

課題名 (タイトル) :

配列情報に基づく遺伝子発現解析

利用者氏名 : 長嶋剛史

所属 : 横浜研究所 免疫・アレルギー総合科学研究センター
細胞システムモデル化研究チーム

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

我々の研究チームでは癌細胞におけるシグナル伝達系とそれによって引き起こされる遺伝子発現を調べるために GeneChip による遺伝子発現量の測定を行ってきた。現在我々が参画しているプロジェクトでは、これを発展させ、より網羅的な計測、新規発現遺伝子やスプライスバリエーションの同定を行うために次世代シーケンサーを用いた遺伝子発現の解析に着手したところである。次世代シーケンサーによって生産されるデータ量は膨大なものであり、一連の実験で数千万から数億におよぶショートリード配列(30-100bp の DNA 配列)に関する情報が得られる。本家急では、このような大量のショートリード配列を利用しサンプルを特徴付ける遺伝子発現量の定量、スプライスバリエーションやゲノム構造変化を同定することが目的である。

2. 具体的な利用内容、計算方法

解析に用いるデータは、複数種類のサンプル(例えば造腫瘍能の高い群と低い群)由来のショートリード配列、典型的には 1 サンプルあたり 50bp の DNA 配列 2000 万本である。計算内容は、ショートリード配列のリファレンス配列へのマッピング(bwa、bowtie、tophat、blat)、遺伝子発現量の計算(rSeq、R)、各種統計量の計算(Per1、R)である。括弧内にあるのが実際に使用したプログラムである。これらの計算は qsub コマンド経由で主に upc と ax を用いて行った。

3. 結果

現在までにある種の癌細胞サンプルにおける遺伝子発現データの解析を行い、サンプル間で発現

量が異なるトランスクリプトのリストを得た。共同研究者によるリストの詳細な検討後、サンプル間の差を説明しうる候補の絞込みを行い、それらについて順次検証実験を行っており、そのうちのいくつかについては結果を得ている。

4. まとめ

一般に次世代シーケンサーと呼ばれる計測機器によって産出される大量のショートリード配列を、複数の細胞種や刺激条件下で比較し、変化のあった転写産物を抽出するために RICC を利用した解析を行っている。今回の解析で得られた結果について共同研究者と検討を重ね、一部について検証実験を開始した。対象とするデータのサイズおよび配列解析に必要な計算量は一研究室で確保できるそれを大きく超えているのが現状であり、データ解析を行う上で RICC は欠かすことのできない重要な計算機資源である。

5. 今後の計画・展望

これまでに取得した配列データおよび今後取得を予定しているものについても同様の解析を行う予定である。そこから得られる結果についても発現変動転写産物の同定、リストの詳細な検討と絞込みの後に検証実験を行い、その結果を順次論文としてまとめる予定である。

6. RICC の継続利用を希望の場合は、これまで利用した状況(どの程度研究が進んだか、研究においてどこまで計算出来て、何が出来ていないか)や、継続して利用する際に行う具体的な内容

RICC を用いて行った解析の結果、興味深い遺伝子群を含む発現変動遺伝子のリストを得ることができた。得られたリストは更に詳細な解析を進めるためのベースになる情報として活用してい

平成 22 年度 RICC 利用報告書

る。今後得られる配列データについても同様の解析を行い、サンプル間の差異を特徴付ける発現変動遺伝子の抽出を行う予定である。

7. 利用研究成果が無かった場合の理由
現在研究発表を行う過程の途中段階のため