若手研究者インターナショナル・トレーニング・プログラム(ITP)

バイオインフォマティクスとシステムズバイオロジーの国際連携教育研究プログラム 報告書

Name：David duVerle
Title：Ph.D. candidate, 1st year
Institute: Bioinformatics Center (Mamitsuka Laboratory), Kyoto University Institute for Chemical Research; Kyoto, Uji, Gokasho 611-0011, Japan
Partner institute: Max Planck Institute for Molecular Genetics (Department Vingron), Ihnestraße 63-73, 14195 Berlin, Germany
Duration: December 2009 - March 2009 (3 months)

Report:

During my stay at the Max Planck Institute for Molecular Genetics, I was able to interact with most members of Professor Vingron's department and learn about each team's focus and recent work. Thanks to frequent group meetings and presentations, as well as seminars and invited speakers, I received a good overview of the methods and problems currently faced in gene regulation and transcriptional regulation, as well as interesting insights into current research in epigenetic regulation and evolutionary genomics.

A presentation by team leader Sebastiaan Meijsing [1] raised my attention to the problem of gene regulation by glucocorticoid receptors (GR) and the existence of a large dataset of GR binding site sequences over the entire genome, made recently available through a ChIP-sequencing assay in three separate cell lines.

Report (Continued)：

Among the methods proposed to help identify the regulation mechanisms linked to GR and their specificity to each cell line, Ho-Ryun Chung, in the department's Epigenetic Regulation group, suggested using an adaptation of the algorithm used in [2] by Rajewsky et al.: a simplified HMM framework using zero-order transition probabilities between hidden states and TFBS position matrices for each state's emission probabilities. In addition to giving a total likelihood score for the sequence parsed, running this markovian model over a candidate sequence made it possible to identify a list of probable factors involved, by examining the posterior probabilities in the final HMM parameters.
The main task consisted of modifying the existing code to better represent background frequencies in the sequence considered, as well as allow a set of training sequences as input (existing code only worked with one sequence at a time, for parsing rather than predictive purposes).

Although the lack of large annotated datasets for promoter sequences at this time makes it difficult to run an extensive evaluation of this tool, its output appears qualitatively close to known results for selected promoter sequences.

Most importantly, the insights brought on by this work are already translating into ideas for novel approaches to the problem of protein cleavage site prediction.

[1] Sebastiaan H. Meijsing et al., DNA Binding Site Sequence Directs Glucocorticoid Receptor Structure and Activity, 2009.
http://www.sciencemag.org/cgi/content/abstract/324/5925/407

[2] N. Rajewsky et al., Computational detection of genomic cis-regulatory modules applied to body patterning in the early Drosophila embryo, 2002.
http://www.ncbi.nlm.nih.gov/pubmed/12398796

While staying at the the Max Planck Institute, I also was able to take advantage of the rich cultural life Berlin has to offer. During the three months of my stay, I visited many of the city's museums and landmarks (Pergamon Museum, Alte Galerie, Schloss Charlottenburg…), attended cultural events (Deutsche Oper Berlin, Staatskapelle Berlin, Berlinale Film Festival…), and many lab-related social events that provided for great opportunities to socialise with members of Professor Vingron's department in a less formal setting.

I feel this stay at Professor Vingron's Department was very beneficial to my research, both by introducing me to various techniques and methods of immediate use to my research at Kyoto University, as well as by giving me a wider exposure to many related domains that I hope to explore more in the future.

Plan (Continued)

Clockwise from top-left: Staatskapelle Berlin, Pergamon Museum, Schloss Charlottenburg & Berlinale Film Festival.