

1. プロジェクト名

RNA-Seq データ取得+ベーシック解析 (データ量 4 G b + 1 4 検体)

2. 作業概略

- 2-1) 作業内容 : 真核生物 RNA-Seq (Poly-A 選択法, Stranded)
- 2-2) 受け入れサンプルまたはデータ形式: total RNA
- 2-3) 生物種 : human
- 2-4) サンプル数 : 1 4 サンプル
- 2-5) リファレンスゲノム: h g 3 8

3. データ取得仕様

- 3-1) 利用機器 : Illumina NovaSeq 6000
- 3-2) 取得リード長: PE150 (150bp×2 paired-end)
- 3-3) 取得データ量: 4 G bases per sample
- 3-4) 取得リード数: 26.7 M reads per sample (13.3M paires)
- 3-5) アプリケーション: 真核生物 RNA-Seq (Poly-A 選択法, Stranded)
- 3-6) 使用試薬: NEBNext Poly(A) mRNA Magnetic Isolation Module  
(Cat No.E7490)  
NEBNext Ultra II Directional RNA Library Prep Kit  
(CAT NO.E7760)

4. 作業手順

- 4-1) BioAnalyzer または同等の機器を用いて、サンプルの品質確認を行う。
- 4-2) クオリティチェックをパスしたサンプルについて、シーケンスライブラリーを作製する。
- 4-3) シーケンスを行う。
- 4-4) プログラム FastQC を用いて、シーケンスリードの品質確認を行う。
- 4-5) プログラム Trimmomatic を用いて、シーケンスリードのトリミングを行う。
- 4-6) プログラム HISAT2 を用いてリファレンスゲノムへのマッピングを行う。
- 4-7) プログラム featureCounts を用いて、各遺伝子の Raw リードカウントを計算する。
- 4-8) プログラム featureCounts を用いて算出した Fragments カウントを元に各遺伝子の発現量 (TPM) を計算する。
- 4-9) サンプル間でデータ傾向の相関評価、クラスタリング解析及び PCA 解析を行う。
- 4-10) TPM および Raw リードカウントを用いたヒートマップを作成する。

## 5. 納品物

- 5-1) raw fastq : \*.fastq.gz
- 5-2) トリミング後の fastq : \*.trimmed.fastq.gz
- 5-3) マッピング後のリードデータ : \*.sorted.bam, \*.sorted.bam.bai
- 5-4) bigWig データ : \*.bw
- 5-5) Raw reads count 値テーブルデータ : \*.txt
- 5-6) TMP 値テーブルデータ : \*.txt
- 5-7) 系統クラスタリングデータ (dendrogram) : .png ファイル
- 5-8) 相関データ (pairs plot) : .png ファイル
- 5-9) PAC クラスタリングデータ : .png ファイル
- 5-10) 作業完了報告書 .pdf ファイル

## 6. 納期

令和5年12月22日(金)

## 7. 納入形式

電磁的記録媒体 (USB メモリまたはハードディスクドライブ) または  
クラウドサーバー経由による納入

## 8. 納入場所

山形県鶴岡市覚岸寺字水上 246-2 C-6

国立がん研究センター・鶴岡研究連携拠点がんメタボロミクス研究室 牧野嶋まで