

bb01/ xPore

เป็น 1100 ปี 150 RNA modifications

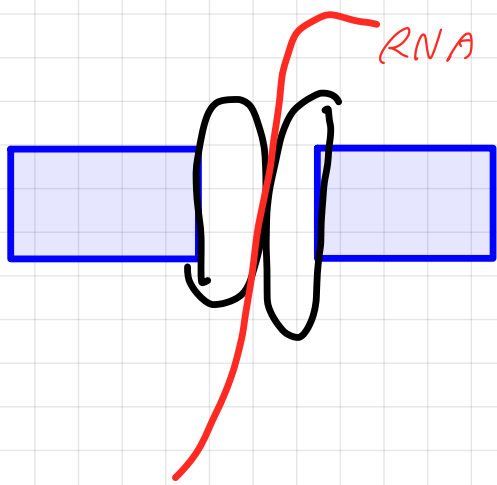
RNA modification

คือ การเปลี่ยนแปลงของ 1 base pair

1. Problem Statement

วิธีคิด RNA Sequencing ในกรณีทั่วไป ใช้วิธี NGS

การนำสาย RNA ที่ 1 base A G C U มาใส่ในเครื่อง NGS



ที่เครื่อง NGS นั้น จะใช้หลักการวัดการเปลี่ยนแปลงของ RNA

ซึ่งมี output ของเครื่อง NGS เป็น raw data

แต่ในทางปฏิบัติแล้ว เราต้องนำ raw data ไปประมวลผลก่อน

เรียกว่า RNA modifications

ซึ่งได้มาจาก raw data เราจะได้ modified RNA

ดังนั้น วิธีคิด ML ในกรณีนี้ เราจะใช้ข้อมูลจาก raw data ของ NGS มาประมวลผลก่อน แล้วจึงนำข้อมูลที่ได้มาใช้ในการวิเคราะห์ RNA modifications

xPore

1. สามารถใช้ทำนายการเปลี่ยนแปลงของ RNA modifications ได้

2. ไม่ต้องการข้อมูลในกรณีอื่น

2.Data Collection and Preparation

ข้อมูล

1. FAST5

เก็บ raw signal (สัญญาณไฟฟ้า) มาเป็น PA
อยู่ใน HDF5 format (binary)

2. FASTQ

เก็บข้อมูล sequencing ที่อ่านแล้ว

Format

- @ Name/ID
- sequence
- Optional info, starting with '+'
- quality

3. FASTA

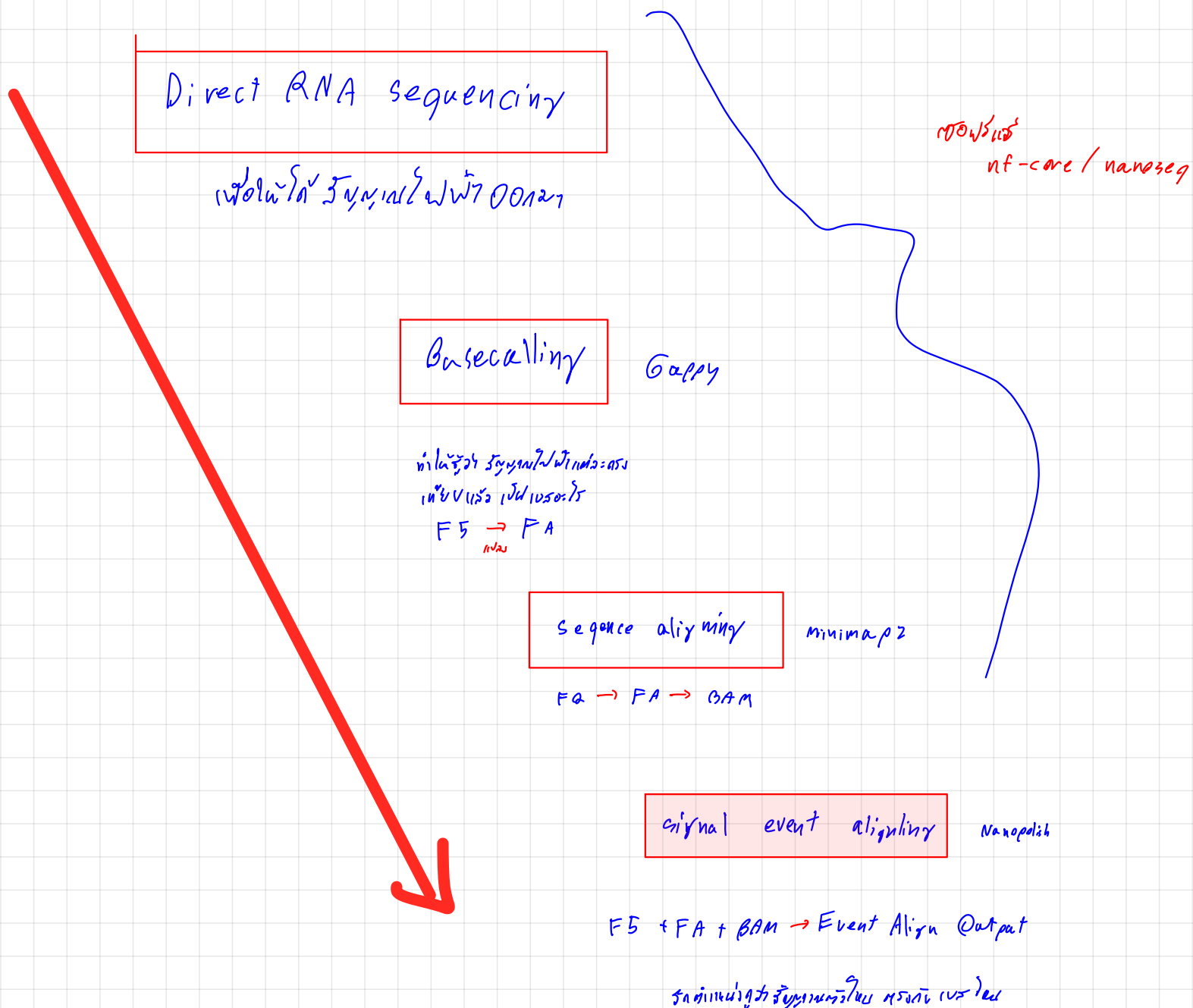
เก็บ sequence ที่อ่านแล้วมาจัดเป็น DNA หรือ RNA

- > ID
- RNA หรือ DNA

4. BAM/SAM

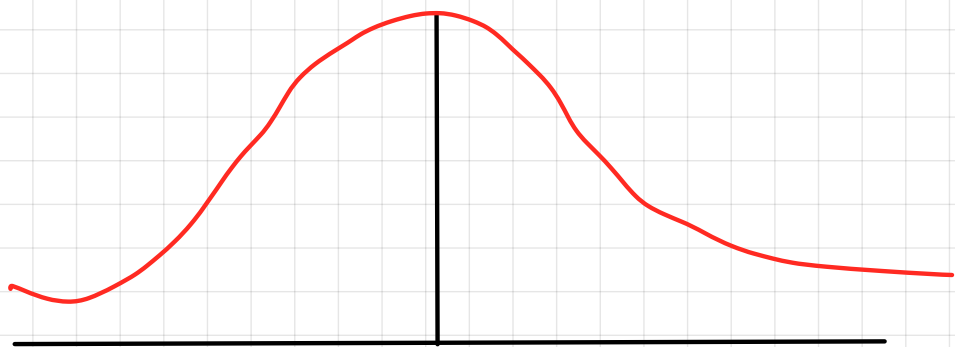
เก็บ alignment ของ FASTA กับ REFERENCE genome

- เก็บ alignment



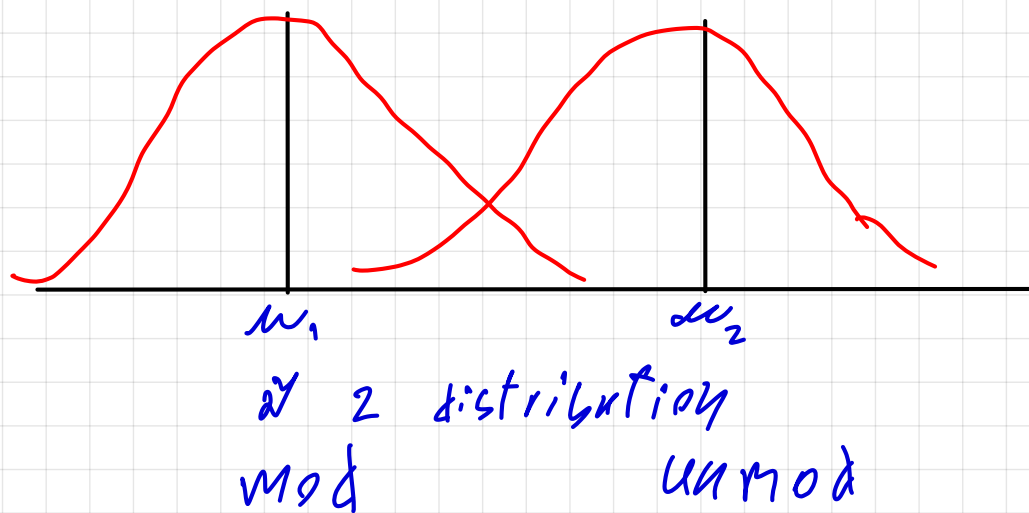
3. Bayesian[Multi-Sample] Gaussian Mixture Modelling

Gaussian (normal distribution)



$$p(x) = \sum_{k=1}^K \pi_k N(x | \mu_k, \Sigma_k)$$

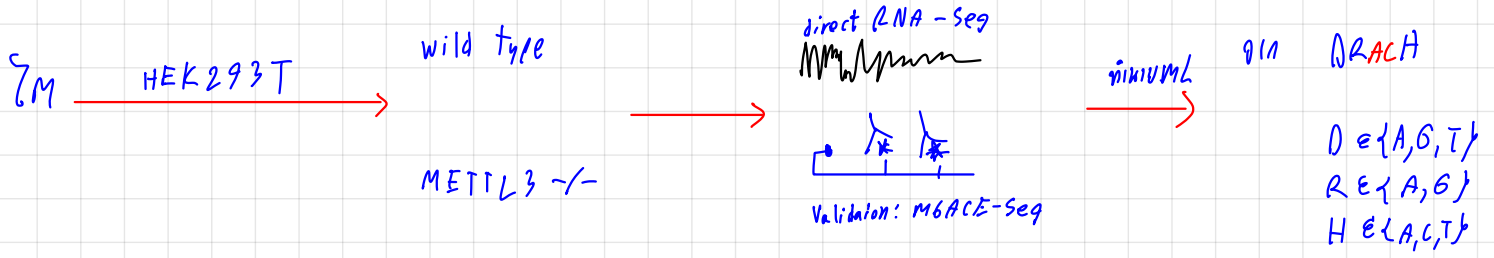
Gaussian (Mixture) or GMM



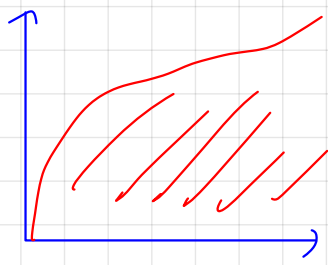
Example: GMM

- in which case
- Frequentist
 - Bayesian

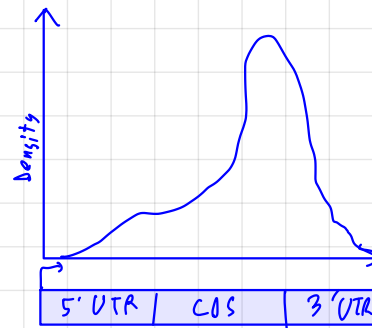
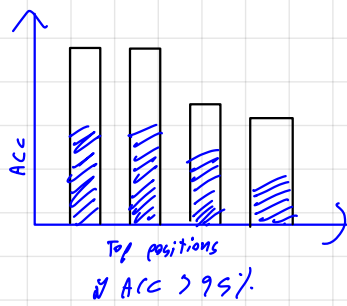
7. $\overline{150591} \quad 151000$



2. Validation: M6A calling



$AVC : 0.86$



Full Dataset
ทำกราฟแสดงที่ขึ้นกับตัวแปรอิสระ

Keys Takeaway

- Validation of ML metrics from: SN
- Comparison
- Applicability