

RNA-seq を用いた呼吸器疾患診断システムの開発

山岸賢司

日大工・生命

【緒論】

インフルエンザをはじめとする呼吸器疾患は、高齢者の死亡原因の上位に位置し、重症化する前に容易に診断できる手法の確立は Active ageing に大きく貢献する。本研究では、血液中のエクソソーム内の RNA の発現パターンを利用し、呼吸器疾患を早期に診断するシステムの開発を目指している。

今年度は、次世代シーケンサーを用いた RNA-seq のデータ解析手法を構築し、4つの呼吸器疾患病態モデルマウス[肺炎, インフルエンザ感染, 慢性閉塞性肺疾患(COPD), 気管支喘息]に対して応用した。

【方法】

呼吸器疾患病態モデルマウスの RNA 塩基配列は、気管支肺胞洗浄液中からエクソソームを単離し、そのエクソソームから抽出した RNA に対し、次世代シーケンサー (Ion PGM™, Life Technologies Inc.) を用いて解析した。病態モデルは、肺炎 4 個体(Lungs1-4), インフルエンザ感染 5 個体(Virus1-5), 慢性閉塞性肺疾患 3 個体(COPD1-3), 気管支喘息 4 個体(Asthma1-4), 正常 14 個体(Mock1-14)を用いた。

【結果】

1. RNA-seq データ解析プロトコルの構築

i. 参照ゲノム配列へのマッピング手法の構築

次世代シーケンスにより解析されたエクソソーム中の RNA を、2つのマッピングソフトウェアを有機的に組み合わせることで、参照ゲノム配列に正しくマッピングする手法を構築した。現在、次世代シーケンサーから得られる RNA の配列データを染色体の参照ゲノム配列上に一義的にマッピングする手法は確立されていない。特に、イントロンを跨ぐエクソン由来の塩基配列を持つリード配列は、イントロン領域がスプライシングにより欠落するため、そのままでは参照ゲノム配列と一致しない。そのため、汎用的に用いられているマッピング方法では、このようなリード配列は、参照ゲノム配列にマッピングされないという問題があった。そこで、ゲノムマッピングソフト TopHat2 を用いて、まずスプライシングを予測し、マッピングされなかったリード配列を、再度別のマッピングソフトウェア Bowtie2 を用いてマッピングした。そして、双方のマッピング結果を組み合わせることで、イントロンを跨ぐエクソン由来の塩基配列を持つリード配列を正しくマッピングできるよう設計した(Figure 1)。

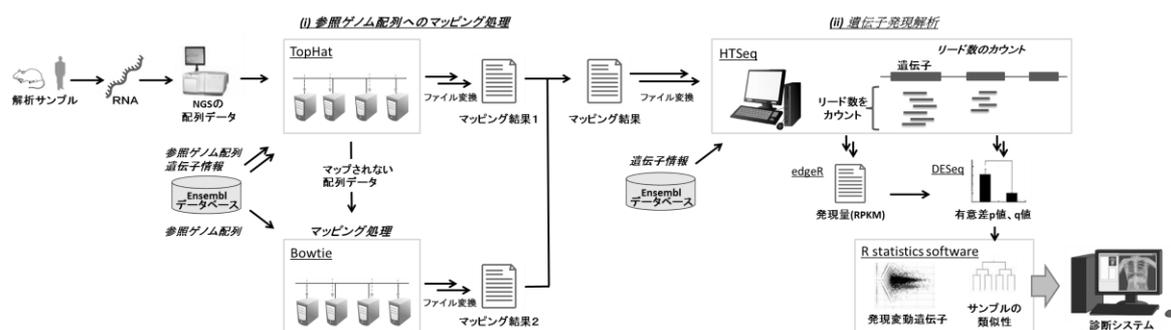


Figure 1. RNA-seq データ解析プロトコル

ii. 遺伝子発現解析手法の構築

各遺伝子の発現量は、各遺伝子にマッピングされたリード配列の数とした。HTSeqを用いて、各遺伝子にマッピングされたリード配列をカウントした。次に、遺伝子ごとにマッピングされたリードの数を比較し、有意差 p 値および q 値を、それぞれの病態モデルごとに分けて DESeq を用いて算出した。この q 値が 0.1 未満の遺伝子を発現変動遺伝子と判断した(Figure 1)。以上の一連の解析において、参照ゲノム配列と遺伝子情報は、Ensembl の mm10 version78 を使用した。

2. マウス呼吸器疾患病態モデルへの応用

呼吸器疾患の病態モデルマウスに対して、次世代シーケンサーにより塩基配列を決定した。決定した塩基配列は、構築した RNA-seq データ解析プロトコルを用いて処理し、病態モデルごとに RNA の発現量が変動する遺伝子を特定した。その結果、肺炎モデルでは、約 30,000 個の遺伝子の中から 24 個の発現変動遺伝子を特定することができた。これらの遺伝子の中には、肺炎モデルに使用した LPS 刺激に応答する NF κ B シグナル伝達経路の遺伝子の存在が確かめられた。他の呼吸器疾患病態モデルでは、インフルエンザ感染モデルは 13 個、気管支喘息モデルは 52 個の遺伝子を特定した。一方で、慢性閉塞性肺疾患モデルでは、 q 値が 0.1 未満となる遺伝子は存在せず、発現変動遺伝子を特定できなかった。

次に、発現変動遺伝子の発現量を入力因子としたクラスター解析を行い、各病態モデルマウスの発現パターンの類似性を解析した。発現変動遺伝子が特定できなかった慢性閉塞性肺疾患モデルについては、 q 値の上位 20 個の遺伝子を入力因子として採用した。その結果、解析に用いたすべてのサンプルで、病態モデルごとにクラスタリングすることができた(Figure 2)。また、上記解析に使用していない肺炎病態モデルと正常モデルの

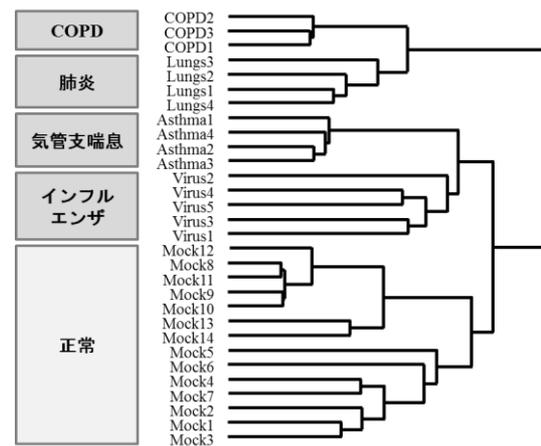


Figure 2. 遺伝子発現パターンを利用したクラスター解析による病態モデルの分類

合計 6 個体のマウスを病態未知のモデルマウスとし、RNA の発現パターンがどの疾患モデルと類似しているか解析した。その結果、肺炎モデルと正常モデルともに、同じ病態と近い階層にクラスタリングすることができた。

以上の結果から、RNA 発現パターンを利用することで、呼吸器疾患を診断するシステムに応用可能であることが示唆された。次年度は、バイオマーカー遺伝子の特定とヒトサンプルへの展開、新たな疾患関連遺伝子の探索手法の開発を行う。

【謝辞】

病態モデルマウスの作成、および次世代シーケンサーによる RNA 塩基配列の決定は、医学部内科学系呼吸器内科学分野 橋本修教授、権寧博准教授らの研究グループにより行われた。

【参考文献】

- [1]RNA-Sequence Analysis by Next-Generation Sequencer for the Early Diagnosis of Respiratory Disease. Y.Fukano, Y.Miyazawa, T.Mikoshi, T.Inoue, S.Maruoka, K.Kuroda, Y.Gon, S.Hashimoto, K. Yamagishi*. J. Comput. Chem. Jpn, 13, 6 332, 2014 (DOI: 10.2477/jccj.2014-054)
- [2]RNA-seq に対する NGS データ解析プログラムの開発。見越大樹, 深野義人, 宮澤由妃, 山岸賢司*. 日本コンピュータ化学会誌. 13 号 6 巻 299, 2014 (DOI: 10.2477/jccj.2014-0049)
- [3]次世代シーケンサーを用いた RNA-seq 解析手法の確立と呼吸器疾患症早期診断システムへの展開。深野義人, 宮澤由妃, 山岸賢司他, 生命医薬情報学連合大会, 2014. 10. (仙台) [Lightning Talk 選出発表]