

Centralizing Bioinformatics *via* High-Performance Computing: Opportunities and Challenges in the -Omics era

In computational biology and bioinformatics, the global market continues to experience double-digit growth and so is the deluge of bigdata, especially with the advances in recent sequencing technologies creating Terabytes of -omics based datasets. Despite this data globalization, the data / algorithmic complexity, heterogeneous data sources and nomenclatures, impractically excessive runtimes, *etc.* make the data exploration and mining a huge challenge. It is thus obvious that comprehensive studies to understanding complex biological organisms require computational processing power and alternative paradigms for efficient data integration, replication and organization.

At UCR's high-performance computing | bioinformatics facility, we have developed a series of parallel-, multi-core-CPU-based open-source pipelines for large-scale -omics data analysis, which enables efficient and parallel analysis of multiple datasets in a short time. Our unit has been contributing to different research projects across the world by providing, **(i)** Support and expertise in programming and advanced data analysis focusing primarily on high-throughput genomics technologies including NGS data (RNA-seq, ChIP-seq, SNP-seq, *etc.*), **(ii)** Virtual server environments, secure and public web portals, and a large suite of open source applications, **(iii)** Instruction of hands-on tutorials and workshops on a wide variety of informatics topics, **(iv)** Custom data analysis and consultation services, and **(v)** Establishment of research collaborations with experimental scientists from different departments.

My lab conducts focused and outcome-oriented research in priority areas of interest as, (i) Systems-based understanding of host-pathogen interactions (e.g. predicting genome-scale protein-protein interaction networks, (ii) Computational genomics to improve health / response to environmental changes (metagenomics, modeling of gene regulatory networks, visualization), (iii) Develop epidemiological models for predicting pathogen spread, and (iv) Develop novel algorithms for genome-scale functional annotation and implementing them into tools / databases to aid the research community. The seminar will cover various aspects of the central methods I have been developing for informatics workbench that includes statistical pattern recognition, Bayesian (*unsupervised learning*), and Artificial Intelligence (*supervised learning*) based data-mining techniques, among other bioinformatics approaches. Support to these projects have come from various state and federal level funding agencies that has helped me to integrate technological advances such as Next-Generation Sequencing and efficient statistical models into my research program.