

小型長鎖シーケンサーを用いた RNA-Seq 解析

Analysis of RNA-Seq with portable long-read sequencer

メディカル情報生命専攻 2018 年 3 月修了 指導教員:鈴木穰 教授 47-166411 勝俣恵理

【キーワード】 ナノポアシーケンサー、RNA-Seq、ロングリード

【背景】

Oxford Nanopore Technologies 社が上市したロングリードシーケンサー MinION は、機器のサイズが小型である、比較的安価なコストで導入できるという特徴を持つ。従来のシーケンサーと比較して容易に導入が可能であること、長鎖解読特性を持つことから、転写産物全体の構造の同定や構造多型の解析に有力な手法となり得る。ところが、MinION を用いた RNA-Seq で実用的な解析はまだ行われていない。本研究では、6 種類の肺腺癌細胞株及び同一個人由来の各 7 臓器から得られた RNA に対して MinION を用いた cDNA-Seq (cDNA を生成・増幅してシーケンス)解析並びに direct RNA-Seq (RNA を直接シーケンス)解析を行った。得られたデータを用いて、発現定量や融合遺伝子の検出、アリル不均衡に発現している遺伝子の探索、ハプロタイプフェージング等の解析を行った。

【手法】

cDNA-Seq において、肺腺癌細胞株は細胞から抽出した total RNA、及び男女 2 検体の 7 臓器から得られた total RNA を用いた。SMART-Seq2 法 (1)を用いて全長 cDNA を合成、増幅した。direct RNA-Seq において、LC-2/ad 細胞株から抽出した polyA+ RNA を用いた。それぞれ MinION でシーケンスを行った。cDNA-Seq では得られたデータのうち品質の高い pass 2D リードを、direct RNA-Seq では 1d リードを解析に用いた。得られたデータは HiSeq シーケンサー (illumina 社)を用いた従来法 RNA-Seq データと比較した。ゲノム多型または変異情報として HiSeq シーケンサーを用いた WES 解析により多型データを検出した。多型データと MinION データを組み合わせると解析を行った。

【結果と考察】

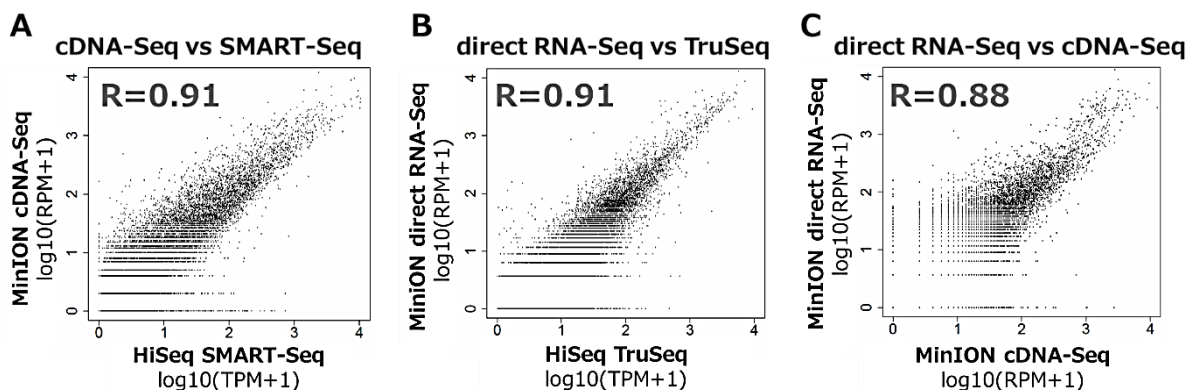
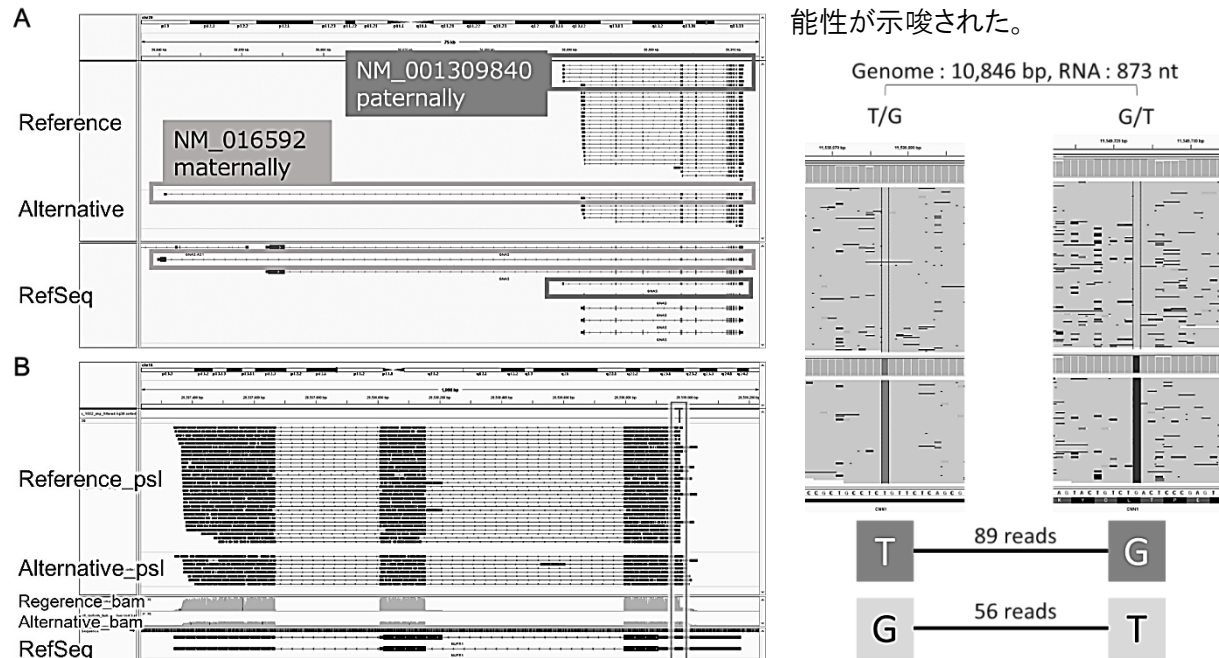


Fig 1. 各プラットフォーム間の遺伝子発現比較

各遺伝子の発現量を散布図で示した。(A) SMART-Seq 法でライブラリ調製した HiSeq データ, MinION での cDNA-Seq データの比較。(B) TruSeq でライブラリ調製した HiSeq データ, MinION での direct RNA-Seq データの比較。(C) MinION での cDNA-Seq と direct RNA-Seq の比較。

MinION での cDNA-Seq, direct RNA-Seq, HiSeq での RNA-Seq の間で遺伝子発現の比較を行った。cDNA-Seq データと SMART-Seq 法でライブラリ調製した HiSeq データの間では相関係数が 0.91 を示した (Fig 1 A)。direct RNA-Seq データと TruSeq でライブラリ調製した HiSeq データの間では相関係数が 0.91 を示した (Fig 1 B)。cDNA-seq データと direct RNA-Seq データの間では相関係数は 0.88 を

示した。同じシーケンサーを用いた cDNA-Seq と direct RNA-Seq のデータ間の相関よりも direct RNA-Seq と TruSeq のデータ間のほうが高い相関を示した。RNA を直接シーケンスする direct RNA-Seq は逆転写や PCR によるバイアスがかかりにくいと考えられている。TruSeq でのサンプル調製は全長ではなく、断片化された cDNA を合成・増幅するため、SMART-Seq に比して PCR バイアスが少ない可能性が示唆された。



左: Fig 2. アリル不均衡に発現している遺伝子

臓器由来 RNA の cDNA-Seq で得られたデータを IGV 上で可視化した。(A) SNP においてリファレンスと同じ塩基を持つリード群に父性(paternal)なアリルが、リファレンスと異なる塩基を持つリード群に母性(maternal)なアリルが発現していた。(B) 女性由来肝臓データの NUPR1 遺伝子。ヘテロ SNP は四角で囲まれた座位で検出された。

右: Fig 3. CNN1 遺伝子におけるハプロタイプフェージング

CNN1 遺伝子にマップされたリードのうち、2 カ所の SNP をまたぐリードの本数と、SNP 上の塩基の組み合わせを示した。ゲノム上で 10,846bp 離れたヘテロ SNP は「TG」「GT」の組み合わせで発現している。

MinION で各臓器の cDNA-Seq を行い、アリル間の発現に偏りがある遺伝子の確認とハプロタイプフェージングを行った。GNAS 遺伝子はアリル間で偏って発現している isoform の存在が知られている (2)。MinION データで GNAS 遺伝子において片アリルにのみ発現している isoform が認められた (Fig 2A)。各臓器において、アリル間で偏って発現することを $p < 0.01$ で認められる isoform が 9-221 検出された。アリル不均衡に発現している例として、女性由来肝臓データにおいて、NUPR1 遺伝子を示した (Fig 2B)。同一 isoform 上で複数のヘテロ SNP をカバーするリードの持つ塩基を調べ、フェージングを行った。CNN1 遺伝子で行ったフェージングでは、「TG」「GT」の組み合わせでを有するリードがそれぞれ 89 本、56 本確認された (Fig 3)。MinION データを用いることで容易にフェージングできることが示唆された。

【参考文献】

- (1) Picelli, S., *et al.* (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171-181.
- (2) Bastepe, M. (2007). The GNAS Locus: Quintessential Complex Gene Encoding Galpha, XLalphas, and other Imprinted Transcripts. *Curr. Genomics* 8, 398-414.