

# Preface

Biological and biomedical studies have entered a new era over the past two decades thanks to the wide use of mathematical models and computational approaches. A booming of computational biology, which sheerly was a theoretician's fantasy twenty years ago, has become a reality. Obsession with computational biology and theoretical approaches is evidenced in articles hailing the arrival of what are variously called quantitative biology, bioinformatics, theoretical biology, and systems biology. New technologies and data resources in genetics, such as the International HapMap project, enable large-scale studies, such as genome-wide association studies, which could potentially identify most common genetic variants as well as rare variants of the human DNA that may alter individual's susceptibility to disease and the response to medical treatment. Meanwhile the multi-electrode recording from behaving animals makes it feasible to control the animal mental activity, which could potentially lead to the development of useful brain-machine interfaces. Embracing the sheer volume of genetic, genomic, and other type of data, an essential approach is, first of all, to avoid drowning the true signal in the data. It has been witnessed that theoretical approach to biology has emerged as a powerful and stimulating research paradigm in biological studies, which in turn leads to a new research paradigm in mathematics, physics, and computer science and moves forward with the interplays among experimental studies and outcomes, simulation studies, and theoretical investigations. In the current collection of papers, which are mini-reviews written by leading experts in their own areas of computational and systems biology, we attempt to summarize and share with the readers some of the most recent thriving developments.

The conference from which this book results was to celebrate the 70th birthday of Qian MinPing, a Professor in Mathematics and Theoretical Biology at Peking University. She is one of the people who foresaw the forthcoming tide of the computational and systems biology more than 20 years ago. "She is an amazing woman" said Mike Waterman, one of the conference attendees. Most contributing authors of the book are her students and are proud of being members in the "Qian-School." Below is a brief biography of Prof. Qian, written by herself.

*The editors of the volume asked me to write something about myself instead of contributing a research article. Reflecting on this request, it finally daunted on me*

*that I am now at the age of telling grandma's story which might be indeed more interesting to the younger generation. The life of Chinese scientists of my generation could be quite colorful and their stories multitudinous. We experienced two wars (The World War II and the civil war between the communists and the nationalists), numerous political movements including "The Cultural Revolution" of more than ten years, and finally the past thirty years of reform era with rapid changes in every aspect of Chinese life.*

*In 1979, after the end of the nightmare like "The Cultural Revolution," absolutely out of my expectation even in dream, I was fortunate enough to be a selected member of the first group of people sent to the US as visiting scholars. When I arrived in the US, I was impressed by the affluent life in America, with great contrast to that in China at that time. However, my strongest feeling was about the expectation and enthusiasm for the future of China and the lately started academic career of myself. That seems to be so long ago and might not be easily understandable by today's youngsters. But it concerns almost everything with the contemporary Chinese history in general and my own family background in particular.*

*My father was a Professor in polymer science educated in England. Growing up under the family influence, I decided in high school that I would devote myself to science. I entered the Peking University in 1956 and studied mathematics. Even though officially I was a student and later a faculty member in a good university, I could hardly fulfill what I wished. We often could only do research in our spare time after "work"; Even for that we were criticized as holding "illegal" seminars on mathematics in evenings and on Sundays. Thus, any opportunity to me for learning was really like food in starvation and water for thirstiness. Even though I was already 40 years old, what I was thinking the most was to take the advantage of the opportunity to study as much as one could, so that I would be able to catch-up scientific development of the world. It would be silly to expect myself making first class contributions to mathematics or science starting that late. The realistic goal to us was to play a connecting-role between the preceding generation, such as Professor Xu, Bao-Lu (Pao Lu Hsu), and the coming generation that entering undergraduate studies at the time. Therefore, the job I assigned myself was to introduce the most important new development in the field of stochastic processes to my students and to guide them to the frontier of scientific research. Still, if one wants to teach the students about modern research, he or she self definitely needs to have experience with such. Hence, even though we had not had much opportunity in doing research before, I, and most others like me in China, tried very hard to get a flavor of original research. Now looking back to my last 30 years, I could say I feel happy and satisfied, since I had made a good decision on my job and I tried my best for it. Many friends have asked me if I have ever thought about staying in the US, and the answer was always "No," since my motto is that "To marry a person who loves you instead of one you admire": China indeed needed me the most at that moment of the history.*

*To me, working with students and enjoying their achievements were really the greatest pleasure. In fact, I have been growing with my students in my academic*

*life: Often before I teach them, I do not really know much more than they do; the only advantage I have is my experience. When I teach a course, no matter whether it is familiar or new to me, I always try to bring something new into the class, and to improve my previous teaching and to adjust it according to the changing situation. I feel very sad if no one among the students really wants to understand what I teach. However, if the students truly enjoy my teaching, especially those I made a special effort, the pleasure to me is as great as I obtain a nice result in research.*

*If I ever had some good influence on my students, I think it is mostly through my enthusiasm and persistence in work and the attitude of pursuing the scientific truth and originality with self-confidence, irrespective of being in a familiar field or in an early stage of other new research. I myself obtained such influence mostly from my farther, my brother, and my teachers, such as P.L. Hsu and Z.P. Jiang, not through their preaching but from their deeds. In fact, my father rarely talked to me on serious topics but teased me as his little girl. His influences on me were from his own hard working days and nights, and life-long enthusiasm toward scientific research. My brother Min Qian, twelve years older, has been my real scientific mentor since my university days. As a faculty member in the same department, he has never taught me any formal courses. But he often talked to me about mathematics, and even physics, in the way like oral exams between teachers and students. I have learnt what it means to thoroughly understand something and how to grasp the essence of a problem this way. Learning from Professor P.L. Hsu's lectures notes on general topology (not on probability and statistics) in a complete new presentation and writing reports to him several times had given me life-time benefit, since that made me understand the importance to always pursue one's own originality in study, no matter in teaching or in research. I have learnt to be a devoted scientist without seeking fame and wealth from Professor Z.P. Jiang.*

*Since 1998, I have become increasingly attracted to the fascinating development of genetics and computational molecular biology. Starting at the age of 59 from knowing almost nothing, I, together with my students, have learnt a great deal of biology. I teach them what I know, but more importantly, how to obtain and understand the new biological knowledge fast by taking advantage of our background of mathematical training. At the meantime, I have learnt from them a great deal through presentations and discussions in seminars. Indeed, working together with students keeps me updated in biology, which, unlike mathematics, is such an extremely rapidly changing field. Together we continuously build connections between the wide range of new theoretical and experimental results.*

*Here I would like to talk a little bit about my feeling for leaning and studying biology as a mathematician. We all see that right now it is a time of tremendous new discoveries in biology. Significant experimental facts and novel data are obtained in biomedical sciences almost every month. Mathematics can play some important roles in explaining and understanding them. Real implications of observations and data are often far from what they seem to be. For instance, huge amount of microarray expression data have been obtained in recent years, and many have public access. Some conclusions have already been drawn, such as which genes or genomic parts are involved for what complex diseases or phenotypes. Still, because*

*the above conclusions are drawn from statistical analysis, rarely a mechanistic understanding is offered on why and how the genes are involved. The questions why so many statistically significant SNPs appear in the genomic desert and how SNPs in regulatory regions affect the phenotypes remain to be elucidated. What we know and understand is still very limited. The situation reminds us of the early 20th century when the great progress in theoretical physics followed a large amount of new observations and data, such as those in X-rays radiations, photoelectrical effect, and electron diffraction. Revolution or paradigm shift in physics eventually led to great applications in technology that have affected the life of mankind ever since. I often wonder whether a deep and thorough theoretical biology will come in the 21st century based on all these recent observations and data. I feel strongly that whatever the final theoretical edifice will be, the statistical genetics and bioinformatics will be integral parts of it.*

*To reach a thorough understanding of something from what one observes, theoretical induction and integration with imagination are not only important but actually also are necessary. I believe that mathematics can play precisely this indispensable role in this respect. However, this may not be just simple use of existing models and methods from existing mathematics. Real-world biological systems are extremely complex, and one needs first to grasp their most essential elements before representing them in terms of mathematical models. Furthermore, one also needs new mathematical concepts, tools, and methods for modeling and integrating simpler components, and to characterize more and more complex systems. Thus, a mathematician coming into contact with biological problems should position him- or herself as a scientist instead of considering him- or herself as merely a mathematical tool and model provider. He or she should learn and study biology together with biologists. The ultimate real truth would be obtained as a whole, rather segmented into biology, chemistry, physics, and mathematics as isolated disciplines.*

This volume consists of 19 chapters and covers a wide spectrum of topics. In Chap. 1, S. Zhong and his colleagues used thermodynamic models to analyze gene regulatory mechanisms. In Chap. 2, Y. Ding reviewed major algorithms for RNA secondary structure prediction, with a focus on ensemble-based approaches that have proved to be advantageous in many applications since they provide complete statistical characterizations of the Boltzmann ensemble of RNA secondary structures. He described applications of an RNA structure sampling algorithm to the rational design of short interfering RNAs for gene silencing by RNA interference and to target identification for microRNAs that play important roles in posttranscriptional gene regulation. In addition to sequence features, incorporation of target mRNA secondary structure is an important consideration in these applications. The microarray technology has developed rapidly and has advanced our knowledge of the genomes of various species and the understanding of complex diseases. Particularly, the oligonucleotide microarrays have received increasing attention in biological and biomedical research. However, many aspects of the oligo arrays have not been thoroughly studied or fully understood, which lead to issues related to the array data quality control. In Chap. 3, W.J. Fu and his colleagues demonstrated that

new developments in these areas of the oligo arrays lead to better understanding of the array mechanism and improvement in the microarray data analysis. In Chap. 4, H. Ge tried to apply the models of stochastic processes into two very active fields now, nonequilibrium thermodynamics and biological signal transduction. Many essential concepts and relations related to classical thermodynamic laws have been put forward and discussed in details. Besides it, stochastic approach is also used to model biological signal transduction pathways and modules. Here, he focused on the phosphorylation and dephosphorylation module and mainly investigate its sensitivity against external signals. It was found that at least to some extent stochastic models could explain the mechanism producing ultrasensitivity better than the corresponding deterministic one. In Chap. 5, J.F. Feng and his colleagues reviewed some of recent progresses in applying Granger causality to recover network structures: gene networks, protein networks, and neuronal networks. Some successful applications are included to demonstrate the power of the approach. Phylogenetic footprinting is one of the most effective approaches for transcription factor binding site identification. In the past decade, many phylogenetic footprinting methods have been developed and have demonstrated their power in predicting binding sites. In Chap. 6, X.M. Li and his colleagues differed from other reviews on phylogenetic footprinting and presented a few representative methods based on whether these methods depend on alignments. They also pointed out a few challenging problems for future directions. In Chap. 7, P. Wang and her colleagues introduced penalized regression-based methods, *space* and *LogitNet*, for constructing genetic interaction or regulatory networks from high-dimensional continuous and binary array data. They also introduced *remMap* for constructing networks using two different types of high-dimensional array data. These methods are illustrated through both simulated and real data examples.

Protein domains are parts of the protein that can function independently of other parts. Thus, domains form the basic units of proteins, and domain–domain interactions are the fundamental causes of protein interactions. Although large amounts of protein interaction data sets from many different organisms are available, our knowledge of domain interactions is limited. Several computational methods have been developed to predict domain–domain interactions from protein interactions and other information including gene coexpression, gene annotation, and domain fusion. In Chap. 8, F.Z. Sun and his colleagues reviewed several computational approaches to achieve this objective, including a maximum likelihood estimation method, a likelihood ratio based method, maximum parsimony, and methods integrating interaction data from multiple organisms, gene annotation, co-evolution, etc.

Now with the developed mathematical theory of irreversible stochastic processes carried out by Min-Ping Qian, Min Qian, and their colleagues at Peking University, it is clear that the irreversible stochastic processes are applicable to many of the interesting open-system phenomena in chemistry and biochemistry. In Chap. 9, H. Qian and his colleagues started with the simple Michaelis–Menten enzyme kinetics from a purely stochastic perspective and then turned to an irreversible Markov process called coupled diffusion, which could be used to model motor protein, fluctuating enzymes in a living cells, and self-regulating genes. They also found that a

bifurcation, saddle-node or pitchfork, occurs in certain coupled diffusion systems while decreasing the rates of jump processes. In Chap. 10, D.F. Wu and her colleagues briefly reviewed the current status of the probability model and the statistical methods in cancer screening and their limitations.

The smallest confidence interval for a given class of intervals was defined to be the intersection of all intervals in the class. If this intersection belongs to the given class, we say the smallest interval exists in the class, and this interval is simply the best in that class. In Chap. 11, W.Z. Wang introduced a general method to construct the smallest one-sided  $1 - \alpha$  confidence interval when there exist nuisance parameters. In Chap. 12, J. Xie and her colleagues introduced the idea of group variable selections in a regression model and applied the method to genomic data. In Chap. 13, inspired by the Granger causality idea in time series, W.Q. Yang and his colleagues extended the notation to static data and applied it to protein data. In Chap. 14, N.R. Zhang reviewed the computational and statistical problems that arise in DNA copy number data and surveyed recent advances in their treatments. In Chap. 15, T.L. Zhang outlined a number of cluster detection approaches and disease mapping approaches.

Treating mRNA transcript abundances as quantitative traits and mapping gene expression quantitative trait loci for these traits has been studied in many species from yeast to human. There has been significant success in finding associations between gene expression and genetic markers. These eQTL studies have been used to identify candidate causal regulators, to construct gene regulation networks, to identify hot spot regions, and to better understand clinical phenotypes. Because of the large number of genes and genetic markers in such analyses, it is extremely challenging to discover how a small number of eQTLs interact with each other to affect mRNA expression levels for a set of (most likely co-regulated) genes. In Chap. 16, J.S. Liu and his colleagues reviewed a few methods for studying eQTL data and outlined a new Bayesian method they recently developed for eQTL mapping. In Chap. 17, H.Y. Zhao and his colleagues first constructed a weighted gene co-expression network and then extracted gene modules from the constructed network based on some topological measure. To interpret the biological meaning of the extracted modules, they used information from Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, and genome-wide location data to study whether each module is enriched for certain categories. Furthermore, they compared the utility between topological overlap and Pearson correlation similarity measures to define modules. Additionally, to study the relationships between modules derived from different expression data sets for the same species, they compared the consistency of gene modules inferred using different expression data sets. Lastly, they performed expression Quantitative Trait Loci (eQTL) analysis to gain a better understanding of the genetic basis of gene modules. In Chap. 18, X.J. Zhang developed a rigorous approach to decode spike trains in a single neuron and an ensemble of neurons with or without interactions. Finally in Chap. 19, Y. Zhang proposed a new method that can accurately approximate the statistical significance of peaks adjusting for multiple testings.

Finally we would like to thank Zhang QianYi, the secretary of the Computational Systems Biology Centre in Fudan University, for her hard work to go through all chapters several times to unify all references and paper style.

Fudan, Warwick/Michigan/California     *Jianfeng Feng, Wenjiang Fu, Fengzhu Sun*

Frontiers in Computational and Systems Biology

Feng, J.; Fu, W.; Sun, F. (Eds.)

2010, XXV, 24 p., Hardcover

ISBN: 978-1-84996-195-0