Quantitative Biology 2022, 10(1): 1–5
https://doi.org/10.15302/J-QB-022-0290

PROFILE

Yanda Li: A never-stopping pioneer, investigator and founder

Xuegong Zhang[1,2,*], Jin Gu[1]

[1] MOE Key Laboratory of Bioinformatics, Bioinformatics Division, BNRIST and Department of Automation, Tsinghua University, Beijing 100084, China
[2] School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China
* Correspondence: zhangxg@tsinghua.edu.cn

Received December 7, 2021

Yanda Li was born in October 1936 in Dongguan, Guangdong, China. His childhood was in war time. He had to move with his parents across several counties and cities, but he managed to accomplish his high-school study in Guangzhou (Fig. 1). He showed deep interest in physics in high school and was admitted to the Department of Electrical Engineering of Tsinghua University in 1954. Because of his excellent performance in college study, he joined the faculty in 1958 before he finished his college study in 1959. Throughout the 60 years from his joining the faculty to his retirement in 2018, Prof. Li has pioneered in a broad range of academic fields and is a founder of several new research fields in Tsinghua University.

Figure 1. Yanda Li in his high-school age. Credit: Yanda Li, unknown photographer.

Yanda Li worked on electronics and automatic control in his early career. In 1959, he participated in the maintenance and operation of China's first computer-controlled machine tool. Later he led a group of graduate students to develop the first computer for controlling drilling machines in China.

In 1978 when China started to open its door to the outside world, Yanda Li passed the tough selection and gained the opportunity to study at Massachusetts Institute of Technology (MIT) as one of the first visiting scholars from Chinese mainland. His research interest switched from electronics to the field of signal processing, and joined the lab of Dr. Alan V. Oppenheim. That was the period when Yanda Li started to do real cutting-edge research. He chose to work in the field of geophysical signal processing for petroleum exploration. He found that phase spectrum has close relations with wave arrival delays of seismic prospecting signals, and invented methods for recovering subsurface structure information from wave delays. The work had many impacts on other fields of signal processing, and was later summarized as the monograph *Signal Reconstruction Theory and Applications* [1].

Yanda Li returned to Tsinghua University in 1981 after two years' study at MIT. He built a research team on petroleum exploration signal processing and collaborated with the major petroleum industry. Soon he and his team realized that the key limitation of seismic signal

© The Authors (2022). Published by Higher Education Press

processing for oil and gas exploration is not in the power of data processing methods, but in the information contained in the signals. While advanced signal processing methods may help to improve the resolution of seismic signals to a certain extent, the increase of information is more important for improving the prediction accuracy on oil and gas reservoirs. He proposed to integrate information from well-logging data with seismic data, and integrate data with geological knowledge. He led his team to develop pattern recognition and neural network methods for integrating multi-modal signals with knowledge, which opened a new approach for the task and formed a new branch of Intelligent Information Processing. The new approaches were very successful in several major petroleum fields in China [2–4]. In 1992, Yanda Li organized the first International Symposium on Intelligent Information Processing (ISIIP'92) in Beijing.

In 1994, the Internet entered China, but was still far from getting widely accessible. However, Yanda Li realized that the Internet would be the future of communication, and predicted that the processing of information on the Internet would be a big challenge and promising application field for intelligent information processing. He started a new research group on the Internet Information Processing. This group later became the seed of a major research lab in the department on multimedia technology and incubated a high-tech company listed at NYSE.

The success in those fields has not stopped Yanda Li from further exploring new research areas. Around 1996, the rapid accumulation of genomics data propelled by the Human Genome Project caught his attention. He talked to many biologists and looked into the information structure of the genome. He found that, on one hand, there are many challenges for intelligent information processing in genomics data, and on the other hand, life science is in nature an information science: the key to understanding biology is to understand how the information underlying life works. At that time, the field of bioinformatics and computational biology had not been a widely recognized research discipline yet in the world, and was not known by most people in China. Yanda Li spent tremendous efforts for several years to promote this discipline in Tsinghua University and in the whole China. He founded the Bioinformatics Institute across the Department of Biology and Department of Automation, which became the first Ministry of Education (MOE) Key Laboratory of Bioinformatics in China in 2002. He sent young faculty member Dr. Xuegong Zhang to visit Prof. Wing H. Wong's lab at Harvard University in 2001–2002 to systematically study computational biology, and invited Dr. Michael Q. Zhang from Cold Spring Harbor Laboratory as a Guest Professor from 2003 to help Tsinghua University build this discipline. In 2007, Yanda Li (Fig. 2) encouraged Xuegong Zhang to organize a month-long full-time Graduate Summer School on Bioinformatics of China (GSSBC'07). It trained about 200 graduate students from ~80 institutions all over China. Many of them nowadays become academic or industry leaders in this field. In 2008, Yanda Li, Xuegong Zhang and Michael Q. Zhang established a Chair Professor Team on bioinformatics in Tsinghua, with members including Michael S. Waterman, Wing H. Wong, Michael Q. Zhang, Jun S. Liu, Tao Jiang, Thomas Gingeras, Fengzhu Sun, Bin Ren and Tim Chen. With all these great efforts, Tsinghua University has become a major base for bioinformatics research and education with world-wide reputation.

Bioinformatics is a very wide discipline. Yanda Li chose two unique angles at the beginning: one is the analysis of the long-range information structure of the human genome, and the other one is the computational analysis of alternative splicing. His work on using signal processing theory and methods to find long-range correlations in the genome [5] was about two decades earlier than the time that long-range interaction and 3D genomics became popular. The database AsMamDB they developed [6] was the first alternative splice database of mammals. And they were also the first to elaborate the impact of very short alternative splicing events on protein structures and functions [7]. These pioneering works had made major contributions in the systematic study of alternative splicing, long before the importance of alternative splicing had received sufficient attention after tiling microarray, and RNA-sequencing technologies proved that the majority of human genes could undergo alternative splicing.

MicroRNAs (miRNAs) are a type of regulatory RNAs that play major roles in the post-transcriptional

Figure 2. Yanda Li giving lecture at GSSBC'07. Photo by Xuegong Zhang.

© The Authors (2022). Published by Higher Education Press

regulation of gene expression. Their importance is widely known nowadays but was less recognized in the early 2000s, until Andrew Fire and Craig Mello won the Nobel Prize in 2006 for their discovery of RNA interference (RNAi) which has a similar mechanism with miRNAs in regulating gene expression. Prof. Li's lab started to work on miRNAs since 2003 and 2004. At that time, computational prediction of miRNAs from genomic sequences was a major approach for finding new miRNAs in a genome as no experimental method was available for high-throughput discovery of new miRNAs. Prof. Li guided students from his lab and Xuegong Zhang's group to work on this challenge and they developed two computational methods for miRNA prediction: miRAlign to identify new miRNAs based on sequence and structure alignment [8], and Triplet-SVM for classification of real and pseudo miRNAs without alignment with known miRNAs [9]. Those methods achieved the best performances at that time for identifying new miRNAs by homology searching and by *ab initio* discovery, respectively, and have made major contributions to the field. Besides, Li's lab has contributed a series of early studies miRNA expression patterns (*e.g.* [10–13]). Many of the students who did those work under Prof. Li's supervision have now become active leaders in the fields, such as Xiaowo Wang and Jin Gu.

Another field that Prof. Li had made major contributions is the study of epistatic effects of multiple SNPs or genes in association studies. In 2002–2005, many researchers rushed into the field of GWAS (genome-wide association studies), but most focused on looking for the association of single loci with disease or other phenotypes. Yanda Li believed that complex human diseases must be associated with the interaction of multiple genes rather than single gene. Collaborating with Prof. Yan Shen's lab at Peking Union Medical College (PUMC), they developed a novel strategy for complex diseases and found the association of an SNP combination pattern in the dopaminergic pathway in paranoid schizophrenia [14, 15]. Yanda Li found the lack of effective methods for association studies with multiple genes. He advised students to develop a method for analyzing multi-locus penetrance variance for association study in complex diseases [16]. Later, he and Rui Jiang co-advised students to develop a new method for epistatic module detection in association studies using a Bayesian model with Gibbs sampling strategy [17], and a method for integrating multiple genomic data to predict-disease causing nonsynonymous SNPs in exome sequencing studies [18].

Yanda Li was among the earliest scientists to propose that biological systems, especially gene regulation and the effect of genes on phenotypes can only be understood from a systematic and network view (Fig. 3). He was the first to propose a mechanism of splicing site competition in determining whether a splice site is alternative or constitutive [19]. The idea of possible competitions among biological elements had far-reaching influences. Partially inspired by the idea, Xiaowo Wang and Zhen Xie, together with Yanda Li and Michael Q. Zhang, introduced the mechanism of competition in modeling miRNA-gene regulation, and proposed a quantitative model for the regulation system using computational and synthetic experiments [20]. Based on the systematic view of biology, Yanda Li and Michael Q. Zhang propelled the development of systems and synthetic biology as a major new research direction. It has become a major new growing point in Tsinghua University's interdisciplinary research both in the Department of Automation, the School of Medicine and the School of Life Sciences.

A unique contribution of Yanda Li to the whole field of biomedical research is the promotion of studies on Traditional Chinese Medicine (TCM). Since the early days of his research in bioinformatics, he believed that bioinformatics and systems biology could provide the answer to many questions in the modernization of TCM. He recruited TCM-trained researcher Shao Li to his team as a postdoc fellow in 2001–2003 and started to explore ways to integrate information science, systems science, biology and bioinformatics for revealing the mystery of TCM. Under the supervision of Yanda Li and Yongyan Wang, Shao Li proposed a new framework for understanding the key concept of Cold and Hot ZHENG (or TCM Syndrome) in the context of the neuro-endocrine-immune network [21]. With continuous efforts of about a decade, they formed a comprehensive theory of TCM network pharmacology that can explain the underlying mechanism of TCM. TCM has a long history of viewing and treating each individual patient as an imbalanced system. The reductionists' "one target, one drug" approach cannot succeed in explaining observations in TCM. To reveal the possible molecular mechanism of TCM drugs that have been proven to be effective in the long period of clinical practice, they proposed the novel concept of "network target, multi-component drug". This revolutionary approach has led to the discovery of the possible network regulation mechanisms of several classical TCM herbal formulae. It not only opened the door for the TCM modernization from the perspective of computational and systems biology, but also pointed out a new direction for inferring complex relationships underlying various types of biological observations [22].

This year has celebrated the 86th birthday of Prof. Yanda Li. Although he officially retired in 2018, he is still active in participating academic activities and

© The Authors (2022). Published by Higher Education Press

Figure 3. Yanda Li giving a lecture in 2011. Photo by Xuegong Zhang.

discussions with young faculty members and students. He proposed that the key to understand complex systems of life from data is to discover underlying relations in seemingly unrelated data [22]. He often advises young researchers not to limit their imagination by the boundary of disciplines, and not to be afraid of thinking on questions and solutions that no one has ever touched on. In the past 60 years, Prof. Li has mentored more than 100 graduate students. Many of them have now become faculty members in many top universities globally, and many others have become industrial leaders in related fields.

Among the many founding contributions in education and research, I would like to especially acknowledge the fundamental contribution of Prof. Yanda Li to the launching of our journal *Quantitative Biology* (QB). The prehistory of QB was that the Higher Education Press asked Prof. Yanda Li to think for starting a new English language academic journal. Yanda Li thought the interdisciplinary research of information sciences and life sciences would be the future and invited Professors Michael Q. Zhang and Chao Tang to act as Editors-in-Chief for this new journal. Both of Profs. Michael Q. Zhang and Chao Tang agreed that all future life science studies should and will be quantitative, and should be able to be described in some form of mathematics. That was how QB was founded. Nowadays, our journal has entered her 10th year. We greatly appreciate the pioneering vision of Yanda Li. We wish Prof. Yanda Li a very healthy and happy life, and all the academic seeds he sowed in the past decades will grow more prosperous in the future.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Xuegong Zhang and Jin Gu declare that they have no conflict of interest.

OPEN ACCESS

This article is licensed by 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

1. Li, Y. and Chang, T. (1991) Signal Reconstruction: Theory and Applications (in chinese). Beijing: Tsinghua University Press
2. Zhang, X. and Li, Y. (1994) The application of artificial neural networks in editing noisy seismic data, chapter 7. In: Soft Computing: Fuzzy Logic, Neural Networks, and Distributed Artificial Intelligence, Aminzadeh, F. and Jamshidi, M. eds. Englewood: Prentice Hall
3. Zhang, X., Li, Y. and Liu, F. (1995) Estimating reservoir's lithological parameters from seismic data using neural network. SEG 65th Annual Meeting, Houston, USA
4. Lu, W., Zhang, X. and Li, Y. (1999) Petroleum reservoir framework prediction by information fusion. In: Proc. 2nd Int. Conf. On Information Fusion, Sunnyvale, USA. pp. 1090–1093
5. Lu, X., Sun, Z. R., Chen, H. M. and Li, Y. D. (1998) Characterizing self-similarity in bacteria DNA sequences. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics, 58, 3578−3584
6. Ji, H., Zhou, Q., Wen, F., Xia, H., Lu, X. and Li, Y. (2001) AsMamDB: an alternative splice database of mammals. Nucleic Acids Res., 29, 260−263
7. Wen, F., Li, F., Xia, H., Lu, X., Zhang, X. and Li, Y. (2004) The impact of very short alternative splicing on protein structures and functions in the human genome. Trends Genet., 20, 232−236
8. Wang, X., Zhang, J., Li, F., Gu, J., He, T., Zhang, X. and Li, Y. (2005) MicroRNA identification based on sequence and structure alignment. Bioinformatics, 21, 3610−3614
9. Xue, C., Li, F., He, T., Liu, G. P., Li, Y. and Zhang, X. (2005) Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. BMC Bioinformatics, 6, 310
10. Gu, J., He, T., Pei, Y., Li, F., Wang, X., Zhang, J., Zhang, X. and Li, Y. (2006) Primary transcripts and expressions of mammal intergenic microRNAs detected by mapping ESTs to their flanking sequences. Mamm. Genome, 17, 1033−1041
11. Gu, J., Fu, H., Zhang, X. and Li, Y. (2007) Identifications of conserved 7-mers in 3′-UTRs and microRNAs in Drosophila. BMC Bioinformatics, 8, 432
12. Wang, X., Zhang, X. and Li, Y. (2008) Complicated evolutionary

© The Authors (2022). Published by Higher Education Press

patterns of microRNAs in vertebrates. Sci. China C Life Sci., 51, 552−559

13. Wang, X., Gu, J., Zhang, M. Q. and Li, Y. (2008) Identification of phylogenetically conserved microRNA cis-regulatory elements across 12 *Drosophila* species. Bioinformatics, 24, 165−171
14. Xu, Q., Jia, Y. B., Zhang, B. Y., Zou, K., Tao, Y. B., Wang, Y. P., Qiang, B. Q., Wu, G. Y., Shen, Y., Ji, H. K., *et al*. (2004) Association study of an SNP combination pattern in the dopaminergic pathway in paranoid schizophrenia: a novel strategy for complex disorders. Mol. Psychiatry, 9, 510−521
15. Sun, X., Jia, Y., Zhang, X., Xu, Q., Shen, Y. and Li, Y. (2005) Multi-locus association study of schizophrenia susceptibility genes with a posterior probability method. Sci. China C Life Sci., 48, 263−269
16. Sun, X., Zhang, Z., Zhang, Y., Zhang, X. and Li, Y. (2005) Multi-locus penetrance variance analysis method for association study in complex diseases. Hum. Hered., 60, 143−149
17. Tang, W., Wu, X., Jiang, R. and Li, Y. (2009) Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy. PLoS Genet., 5, e1000464
18. Wu, J., Li, Y. and Jiang, R. (2014) Integrating multiple genomic data to predict disease-causing nonsynonymous single nucleotide variants in exome sequencing studies. PLoS Genet., 10, e1004237
19. Xia, H., Bi, J. and Li, Y. (2006) Identification of alternative 5′/3′ splice sites based on the mechanism of splice site competition. Nucleic Acids Res., 34, 6305−6313
20. 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
21. Li, S., Zhang, Z. Q., Wu, L. J., Zhang, X. G., Li, Y. D. and Wang, Y. Y. (2007) Understanding ZHENG in traditional Chinese medicine in the context of neuro-endocrine-immune network. IET Syst. Biol., 1, 51−60
22. Li, S. Li, Y. Zhang. P., Gu, J. and Jiang, R. (2021) Principle, method and application of relationship inference based on biological network. Sci. Sin. Inform. (in Chinese), https://doi.org/10.1360/SSI-2021–0243

© The Authors (2022). Published by Higher Education Press