

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 応募書類

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
Plan : <u>滞在の目的</u> 今回、私が滞在希望するボストン大学の DeLisi 研究室では主にハイスループットデータに対する統計的手法を用いた網羅的なプロテオミクス・ゲノミクス解析の研究を行っている。現在ではヒトの病気関連遺伝子をターゲットとした SNP 解析や発現プロファイル解析が大々的に行われている。状態の異なるサンプルから得られる比較データより、各状態において統計的に有意に関わっている遺伝子セットを抽出するための方法論、またそのリストがどのような生物学的特徴によって overrepresent されているかを、a priori を用いて解析するといった研究が行われている。例えば発現解析においては、プロファイルから推測される相関の強い遺伝子リストに対し Gene Ontology や KEGG Pathway, PPI 等の既知ネットワークモデル上での共起特異性を計算する Gene Network Enrichment Analysis (GNEA) が広く使われている。DeLisi 研究室では、これらの遺伝子セット内で見出される多重ネットワーク情報を視覚化し、トポロジーや様々なエビデンスデータに基づいたサブネットワーク (モジュール) の抽出、アノテーション、再構築を行う手段として、VisANT の開発を行っている。その利用範囲は幅広く、上に述べた発現データのみならず、その他の実験から得られた相互作用データ (Two-hybrid system, Affinity technology 等)、計算機予測された相互関連情報(Domain fusion, Phylogenetic profile 等)といった数多くのデータタイプに対して適用可能である。これらは更にメタグラフとしての表現が可能であり、階層的なネットワーク解析を行うためのプラットフォームとして他には無い有用性を備えている。 現在、私が取り組んでいる研究テーマ (下記参照) においてこのようなネットワーク表現を用いて解析を行う事は、まだまだ事例としては少なく新規性のあるアプローチであると考えている。解析対象である antigenic-variation に関わる遺伝子データは今後も増え続けると思われ、そのためにロバストなネットワーク解析法を確立するのに大きな進展となることが期待される。また将来的には、これらの結果を antigenic variation database: varDB へフィードバックすることにより、ウェブインターフェース上で同様の解析が行えるよう利用者にとってデータベースの有用性を広げることを考えており、そのために VisANT の開発に携わる方々と今後の協力体制を築くことも目的である。 本研究室とはこれまでに共同ワークショップを開催しており、今まで以上に深い研究テーマへの理解と交流関係を築くことが第一と考えている。 <u>研究計画</u> 私は現在、ヒトや家畜動物寄生のバクテリアや原虫に見られる抗原変異(Antigenic variation)の解析を行っている。各生物種が抗原変異を引き起こしている遺伝子群を対象としており、これまでに varDB のデータを用いた配列解析を行ってきた。抗原変異は宿主免疫反応を回避するメカニズムとして寄生

生物に備わっている特徴であるが、それらの抗原遺伝子の発現制御やバリエーション発生の仕組みについて詳しくは解明されていない。各生物種は数十から千以上もの抗原変異性のタンパク遺伝子セットを持っており、配列上には主に膜貫通領域などの基本構造に関わる **conserved region** と抗原部位となる **hyper/semi-variable regions** が含まれている。バクテリアや原虫の膜タンパクの配列バリエーションは相同組み替えや遺伝子変換によって引き起こされるものがあり、それらの多くはモザイク様構造を持っている。このような配列セグメントの頻繁なインデルによりこれらの遺伝子が持つ進化的な関係は複雑になり、従来の二分木表現による系統解析は意味を成さない。そこでこれらの遺伝子上に存在する **polymorphic block** をノードとし、共通の（類似度の高い）ブロック同士をエッジで繋いだグラフで表現することにより、ゲノム内もしくは同一種間（異株間）内での相互関係をより **informative** に捉えられると考えている。これらのグラフを基に、個々の遺伝子の染色体ポジションや発現状態等の付加情報を組み合わせる事により、抗原変異性タンパク遺伝子セットから生物学的に重要なサブクラスターを抽出出来ることが期待される。

そこで以下の3点を目標として研究を行っていきたい。1) 抗原変異性タンパク遺伝子の **hyper/semi-variable region** の探索、2) **polymorphic block** セットを用いたネットワークの構築、3) ViaANT を使った **overrepresentation analysis** およびサブグラフの抽出。**Hyper/semi-variable region** は MAFFT と GBlocks を用いた方法により比較的高い精度で網羅的に抽出出来ると予想されるが、**hyper/semi-variable region** 内の **polymorphic block** を検出する方法は厳密には無く、まずはいくつかの確率モデルを用いた手法を研究する。

滞在期間中は DeLisi 研究室の定期ミーティングに参加し、お互いの研究内容に触れ合う機会を増やすだけでなく、積極的に問題点について話し合う等して研究レベルのコミュニケーション能力を養うよう心掛けたい。

以上の計画を踏まえて、今後の研究活動に役立つ多くの経験を残せるよう、3ヶ月の間精力的に研究生活に従事していきたい。