

次世代シーケンサーデータ依頼解析サービスの現状

森原 なぎさ (広島大学 技術センター / 自然科学研究支援開発センター)

MORIHARA Nagisa : Introduction to next-generation sequencing data analysis services

Natural Science Center for Basic Research and Development (N-BARD) at Hiroshima University provides library preparation and data analysis services for next-generation sequencing as a shared facility. This report presents the current status of our data analysis services.

1. サービスについて

広島大学の自然科学研究支援開発センター 機器共用・分析部(霞)では、共用施設として次世代シーケンサー用のライブラリ作製と次世代シーケンサーデータ解析サービスを提供している。各1名ずつの担当者で行っているため年間件数は限定的だが、今回はデータ依頼解析サービスの現状について報告する。詳細なサービス内容は当部のウェブサイト(報告書下部参照)に記載しているが、料金はサンプル単位ではなくプロジェクト毎に設定しており、学内者は15,000円、学外者は37,500円となっている。解析内容は限定しておらず、希望の解析が実行可能かを確認した上で依頼を受ける形をとっており、ツールの指定にも対応する。納期については3週間~としているが、過去に経験のあるアプリケーション等の場合は1週間程度となることもある。

2. 有償ツールと無償ツールの比較

サービス開始の2013年当初から主に変異解析・RNA-seq・パスウェイ解析においてトータル5種類の有償ツールを使用してきたが、「次世代シーケンサーDRY 解析教本改訂第2版」著者清水厚志・坊農秀雅(秀潤社)の「発現解析」「疾患ゲノム解析」の節を参考に、Linuxへ変異解析及びRNA-seqの解析環境を導入することで、全ての有償ツールライセンス契約を2023年1月に終了した。これにより継続的なプロジェクトをフリーツールで再解析し、比較する必要があったためその結果を報告する。

(1) 変異解析

ヒトターゲットキャプチャーライブラリ(IDT社のLotus kit)から1サンプル選んでgermline(生殖細胞)変異解析を行った。有償ツールのパラメーターはデフォルト値を使用し、無償ツール(BWA-GATK)は書籍に

従った。結果として有償ツールではミスアライメントと思われる変異が所々検出され、Unexpected Pairsを示すリードがより多く見られた。検出変異数は有償ツール:294個、BWA-GATK:116個であった。ミスアライメントと思われる変異に加え、利用者の希望でgermline変異解析ではあるが頻度閾値を10%以上に設定したことで有償ツールの方では変異数が多くなっているが、BWA-GATKの条件に寄せてquality>30, frequency>30でフィルターすると120個まで減り、共通変異は108個となり結果がほぼ一致した。

(2) RNA-seq

マウス肺細胞RNA-seq 6サンプル(コントロールn=3, ターゲットn=3)で解析を行った。有償ツールのパラメーターは各社の推奨またはデフォルトを適用し正規化はTPMを使用した。無償ツールは書籍に従いSTAR, RSEM, RのDESeq2パッケージ(検定はWald-test)を使用した。検出された発現変動遺伝子数は図1の通りで、いずれもFoldChangeの閾値は2としているが、補正後p-valueの閾値は有償ツールでは0.05に設定している一方で、STAR-RSEMでは発現変動遺伝子が多く検出される傾向があるため0.01を適用している。

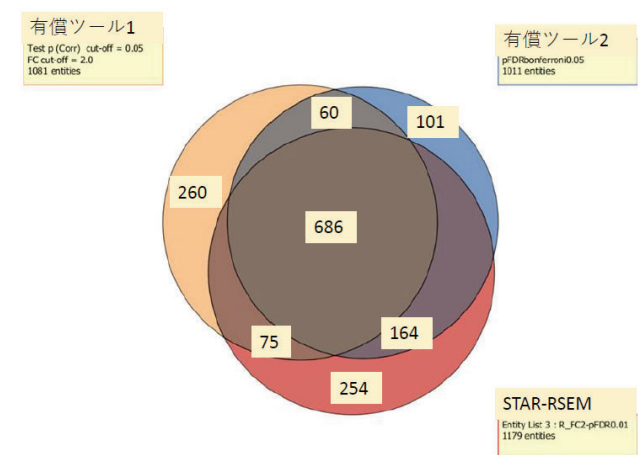


図1. 各ツールの発現変動遺伝子数ベン図

図1の発現変動遺伝子リストを使用し、Metascape (<https://metascape.org/>)によって機能解析を行った。各リストで多少異なっているが、炎症応答、サイトカイン生成制御、細胞活性など、概ね同じ機能が有意差を持つ結果となった(図2)。

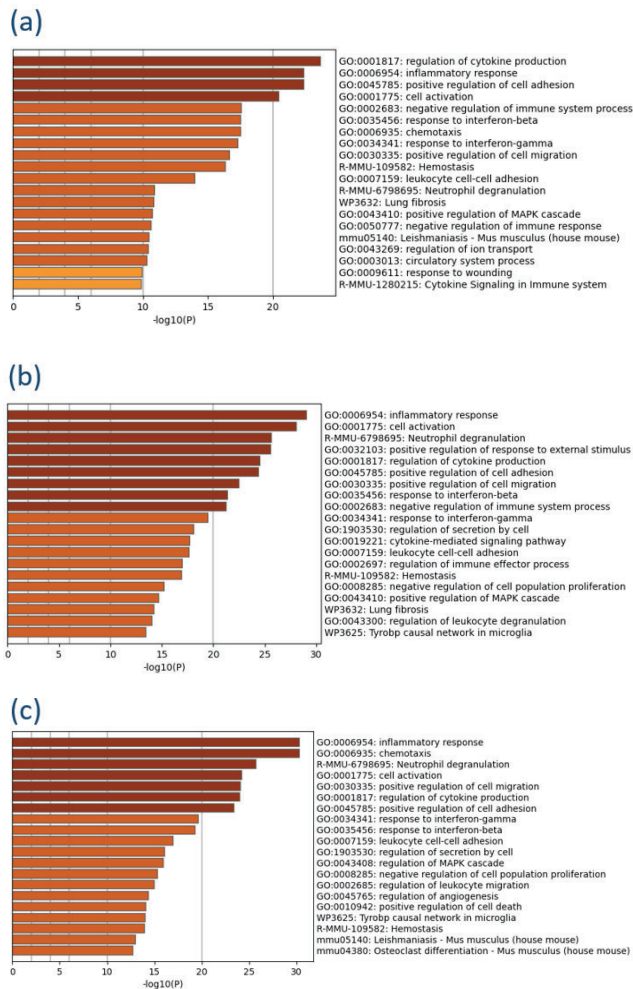


図2. 各発現変動遺伝子リストを使用した Metascape 解析結果 (a)有償ツール1 (b)有償ツール2 (c)STAR-RSEM

3. 依頼実績

依頼件数は年間15件程度となっている。2022年度一番多かったアプリケーションはRNA-seqで5件あったが、2023年度は今のところ3件と減少してきており、受託企業より発現量のテキストファイルの形で納品されることが多いこと、iDEPやRaNA-seqなどのウェブツールも充実してきていることが原因と考察している。

2023年度は10x社のChromiumX(シングルセル解析装置)を導入したことが影響し、最も多かったアプリケーションはシングルセル解析で5件となった。続い

て16Sメタゲノム解析が4件であり、その他TCRレパトア、変異解析、NGSではないがマイクロアレイ解析などの依頼が少数ずつあった。SRA等へのシーケンスリード登録、GSEA用のファイル整形など簡単な相談であれば無料で対応している。

自身のRやPythonなどの知識は主に独学のため、一から全てのコードを組み立てるのに時間がかかる。マニュアルがあるような特定のツールを使わない解析が必要な場合、生成AIに大枠のコードを書いてもらい、それを修正するといった手順を取ることがある。直近の例では、複数の株間で共通する遺伝子をパスウェイ経路に色分けする作業を効率化するため、複雑な組み合わせパターン抽出やカラーコード付加などについての基本コード作成を生成AIに任せました。

4. 今後の展望

追加解析を希望されるケースにも長期的なサポートを提供していることや、人員数が限定されていることから一度に多くの依頼を受けることは避けており、今のところ学内に対してもサービスの周知を積極的に行っていない。しかし当施設で運用している唯一の次世代シーケンサーPromethIONのデータ解析支援については、利用率向上のためこちらから働きかけていく予定である。

依頼ではなく自前PCでの解析要望が最も多い16Sメタゲノム解析に関して、Qiime2ソフトウェアのインストールから一連の解析プロセスをGitHubに記載しHPにリンクすることを考えている。特に解析環境の構築がネックになっている場合が多く、より簡単な方法を提案することに注力する。研究者自身がデータ解析をすることは重要であり、これまででもご持参いただいたコンピューターでのサポートを行っていたが、引き続き重視していく。

参考

自然科学研究支援開発センター 機器共用分析部
次世代シーケンサーデータ依頼解析サービス HP
QRコード

