

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)

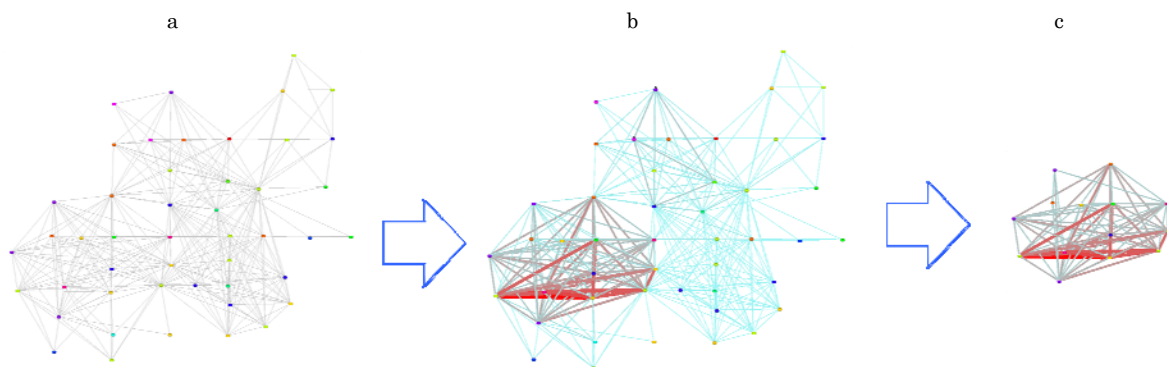
バイオインフォマティクスとシステムズバイオロジーの国際連携教育研究プログラム 報告書

Name: 平糠 和志
Title:抗原変異性タンパク遺伝子グループのネットワーク解析
Institute: 京都大学化学研究所バイオインフォマティクスセンター
Partner institute of your choice : Department of Bioengineering, Boston University
Duration of your choice: 8月31日～11月30日
Report : [研究生生活] 9月の頭から3ヶ月間、Boston University の Department of Bioengineering, DeLisi 研究室に滞在し、様々な経験をすることが出来た。研究室のメンバーとは研究に関するやり取りだけでなく、日々の生活においても大変お世話になり、有意義に研究生生活を送った。 ボストン大学は全米4位の規模を誇る歴史ある大学であり、大学建物の外の道や店には常に学生達がひしめき合っていた。Charles River 沿いに立ち並んだ Charles River Campus を貫く沿線通りから1ブロック離れた通りにある Life Science and Engineer Building に DeLisi 研究室はあり、その一階にある学生部屋のスペースを研究の場として提供させて頂いた。学生部屋には私以外に4人の学生と3人のポスドクがおり、初日から皆とても親密に接して頂いた。毎週水曜に行われている定期ミーティングでは、自分の研究成果の発表や研究に関連する論文の紹介を行うだけでなく、外部からゲストを呼んでセミナーを行ったり、Bioinformatics Graduate Program に属する他の研究者の発表を聞いたり、研究室外でもコミュニケーションを取る機会に恵まれた。特に滞在中には Bioinformatics Graduate Program の創立10周年記念の行事や、Smith-Waterman Algorithm で有名な Temple F. Smith 教授の退任パーティがあり、ボストン大学自体がバイオインフォマティクスの主要な研究拠点となっていることが感じられた。 DeLisi 研究室の学生の中にはマサチューセッツ工科大学の実験チームに定期的に参加し、自ら積極的に共同研究を行っている者がいた。また Department of Mathematics に属する学生や教官が DeLisi 研究室のミーティングに毎週参加したりと、互いの専門分野の知識を持ちこんで何かを模索しようとしている姿は大変印象的であり、私が所属するバイオインフォマティクスセンターもこういった環境に恵まれているものの、やはり学生達自らが積極的に活かそうとする場面が少ないように感じた。これらのことは今後自身の研究活動においても非常に大事なことであると認識し、普段から色々な人脈を通して可能性を見つけていきたい。
 
写真左：10 th Anniversary of Bioinformatics Program の様子。Life Science and Engineer Building の1Fロビーで学生達がポスター発表をしている。 写真右：毎週水曜の定期ミーティングの様子。プレゼン発表の前に研究の状況を聞かれ、そのまま30分以上の議論に発展することもしばしばあった。

[研究成果]

DeLisi 研究室では主にヒトの病気関連遺伝子をターゲットとした解析が行われているが、マイクロアレイデータを扱うグループと SNP データを扱うグループに大きく分かれている。後者はポストドクの方が精力的に研究されており今回は残念ながら詳しく解析内容を聞く時間が取れなかったが、前者においては **Gene Set Enrichment Analysis (GSEA)** という解析手法に興味を持ち詳しく教えて頂くことが出来た。マイクロアレイデータから発現量差のスコアを計算し、発現量が有意に変化している遺伝子セットを抽出する方法はこれまでに多くの研究が行われてきたが、それらの遺伝子セットが実際にどのような生物学的特徴によって **overrepresent** されているかを確かめるための統計的手法が **GSEA** である。具体的には **GO** や **Chromosomal location** のような各遺伝子に付随する特徴に対して **Fishe's exact test** を行なったり、**KEGG Pathway** や **PPI** 等を用いて各遺伝子の最短パスや結合頻度などの繋がりを解析するものである。

私が現在研究対象としている抗原変異性タンパクファミリーは種によっては 1000 遺伝子近く存在しており、その中からサブグループを抽出し解析することは興味深いものであるが、まずこれらの遺伝子ファミリーに対してマイクロアレイデータの利用は有用ではないという問題があった。これは抗原変異性タンパクファミリーの発現が **mutually exclusive expression** によるものであり、発現量を指標にしたサブグループの抽出は解析目的に適さない。そこで抗原部分のバリエーションが個々の遺伝子間で起こる相同組み替えによって引き起こされていることに注目し、発現ネットワークの代わりに相同組み替えネットワークを構築することで、より関連性のあるサブグループを抽出する方法を試みた。構築したネットワークは **VisANT** プログラム上で表示し、重みによる色づけ、カットオフによるサブネットワークの切り出し、他の生物学的情報の関連付けを行った。



A constructed network visualized in VisANT. (ex. *P. falciparum* 3D7 var family gene)

まず各 **Pathogen** ごとに抗原変異性タンパク遺伝子ファミリーをクエリーとし、**hyper-variable region (HVR)** 間に見つかる共通アミノ酸ブロックに注目した。各遺伝子上に存在する位置特異的 **HVR** をノードとし、**gene conversion** によってアミノ酸ブロックが **share** されていると考えられる **HVR** 同士をエッジで繋いだ。このとき同ポジションの **HVR** からのみ検出されるアミノ酸ブロックは除外している。次に同一遺伝子上に存在する **HVR** を一つのノード (メタノード) とし、各エッジをメタエッジとしてまとめたネットワークを作成した (上図 a)。この時、各遺伝子間における **shared block** の数や **p-value** に基づきメタエッジに重みを付けた上で (b)、メタエッジにカットオフを設定しサブネットワーク (クラスター) を抽出することが出来た (c)。このようにネットワーク構造を使う利点と

して、系統樹のように一意の距離によって階層分類したクラスターには、重要なトポロジー情報が失われているのに対し、サブネットワークのまま抽出されたクラスターはその後の解析に有用な情報もしくは検定対象を含むことが出来ることである。

[問題点・今後の課題]

当初、相同組み替えブロックの検出に確率モデルを用いた方法を適用しようと考えていたが、従来の方法では HVR 内で十分な検出力が期待できず残念ながら断念した。現在は単純にアミノ酸ブロックの数え上げを行い頻度行列を用いて p-value を計算しているが、この点についてもメタエッジの重み付けをする際に影響してくるので更なる考察が必要である。また本来の目的はサブグループの抽出ではなく、それらのセットに対する生物学的特徴の解析である。今回は十分な時間が得られなかったものの、これからの研究対象として重点的に解析を行っていききたい。

また今回、VisANT 上でネットワークの書き出しに必要な VisML(VisANT XML)の自動生成プログラムを作成した。現在 varDB にはマルチプルアライメントを行う MAFFT プログラム、HVR の切り出しを行う GBlock プログラムが組み込まれているが、それらの結果から更に VisML を生成しユーザーが容易にネットワーク解析、およびクラスターの抽出を行えるよう、VisANT プログラムの varDB へのプラグイン化を構想している。今回はその件についても VisANT の開発者と話し合う機会もあり、今後も精力的にコンタクトを取っていききたいと考えている。

[謝辞]

この度は、国際交流プログラムという大変貴重な機会を与えて下さった金久實教授、留学前から帰国までサポート頂きました馬見塚拓教授、滞在先において面倒を見てくださった Charles DeLisi 教授、そしてプライベートでも有意義な時間を与えてくれた同研究室のメンバーに心から感謝申し上げます。

