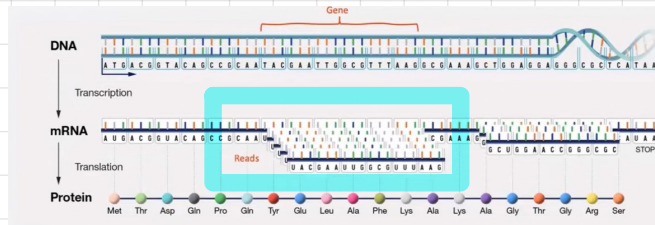


xPore: An AI-Powered App for Bioinformaticians

1. Problem Statement

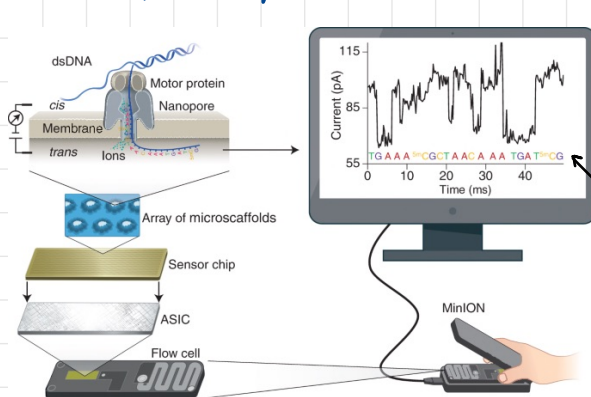
▷ ศึกษาจาก "read ของ mRNA"

▷ ตรวจสอบ m^A / Base A ที่เปลี่ยนไป



เครื่องมือ

ในงานวิจัยใช้ **Nanopore sequencing** โดย **MinION**



สิ่งที่บ่งบอก ลำดับเบส คือ ความต้านทาน

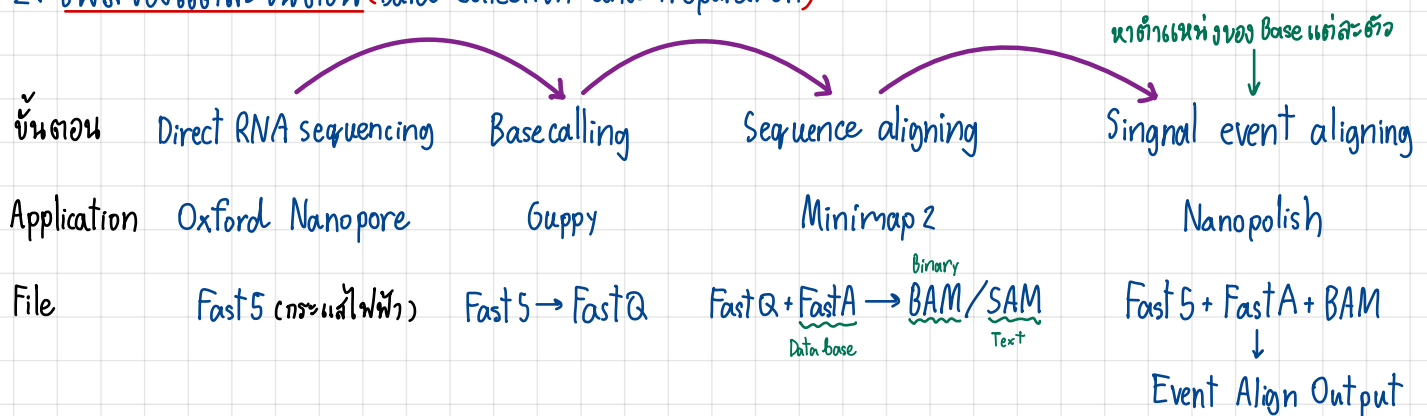
นำกระแสที่ได้ ไปเข้า **Base calling**

ทำให้แปลงเป็นตัวหนังสือ

ข้อดี : ขนาดเล็ก , Real Time , ใช้ mRNA

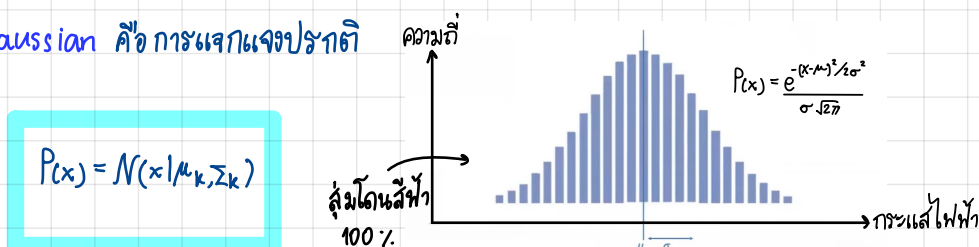
* Objective : ตรวจสอบ Base ที่แตกต่างจาก Data base ของ read mRNA นั้น

2. ไฟล์ของแต่ละขั้นตอน (Data Collection and Preparation)



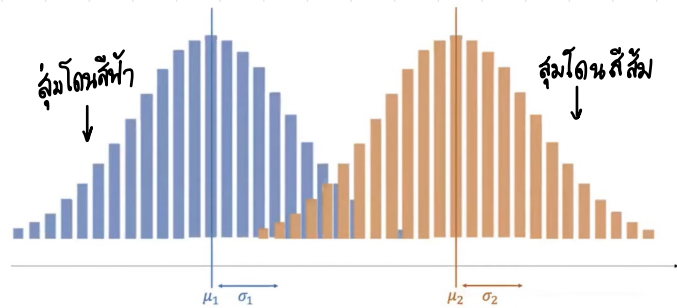
3. Bayesian Multi-Sample Gaussian Mixture Modelling (GMM)

▷ Gaussian คือ การแจกแจงปกติ



▷ Gaussian Mixture Modelling

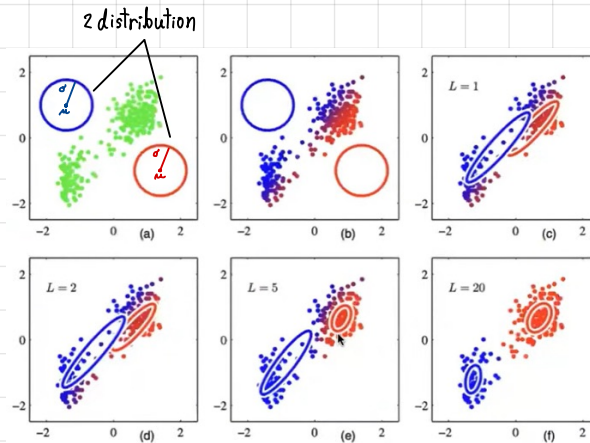
$$P(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x | \mu_k, \Sigma_k)$$



• หลักการของ GMM Inference

ตัวอย่าง //

Iterative algorithm

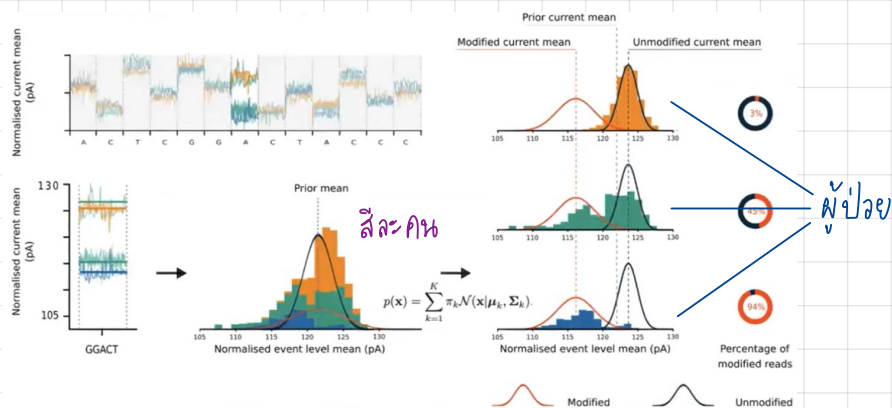


"ทำให้รู้ว่าแต่ละจุดมาจาก วงกลมไหน"

▷ Bayesian : วิธีใช้ประมาณค่า μ_k, σ_k, π_k

ข้อดี → ถ้าใส่จำนวน distribution Bayesian จะให้แม่นยำมากแทนไม่จำเป็นต้องครบ ทำให้รู้ว่ามันมี distribution จริงๆ

▷ Bayesian Multi-sample Gaussian Mixture Model



สรุปงานวิจัย

1. มี distributions ได้ 2 อัน (Unmodified, modified)
2. ใช้หลายตัวช่วยได้พร้อมกัน
3. ใช้ Gaussian
4. หา modification rate จาก raw input ได้
- ★ 5. ทำให้ได้เร็ว → Parallelisation

4. Evaluation

▷ หลังจากได้ผลมา

▷ การวิเคราะห์ : ML Metrics → ROC curve, Precision-Recall curve, Accuracy

Analysis → Domain specific evaluation, Effects of the data size

- ▷ ข้อสรุปประสงค์
1. xPore สามารถระบุได้ว่า m⁶A ตำแหน่งไหนที่น่าสนใจใน Top position ว่ามีความชัดเจนสูง confirm ความรู้เก่า
 2. ต้องหาได้ว่าตำแหน่งที่ modified ไป มี modification rate เท่าไร ก็ read ที่ %

เพิ่มความ Impact → ลองกับ full dataset ให้ user เชื่อใจ
→ ใช้กับ dataset อื่น (อวัยวะอื่น)
→ ใช้กับคนจริง

* Evaluation ใช้เวลานานสุด เพราะ "dataset เยอะ"

5. Visualization and Presentation

▷ ต้องทำให้เข้าใจง่าย จัดเรียงข้อมูลให้เป็นระเบียบ

▷ เผยแพร่ลง Online ให้คนอื่นใช้

▷ รับ comment มาแก้