The Computational Molecular Biomedicine group is pursuing N major research topics as well as collaborative projects in diverse biomedical research areas. Most projects combine broad knowledge of molecular biology with state-of-the-art mathematical modeling and machine learning techniques.

1. Proteomics data analysis method and software development

The goal of this project is to develop high-performance, high-throughput analysis methods and software tools for proteomics datasets. We are particularly interested in applying machine learning techniques on both CPU and GPU architectures to improve peptide identification and quantification from mass spectrometry data. Example application areas include identification of (i) crosslinked peptides whose mass spectra are generally 3-4 times more complex than those found in typical proteomics studies and (ii) novel peptide sequences and modifications that currently do not exist in reference databases.

Parts of this project will be performed in collaboration with the Systems Biology Center at Chulalongkorn University (CUSB), whose characterizations of novel proteomes will greatly benefit from improved *de novo* identification of unknown peptides and proteins, and Dr. David W. Speicher’s proteomics research laboratory at the Wistar Institute in Philadelphia, PA USA, where zero-length crosslinking is being routinely utilized to probe the structure of red cell membrane proteins and other complex systems.

1. Integrative analysis of high-throughput biomedical datasets

For this project, our goal is to identify biological bases behind diseases or other complex biological systems through computational analyses of high-throughput biomedical datasets. This can be either hypothesis-driven, where publicly available datasets are used to support/disprove a hypothesis, or data-driven, where emergence of new datasets or technology dictates research direction. Therefore, topics of interest will usually be chosen based on the availability of public databases or via collaborations with experimental researchers from inside and outside of the university.

1. Mathematical modeling and simulation of chromatin folding process

Genomes are usually thought of as 1-dimensional sequences of genetic codes. However, they are located on chromosomes which fold in 3-dimensional space to form loops, topologically associated domains, and higher-order structures. These structures can have great impacts on gene expression regulation and evolution. For example, genes located in the same domain may be co-regulated by sharing the same transcription factor molecules. Chromosome packing can also limit the accessibility of genes in certain regions to transcription machineries and consequently control their activity.

In this project, theoretical behaviors of popular models will be examined in conjunction with chromosome conformation capture datasets from various organisms to gain further insight into the folding process of chromosomes. Parts of this project will be performed in collaboration with Dr. Wataru Iwasaki’s laboratory at the University of Tokyo and Dr. Ken-ichi Noma’s laboratory at the Wistar Institute.

1. Visualization of biomedical data using virtual reality

Even though we can perceive three dimensions, our visualization of data has largely been limited to 2-dimension media such as monitor screen or a piece of paper. With the emergence of affordable virtual reality technology, we can now explore the possibility of representing structural and non-structural biomedical data in 3-dimensional space and subsequent visualization in virtual reality. A primary goal of this project is to develop novel visualization methods and tools that facilitate comprehensive integration, visualization, and interpretation of complex biomedical datasets.

1. Curation and interpretation of research trends and other statistics in scientific literatures

To conduct effective research, one must rigorously study the literatures and choose the appropriate experimental protocols and data analysis tools. However, new research reports and academic software get published at ever-increasing rates that can overwhelm even the most seasoned researchers. For example, given the sheer number of software tools for processing RNA-seq data, it is not possible to select the most appropriate tools without examining the original articles for detailed descriptions and other articles that employed the tools for usage examples. Furthermore, some popular, highly cited tools may have since been replaced by emerging, better-performing alternatives. In this project, our first goal is to develop a curator that collects published articles describing software tools, classifies their purposes (e.g. what they need as input data, what they output), identifies their usage trends, and recommends the appropriate choices for a user's need. More complicated applications of such curation system will be explored over time.