Quantitative Biology 2020, 8(2): 177–186
https://doi.org/10.1007/s40484-019-0182-9

MEETING REPORT

International Workshop on Applications of Probability and Statistics to Biology, July 11–13, 2019

—In Honor of Professor Minping Qian's 80th Birthday

Minghua Deng[1,2], Jianfeng Feng[3,4,5,6,7], Hong Qian[8], Lin Wan[9,10], Fengzhu Sun[11,*]

[1] Center for Quantitative Biology, Peking University, Beijing 100871, China
[2] LMAM, Center for Statistical Science, School of Mathematical Sciences, Peking University, Beijing 100871, China
[3] Institute of Science and Technology for Brain-inspired Intelligence, Fudan University, Shanghai 200433, China
[4] School of Mathematical Sciences, Fudan University, Shanghai 200433, China
[5] Centre for Computational Systems Biology, Fudan University, Shanghai 200433, China
[6] Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence at Fudan University, Ministry of Education, Shanghai 200433, China
[7] Department of Computer Science, University of Warwick, Coventry, CV4 7AL, United Kingdom
[8] Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA
[9] NCMIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[10] School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[11] Quantitative and Computational Biology Program, University of Southern California, Los Angeles, CA 90089, USA
* Correspondence: fsun@usc.edu

Received August 2, 2019

The International Workshop on Applications of Probability and Statistics to Biology (APSB) was successfully held in Shanghai, China, July 11–13, 2019. The workshop was hosted by the Institute of Science and Technology for Brain-inspired Intelligence (ISTBI) at Fudan University, and in honor of the 80th birthday of Prof. Minping Qian of Peking University. Most of the twenty eight speakers were former students or close collaborators of Prof. Qian; and there were over eighty participants from all over China and United States. The conference featured four keynote talks by Profs. Minping Qian, Jianfeng Feng, Jianhua Guo, and Michael Q Zhang and 24 talks by others. While Prof. Qian had trained many theoretical probabilists, the organization committee decided to focus on the applications of probability and statistics to biology due to Prof. Qian's more recent research interest that has spanned the past 20 years. The topics of the talks covered dynamics modeling and analyses of data from single cells, gene expression and genomic sequence analysis, molecular networks, time series, molecular and brain imaging, meta-genomics, genome wide association studies, and clinical trials. Figure 1 shows the attendees of the workshop and Prof. Qian's keynote talk. Figure 2 shows some snapshots from the birthday celebration.

In the present report, we give a brief introduction of Prof. Qian's contributions to research and education, followed by summaries of the talks.

A BRIEF BIOGRAPHY

Minping was born in 1939 in an academic family. Her father, Baojun Qian, was a famous professor of polymer science. In 1951, he was invited to establish East China College of Textile Technology (now Donghua University) and became its Vice President, President and honorary President. Minping's brother, Min Qian, was a well-known mathematical physicist of Peking University, China. He contributed greatly to the theory of Markov

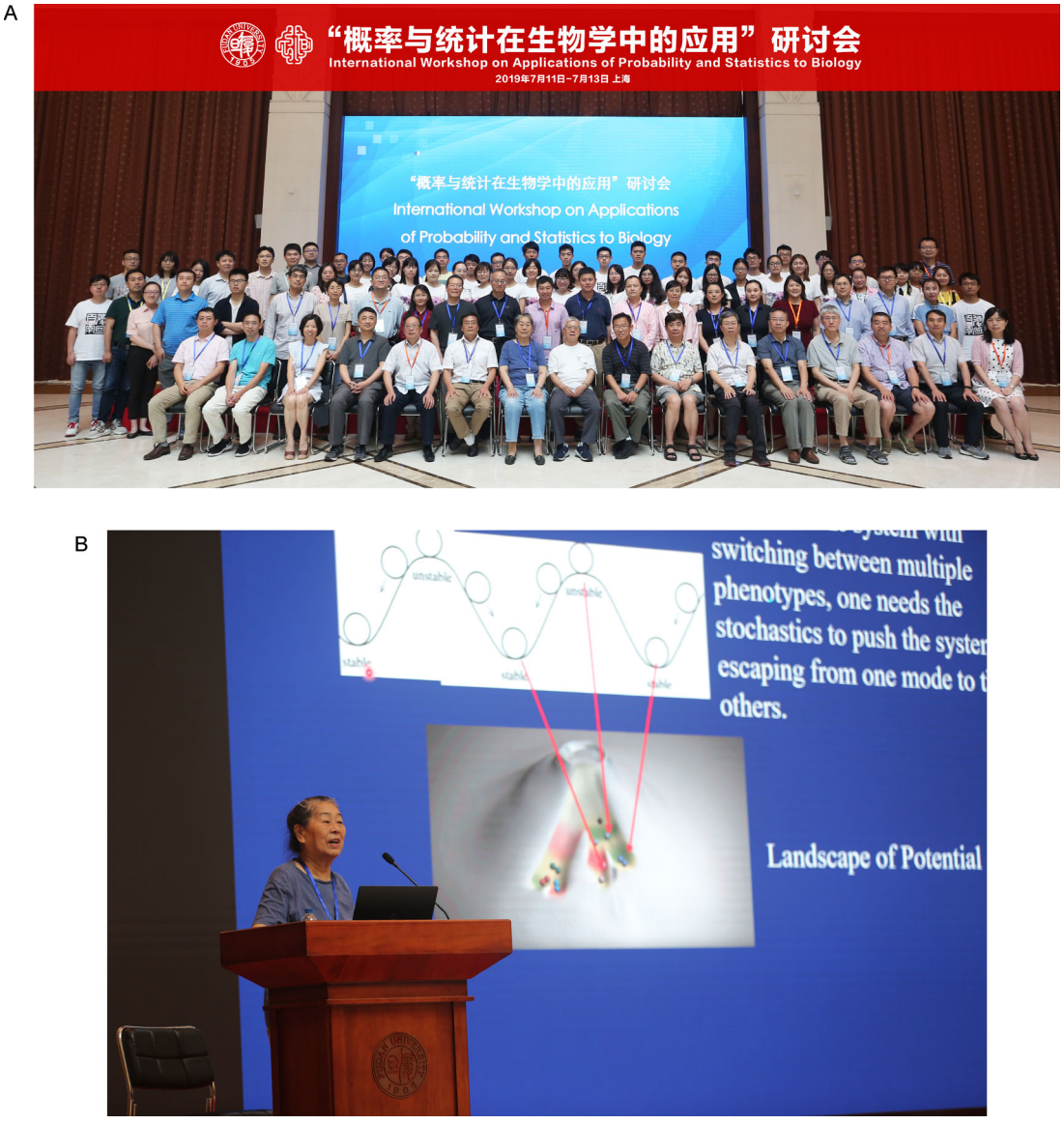

Figure 1. (A) The attendees of the workshop. (B) Prof. Qian's keynote talk.

processes and its applications to physics and biophysics. The brother and sister jointly pioneered the nonequilibrium theory of Markov processes and later developed a set of mathematical models and computational methods for chemistry and biology from a stochastic perspective. Minping married Guanglu Gong in 1964, a professor of mathematics first at Peking University and then at its neighboring Tsinghua University, China. This year marks their 55th year of happy marriage. In addition to their happy life together, they have collaborated extensively in research in probability theory and its applications.

Minping studied in the Department of Mathematics of Peking University as an undergraduate student majoring in mathematics in 1956 and later became a faculty member of the same department moving through the ranks from an assistant, lecturer, associate, to full professor until retirement in 2004. She later worked as a volunteer in both the School of Mathematical Sciences of Peking University and the Center for Systems and Synthetic Biology of Tsinghua University collaborating actively with Dr. Michael Zhang there. As most scientists of her generation, she did not receive a well-rounded education and her research was interrupted by the “Cultural Revolution”. She could not do much research until she was 40. Her academic career changed in 1979 when she was luckily selected as one of few visiting scholars to Washington University at St Louis and University of California at Berkeley in US. She became

Figure 2. Snapshots from the birthday celebration.

an instructor later on and had a chance to continue to work and live in US. However, she determined to go back to China, since China urgently needed people to educate the younger generation of modern science. Without any hesitation, she came back to Peking University as a faculty member immediately after the end of her visiting program to US and she has been happy about that decision ever since. She later visited many universities including Washington University at St Louis and UC Berkeley again, University of Southern California, University of Minnesota, UCLA, University of Maryland, Michigan State University, Yale, USA, etc.

Although her academic career started long after the conventional prime time for mathematicians, her accomplishments in the investigations of stochastic circulation, entropy production, reversibility and nonequilibrium theories of Markov processes, neural networks, machine learning, and computational biology and bioinformatics are impressive. She has been on the forefront of research ever since she started her academic career in 1979 until

the current day at the age of eighty. Over a third of her publications were published after she retired in 2004. She entered the field of computational biology at the age of 60 and still made highly significant contributions. Most importantly, she trained a large number of researchers in applied probability and applications to various fields including neural science and molecular biology.

Minping does not care about fame nor money. All she cares about are knowledge, science and helping others. After retirement, she continues to train students and do research collaborating with others, but primarily with Minghua Deng and Daquan Jiang at Peking University and Michael Zhang at Tsinghua University. Dr. Xuegong Zhang, head of the Bioinformatics group at Tsinghua University, recalled the scene that he offered to give some monetary compensations to Prof. Qian for her contributions to research at Tsinghua University, she categorically declined and felt that it was an honor and pleasure for her to be involved in research.

Minping deeply and sincerely cares about students and others. Her help changed many people's life in the positive direction. She paid particular attention to underprivileged students who were from rural areas, were extremely poor and had no resources. Her motto was "if I do not help them, who will?" For examples, she personally recommended Fengzhu Sun, currently a professor of computational biology, to USC as a graduate student, and Minghua Deng, currently a professor of probability and statistics at Peking University, to USC as a postdoctoral fellow. Both of them came from very poor families. Similarly, Daquan Jiang of Peking University, Lin Wan of Chinese Academy of Science, and Kui Zhang of Michigan Technological University, all came from rural areas, received great help from Prof. Qian during their career. Without Minping's help, their life could have been quite different. These are just a few examples of Minping's generosity and kindness in helping others.

Minping's contributions to education and training of students

Minping made it her goal to train students to their full potential in 1979 and she successfully accomplished this goal over 40 years. She will continue to train students in the future. Over the years, she has trained 27 Master students and 20 PhD students who are active in academia, industry and government. In addition, a large number of undergraduate students performed research and wrote theses under her direction.

She and Guanglu Gong jointly wrote two excellent textbooks in Chinese: Stochastic Processes [1] and Applied Stochastic Processes [2], which have been used in many universities and have played important roles in the education of stochastic processes throughout China. Jointly with Daquan Jiang and Min Qian, they wrote a Springer monography "Mathematical Theory of Nonequilibrium Steady States" [3] as a Lecture Notes in Mathematics. She also jointly edited a book with Jianfeng Feng and Jurgen Jost titled "Networks: From Biology to Theory" [4]. A book jointly edited by Jianfeng Feng, Wenjiang Fu and Fengzhu Sun, "Frontiers in Computational and Systems Biology" [5], was dedicated to her 70th birthday in 2009.

In the late 1990s, she began to be interested in computational biology. She was the earliest Chinese mathematician who ventured into the field and made significant contributions. She organized the first summer school in computational molecular biology and bioinformatics for students from all over China. She invited Michael Waterman at USC to give a lecture at Peking University on his return from a meeting in Inner Mongolia in 1997 and a close friendship between them was soon developed. Through many students who Minping recommended to USC, Michael Waterman and his USC colleagues trained a large cohort of Chinese students and postdoctoral fellows. Michael Waterman played a pivotal role in the development of computational biology inside China by being a guest professor in Tsinghua University and Fudan University for over 10 years. He was recognized by the Chinese Friendship Award and became a foreign member of the Chinese Academy of Sciences. Minping recommended over 20 students and postdoctoral fellows to USC and many of them are now active players in the field of computational biology including Drs. Minghua Deng at Peking University, Xiaoman Li at University of Central Florida, Xiaotu Ma at St Jude Children's Research Hospital, Fengzhu Sun at University of Southern California, Lin Wan at Chinese Academy of Sciences, Xiting Yan at Yale University, Kui Zhang at Michigan Technological University, and Yu Zhang at Penn State University. Many others work in pharmaceutical and biotechnology companies. In addition, she recommended many students to other universities including Stanford, Yale, and UC Berkeley. Many of them are leading investigators in the field such as Zhengqing Ouyang at Jackson Laboratory for Genomic Medicine, Jun Xie at Purdue University, Yuping Zhang at University of Connecticut, Sheng Zhong at UC San Diego, etc.

She taught many courses during her career at Peking University including stochastic processes, probability theory, advanced probability theory, applied stochastic processes, etc. She was always highly enthusiastic about the courses she taught and tried to teach the materials with new perspectives. Instead of just presenting the material to the students, she taught the students how to think and how to apply the knowledge in real world problems. In special topics courses, she asked many pertinent and deep

scientific questions that greatly benefitted the students. Her enthusiasm and teaching influenced a large number of students in their careers.

Minping's contributions to research

Minping strongly believes that in order to train students most effectively, one has to be at the forefront of research. Therefore, she has been active in pursuing scientific research of her interest ever since 1979. From early 1980s, Minping Qian, Min Qian and Guanglu Gong, together with their students, published several highly influential papers on the circulation theory of Markov chains [6,7], the concept of entropy production [8,9] and its relationship to reversibility of Markov processes [8]. They mathematically proved their asymptotic behaviors in systems such as Einstein-Green Kubo's relation [10], fluctuation and dissipation theory, and spectrum of Markov chains [10]. They also investigated the nonequilibrium theory of Markov processes [3,11]. In the 1990s, jointly with Dayue Chen and Jianfeng Feng, she investigated the hierarchical structure of metastability (attractors) using large deviation theory for Markov processes and stochastic Ising models [12,13]. Minping and her student Yulin Li also worked out the corresponding results for diffusion processes [14]. Such theories have been used to analyze various concrete biophysical problems, including ion channels, enzyme dynamics and single molecule dynamics [15–22].

From the early 1990s, Minping and Guanglu realized the importance of both theories and applications of machine learning and neural networks long before these topics became popular inside China. They led a group of students, in particular, Jianfeng Feng, Haitao Fang, and Yong Liu, who investigated the mathematical and computational issues related to neural networks and simulated annealing [24–29].

Minping entered the field of computational biology at the age of 60. As an applied mathematician, she did not know too much about biology at the time. She immersed herself deep into biology by reading biology books, attending biology lectures, seminars, and working closely with biologists. She supervised 8 PhD students and 19 Master students in this field with thesis topics ranging from gene expression, promoter detection, gene identification, sequence assembly, molecular motors, to single-cell dynamics and data analyses [30–36].

We would like to conclude this section with a quote from Michael Waterman of USC, founder of the field of computational biology, who could not make to the workshop. "*I met Minping Qian in 1997 at Peking University and was amazed by her intelligence and her sparkling energy. I still am! Minping, what a contribution you have made with your research and the army of brilliant students you mentored and released to the world! And we all know that you are just beginning.*"

SUMMARY OF TALKS

Single cell dynamics and data analysis

With the development of sequencing technologies, it is now possible to obtain gene expression data at the single cell level. Traditional bulk sequencing can only reveal the average effects of a large number of cells. With single cell sequencing, it is possible to investigate the stochasticity of gene expression. However, there are many challenges related to single cell data analysis. Prof. Minping Qian of Peking University delivered a plenary speech on the single cell dynamics and its data analysis. Prof. Qian addressed the stochasticity of single cell dynamics which raised grand challenges in quantitative descriptions of single cells. She reviewed the meta-stability theory developed during later 1980s and early 1990s by Dayue Chen, Jianfeng Feng and herself. The meta-stability theory provided a solid foundation for mathematical modeling of single cell dynamics. Prof. Qian reported the important properties of single cell dynamics revealed by their mathematical theory, which were further used to reveal the topological data structures of real single-cell data. Prof. Qian concluded by her recent joint work with Michael Zhang's group at Tsinghua University on applications of mathematical modeling and deep learning method to single-cell data analysis.

Followed Prof. Qian's talk, Dr. Lin Wan of the Academy of Mathematics and Systems Science of Chinese Academy of Sciences reported his recent work on visualization and reconstruction of cell developmental trajectories based on single-cell RNA sequencing data. He addressed the computational challenges raised by the high-dimensionality and heterogeneity of single cell data. He reported three algorithms (DensityPath, TSEE and SCTree) his group recently developed for the visualization, reconstruction, and statistical testing of the intrinsic structures of cells embedded in the high-dimensional noisy single cell data. Then Dr. Xiting Yan of Yale University reported recent progresses on the statistical modeling of single cell genomics data. She emphasized that due to the prevalence of dropouts, the ultra-high dimension of the data, and the hierarchical data structure, statistical analysis of the single cell RNA sequencing data often faces various challenges. She then reported two statistical models developed by her group for data imputation and gene expression association analysis to help overcome some of these challenges. They conducted comprehensive evaluations of the two models on both simulated data and real data sets, and demonstrated good

performances of their methods when compared to previously published data imputation methods.

Mathematical and probabilistic models in biology

On the frontiers of the theory of probability and its applications to biology involves many particles, for example molecules, with spatial random movements as well as biochemical reactions. As a generalization of Ising model from physics, a fundamental model along this line is the interacting particle system (IPS) whose macroscopic, deterministic limit is a nonlinear partial differential equation (PDE) of reaction diffusion type. Models as such have found wide applications in biophysics of transcription and translation along a DNA. Dayue Chen from Peking University reported his work on the ergodicity of the invariant Bernoulli product measure of exclusion processes on a tree; and using this result he gave the speed of a tagged particle.

Cellular gene expression kinetics as a stochastic process in terms of gene activation, mRNA and protein numbers in a single cell is now well established. The gene expression in a single cell exhibits a bursting behavior. In recent years, there is a growing interest in periodic oscillatory patters in single cell gene expression. Based on a widely accepted stochastic kinetic representation of the central dogma, Chen Jia of Wayne State University simplified the model into a stochastic switching ordinary differential equation (ODE) and showed analytically that it can have a characteristic frequency in the kinetic system. This work provides a sound theoretical basis for the stochastic oscillation in single cell gene expression. Dr. Chunhe Li of Fudan University developed a mechanistic, biochemical model based on the widely accepted stochastic kinetic representation of the central dogma, with some simplification. He showed the nonequilibrium landscape for a simple stochastic gene regulatory kinetic system. Furthermore, he presented a method, called minimum action principle, for computing the most probable transition path for phenotype switching between different cell states, which can be visualized as a barrier-crossing process between two basins in the landscape.

Gibbs' free energy function gives the landscape for studying the dynamics of biological macromolecules. But this theory is only applicable to equilibrium matter; not to living cells under a chemostat. Hong Qian of University of Washington presented a new theory based on a stochastic kinetic formulation of any complex chemical kinetic network such as in a single cell, with N biochemical species and M stochastic elementary reactions in a small volume V. He showed that in the macroscopic limit of V tends to infinity, the theory recovers the traditional chemical kinetics with rate equations. More interesting and surprisingly, by doing just mathematics, he was able to show the asymptotic probability defines an entropy function which becomes the Gibbs function when V tends to infinity. This theory, therefore, provided a mathematical foundation for a cellular, emergent landscape in terms of chemical kinetics.

Hodgkin-Huxley (HH) model for neuronal action potential has been one of the most significant dynamic, mechanism based mathematical models in biology. There is a growing body of research that now uses stochastic dynamics to represent biological processes at the cellular level. Xuejuan Zhang of Zhejiang Normal University presented a new computational method for simulating the stochastic HH model in which voltage-gated ion channel kinetics is represented by a Markov jump process, which in turn determines a stochastic membrane current and voltage. The continuous membrane potential $V(t)$ is coupled to discrete channel kinetics. This type of model is known as piecewise deterministic Markov process (PDMP), also called stochastic switching ODEs. The transition time in the stochastic simulation algorithm has to be determined through an implicit equation; Zhang proposed an approximate simulation algorithm that can efficiently and accurately sample the time evolution of the hybrid stochastic dynamics.

Molecular sequences and networks

Molecular sequence and network analyses are at the core of computational biology. Extensive research have been carried out for analyzing molecular sequences and networks. However, significant challenges remain to deal with the large volumes of molecular data. Dr. Michael Q. Zhang of University of Texas at Dallas presented an insightful keynote talk on open mathematical problems in computational biology. He covered a wide range of topics including random matrix theory in modern mathematics, Maxell-Boltzmann distribution and Ising model in physics to molecular sequence analysis and single-cells in biology. He pointed out the closed connections among the different subjects and projected that the integration of these subjects will yield important insights in biology. Dr. Jianhua Guo of Northeastern Normal University gave another keynote talk on detecting communities in complex networks under the stochastic block model (SBM). Using a split-likelihood framework, his group developed a highly accurate yet computationally efficient algorithm for community detection. They also developed a modified version of the SL algorithm, called the conditional split likelihood algorithm (CSL) to deal with networks with hub nodes or those with substantial degrees of variation within communities. They applied their algorithm to analyze a variety of

networks from human social networks to molecular networks.

Dr. Xuegong Zhang of Tsinghua University presented a new statistic D_2^R for detecting repeat regions in both long genomic sequences and NGS reads data. The new statistic was based on alignment-free dissimilarity measures d_2^* and d_2^S that were joined developed between Dr. Zhang's group and Drs. M. Waterman and Fengzhu Sun of University of Southern California. Based on the new statistic, the authors developed an algorithm of linear time and space complexity for detecting most types of repetitive sequences in multiple scenarios, including finding candidate CRISPR regions from bacterial genomic or metagenomics sequences.

Dr. Minghua Deng of Peking University introduced several network inference methods from his group for compositional data. Compositional data is a special data type with only relative abundance recorded such as microbial abundance in metagenomes, keyword frequency in text mining, mineral content in geology, etc. They proposed to infer the network by estimating the precision matrix among the latent absolute variables, and they developed CD-Trace and CDTr to estimate the precision matrix for compositional data, in which the estimations were obtained by optimizing the lasso penalized d-trace loss. The proposed methods outperform previous methods on simulation data and real data.

Dr. Fengzhu Sun of University of Southern California presented recent research from his group on metagenomics with concentrations on the identification of phages from metagenomic short reads and phage-host interaction identification using alignment-free sequencing comparison measures. He also showed new integrative approaches for predicting phage-host interactions by considering CRISPR matches and alignment between phages and their hosts, as well as a novel two-layer network between phages and bacterial hosts. Applications of the integrative approach yielded important insights into phage-host interactions in both human gut and marine environments.

Gene expression and epigenomics

Cross-sectional or longitudinal gene expression data from microarrays or next generation sequencing are abundant. How to extract biological information from both public and private data to understand biological and biomedical problems to benefit society is challenging. Dr. Lei Li of the Academy of Mathematics and Systems Science of Chinese Academy of Sciences presented a novel dual eigen-analysis method from his group by sorting elements of the dual eigen-vectors for gene expression data matrix. He showed that the sample-and gene-eigenvectors correspond respectively to the macro-and micro-biological information. They applied the dual eigen-analysis method to a gene expression profiles of multi-tissues from outbred mice fed with a high-fat diet (HFD) and the results imply that HFD influences the hepatic function or the pancreatic development as an exogenous factor. Dr. Lin Hou of Tsinghua University introduced a mediation analysis method from her group to map expression quantitative trait loci (eQTLs) that can explain regulatory mechanisms of trans-eQTLs. Motivated from the observation that trans-eQTLs are more likely to associate with more than one cis-gene than randomly selected SNPs in the GTEx dataset, they proposed to identify trans-eQTLs that are mediated by multiple mediators with increased power in both simulations and real data analysis.

Longitudinal data are becoming increasingly common in both genomics and metagenomics. Statistical and computational methods for analyzing such data are underdeveloped. Dr. Yuping Zhang of University of Connecticut introduced several unsupervised learning methods to analyze longitudinal genomic data, including principle trends analysis (PTA), joint principle trends analysis (JPTA) and lagged principle trends analysis (LAPTA). These methods can extract principal trends of time-course gene expression data from a group of patients, and identify genes that make dominant contributions to the principal trends under diverse research scenarios with or without prior biological knowledge. Probabilistic models for biological systems come from two complementary perspectives: data-driven statistical models and mechanistic stochastic dynamic models. Dr. Wei Lin of Fudan University presented ideas representing one of the current excitements in data science in "learning differential equations" and "learning natural laws", on how to construct the latter from the former. More specifically, he determines causation interactions among a large group of dynamic variables based on the method of causal inference from statistics. He then introduced Takens' delay embedding theorem that allows one to reconstruct the high-dimensional dynamics on an attractor from data on short-term longitudinal data. This idea shares remarkable similarity with Onsager's regression hypothesis on statistical fluctuating systems with ergodicity. The method can be used to analyze longitudinal gene expression data.

RNA structure and interaction play important roles in post-transcriptional and translational regulation. Dr. Zhengqing Ouyang of the Jackson Laboratory for Genomic Medicine reported recent progresses from his group on the analysis of RNA structure and regulation. He introduced a statistical approach for the inference of RNA structures at the transcriptome level — the so-called RNA structurome. He also presented a new method to model protein-RNA association strength and predict the functional targets of RNA binding proteins based on CLIP-

seq. He further described their research on determining RNA translation efficiency and start-site selection using ribosome profiling (also named Ribo-seq). Their methods and analysis have been applied to large-scale datasets and disease studies including cancer.

Genome-wide association studies (GWAS)

It is now routine to cost effectively genotype all the common single nucleotide polymorphisms (SNP) or even to sequence the whole genome of a large number of individuals. Despite the large number of available genome wide association studies for many phenotypes of thousands to tens of thousands individuals, the SNPs can only explain a small fraction of heritability for most phenotypes. Further development of statistical and computational methods are needed to make full use of the genotype, sequence and phenotype data. Linking genotype to phenotype continues to be a dominant problem in genetics. Dr. Xiaofeng Zhu of Case Western Reserve University reported their study on analyzing rare variants using families in large whole genome sequencing data. Different from existing statistical methods for analyzing rare variant associations that are mainly focused on weighting rare variants using genome annotation, his group developed a method in which traditional linkage information from family data is employed to help prioritize rare variants, which is independent from genome annotation. He illustrated that family data can improve statistical power to detect rare variant associations on real whole genome sequencing data. Dr. Kui Zhang of Michigan Technological University reported a statistical method, called MF-TOWmuT, for testing an optimally weighted combination of common and rare variants with multiple traits using family data. The MF-TOWmuT method was used to detect association of multiple phenotypes and multiple genetic variants in a genomic region using an optimally weighted combination of variants with family samples. Their method can be applied to both qualitative and quantitative phenotypes and both rare and common variants.

Dr. Jun Xie of Purdue University presented a powerful statistic for testing the association of a phenotype with sets of multiple variants. Her group developed a fast algorithm for efficient and accurate p-value calculation based on the new statistic. The algorithm can incorporate other covariates such as clinical and demographical data and confounding factors, and can deal with high correlations among variants. She illustrated the advantages of their methods in aggregating individual genetic variant effects, reducing the burden of multiple testing, and improving power to detect weak genetic effects.

Estimating the genetic heritability of a phenotype using genome wide SNPs is highly significant, whole the models in the current literature are usually mis-specified. How reliable is the estimated heritability under the mis-specified model? Jiming Jiang of University of California at Davis presented recent joint work with Dr. Hongyu Zhao of Yale University on the behavior of the restricted maximum likelihood (REML) estimator for heritability under a mis-specified linear mixed model (LMM). They established the consistency of the REML estimator of the variance of the errors in the LMM, and convergence in probability of the REML estimator of the variance of the random effects in the LMM to a certain limit. The asymptotic results were fully supported by the results of both simulation studies and real data analysis.

Cancer genomics and metagenomics

The past decade has witnessed a rapid progress of our understandings on the genetics of cancer and its progression. Probability and statistics modeling played a pivotal role in the dissection of general patterns from big cancer genomic datasets and continue to be of central importance in precision personalized medicine. Dr. Xiaotu Ma of St Jude Children's Research Hospital introduced cancer genomics from a probability and statistics perspective. Starting from functional classification of genes into oncogenes and tumor suppressor genes, he demonstrated the importance of comprehensive analysis of different mutation types for individual cancer genomes. He also introduced tumor purity analysis, which in turn leads to the concept of ploidy and clonality connected to tumor evolution under treatment pressure. Understanding tumor ploidy and clonality has profound implications in early detection and disease monitoring.

The human microbiome is the collection of microbes inhabiting the human body. Human microbiomes have been shown to be associated with complex diseases and certain types of cancer. Dr. Jianxin Shi of US National Cancer Institute presented a large-scale case-cohort study to estimate the overall contribution of oral microbiome to the risk of developing cancers. His group developed two statistical methods, one based on a linear mixed model and the other based on a high-dimensional Cox proportional hazard model to estimate the contributions of oral microbiome to certain cancer. Preliminary results and implications analyzing the case-cohort oral microbiome study were discussed.

Brain image, disease treatment and clinical trial design

Brain imaging including fMRI and DTI are now routinely used to understand human brain. Dr. Jianfeng Feng of

University of Warwick and Fudan University presented a keynote talk about ongoing large scale project of constructing a whole brain-mind machine (WBM) based upon the known data such as fMRI and DTI and models such as the integrate and fire and Hodgkin-Huxley type model. The authors developed a novel mathematical approach termed mesoscale data assimilation to link spiking data with BOLD signals and then assimilated the whole brain-mind machine into our human brains in the sense that both BOLD signals are identical at the resting state. Finally, the GPU implemented WBM with 2 billion spiking neurons is implemented together with FPGA boards with scalable structures both at the cellular, network and board scales.

Reproducibility of discoveries is frequently questioned in large scale studies including brain wide association studies (BWAS) and genome wide association studies (GWAS). Dr. Yinglei Lai of George Washington University presented some recent work on new measures for reproducibility. Reproducibility between two studies was traditionally evaluated using the widely used dice similarity coefficient (DSC) that measures the fraction of discoveries in both studies among the set of discoveries in either one of the studies. He and his collaborators developed two new reproducibility measures including discovery reproducibility (DR) that evaluates how reproducible the reported discoveries are and study reproducibility (SR) that evaluates the consistency between two studies. The new measures were used to investigate reproducibility of the UK Biobank and IMAGEN datasets and the authors found that the data could be highly reproducible.

Applying the knowledge from basic science to the benefit of treating patients is the ultimate goal of scientific research. Dr. Shuoyan Wang of Fudan University investigated the oscillatory behaviors of deep brain local field potentials and their involvement in the neurophysiological and neuropathological functions. His group revealed distinct rhythms of theta, alpha, high beta and high gamma oscillations in dystonia, neuropathic pain, and Parkinson's disease. They developed an approach to adaptively identify synchronization level to dynamically capture the dynamic neural states of multiple neural oscillations. The knowledge could be useful for developing the state-dependent intelligent deep brain stimulation to treat human diseases.

To evaluate the benefits of certain treatments of diseases, careful design of clinical trial is needed. Dr. Weizhen Wang of Wright State University introduced a two-stage clinical design with consideration of both response and toxicity, and gave an optimal test procedure on the response rate and the nontoxicity rate. He showed that the power function for each test in a large family of tests is non-decreasing in both rates. He also provided optimal two-stage designs with the least expected total sample size and the optimization algorithm.

ACKNOWLEDGEMENTS

We would like to thank Michael Waterman at USC for proofreading the report and all the speakers for their presentations and contributions to the report.

REFERENCES

1. Qian, M. P., and Gong, G. L. (1997) Theory of Stochastic Processes. Peking University Press, (In Chinese)
2. Gong, G. L. and Qian, M. P. (2004) Applied Stochastic Processes. Tsinghua University Press, (In Chinese)
3. Jiang, D. Q., Qian, M. and Qian, M. P. (2004) Mathematical Theory of Nonequilibrium Steady States. Springer
4. Feng, J. F., Jost, J. and Qian, M. P. (2007) Networks: from Biology to Theory. Springer
5. Feng, J. F., Fu, W. J. and Sun, F. Z. (2010) Frontiers in Computational and Systems Biology. Springer
6. Qian, M. P. and Qian, M. (1982) Circulation for recurrent Markov chains. Probab. Theory Relat. Fields, 59, 203–210
7. Gong, G. L. and Qian, M. P. (1998) The symmetry of diffusions and the circulations of their projection processes. Sci. China Ser. A Math., 41, 1017–1022
8. Qian, M. P., Qian, M. and Gong, G. L. (1991) The reversibility and the entropy production of Markov processes. Contemp. Math., 118, 255–261
9. Gong, G. L. and Qian, M. P. (1997) Entropy production of stationary diffusions on non-compact Riemannian manifolds. Sci. China Ser. A Math., 40, 926–931
10. Chen, Y., Chen, X. and Qian, M. P. (2006) The Green-Kubo Formula, autocorrelation function and fluctuation spectrum for finite Markov chains with continuous time. J. Phys. Math. Gen., 39, 2539–2550
11. Qian, M. P. and Jiang, D. Q. (2017) Non-equilibrium stochastic dynamics. Scientia Sinica Mathematica, 47, 1703–1716
12. Chen, D. Y., Feng, J. F. and Qian, M. P. (1996) Metastability of exponentially perturbed Markov chains. Sci. China Ser. A Math., 39, 7–28
13. Chen, D. Y., Feng, J. F. and Qian, M. P. (1997) The metastable behavior of the three-dimensional stochastic Ising model I. Sci. China Ser. A Math., 40, 832–842
14. Li, Y. L. and Qian, M. P. (1998) Hierarchical structure of attractors of dynamical systems. Sci. China Ser. A Math., 41, 1128–1134
15. Deng, Y. C., Peng, S. L., Qian, M. P. and Feng, J. F. (2003) Identification transition rates of ionic channels via observations at a single state. J. of Physiscs A: Math & General, 36, 1195–1212
16. Qian, M., Qian, M. P. and Zhang, X. J. (2003) Fundamental facts concerning reversible master equation. Phys. Lett . A 309, 371–376
17. Zhang, Y. P., Qian, M. P., Ouyang, Q., Deng, M. H., Li, F. T. and Tang, C. (2006) Stochastic model of yeast cell cycle network. Physica. D, 219, 35–39

18. Jia, C., Liu, X. F., Qian, M. P., Jiang, D. Q. and Zhang, Y. P. (2012) Kinetic behavior of the general modifier mechanism of Botts and Morales with non-equilibrium binding. J. Theor. Biol., 296, 13–20
19. Zhou, D., Wu, D. M., Li, Z., Qian, M. P. and Zhang, M. Q. (2013) Population dynamics of cancer cells with cell-state conversions. Quant. Biol., 1, 201–208
20. Jia, C., Jiang, D. Q. and Qian, M. P. (2014) An allosteric model of the inositol trisphosphate receptor with nonequilibrium binding. Phys. Biol., 11, 056001
21. Jia, C., Qian, M. P. and Jiang, D. Q. (2014) Overshoot in biological systems modelled by Markov chains: a non-equilibrium dynamic phenomenon. IET Syst. Biol., 8, 1–8
22. Jia, C., Qian, M. P., Kang, Y. and Jiang, D. Q. (2014) Modeling stochastic phenotype switching and bet-hedging in bacteria: stochastic nonlinear dynamics and critical state identification. Quant. Biol., 2, 110–125
23. Liang, Z. Y., Li, G., Wang, Z., Djekidel, M. N., Li, Y., Qian, M.-P., Zhang, M. Q. and Chen, Y. (2017) BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. Nat. Commun., 8, 1622
24. Qian, M. P., Gong, G. L. and Clark, J. W. (1991) Relative entropy and learning rules. Phys. Rev. A, 43, 1061–1070
25. Albeverio, S., Feng, J. F. and Qian, M. P. (1995) Role of noises in neural networks. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics, 52, 6593–6606
26. Fang, H. T., Gong, G. L. and Qian, M. P. (1997) Annealing of iterative stochastic schemes. SIAM J. Contr. Optim., 35, 1886–1907
27. Fang, H. T., Qian, M. P. and Gong, G. L. (1997) An improved annealing method and its large Time Behavior. Stochastic Processes and Their Appl., 71, 55–74
28. Gong, G. L., Liu, Y. and Qian, M. P. (2001) An adaptive simulated annealing algorithm. Stochastic Process. Appl., 94, 95–103
29. Gong, G. L., Qian, M. P. and Xie, J. (2001) Reversible algorithm of simulating multivariate densities with multi-hump. Sci. China Ser. A Math., 44, 357–364
30. Xu, M. X., Han, W., Qian, M., Ma, X., Ding, P., Wang, Y., Xia, D., Rui, M., Wang, L., Zhang, Y., *et al.* (2004) Last intron of the chemokine like factor gene contains a putative promoter for the downstream CKLF super family member 1 gene. Biochem. Biophys. Res. Commun., 313, 135–141
31. Yan, X., Deng, M., Fung, W. K. and Qian, M. (2005) Detecting differentially expressed genes by relative entropy. J. Theor. Biol., 234, 395–402
32. Duan, S. J., Wan, L., Fu, W. J., Pan, H., Ding, Q., Chen, C., Han, P., Zhu, X., Du, L., Liu, H., *et al.* (2009) Nonlinear cooperation of p53-ING1-induced bax expression and protein S-nitrosylation in GSNO-induced thymocyte apoptosis: a quantitative approach with cross-platform validation. Apoptosis, 14, 236–245
33. Wan, L., Li, D., Zhang, D., Liu, X., Fu, W. J., Zhu, L., Deng, M., Sun, F. and Qian, M. (2008) Conservation and implications of eukaryote transcriptional regulatory regions across multiple species. BMC Genomics, 9, 623
34. Wan, L., Sun, K., Ding, Q., Cui, Y., Li, M., Wen, Y., Elston, R. C., Qian, M. and Fu, W. J. (2009) Hybridization modeling of oligonucleotide SNP arrays for accurate DNA copy number estimation. Nucleic Acids Res., 37, e117
35. Wang, Q., Peng, P. C., Qian, M. P., Wan, L. and Deng, M. H. (2012) Hybridization and amplification rate correction for affymetrix SNP arrays. BMC Med. Genomics, 5, 24
36. Wang, F. G., Chen, R., Ji, D., Bai, S. N., Qian, M. P. and Deng, M. H. (2013) Adjustment method for microarray data generated using two-cycle RNA labeling protocol. BMC Genomics, 14, 31

