

課題名 (タイトル) :

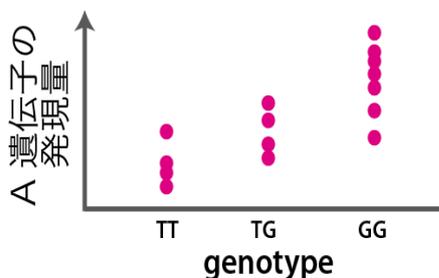
全ゲノム eQTL 解析

利用者氏名 : ○宮 冬樹*, 森園 隆*

所属 : *横浜研究所 ゲノム医科学研究センター 統計解析・技術開発グループ 情報解析研究チーム

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

これまでヒトゲノム中で 1%以下の頻度を含め 5 千万以上の 1 塩基変異 (SNV) が見つかっており、その約半数以上は近傍遺伝子等の発現量に影響を及ぼす expression Quantitative Trait Locus (eQTL) を有するという報告もある。例としては図のように、ある SNV の遺伝子型が TT のホモ、TG のヘテロ、GG のホモを持つヒトがそれぞれいた場合に (点が一人を示す)、その各個人の genotype ごとに近傍遺伝子 A の発現量を調べると相関がある (G アレルを持つヒトほど A 遺伝子の発現量が高い) というように genotype と phenotype (発現量) に関係があるものを eQTL があるという。



実際にこれまで疾患に関与する変異を探索するゲノムワイド関連解析 (GWAS) において、疾患との関連が見つかった変異においても eQTL を有することで疾患とリンクしているものが多数見つかってきている。しかしながら、これまでの eQTL 解析は解析の規模 (変異数や発現を調べた遺伝子数など) が小さかったり、あるいは計算に膨大な時間がかかることから近傍の遺伝子とその近傍の変異の関係しか見られなかったり (これを cis-eQTL と呼ぶ) と限られており、ある変異箇所の遺伝子型と近傍のみならず全ゲノムレベルの各遺伝子と関係がある eQTL (これを trans-eQTL と呼ぶ) の解析は限定的であった。

このような背景があることから、我々は既存の eQTL 解析以上のスケールで、ゲノムワイドかつ網羅的に、日本人を対象として新規 trans-eQTL の探索・解析を行うこととした。具体的には、genotype データについては、これまでの HapMap project 等で得られている変異情報と、さらに最近公開された 1000 Genomes Project の変異情報を全て統合して用い、発現データについては新規に次世代シーケンサーを用いて情報を収集解析した。また、これまでの変異解析チップを用いた手法では収集が難しく、今回次世代シーケンサーデータを用いることで初めて可能となる insertion (挿入)・deletion (欠失) と発現との eQTL 解析、および新規転写産物と SNV との eQTL 解析も行うこととした。

2. 具体的な利用内容、計算方法

計算方法としては、genotype と発現の間に関係がないかを線形回帰分析 (linear regression analysis) および permutation test を用いた多重検定補正法による統計解析手法を用いる。遺伝子数が約 3 万 (タンパク質をコードしていない ncRNA 等を含む)、日本人における頻度 5%以上の変異数が約 600 万、permutation test によるランダム入れ替え回数が 1 万回で、組み合わせの総計算回数だけで 3 万 x 600 万 x 1 万、すなわち 1800 兆という膨大な数の計算を行う。

線形回帰分析は既存の plink と呼ばれる C++言語で書かれた実行プログラムを改変したのを用い、多重検定補正部分は C 言語で書かれた自作のプログラムで実行する。

3. 結果

2013 年 2 月までの段階でほぼ予定していたプログラムの実行が終了した。現段階で 5,000 を超える trans-eQTL が検出されており、これは他の

研究者によりこれまで論文で報告されてきた数と比較し、有に 10 倍以上の規模である。また、ゲノム上の 1 つの変異が複数の遺伝子と関連が示唆される eQTL も検出されており、ゲノムワイドに大規模に解析することで、これまで以上に複数の遺伝子を跨いだ eQTL が多数検出されてきていると見られる。さらに次世代シーケンサーのデータを用いることで初めて可能となった全ゲノムレベルの挿入・欠失における eQTL 解析においても数百個の統計的に有意な eQTL が検出されており非常に興味深い。また、過去に eQTL 解析とは独立して疾患関連変異として報告されていた変異について、我々の eQTL 解析結果で同変異を探索すると eQTL を有しているものがあり、これらは疾患関連変異として報告された時には明らかとなっていなかった遺伝子と変異との関連を明確に示す証拠となり、どの変異がどの遺伝子に影響を与えているかについて、我々の eQTL データを見ることで明らかにできるものが存在することを確認できた。

4. まとめ

当初の予定していたプログラムの実行は無事完了しつつある。世界的に見て、これまで報告されてきた規模を遥かに超える大量の eQTL を検出することが出来た。これまでに報告された膨大な疾患関連変異について、今回行った eQTL データを調査することで遺伝子と変異の関連を強く示唆できるデータが得られることを確認した。一部途中経過の段階で学会にて成果を発表した。

5. 今後の計画・展望

近日中に全てのプログラム実行が終了予定となっており、今後より詳細について解析予定である。これまでの既報の疾患関連変異のみならず、今後明らかとされる疾患関連変異についても、今回行った eQTL データを参照することで特定の遺伝子との関連を調査することができ、極めて有用性が高いと考えられるため、今後結果をデータベース化して公開することを検討している。本解析の成果は近日論文として投稿予定である。

平成 24 年度 RICC 利用研究成果リスト

【論文、学会報告・雑誌などの論文発表】

該当なし。

【国際会議などの予稿集、proceeding】

該当なし。

【国際会議、学会などでの口頭発表】

1. 宮 冬樹, 阿部 哲雄, 森園 隆, 秋山 真太郎, 久保 充明, 角田 達彦; 次世代シーケンサーを用いた日本人における全ゲノム網羅的 eQTL 解析 (Genome-wide eQTL analysis using next generation sequencer in Japanese population), 日本人類遺伝学会 第57回大会, 東京 (2012).
2. Miya F, Morizono T, Abe T, Kubo M, Tsunoda T; Massive genome-wide eQTL analysis in Japanese population. The 35th Annual Meeting of the Molecular Biology Society of Japan, Fukuoka (2012).

【その他】

- ・上記 1. の学会にて「日本人類遺伝学会 第 57 回大会 大会最優秀ポスター賞」を受賞した。