

E-SAN THAILAND CODING & AI ACADEMY

โครงการวิจัยโมเดลระบบนิเวศการเรียนรู้ที่บูรณาการ CODING & AI สำหรับเยาวชน
Model of Learning Ecosystem Platform integrate with Coding & AI for Youth

โครงการย่อยที่ 6
การพัฒนาเยาวชนเพื่อเข้าสู่วิชาชีพขั้นสูงด้าน Coding & AI
ร่วมกับ Coding Entrepreneur & Partnership: Personal AI

xPore

AI-Powered App for Bioinformaticians

นักชีววิทยา สาขาวิชาคห

ผศ. ดร.นฤมล ประภานวนิช
โครงการย่อยที่ 6

E-SAN THAILAND CODING & AI ACADEMY

โครงการวิจัยไมโครสโคปการเรียนรู้กับฐานการ CODING & AI สำหรับเยาวชน
Model of Learning Ecosystem Platform integrate with Coding & AI for Youth

ARTICLES
<https://doi.org/10.1038/s41587-021-00949-w>

Scopus metrics
78 99th percentile
Citations in Scopus
9.61 Field-Weighted citation impact

nature biotechnology

Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore

Ploy N. Pratanwanich^{1,2,3,✉}, Fei Yao^{1,1}, Ying Chen^{1,1}, Casslynn W. Q. Koh^{1,1}, Yuk Kei Wan^{1,1}, Christopher Hendra^{1,4}, Polly Poon¹, Yeek Teck Goh¹, Phoebe M. L. Yap¹, Jing Yuan Chooi⁵, Wee Joo Chng^{5,6,7}, Sarah B. Ng¹, Alexandre Thierry⁸, W. S. Sho Goh^{1,9} and Jonathan Göke^{1,10,✉}

RNA modifications, such as N¹-methyladenosine (m¹A), modulate functions of cellular RNA species. However, quantifying differences in RNA modifications has been challenging. Here we develop a computational method, xPore, to identify differential RNA modifications from nanopore direct RNA sequencing (RNA-seq) data. We evaluate our method on transcriptome-wide m¹A profiling data, demonstrating that xPore identifies positions of m¹A sites at single-base resolution, estimates the fraction of modified RNA species in the cell and quantifies the differential modification rate across conditions. We apply xPore to direct RNA-seq data from six cell lines and multiple myeloma patient samples without a matched control sample and find that many m¹A sites are preserved across cell types, whereas a subset exhibit significant differences in their modification rates. Our results show that RNA modifications can be identified from direct RNA-seq data with high accuracy, enabling analysis of differential modifications and expression from a single high-throughput experiment.

downloads 27k

makeagif.com

Logos: CU CHULALONGKORN UNIVERSITY, Genome Institute of Singapore (GIS), NUS National University of Singapore

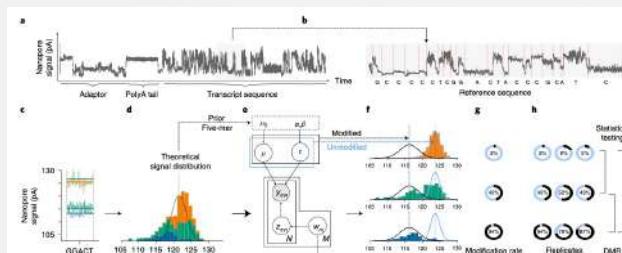
xPore App: RNA modification software xPore uses machine learning to analyze nanopore sequencing data for RNA modifications. It integrates with the E-SAN Thailand Coding & AI Academy platform.



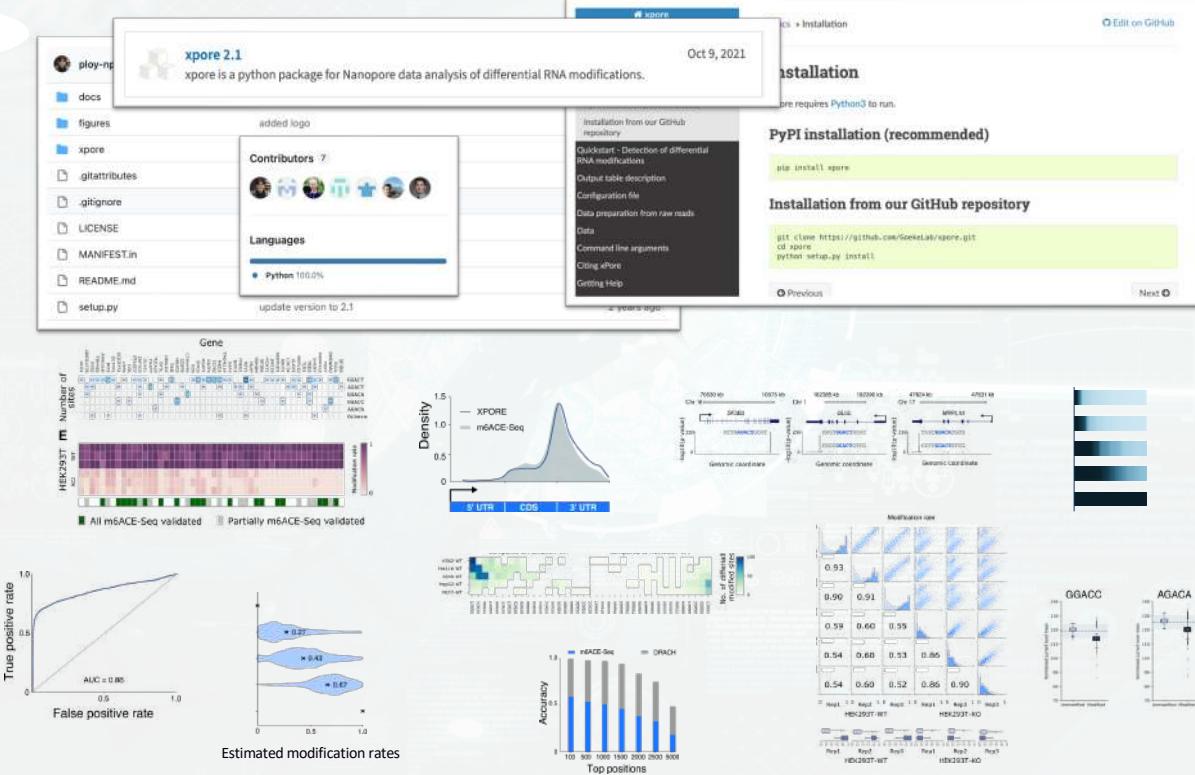


E-SAN THAILAND CODING & AI ACADEMY

โครงการวิจัยไมโครสโคปนีโอศึกษาเรียนรู้กับฐานการ CODING & AI สำหรับเยาวชน
Model of Learning Ecosystem Platform integrate with Coding & AI for Youth



XPORE





E-SAN THAILAND
CODING & AI ACADEMY

โครงการวิจัยโมเดลระบบปั้นเวศการเรียนรู้กับบูรณาการ CODING & AI สำหรับเยาวชน
Model of Learning Ecosystem Platform integrate with Coding & AI for Youth

Outline



การพัฒนาเยาวชนเพื่อเข้าสู่วิชาชีพขั้น
สูงด้าน Coding & AI ร่วมกับ Coding
Entrepreneur & Partnership:

Personal AI

1 Problem Statement

รับฟังว่า: ผู้คน software นี้จะเป็นอย่างไร

2 Data Collection and Preparation

รับฟัง: ภาระของ Data

3 Bayesian [Multi-Sample] Gaussian Mixture Modelling



Machine learning technic

4 Evaluation

รับฟัง: เสียง

5 Visualization and Presentation

รับฟัง: ภาพ

6 Future Work



CODE COMBAT

Google

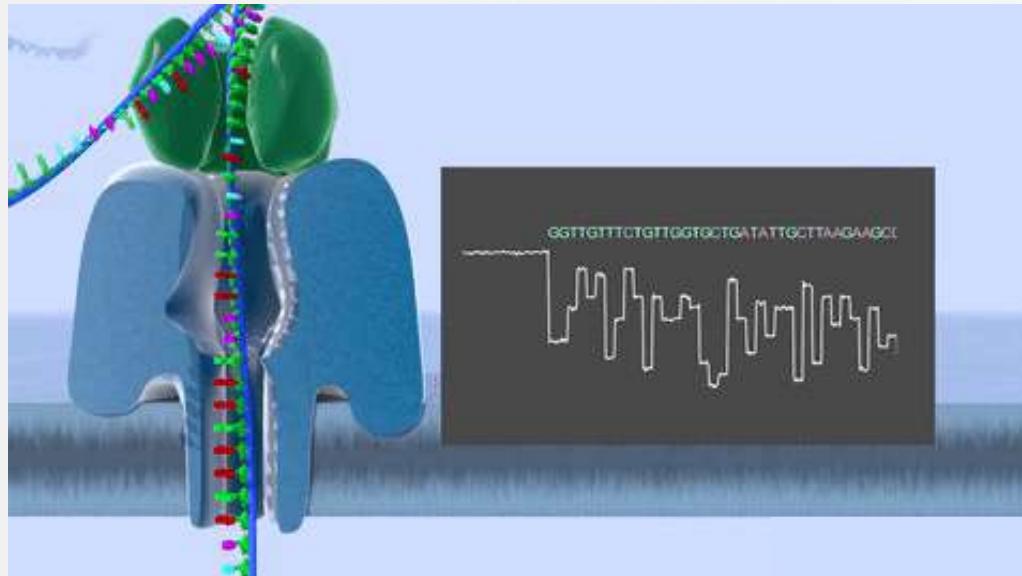


E-SAN THAILAND
CODING & AI ACADEMY

โครงการวิจัยโมเดลระบบปั้นเวศการเรียนรู้กับฐานการ CODING & AI สำหรับเยาวชน
Model of Learning Ecosystem Platform integrate with Coding & AI for Youth

1. Problem Statement : ปีช. , goal คืออะไร ทำ ให้เข้ากับงานวิจัย

ผู้ใช้งานต้อง
ในส่วนที่ต้องการ
User
dev



Data
Scientist



Bioinfomatician

Biologist

- **Nanopore Sequencing** คือ เทคโนโลยี
 - **RNA Modification** มีบทบาทสำคัญมาก ไม่ว่าจะเป็น คุณภาพของ RNA หรือความต้องการของmotivation ก็
 - **Inputs & Outputs** ภาคภาษาพื้นเมือง
 - **Research Objectives**
- พื้นที่ identify → After that, identify →

Central Dogma

7. ศึกษา domain knowledge กีโนทิวัลรีส一站 / ศักยภาพทางด้าน

DNA ต่อไปนี้คือตัวอย่าง AUCG

ซึ่งใน RNA ที่ว่าด้วย DNA ใน cell

DNA ต่อไปนี้คือตัวอย่าง mRNA ที่ว่าด้วย RNA

Transcription

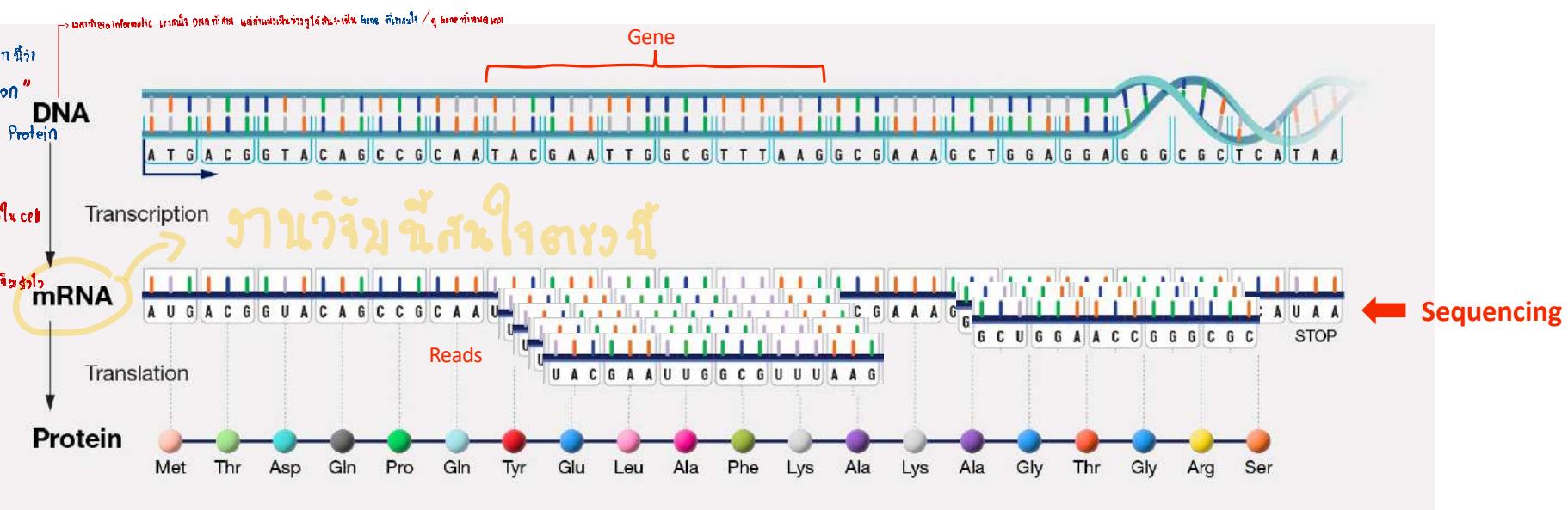
ผลลัพธ์คือ mRNA ที่ต่อไปนี้คือตัวอย่าง Protein
ซึ่งเป็นตัวอย่าง RNA

Translaction

Protein ที่ต่อไปนี้คือตัวอย่าง Protein ใน cell

โดย mRNA ต่อไปนี้คือตัวอย่าง RNA

DNA mRNA protein ต่อไปนี้คือตัวอย่าง Protein



mRNA คืออะไร Data

คือที่ทางเดินไปถึง ทางเดินทางของ基因 / → cell "Gene Expression"

การสืบทอด DNA / ชีววิทยา → ทาง Gene ทางเดินทาง ที่เดินทาง ไป transcript ทำให้ เกิดมหัศจรรย์ transcript ทาง / ทาง

ทางเดินทาง mRNA

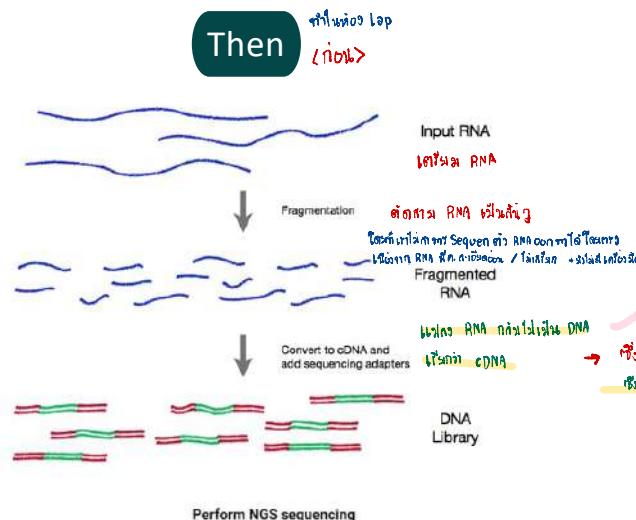
ทางเดินทาง mRNA : ทางเดินทาง Read บน mRNAs ทาง / ทาง ทาง Base ทางเดินทาง กันได้

Source: <https://www.genome.gov/genetics-glossary/Central-Dogma>

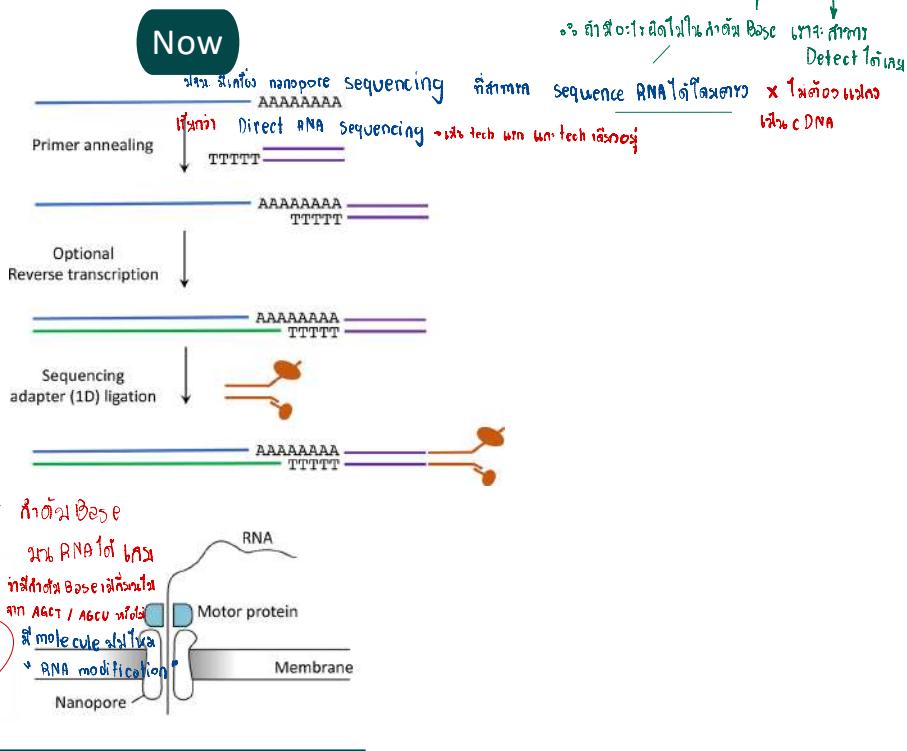
โครงการวิจัยไมโครทรานส์ฟอร์เมอร์บูรณาการ CODING & AI สำหรับเยาวชน
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

> ต่อไปนี้ mRNA ไม่ใช่ sequencing เท่ากัน ให้เขียนคำศัพท์ ต่อไปนี้คือตัวอย่าง mRNA คืออะไร: Gene กับ mRNA / ช่องทาง กตัญญู กันไปไหน กันมา กันว่า

RNA Sequencing



Direct RNA Sequencing



โครงการวิจัยโน้ตเดลรุ่บบົນຄາຣເຮັດວຽກ CODING & AI ສໍາຫຼັບເຍວະນ
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

ກອກອອນໄລນ໌ ແລະ ສໍາເລັດ sequence → ຕອດເລັດດັກຕະຫຼາດການໄໝຂ່າຍ້າ ກ່າວໃນ; ສໍາເລັດໄມ້ເຄີຍ ກົກລວງລວມ molecule ຕັ້ງ



CODE COMBAT

Google

The Asia Foundation

DMAP
DIGITAL MAKERS
ASIA PACIFIC

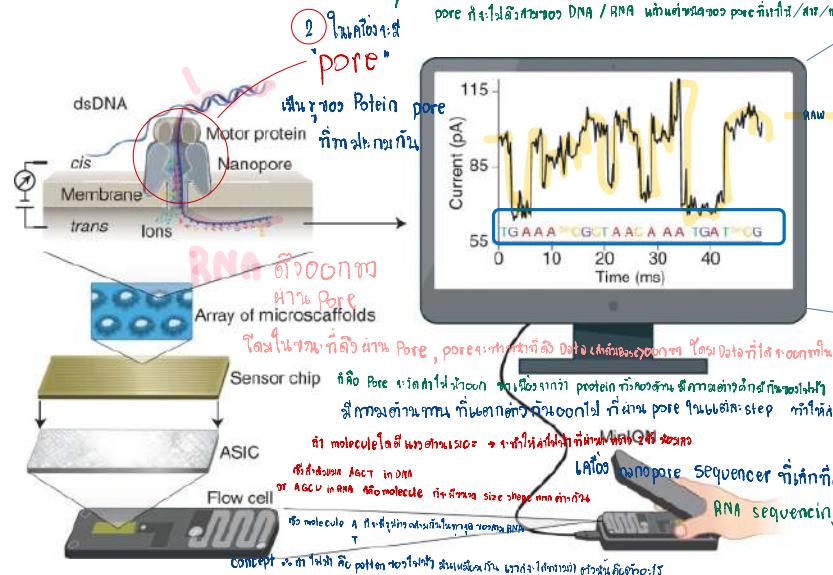
THAI
PROGRAMMER



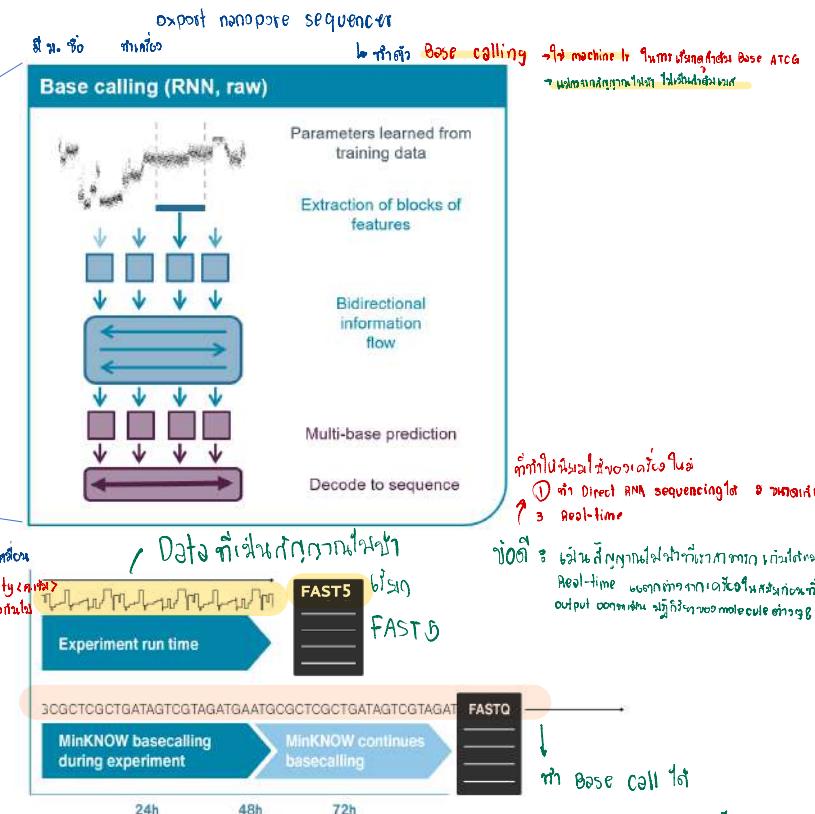
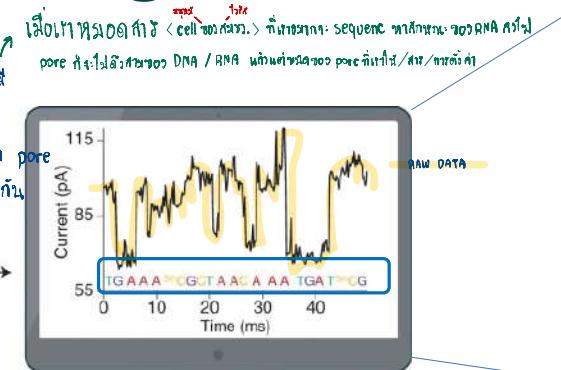
E-SAN THAILAND
CODING & AI ACADEMY

RNA Sequencing

① ໜ້າດັກ RNA ເຊື້ອໄຫວ້ວ



ກ. ກຳງານໂຄງເຄີຍ



Ref: Yunhao Wang, et al., "Nanopore sequencing technology, bioinformatics and applications", *Nature Biotechnology* (2021)

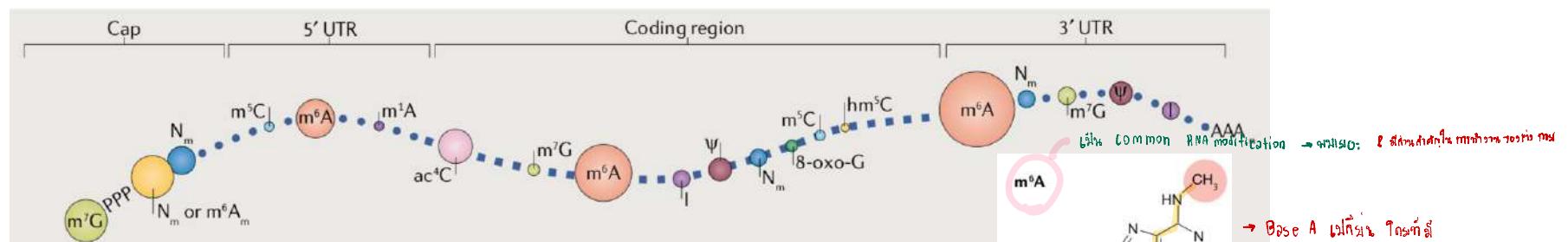
ໂຄງການວັດຍໂດລະບົບເນື່ອເວົາການເຮັດວຽກ CODING & AI ສໍາຮັບເຍວະນ
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

RNA ជាបានុញ្ញនៃការបង្កើត និងការបញ្ចូលព័ត៌មាន ដូចជា AGCU, ឬបានុញ្ញនៃការបង្កើត និងការបញ្ចូលព័ត៌មាន molecule នៅក្នុងបន្ទាន់រីន្ទាន់ ហើយវាត្រូវបានការពារឡើង

គឺជាសាស្ត្រកើស OH group ដែលបានការពារ → mRNA modify រាយការណ៍ និងការសែនសំខាន់ខាងក្រោមនេះ នឹងធ្វើឱ្យបានអាមេរិកសាខាដែលបានបង្កើតឡើង



RNA modifications



Ref: Zaccara, Sara, Ryan J. Ries, and Samie R. Jaffrey. *Nature Reviews Molecular Cell Biology* (2019)

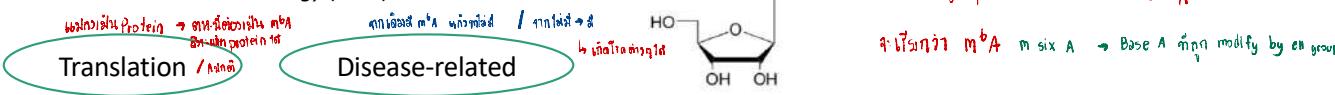
គឺជាការកែតាំង:
សំរាប់

Splicing

RNA Instability

អារិកស្រួលព័ត៌មាន → ការបង្កើត និងការបញ្ចូលព័ត៌មាន
និងការកែតាំង

Disease-related



Single-base-resolution CLIP-based detection methods

Use antibodies to induce truncations or mutations at m6A sites during reverse transcription. → គឺ Enzyme នៃការបង្កើតការបញ្ចូលព័ត៌មាន Sequencer និង short read



m6ACE-Seq
Ref: Koh, Casslynn WQ, Yeeck Teck Goh, and WS Sho Goh. *Nature Communications* 10.1 (2019)

ការបង្កើត RNA ការបញ្ចូលព័ត៌មាន sample ដែលត្រូវការពារ នៅក្នុង m6A (ការបង្កើត) → នៅក្នុងការបង្កើត និងការបញ្ចូលព័ត៌មាន sequenc → Detect នៅក្នុង m6A នៃ RNA និងការបង្កើត Enz ក្នុង

កម្រស់ការបង្កើត និងការបញ្ចូលព័ត៌មាន
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

nanopore sequencer ការបង្កើត

នៅក្នុងការបង្កើត Detect RNA modification នៅក្នុងការបង្កើត និងការបញ្ចូលព័ត៌មាន តាមក្នុងការបង្កើត

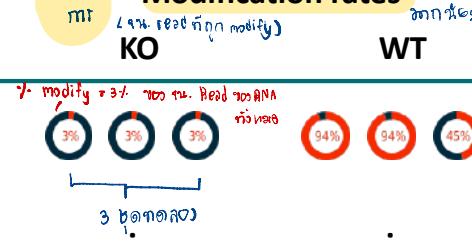
Output Table

๐ សັ່ງທີ່ koh ອາກເຫັນຊາກກາ ມີຄື່ອງ ກວດວິວໄດ້

ອາກເມື່ອໃຈ່າຍ ທີ່ຢູ່ນາມາ ດັກ ກໍານົວ ຕາມ. ທີ່ສຳນຸ ແລກຕ່າງ
ສື່ວົວ ຕາມ modify ຂອງນີ້ ພັດຕັກກ່າວໄປໄວ ແລະ ດີນກົງກົດ ບໍ່ໄດ້

- ① ອາກເສດຖະກິນ ໃນຖາງ position
- ໧. Genomic position
- ສໍາພັບໃນຕູກ ອາກເທົ່ານີ້ ດັກກ່າວໄປໄວ ເພື່ອສົ່ງໄດ້
- or ຖາງ ຮູ່ position ຢັ່ງ transcriptome-wide
- ຄວາມຮັບຮັດຂອງນັບ Gene & ພັດຕັກposition
- ແລະ ພັດຕັກໃນຕູກ 5-mer

Genomic positions	ບ້ານຕ່າ ໃນຕູກ ຕອນເຄື່ອນໄຫວ ໂດຍເກົ່າ 5-mer ທີ່ກີ່ ດັກ molecule		Differential modification rates	
Transcriptome-wide			W _{WT} - W _{KO}	P-value
NNANN	(ສັງເກດທີ່ໄດ້ຕູກການ)	→ 1-101 ດັກກ່າວທີ່ຄ່າກົງ	0.81	Most sig
NNCNN	3-mer	ສໍາເກົນນັ້ນ ບໍ່ມີກົງ ນີ້ ທີ່ກີ່	0.42	
NNGN	1-KO 2-WT ສະວັດທີ 1 ເມືອງກົດ	ອາກເກາກງົງ ຂາກ ລົງ (RNA ດາວວັດທີ 1 ເມືອງກົດ)	- 0.01	Least sig
NNTNN	ດັກກ່າວທີ່ 1-101ກົງ modify ສົ່ງຈ່າຍໃນ			



ໂຄງການວິຈัยໂນເດລຮບບົນເວສາກເຮີຍບຸກົມາ ສຳຮັບເຍວະນ
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

Knock OUT ຂີ່ໂຄໂກ: 101 Gene ທີ່ໄດ້ເກີດ m^bA ຕົ້ນໄວ ຂີ່ໂຄໂກ KO 1-101 m^bA → clean, remove m^bA out of RNA ໃບຕົ້ນກົງ ສໍາnoise ເກີດໃນ ເຊັ່ນຕົ້ນກົງກົງ

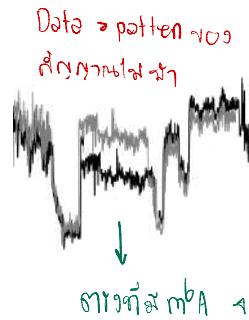
WT ຊະກິດ ຮ່າວເສັ່ວ ກາຊອນນິກ / 101 Gene ຕົ້ນໄວໂຄໂກ

ເລື່ອຕົ້ນກົງໃຫ້ມີກົມາເຊີ່ງເກົ່າ ອາກສື່ໄວ້ດັກກ່າວທີ່ມີກົງເກົ່າກົງເກົ່າກົງ → sig. m^bA.

Research Objectives

input
RNA
GGACT
GGm6ACT

Sequence
Nanopore
Sequencing



review Reseach
Signal-level modification detection methods



MINES
Super

detect

- m6A only
- Training data required.

Tombo

detect

- All modification types. → ไม่ต้องใช้ตัวอย่าง
- No training data required.

Goal
XPORE



Locate modified positions

Quantify fraction of modified reads -- modification rate

→ ร้อยละของ RNA ที่มีการเปลี่ยนแปลง

→ คำนวณความเร็วของการเปลี่ยนแปลง RNA ที่ต่างกันในตัวอย่างต่อตัวอย่าง $\rightarrow \text{sample A} / \text{sample B}$
← ตัวอย่างต่อตัวอย่าง $\rightarrow \text{sample A} / \text{sample B}$

← สำหรับการวิจัย





CODE COMBAT

Google

The Asia Foundation

Let's Code Thailand

DMAP
DIGITAL MAKERS ASIA PACIFIC

THAI
PROGRAMMER

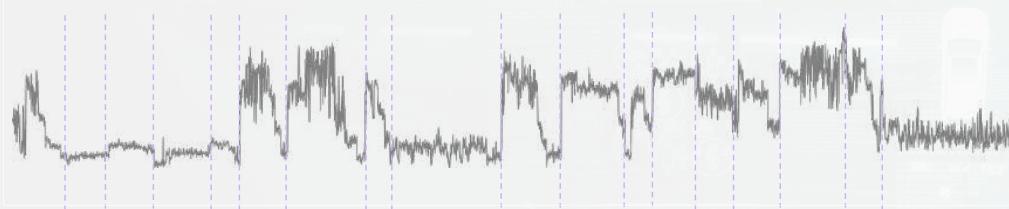


E-SAN THAILAND
CODING & AI ACADEMY

โครงการวิจัยไมโครสโคปนีโอศึกษาเรียนรู้กับบูรณาการ CODING & AI สำหรับเยาวชน
Model of Learning Ecosystem Platform integrate with Coding & AI for Youth

2. Data Collection and Preparation

G T A C T C G G A C T A C C C G C



- Nanopore Raw Signal Data
- Sequencing Data
- Genome Browser
- Nanopore Data Pipeline

ໂຄງການເກົ່າ
nanopore sequencer

FAST5

- Raw signal - Sequencing output
↓ ທີ່ສະໜັບ ແລ້ວໄດ້ ຄົນໃຈແຜນເອົາຂຶ້ນ ສະເພິ່ງ
- Intensity level (pA) ແລ້ວມານີ້ > < Current intensity >
- HDF5 format (binary), storing large and complex data

```
HDF5 "GISPC936_20181120_FAK27249_MN18749_sequencing_run_SHO_20112018_Empty
GROUP "/" {
    ATTRIBUTE "file_version" {
        DATATYPE H5T_IEEE_F64LE
        DATASPACE SCALAR
        DATA {
            {0}: 0.6
        }
    }
    GROUP "PreviousReadInfo" {
        ATTRIBUTE "previous_read_id" {
            DATATYPE H5T_STRING {
                STRSIZE 38;
                STRPAD H5T_STR_NULLTERM;
                CSET H5T_CSET_ASCII;
                CTYPE H5T_C_S1;
            }
            DATASPACE SCALAR
            DATA {
                {0}: "ac7312ce-d058-4382-a6c6-8471302869b9"
            }
        }
        ATTRIBUTE "previous_read_number" {
            DATATYPE H5T_STD_U32LE
            DATASPACE SCALAR
            DATA {
                {0}: 976
            }
        }
    }
    GROUP "Raw" {
        GROUP "Reads" {
            GROUP "Read_984" {
                ATTRIBUTE "duration" {
                    DATATYPE H5T_STD_U32LE

```

```
DATA {
    {0}: 12639754
}
} FAST5 ອັນໄດ້ຮັບໄດ້ຮັບ 1 read ອົບສາມ RNA
DATASET "Signal" {
    DATATYPE H5T_STD_I16LE
    DATASPACE SIMPLE { ( 76256 ) / ( H5S_UNLIMITED ) }
    DATA {
        {0}: 595, 492, 497, 502, 500, 499, 514, 495, 515, 512, 531,
        {11}: 529, 515, 483, 497, 529, 518, 521, 524, 525, 523, 514,
        {22}: 519, 517, 512, 520, 522, 519, 521, 517, 535, 514, 585,
        {33}: 537, 527, 512, 521, 528, 530, 530, 529, 529, 521,
        {44}: 527, 515, 537, 522, 512, 485, 480, 481, 478, 465, 467,
        {55}: 472, 476, 463, 469, 476, 454, 458, 446, 468, 471, 470,
        {66}: 466, 468, 467, 466, 468, 458, 466, 467, 464, 465, 467,
        {77}: 465, 459, 476, 478, 477, 460, 486, 470, 485, 486, 468,
        {88}: 475, 470, 472, 472, 468, 456, 457, 452, 448, 440, 440,
        {99}: 473, 470, 454, 442, 448, 449, 455, 461, 443, 455, 448,
        {110}: 449, 444, 462, 456, 461, 459, 467, 459, 461, 458, 472,
        {121}: 461, 463, 467, 456, 471, 468, 471, 475, 467, 466, 471,
        {132}: 477, 459, 473, 482, 466, 477, 470, 461, 464, 452, 454,
        {143}: 457, 468, 457, 466, 472, 474, 441, 456, 478, 467, 444,
        {154}: 442, 455, 451, 456, 470, 469, 473, 479, 478, 468, 472,
        {165}: 462, 466, 458, 435, 436, 464, 467, 455, 462, 463, 471,
        {176}: 455, 459, 446, 460, 442, 453, 465, 465, 488, 465, 478,
        {187}: 467, 475, 483, 512, 582, 539, 521, 586, 521, 523, 516,
        {198}: 518, 511, 514, 518, 538, 516, 528, 503, 503, 510, 524,
        {209}: 529, 526, 513, 504, 469, 476, 472, 470, 468, 476, 476,
        {220}: 476, 471, 459, 457, 432, 443, 472, 466, 477, 467, 471,
        {231}: 470, 474, 449, 468, 456, 457, 460, 459, 459, 456, 469,
        {242}: 457, 469, 475, 468, 465, 465, 463, 446, 455, 458, 461,
        {253}: 456, 448, 446, 462, 444, 444, 464, 462, 469, 479, 471, 502.
```

ໂຄງການວິຊາໄລຍະ ເລະ ໂຮງງານນີ້ຖືກປຸງປັດການ CODING & AI ສໍາເຫຼັບເຍວ່ານ
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH



FAST 5 և ՀԱՅԻ ՀԱՐՑԻ ՏԵՇԱՐՄԱՆ ՀԱՅՈՒԹ ՀԱՅՈՒԹ ՀԱՅՈՒԹ ՀԱՅՈՒԹ ՀԱՅՈՒԹ

FASTQ

អតស់អង្គភាព និង ពាក្យារា
call base រួច ដាច់ កំណត់ពាក្យារា RNA
ទីម៉ាទ្រីម៉ាទ្រី

ເກີນຄໍາລົມ Base ທີ່ຢູ່ໃນ cell ແກ່ວ້າ ພາກ Base cell

→ ແລະກົງ ກ່າຍການໄມ້ນຳ ທາເລັນ ກໍາຕໍ່ມ່ວນດີ ດີເຊີ້ນ ML , software ທີ່ຈະສ່ວນອາຫຼືກາແກ່

- Basecalled sequence
 - Text format:
 - Name/ID, starting with "@"
 - Sequence កីឡា Basecalled មានការរំលែកនៅក្នុង pattern នៃបច្ចេកទេស ដើម្បីការអនុវត្ត
 - Optional info, starting with "+"
 - Quality of the sequence, encoding the probability error

RNA 1 Read

โครงการวิจัยโมเดลระบบปั๊วีศึกษาเรียนรู้ที่บูรณาการ CODING & AI สำหรับเยาวชน
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH



CODE COMBAT

Google

The Asia Foundation

Let's Code Thailand

DMAP
DIGITAL MAKERS
ASIA PACIFICTHAI
PROGRAMMERE-SAN THAILAND
CODING & AI ACADEMY

↗ BC- ML, DL စာရွေးဆောင် အကိုက် ပုံမှန်စံ ပေါ်

FASTA

စာချုပ် Base ၏ စံအားလုံးများ၊ DNA တဲ့ နှင့် ပုံမှန်

- Reference sequence
- Text format:
 - Sequence ID, starting with ">", optionally followed by other attributes
 - Sequence

နိုင်ကြော် ရှာ အားလုံး cDNA ပေါ် ၍ နှင့် sequence ပေးပါ

Gene 0.1

```
>ENST00000480901.1 cdna chromosome:GRCh38:17:47828308:47831525:-1 gene:ENSG00000159111.12 gene_biotype:protein_coding transcript_biotype:retained_intron gene_symbol:MRPL10 desc  
ription:mitochondrial ribosomal protein L10 [Source:HGNC Symbol;Acc:HGNC:14055]  
TTCTTCGGTGGAGATGGCTGGGCCGTGGCGGGATGCTCGCGAGGGGTCTCTGCC  
AGCCGGGTAAGGAGTGCCCCAGGTCTCACGCCGTGCTTGGGGCGCTCTAGTCC  
ATCTGCCCTCTACTACTGATTCTCCCATAACTCTGACCCAGCTAGATCCGTGC  
CTCCTTACCCCGTCCAGTTCTGTGACTCGACTGGCCGGCTGCCTACCCCTCCAGACTGT  
CCCTATGGCTCCAAGGCTGTTACCGGCCACCGCTGTGATGCACTTCAAGCGCAGAA  
GCTGATGGCTGTGACTGAATAATATCCCCCGAACACAGGCCATCCACCCATCATGCC  
ATCTCTCCAGCCCCAACAGGAGGTAAGGAGGAATTGGGTACATGTCAGTTGGGT  
GGGATGGTGGATTAAAGTAACTTGTCTGCCATAGTGAAGTAGGACACTCAGCCATT  
GTCATGCACGTCAATTTCAGTTGACTGCCTGATCCAGATAATTAAAGTGAATCC  
CACTTGATTCTGTATTGGCTTGGCTCTGGATTGG
```

โครงการวิจัยโมเดลระบบป้องกันการเรียนรู้ที่บูรณาการ CODING & AI สำหรับเยาวชน
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH



BAM / SAM ໄທ່ນັ້ນຄັ້ງສົ່ງຂອງກາ

- Alignment results (FASTQ aligned with FASTA) ເນື້ອງວ່າ ຄະ. Read ອອກໄພ ແຫ່ງເກີນ Gene ຕັ້ງໄ

result Ali call • BAM – Binary / SAM – Text

โครงการวิจัยโมเดลระบบปั๊วีค์การเรียนรู้ที่บูรณาการ CODING & AI สำหรับเยาวชน
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

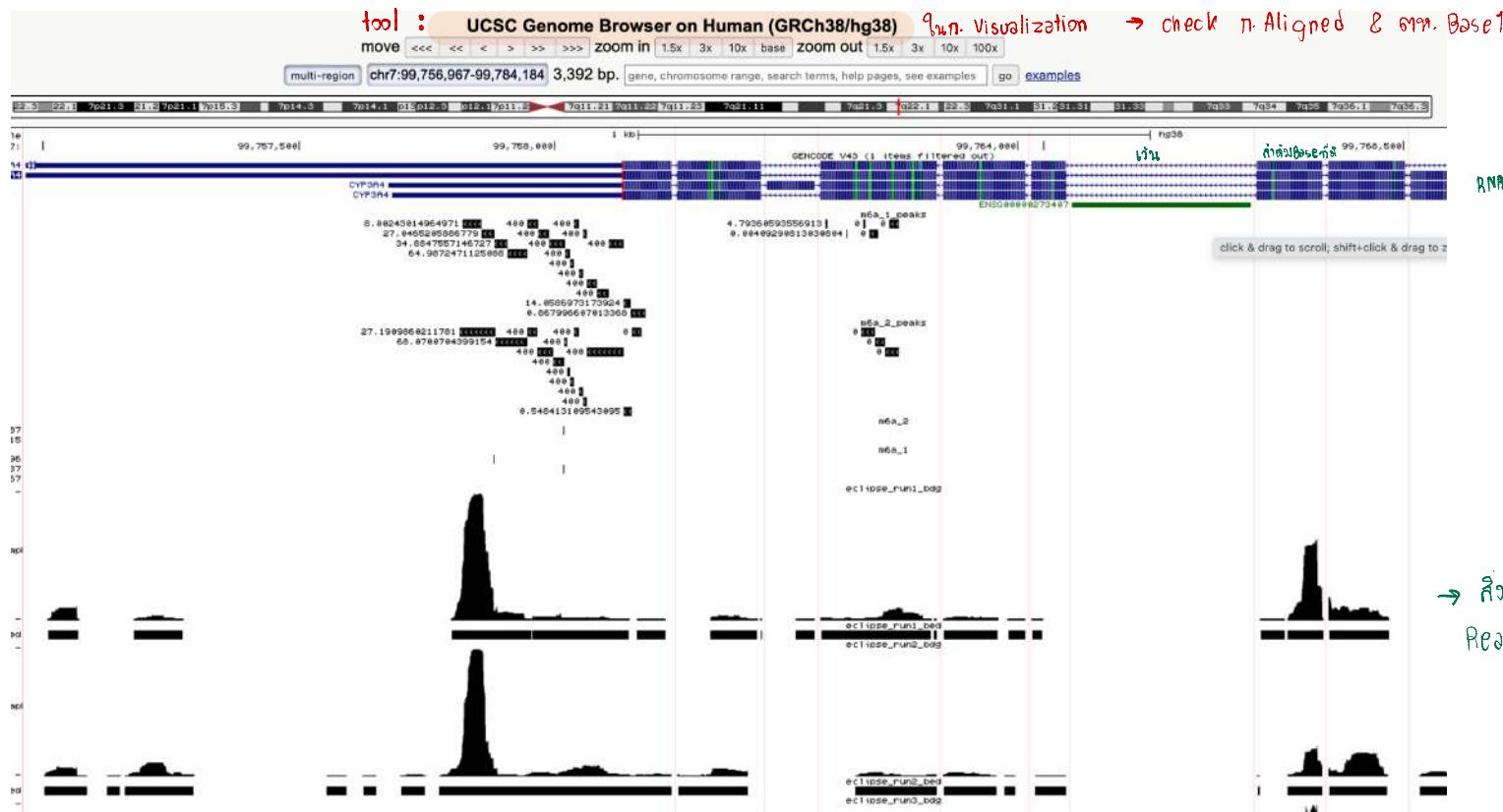


CODE COMBAT

Google

The Asia Foundation

Let's Code Thailand

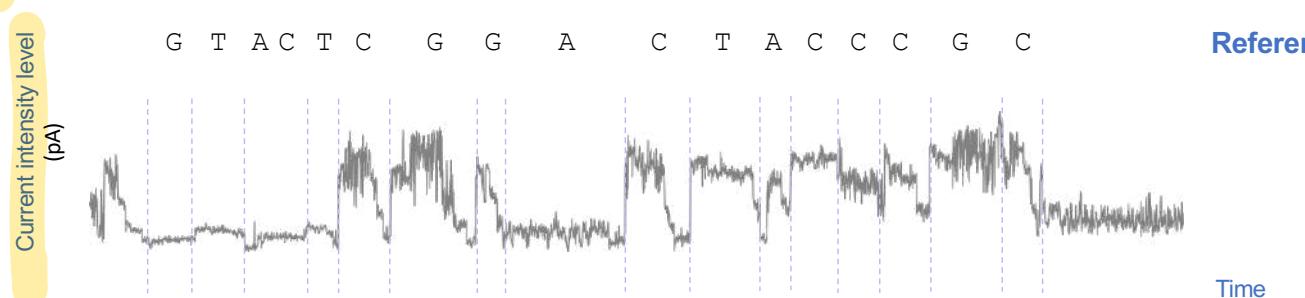
DMAP
DIGITAL MAKERS
ASIA PACIFICTHAI
PROGRAMMERE-SAN THAILAND
CODING & AI ACADEMY

โครงการวิจัยໂນໂລຣະບບນິເວຄາຮັບຮູ້ກໍບຸຽນຂາຍ
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

↑ Data from nanopore to raw signal នៅទី មេការ Data analysis
↑ តាមអង្គភាព

Nanopore pre-processing pipeline for signal-level data analysis

Event Align Output



Reference

1st

software package nf-core/
nanoseq

Direct RNA sequencing

Software : Oxford Nanopore

ផ្លូវការការណ៍ដោយ

ពិនិត្យសកម្មភាពនៃវត្ថុ → ការតើលិខិត

Basecalling

Guppy (ML)

FAST5 → FASTQ

Sequence aligning

Minimap2

FASTQ + FASTA → BAM

Signal event aligning

Nanopolish

FAST5 + FASTA + BAM → Event Align Output
↑ ផ្សាយតាម Ali → ការរាយ Read នៃបីប្រភេទ Gene or RNA នៃបីប្រភេទ Ref

កម្រការវិចិត្តិមិត្តភកសារការិយេប្បុរិបាកសម្រាប់ CODING & AI សំខាន់យោង
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH



<https://xpore.readthedocs.io/en/latest/>

Data preparation from raw reads

1. After obtaining fast5 files, the first step is to basecall them. Below is an example script to run Guppy basecaller. You can find more detail about basecalling at [Oxford nanopore Technologies](#):

```
guppy_basecaller -i </PATH/T0/FAST5> -s </PATH/T0/FASTQ> --flowcell <FLOWCELL_ID> --kit <KI>
```

2. Align to transcriptome:

```
minimap2 -ax map-ont -uf -t 3 --secondary=no <MMI> <PATH/T0/FASTQ.GZ> > <PATH/T0/SAM> 2>> <  
samtools view -Sb <PATH/T0/SAM> | samtools sort -o <PATH/T0/BAM> - &>> <PATH/T0/BAM_LOG>  
samtools index <PATH/T0/BAM> &>> <PATH/T0/BAM_INDEX_LOG>
```

3. Resquiggle using [nanopolish eventalign](#):

```
nanopolish index -d <PATH/T0/FAST5_DIR> <PATH/T0/FASTQ_FILE>  
nanopolish eventalign --reads <PATH/T0/FASTQ_FILE> \  
--bam <PATH/T0/BAM_FILE> \  
--genome <PATH/T0/FASTA_FILE> \  
--signal-index \  
--scale-events \  
--summary <PATH/T0/summary.txt> \  
--threads 32 > <PATH/T0/eventalign.txt>
```

