for drug discovery and development [34]. Masison *et al.* proposed a modular computational framework for medical digital twins and demonstrated it by a respiratory fungal infection case [35]. These proposals focus on modeling the electrophysiological or metabolic functions of the target organs and their associations with patient outcomes. These are important aspects of the dLife system we first proposed in 2019, which aimed to model all the biological processes of life that transform us from molecules (DNAs and RNAs) to the whole human beings [36]. dLife is a digital twin not only of the human body, but also of all biological processes that are carrying on with this body at molecular, cellular, and system levels.

Such a digital life system or dLife is an ambitious scientific goal that requires long collaborative research and innovation of multiple disciplines: new theories and technologies need to be developed for comprehensive multi-scale data sensing, data integration, data perception, knowledge discovery and knowledge presentation. A basic informatic framework or “operation system” needs to be designed for dLife models and their implementations. Special large AI models capable of learning complex relations buried in massive multimodality biological and medical data need to be developed. New mathematical tools and theories may even be needed to capture and represent complex biological laws.

None of these works will be trivial or easy because of the complexity of life and the lack of first-principles behind most biological mechanisms. The recent breakthrough of large language models like GPTs in natural language processing and understanding suggests a possible roadmap for pursuing the seemingly infeasible ambitious goal. It will be important to find or define appropriate “proxy tasks” to help developing the fundamental theories, frameworks and technologies. Such proxy tasks should be simple enough so that it is possible to invent dLife solutions based on current technologies and data. But they should be sophisticated enough so that they can reflect the multifaceted

complexity of real biological and medical tasks, and can showcase the advantage of the dLife system. Representative real application scenarios of AI precision medicine are also crucial for motivating dLife system research, guiding the selection of proxy tasks and driving the technology development.

## ACKNOWLEDGEMENTS

This work was partially supported by the National Natural Science Foundation of China (NSFC) (Nos. 61721003 and 62250005), the National Key R&D Program of China (No. 2021YFF1200900), and Tsinghua-Fuzhou Institute for Data Technology (No. TFIDT2021005).

## COMPLIANCE WITH ETHICS GUIDELINES

**Conflicts of interest** The authors Xuegong Zhang, Lei Wei, Rui Jiang, Xiaowo Wang, Jin Gu, Zhen Xie, and Hairong Lv declare that they have no conflict of interest or financial conflicts to disclose.

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

## OPEN ACCESS

This article is licensed under the CC BY under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

- Loukas, M., Hanna, M., Alsaiegh, N., Shoja, M. M. and Tubbs, R. S. (2011) Clinical anatomy as practiced by ancient Egyptians. *Clin. Anat.*, 24, 409–415
- Ford, B. J. (1982) Bacteria and cells of human origin on van Leeuwenhoek's sections of 1674. *Trans. Am. Microsc. Soc.*, 101, 1–9
- Kligfield, P. (2016) The Bicentennial of the Stethoscope: 1816 to 2016. *Am. J. Cardiol.*, 118, 1601–1602
- Mould, R. F. (1995) The early history of X-ray diagnosis with emphasis on the contributions of physics 1895–1915. *Phys. Med. Biol.*, 40, 1741–1787
- Fye, W. B. (1994) A history of the origin, evolution, and impact of electrocardiography. *Am. J. Cardiol.*, 73, 937–949
- Stone, J. L. and Hughes, J. R. (2013) Early history of electroencephalography and establishment of the American Clinical Neurophysiology Society. *J. Clin. Neurophysiol.*, 30, 28–44
- Edler, I. and Hertz, C. H. (2004) The use of ultrasonic reflectoscope for the continuous recording of the movements of heart walls. *Clin. Physiol. Funct. Imaging*, 24, 118–136
- Beckmann, E. C. (2006) CT scanning the early days. *Br. J. Radiol.*, 79, 5–8
- Ai, T., Morelli, J. N., Hu, X., Hao, D., Goerner, F. L., Ager, B. and Runge, V. M. (2012) A historical overview of magnetic resonance imaging, focusing on technological innovations. *Invest. Radiol.*, 47, 725–741
- Riley, J. C. (2005) Estimates of regional and global life expectancy, 1800–2001. *Popul. Dev. Rev.*, 31, 537–543
- Ellis, T. H. N., Hofer, J. M. I., Timmerman-Vaughan, G. M., Coyne, C. J. and Hellens, R. P. (2011) Mendel, 150 years on. *Trends Plant Sci.*, 16, 590–596
- Dahm, R. and Discovering, D. N. A. (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.*, 122, 565–581
- Thess, A., Hoerr, I., Panah, B. Y., Jung, G. and Dahm, R. (2021) Historic nucleic acids isolated by Friedrich Miescher contain RNA besides DNA. *Biol. Chem.*, 402, 1179–1185
- Hartley, H. (1951) Origin of the word “protein”. *Nature*, 168, 244
- Watson, J. D. and Crick, F. H. C. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171, 737–738
- Wu, R. (1970) Nucleotide sequence analysis of DNA. I. Partial sequence of the cohesive ends of bacteriophage lambda and 186 DNA. *J. Mol. Biol.*, 51, 501–521
- Collins, F. S., Morgan, M. and Patrinos, A. (2003) The Human Genome Project: lessons from large-scale biology. *Science*, 300, 286–290
- Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., *et al.*, (2017) The Human Cell Atlas. *eLife*, 6, e27041
- HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature*, 574, 187–192
- Chen, S., Luo, Y., Gao, H., Li, F., Chen, Y., Li, J., You, R., Hao, M., Bian, H., Xi, X., *et al.* (2022) hECA: the cell-centric assembly of a cell atlas. *iScience*, 25, 104318
- Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V. and Roth, G. A. (2022) The global burden of cardiovascular diseases and risk: a compass for future health. *J. Am. Coll. Cardiol.*, 80, 2361–2371
- Siegel, R. L., Miller, K. D., Fuchs, H. E. and Jemal, A. (2022) Cancer statistics, 2022. *CA Cancer J. Clin.*, 72, 7–33
- Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J. and Rajpurkar, P. (2023) Foundation models for generalist medical artificial intelligence. *Nature*, 616, 259–265
- Committee on the Learning Health Care System in America & Institute of Medicine. (2013) Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington

- (DC): National Academies Press
25. Li, J., Qiu, Z., Zhang, C., Chen, S., Wang, M., Meng, Q., Lu, H., Wei, L., Lv, H., Zhong, W., *et al.* (2022) ITHscore: comprehensive quantification of intra-tumor heterogeneity in NSCLC by multi-scale radiomic features. *Eur. Radiol.*, 33, 893–903
  26. Yuan, Y., Liu, B., Xie, P., Zhang, M. Q., Li, Y., Xie, Z. and Wang, X. (2015) Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proc. Natl. Acad. Sci. USA*, 112, 3158–3163
  27. Wei, L., Yuan, Y., Hu, T., Li, S., Cheng, T., Lei, J., Xie, Z., Zhang, M. Q. and Wang, X. (2019) Regulation by competition: a hidden layer of gene regulatory network. *Quant. Biol.*, 7, 110–121
  28. Huang, H., Liu, Y., Liao, W., Cao, Y., Liu, Q., Guo, Y., Lu, Y. and Xie, Z. (2019) Oncolytic adenovirus programmed by synthetic gene circuit for cancer immunotherapy. *Nat. Commun.*, 10, 4801
  29. Wei, L., Li, S., Zhang, P., Hu, T., Zhang, M. Q., Xie, Z. and Wang, X. (2021) Characterizing microRNA-mediated modulation of gene expression noise and its effect on synthetic gene circuits. *Cell Rep.*, 36, 109573
  30. Demirel, H. O., Ahmed, S. and Duffy, V. G. (2022) Digital human modeling: a review and reappraisal of origins, present, and expected future methods for representing humans computationally. *Int. J. Hum. Comput. Interact.*, 38, 897–937
  31. Campbell, M. and Jovanovic, M. (2022) Digital self: the next evolution of the digital human. *Computer*, 55, 82–86
  32. Morris, R. W., Bean, C. A., Farber, G. K., Gallahan, D., Jakobsson, E., Liu, Y., Lyster, P. M., Peng, G. C. Y., Roberts, F. S., Twery, M., *et al.* (2005) Digital biology: an emerging and promising discipline. *Trends Biotechnol.*, 23, 113–117
  33. Corral-Acero, J., Margara, F., Marciniak, M., Rodero, C., Loncaric, F., Feng, Y., Gilbert, A., Fernandes, J. F., Bukhari, H. A., Wajdan, A., *et al.* (2020) The “digital twin” to enable the vision of precision cardiology. *Eur. Heart J.*, 41, 4556–4564
  34. Subramanian, K. (2020) Digital twin for drug discovery and development—the virtual liver. *J. Indian Inst. Sci.*, 100, 653–662
  35. Masison, J., Beezley, J., Mei, Y., Ribeiro, H., Knapp, A. C., Sordo Vieira, L., Adhikari, B., Scindia, Y., Grauer, M., Helba, B., *et al.* (2021) A modular computational framework for medical digital twins. *Proc. Natl. Acad. Sci. USA*, 118, e2024287118
  36. Mussomeli, A., Parrott, A., Umbenhauer, B. and Warshaw, L. (2020) Deloitte. *TechTrends*, 2020, 70–71