Journal of
Theoretical & Computational Science

Editorial   **Open Access**

Computational Challenges in Personalized and Precision Genomic Medicine

Peng Wang*

*Senior Engineer, Qualcomm Technologies, Inc., 5775 Morehouse Drive, San Diego, CA, USA

In the scientific literature and daily media press, we are hearing more and more news about a subject called personalized, or precision, genomic medicine. The term "personalized" refers to the prospect that genomic data from individual patients may facilitate rational treatment decisions that are tailored to each and every patient. The term "precision" refers to the prospect that genomic data may provide enhanced molecular resolution, mechanistic clarity, and improved quality of clinical treatment. Although we can foresee that personalized genomic medicine (or in short, personalized medicine, or PM) may eventually penetrate into many fields and sub-fields in medicine, oncology at this moment naturally is the best-fitting candidate into this category due to the genomic nature of cancer. Most cancers originate from a mixture of abnormal activities of various oncogenes and/or tumor suppressor genes. Characterization of these genes on all major cancer types can greatly benefit oncological research as well as clinical practice.

In the past few years, the speed of advancement of PM was very inspiring and astonishing, thanks to the widespread use of next generation sequencing (NGS) technology. Like any other subject with explosive growth, PM has to face numerous problems and challenges. In this article, I'd like to specifically share with our audience about the computational challenges which fall into the scope of this journal. These challenges have to be tackled in order to push PM onto the next level and require a lot of synergetic and harmonic efforts within the scientific community. In my humble opinion, there are four major obstacles sitting on the road of computation in PM. First, there is a strong demand on computing power in data processing, storage, and transfer, and this demand is outpacing our informatics capacities. According to Moore's law, Kryder's law, and Butter's law, costs are halved every 18, 12, and 9 months for data processing, data storage, and data transfer, respectively. However, that number has become a mere 5 months for genome sequencing for the period of 2007-2011. Therefore the gap between our ongoing bioinformatics need for computational capacity and our actual computational power grows exponentially. For example, in modern de novo sequencing, researchers have to align billions of short DNA sequence strings into a full genome, itself consisting of three billion base pairs for humans. Even with the best algorithm developed by specialized computational biologists, the assembly into the full genome in de novo sequencing requires two days on a 500-node supercomputer capable of processing 10 terabytes of raw sequencing data every day. Such a computer center can cost millions or even tens of millions of dollars for operation and maintenance, and therefore may only be available and affordable to large pharmaceutical companies, government research centers, and a handful of universities. In data storage, we can anticipate that at some point in the near future the amount of data generated from genome sequencing will surpass our storage capacity. And in data transfer, our current technology will no longer to be able to scale and support the humongous data flow in this area. Due to the explosive nature of growth of computational need in genome sequencing, there is an urgent need of collaborations between computer scientists and computational biologists, with the former providing scalable and affordable solutions in ultra-big-data processing, storage, and transfer, and the latter providing more innovative and efficient approaches/

algorithms for reduction of need of computing power. In addition, government and private-sector funding is also indispensable. Second, there is a need for the development of sustainable and standardizable bioinformatics analysis pipeline. Because of the speedy development and change in sequencing technologies, there are frequent updates of the computational solutions either installed on the sequencing machine or provided as standalone applications, which has become a major challenge in PM. Data processing and analysis in bioinformatics, especially clinical PM, is highly pipelined with each part built with its own special knowledge and methodology. Any upstream change (e.g. data format) may significantly disrupt downstream operability and compatibility. Due to the stringent requirement imposed on clinical applications, bioinformatics software systems must be comprehensively validated to warrant its operability, reproducibility, and quality. And for this reason, bioinformatics software is very challenging with respect to integration. Furthermore, even within a specific part of the pipeline, there may exist many different methodologies in data analysis. For example, different bioinformatics algorithms and tools can be applied in the detection of germline or somatic mutations. At this point, we can feel that the lack of industry-wide standard and consensus in PM and bioinformatics is impeding a more rapid progression and easier adoption of this technology. Looking back in history, the two most classical examples showing the great importance and benefit of industry standards are the establishment of W3C standards in the internet development and wireless communication standards in the wireless industry. Third, it might be a good time for us to balance between a data-based approach and a theory-based approach. Nowadays the buzz word of "big data" has almost permeated into every aspect of our lives. In particular, the commercial aspect has instilled us with the thought that the more data we have the more powerful we are. Perhaps we should calm down a little bit and think more about the fundamental theories, which echoes the purposes of this journal. Theory-driven approach can capitalize on the vast amount of knowledge present in the scientific community and can significantly complement and strengthen data-driven approach. For example, systems biology is an emerging discipline which utilizes computational and mathematical modeling to address questions in complex biological systems. In particular the concept has been widely used in biological and biomedical contexts since 2000. One of the outreaching aims of systems biology is the development of physiochemical modeling tools that can be used to describe the human signal transduction networks from a mathematical point of view. These models and tools can serve as hypotheses generation engines and provide suggestions in signal reprogramming

Corresponding author: Peng Wang, Senior Engineer, Qualcomm Technologies, Inc., 5775 Morehouse Drive, San Diego, CA, USA, Tel: 1-607-342-8522; E-mail: pw96@live.com

and targeted therapies even in the presence of model uncertainty. Last but not least, PM calls out for the need for highly interdisciplinary talents who possess not only a deep knowledge in biology and medicine but also a deep understanding of complex mathematical theories as well as a practical grasp of computing hardware/software skills. These domain knowledge is needed for one to manage the empire of the PM system and framework, digest and truly understand the information generated from this framework, and contribute to and innovate in this field of endeavor. Our biologists, computational scientists, software engineers, etc. need to dive into each other's field and become true interdisciplinary talents.

To sum up, we are faced with numerous challenges, obstacles, and uncertainties in computation in PM. In this article, I have outlined four major issues in my humble opinion and feel that these issues must be paid attention to in order to bring the research and development of PM onto the next level. Despite of these challenges and obstacles, PM has begun to witness its own success and fruition after years of hard work by the scientists, researchers, physicians, and industry partners. With a close collaboration between the scientific and the clinical community and proper financial support from the public and private sectors, we hope that the day that PM finally wins against cancer and benefits the majority of our patients is not too far way.