

**E-SAN THAILAND CODING & AI ACADEMY**

โครงการวิจัยโมเดลระบบสนับสนุนการเรียนรู้ก้าวหน้าทาง 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**

**ผศ. ดร.นฤมล ประภานวณิช**  
โครงการย่อยที่ 6

The background features a futuristic, glowing blue interface with various data visualizations, charts, and digital elements.

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

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

**E-SAN THAILAND CODING & AI ACADEMY**

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

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

Ploy N. Pratanwanich <sup>1,2,3</sup>, Fei Yao<sup>1,1</sup>, Ying Chen<sup>1,1</sup>, Casslynn W. Q. Koh<sup>1,1</sup>, Yuk Kei Wan<sup>1,1</sup>, Christopher Hendra<sup>1,4</sup>, Polly Poon<sup>1</sup>, Yeek Teck Goh<sup>1</sup>, Phoebe M. L. Yap<sup>1</sup>, Jing Yuan Chooi<sup>5</sup>, Wee Joo Chng<sup>5,6,7</sup>, Sarah B. Ng<sup>1</sup>, Alexandre Thierry<sup>8</sup>, W. S. Sho Goh<sup>1,9</sup> and Jonathan Göke <sup>1,10</sup>

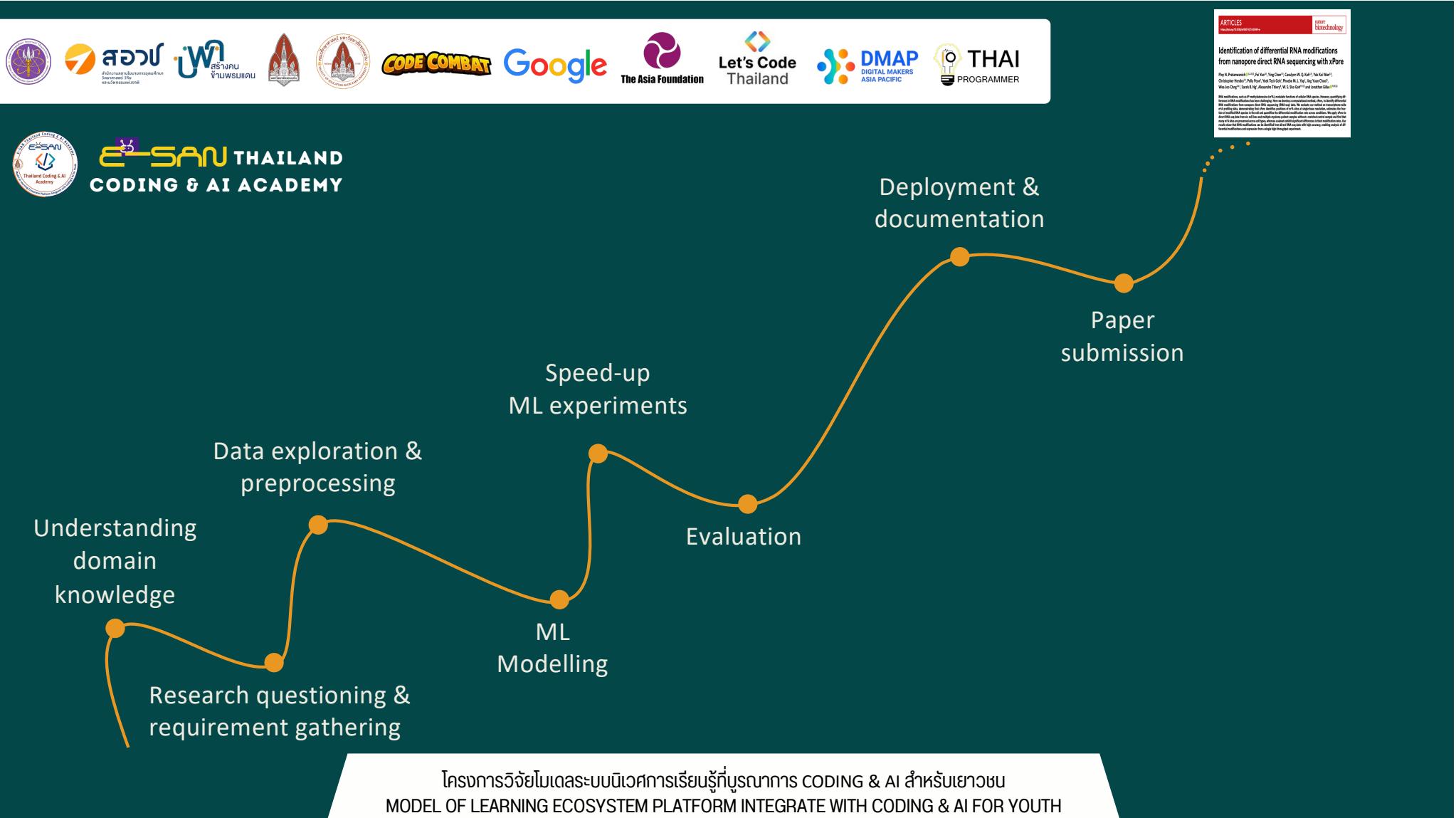
RNA modifications, such as *N*<sup>1</sup>-methyladenosine (m<sup>1</sup>A), modulate functions of cellular RNA species. However, quantifying differences in RNA modifications has been challenging. Here we develop a computational method, xPore, to identify differential RNA modifications from nanopore direct RNA sequencing (RNA-seq) data. We evaluate our method on transcriptome-wide m<sup>1</sup>A profiling data, demonstrating that xPore identifies positions of m<sup>1</sup>A 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 m<sup>1</sup>A 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

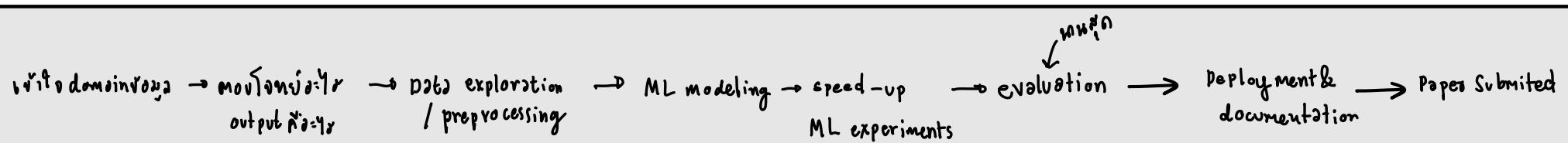
The background features a complex collage of abstract infographics, data visualizations, and digital interface elements, including a 3D model of a brain, a DNA sequence, a waveform graph, and various scientific and technological icons.





# Xpore

App nAI  $\xrightarrow{\text{dev}}$  អេកម្មោងនៃសារព័ត៌មាន  
សំណើលើស RNA modification



Data from nanocore sequencing  
↓  
output files  
use ML to predict RNA modification

**E-SAN THAILAND CODING & AI ACADEMY**

โครงการวิจัยไมโครสโคปนี้เป็นการเรียนรู้ที่บูรณาการ Coding & AI สำหรับเยาวชน

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

**Bioinformatician**

**Data Scientist**

**Biologist**

**xpore**

Oct 9, 2021

Installation

PyPI installation (recommended)

Installation from our GitHub repository

git clone https://github.com/GeekLab/xpore.git  
cd xpore  
python setup.py install

Previous Next

**xpore 2.1**

xpore is a python package for Nanopore data analysis of differential RNA modifications.

Contributors 7

Languages Python 100.0%

added logo

Installation from our GitHub repository

Quikstart - Detection of differential RNA modifications  
Output table description  
Configuration file  
Data preparation from raw reads  
Data  
Command line arguments  
Citing xPore  
Getting Help

update version to 2.1

Gene

Number of sites

HEK293T WT m6ACE-Seq

Modification rate

Density

Genomic coordinate

True positive rate

False positive rate

AUC = 0.86

Estimated modification rates

No. of artificial modifications

m6ACE-Seq DRACH

Accuracy

Top positions

HEK293T WT HEK293T KO

Rep1 Rep2 Rep3 Rep1 Rep2 Rep3

HEK293T WT HEK293T KO

Rep1 Rep2 Rep3 Rep1 Rep2 Rep3

GGACC AGACA

Normalized coverage

Unsmoothed Modified

Normalized coverage

Unsmoothed Modified



CODE COMBAT

Google



DMAP  
DIGITAL MAKERS  
ASIA PACIFIC



E-SAN THAILAND  
CODING & AI ACADEMY

โครงการวิจัยโมเดลระบบป้องกันการเรียนรู้ที่บูรณาการ 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

## 2 Data Collection and Preparation

## 3 Bayesian [Multi-Sample] Gaussian Mixture Modelling

## 4 Evaluation

## 5 Visualization and Presentation

## 6 Future Work

## 1. Problem Statement

- Nanopore Sequencing
- RNA Modification
- Inputs & Outputs
- Research Objectives



CODE COMBAT

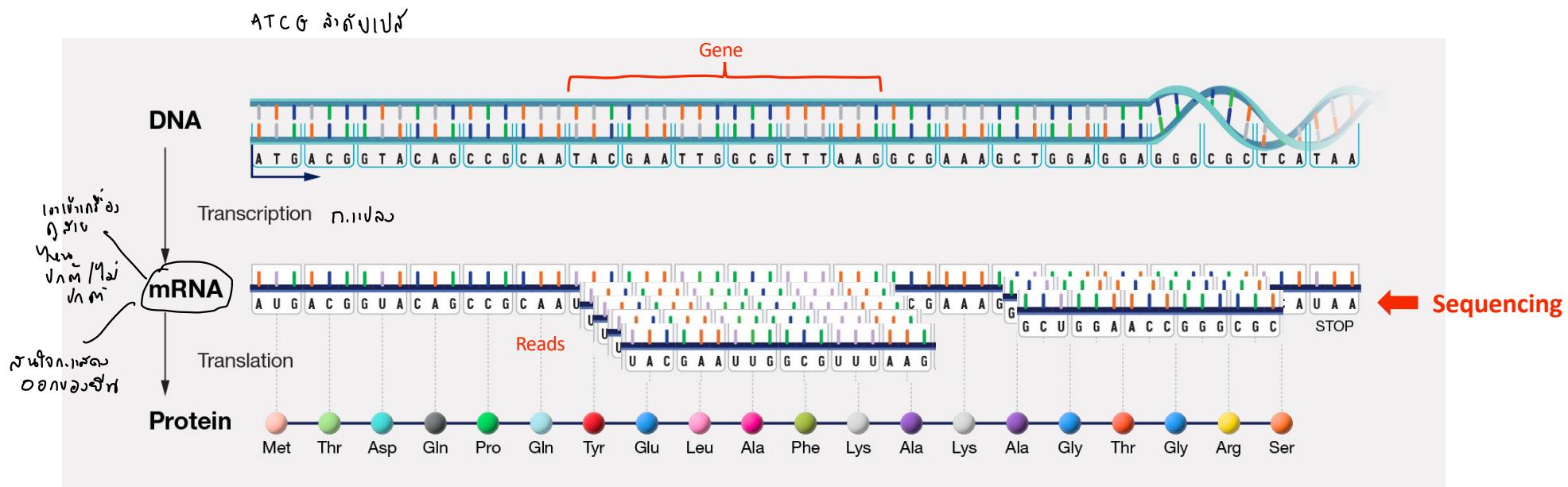
Google



E-SAN THAILAND  
CODING & AI ACADEMY

คณ. bioinformatic

# Central Dogma

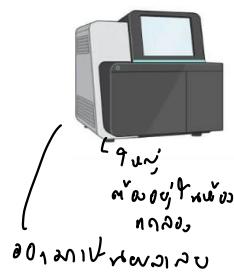
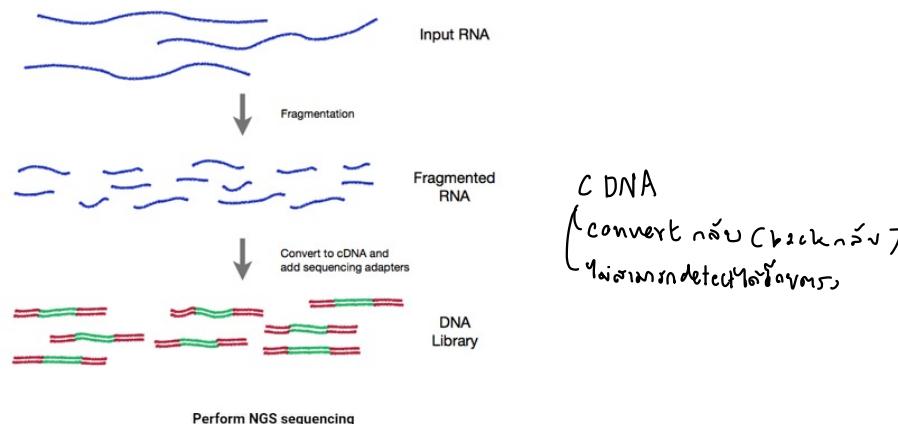


Source: <https://www.genome.gov/genetics-glossary/Central-Dogma>

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

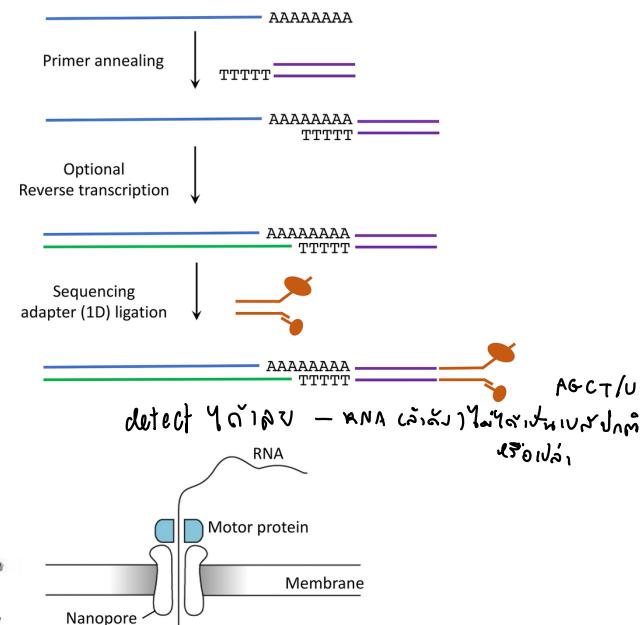
# RNA Sequencing

Then



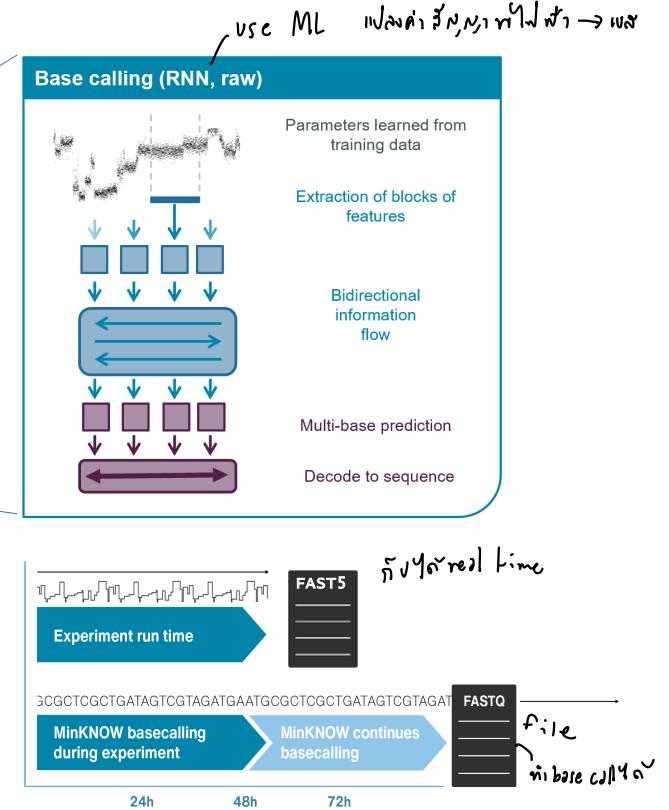
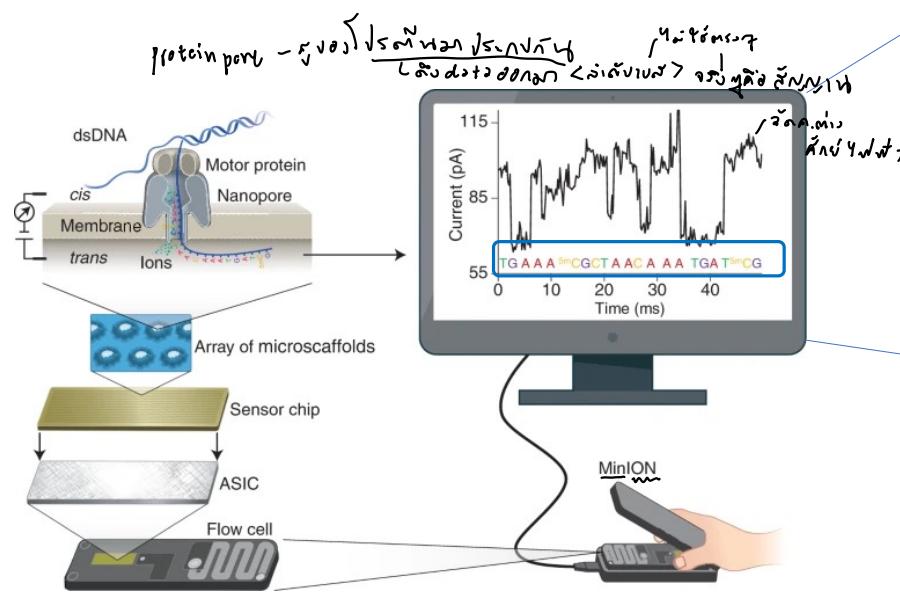
## Direct RNA Sequencing

Now



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

# RNA Sequencing



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

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



CODE COMBAT

Google

The Asia Foundation

Let's Code Thailand

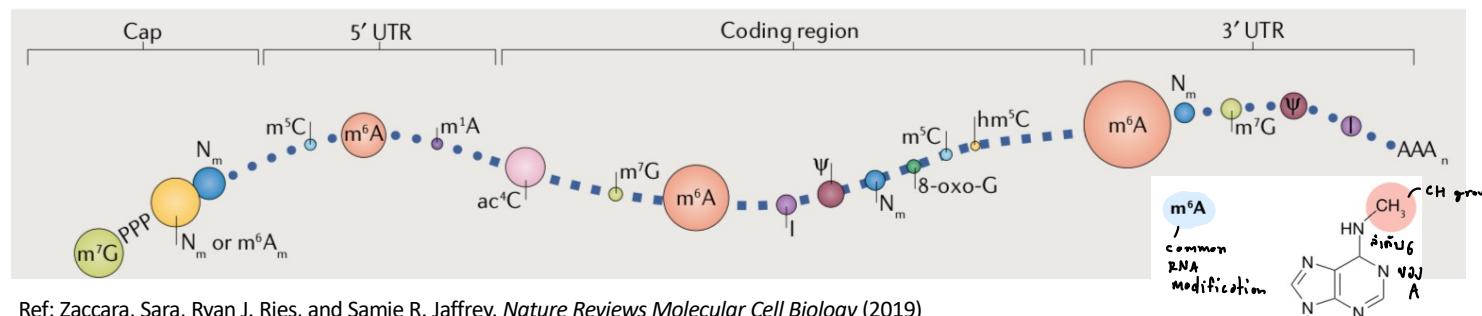
DMAP  
DIGITAL MAKERS ASIA PACIFIC

THAI  
PROGRAMMER



E-SAN THAILAND  
CODING & AI ACADEMY

# RNA modifications / ອົດໄວປັນ



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

Splicing

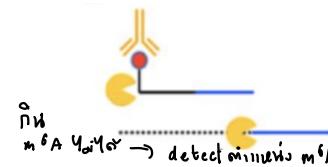
RNA Instability

Translation

Disease-related  
ຮອດທີ່ກິຈນຳ

## Single-base-resolution CLIP-based detection methods

Use antibodies to induce truncations or mutations at m6A sites during reverse transcription.



m6ACE-Seq

Ref: Koh, Casslynn WQ, Yeek Teck Goh, and WS Sho Goh. *Nature Communications* 10.1 (2019)

ໂຄງການວັດຍໂນເດລະບົມືເວສາກເຮັດນັ້ກູບນາກ CODYING & AI ສໍາເຮັບເຍາວຊາ  
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH



CODE COMBAT

Google



E-SAN THAILAND  
CODING & AI ACADEMY

# Output Table

อ่านค่าอย่างไรที่ position

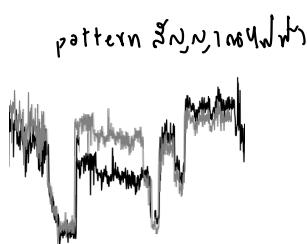
Genomic positions	5-mer	Modification rates				Differential modification rates	P-value
		sample 1 KO	sample 2 WT (วัตถุ)	W <sub>WT</sub> - W <sub>KO</sub>	หัวใจที่มีการเปลี่ยนแปลงที่สุด?		
NNANN		 m'A 001	 100%	0.81	Most sig		
...	...	...	...	...	...	...	...
NNCNN		...	...	...	...	...	...
...	...	...	...	...	...	...	...
NNGNN		 m'A 001	 100%	0.42	...		
...	...	...	...	...	...	...	...
NNTNN		 m'A 001	 100%	- 0.01	Least sig		
Transcriptome-wide	...	...	...	...	...	...	...

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

# Research Objectives

( RNA  
GGACT  
GGm6ACT

Nanopore  
Sequencing



## XPORE

សរុបតាមលក្ខណៈសម្រាប់ស្នើសុំ  
Locate modified positions

Quantify fraction of modified reads -- modification rate \*

ភាគមិនមែនសម្រាប់ស្នើសុំ All reads → modified rate

## Signal-level modification detection methods



MINES

} detect m6A but need training data

- m6A
- Training data required.

Supervised

Tombo

សម្រាប់ស្នើសុំសម្រាប់ស្នើសុំ

nanopore  
អេឡិចក្រចន = អេឡិចក្រចន

- All modification types.
- No training data required.

Unsupervised



## 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

# Data Collection & Preparation

## ○ Data collection

- FAST5 บีกิปົນ Raw signal - sequencing output    FAST5 1 file → 1 Read RNA → ດາວໂຫຼວງ gene

Intensity level (pA)

HDF5 format (binary), storing large and complex data

ສັນຕິພາບ (ມົດລົງ)

- FASTQ Basecalled sequence

txt format

- Name/ID, starting with "@"
- Sequence
- Optional info, starting with "+"
- Quality of the sequence, encoding the probability error

(ຍັກລົກປຶກ ຂຶກແກ່ລູນ ມາຈາກສັນນັກໜຸ້າ / ພອກໜຸດອານຸ. basecalled)

- BAM (binary) / SAM-Text Alignment results (FASTQ aligned w/ FASTA)

ເກີບລະຄອ. Alignment b/w FASTQ & FASTA 'ອຸປະນາມ read ໃນອົງການ ເພື່ອກັນ gene ມີຄູນ'

myo / vimentin

reference sequence

ຍົກລົກ ອຳນວຍມາປັບປຸງ (Reference Database)

txt format

- Sequence ID, starting with ">", optionally followed by other attributes
- Sequence

# FAST5

- Raw signal - Sequencing output
- Intensity level (pA)
- 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
                    DATASPACE SCALAR
                    DATA {
                        (0): 12639754
                    }
                }
                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, 510, 521, 524, 525, 523, 514,
                        (22): 519, 517, 512, 520, 522, 519, 521, 517, 535, 514, 505,
                        (33): 537, 527, 512, 521, 528, 523, 530, 530, 529, 529, 521,
                        (44): 527, 515, 537, 522, 512, 485, 480, 481, 478, 465, 467,
                        (55): 472, 476, 463, 469, 476, 454, 450, 446, 468, 471, 470,
                        (66): 466, 468, 467, 466, 466, 458, 466, 467, 464, 465, 467,
                        (77): 465, 459, 476, 470, 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, 470, 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, 502, 539, 521, 506, 521, 523, 516,
                        (198): 518, 511, 514, 518, 530, 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, 464, 462, 469, 479, 471, 502.
                    }
                }
            }
        }
    }
}

```

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



ສາທາລະນະ  
ລາວ



## FASTQ

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

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

# FASTA

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

```
>ENST0000480901.1 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]  
TTCTTCGGTGGAGATGGCTGCGGCCGTGCCGGGGATCCTGGAGGGGGTCTCCCTCCCC  
AGGCCGGTAAGGAGTGCCCCAGGTCTCACGCCGTGCTTGGGCCGCTCTAGTCCTC  
ATCTGCCCTCTACTACTGATTCTCCCATAAATCTCTGACCCCAGCTAGATCCTGGC  
CTCCTTACCCCGTCCAGTTCTTGACTCGACTGGCCGGCTGCCAACCTCCAGACTGT  
CCGCTATGCCCTCAAAGGCTGTTACCCGCCACCGCTGTGATGCACTTCAAGGCCAGAA  
GCTGATGGCTGTGACTGAATAATATCCCCCGAACCCAGCCATCACCCATCATGCCCTGCC  
ATCTCTCCAGCCCCCACAGGAGGTAAGGAGGAATTGGTACATGTCATTGGTGGT  
GGGATGGTGGATTAAGTAATCTTGCTCTGCCATAGTGAAGTAGGACACTCAGCCATT  
GTCATGCACGTCAATTTCAGTTGACTGCCTGATCCAGATTTAAAGATGAAATCCG  
CACTTGATTCTGTATTGGCTTGGCTCTGGATTGG
```

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

# BAM / SAM

จัดทำโดย

- Alignment results (FASTQ aligned with FASTA)
- BAM – Binary / SAM – Text

```
2fdda8fc-621a-415e-baaaf-1b088a91302d      0          ENST00000351111.6      13      13      7S11M1D13M1I6M1I5M1D6M1I18M1I14M1D2M1I16M2D19M1D1M1D31M2D12M3D9M1I17M4D7M1D11M2D27M3D36M
1D3M1D15M1D10M1I13M1I13M1I28M1D29M1D31M1D13M1I3M2D10M1I12M1D7M2D7M1I7M1I16M2I12M1D28M2D21M2D10M1D10M3D25M2D4M1D8M1I29M2I13M1D7M1D10M3D16M2I15M1D16M1I17M1D25M3D2M1I3M1I
13M1I2M3D8M2D15M2D7M3I14M4I5M1D51M1I10M1D8M1D13M2I12M1D21M6I1D29M1D14M1I5M3D29M2I20M1D10M1D14M2D38M1I18M3I2M1D15M1D11M1D2M1D15M1D12M1I6M2I21M1D4M3
D26M2D16M3D10M1D12M1I7M2D8M13S      *      0      0      CAACATCGCGGCCGTGGCGGGATGCTGTGGGAGGGTCACTCTGCCAGGGCCGGCGCTGCCTACCCCTCCAGACTGTCCGCATCGGCTCCAAGGCTGCTACGCCACCCTCGTGTGAT
GCCCATCAGCGGCAGAACGCTGATGGCTGTGACTGAATATCCCCCGACAGGCCATCTTCCCTCCAGCAGGGGGATAGGCTCATCGGCCATCGGCCATTTCGCCAGATAGCAGCAGTTCAGGACACAACCAGAATGATAAGCGTTGCCAGAAATGTGGCTTGAG
TGCAGAGGGACAAGTTTCTGCGCACCCAGCTTGCAGAAACACAAGATCTGATGAAGGTTCCCAACCCAGGCTCTGAGGATTCCAAGTACCAAATCTGCTGCCCTTTTGCCCCACTATCGCTGGTCAATGAAGGCCGGGCTCAAAGATGG
TCACGGATCTTAAAGACTGTGCCATCACCTCGCGCTGATGGTGGGGCATTGATGACACCATTCACTGAGCCGCTTTACTCTGCTGCCCTGGTCAGGAGCTGTAGGAGGCCCTCACCTGCCCACAGCCAGACCCACTCCCTGCTCACGACCCAG
CCCCACCCTGACCACCCCTGTTGGACCACTACAGAGAGCAACGCGCGAGAAGGATTGTCATGCGCCAATGGTCAGATCTGACACTGGTCCGGACTTGCAGCAGCCCTGGCTGCCAGCCCTCGCATAAAATACATTGTTATTGGCTGTCTCCCTGGCAG
CTCGTGGAAAGAACATTGTCGAGGAGTGTTCGACTTGGTCACTAGAGGTAATGATATTGTCGTTAGGTAATGGCAGCTGGGAGATGCAAGGATTCCAGATGTCAGTCAGCAGCCAACTCTCGCTTAGTTCCCTATTGGGACGTGATATGGGAGCAAAGACTTCATTA
CCCCAGGTCCAAGGGCAGAGAGAGAGATGGGCTCTGAAAAGATAGGGCTCTGGCTCTGGCTGGCCAGACGCTGGCCTCCCTGGCCACAAACCTATCCCATCGCAGCTGAGTGGAGATGGCACACCTGGTGGAGC
CTGTTGTCGAGGCTTGTGTTGAGTGTCTGCTCATGTTAGTTAGTCCCTACTGCTGACCACTGTACCCATATCACAGAGAAGGAGCAGAGAAATTAAAGTGGCTGCTCCACAAAGGTATGCACTGAGTTAGTGAAGTGGCAGACTACGGGACTTGAACCAAAGCTCTGCTTGAAGTGG
GTCCTGAATTTTCACTAGAGCTCACCACTTCAGGTTACCCAGAAGTGGGCCATCCAGGTGTCGTTAGTCCACCCCTGAGGTGTCGAAAGTTGGGAGCATTTGTTATAAAATGAATTTAAAAAAC %$$$$$%&(+),+),-(-.+/.0
))&..,%&.0/&$&,&(*(/. )%'5-,-+,)&((-(*))2+-,-+-,354688.+'*1)(*0-,0,,(')$',+**->(3--))+(-.23(8+&*0%&'((+,-,-2++,.+/++,,2231.:42($&'(+/4.++')-)*)),,07110-%&(+.+
4783-++&(*8*(+++.0.1*,,,)(.'/.**1(+-((&8+)*,+-+,*,**).&,+)*2-0.1.,)&()7,-,-,+)+1+)&%&%&(*+**)*0+.*''),*)**,-2,++,&&,(3/-,-/-,-/1)+,+,++-.(197/.+++)3)*-+*
-0)*+2-',(,1)*0)+,+/-4,),*3(42,+1.++-0-22'+,)-02172(*&(%&*&))+,2***$&+,,,+0)+/-*,5,*+1.*'(*,1,,%)+(-0,-.-/51,*$&-$-&*,.1/-(*/&+),.'')-(-(%'),(.0(( ))&**1.1.*+(*
,,)-0/-(*$&'(-8-(*9-909://420/,.)*.1+*8$'*%+..))(-002/))(&(/-/4.)*'/51-.,.611/0/.,.922('#'%$')*0,/61*,4,7-,1..2/-...*,-*-*84,-*)++,+1)5+23,))(((*1*1,+,+)
+/333,+++,))((,-*3+6(&((*,03:5-,-$(,.706-792,-*)))(25)/**+,+((8.~-2.240/132)(',&,0.~-..**)(8'.+.-2%)**+,%*%&*0(+)*0-0*-!('11')+52/0*'),,0,+*&(%$&*&*),,,,((
204.,+,*8&,(*261/3*,./**+,25-+*,,-)-8%,,22/13-03)(2.#(((-'-(),-,,'(+!)'/..(%$&*'+()&12(0.,-,-,!/8%''''&((()-,))341,)('(01)*-0.,*66.%)(*0-.+!&%$()''((1))&/,(
('(&(( )),&,$1+(*0.,.-*-&*,,'0-&,(./4+**-1.-*,%,-,,/0+*&((('%,./-2+&((/.9431/,.++'-%(*(*/*+*'/,-,))&($(*,'$%&(*+**&/2,)),,'(1))**-0)'*)&!'%.(/('(&)*&*+/-,('
'&12,./3.)(**-/,(*'$+,*.+*575645810.,-))'*'&(*(&%&4,+5./0-..#%)+**+'.(((*-,(')&*0.+10/)+)'%$'!(*+**)&$(%$&*%(%))))%(-0..01,+(+)*'/(%&-.,,)*),',//0-.4-
'+),'$(+)((.+)+'&())*66.,/-1+**)+*(^(&%,+5.*',/((+(-2/3+)+$ NM:i:212 ms:i:1794 AS:i:1794 nn:i:0 ts:A:+ tp:A:P cm:i:50 s1:i:490 s2:i:47d
e:f:0.1026
```

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

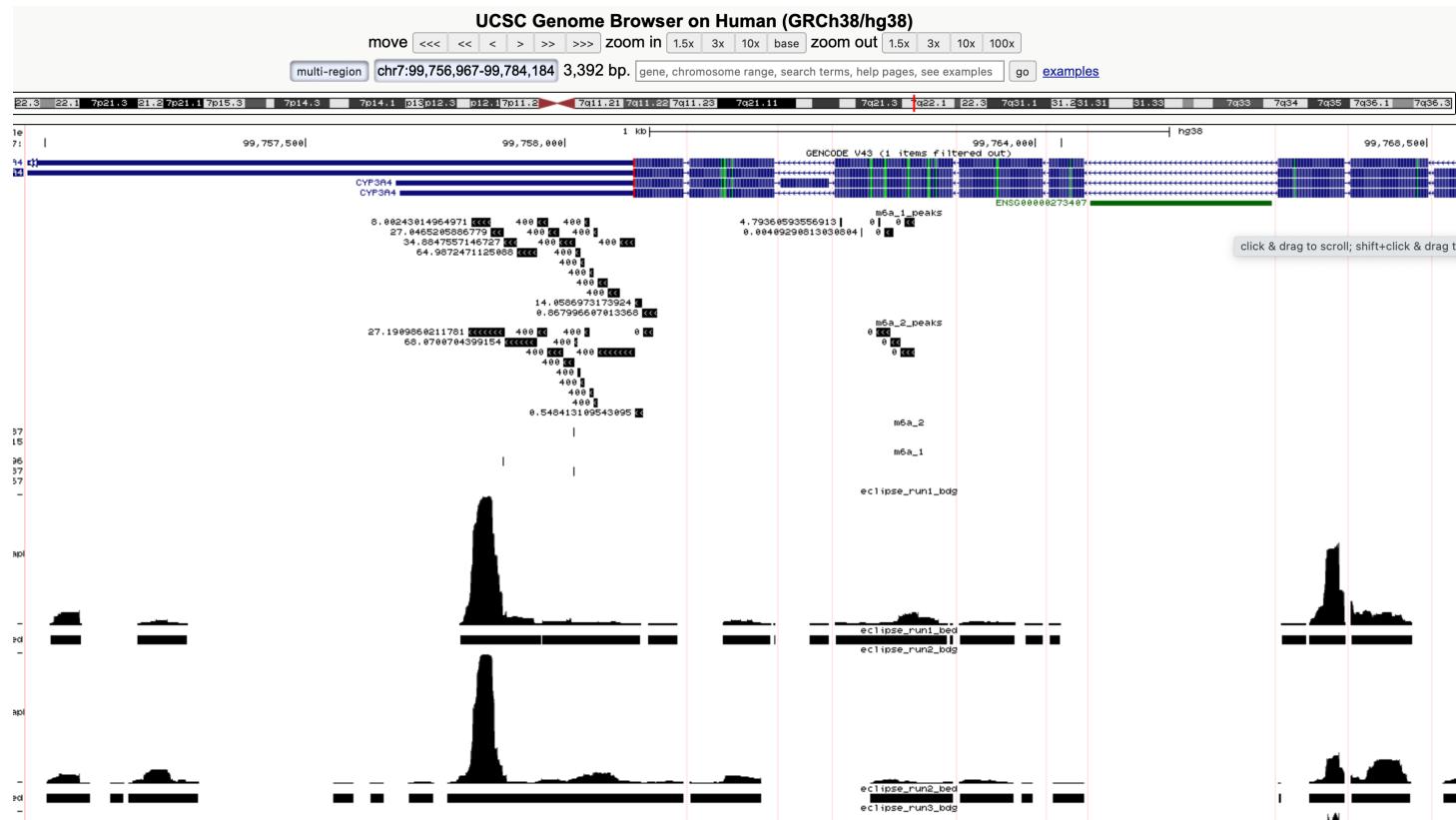


CODE COMBAT

Google



ESAN THAILAND  
CODING & AI ACADEMY



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

- tool : Genome Browser
- list of von segment ของตัวอ่าน som.read ↑ ↓
- ก็อชเชค aligned result ต้องมีอยู่
- can visualize RNA ได้ -> แสดง ไฟล์ผลลัพธ์ที่ดีที่สุด

# O preprocessing



**CODE COMBAT**

**Google**

The Asia Foundation

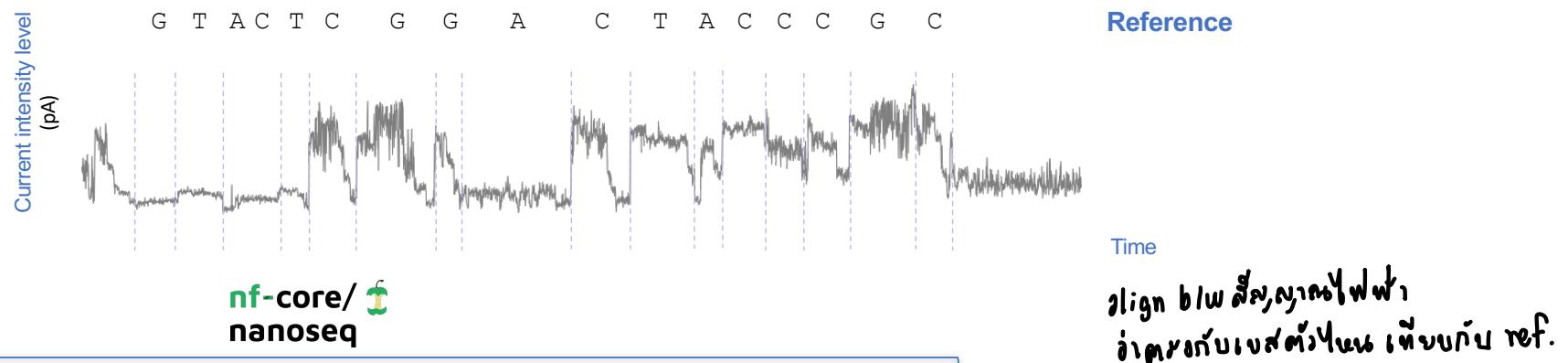
Let's Code Thailand

**DMAP**  
DIGITAL MAKERS ASIA PACIFIC

**THAI**  
PROGRAMMER

E-SANU THAILAND  
CODING & AI ACADEMY

## Nanopore pre-processing pipeline for signal-level data analysis



บุ่ง่าส์จะงานเป็นตัวไนน์ G-T-A-C

Direct RNA sequencing

Oxford Nanopore

บุ่ง่าส์จะงานเป็นตัวไนน์

ใน current Intensity level

Basecalling

Guppy

FAST5 → FASTQ

Sequence aligning

Minimap2

FASTQ + FASTA → BAM

Signal event aligning

Nanopolish

FAST5 + FASTA + BAM → Event Align Output

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

file ของพอก  
ที่ align แล้ว  
map core/RNA = ref.

<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](#):

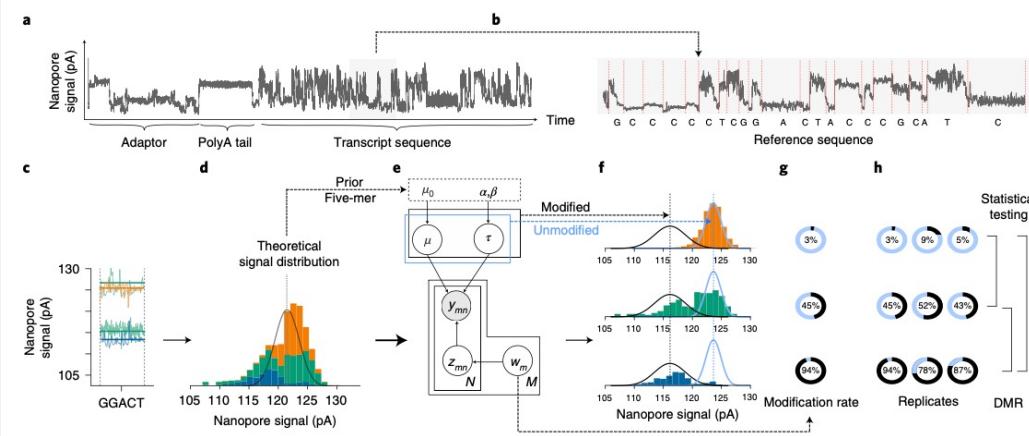
```
nanopolish index -d <PATH/T0/FAST5_DIR> <PATH/T0/FASTQ_FILE>  
nanopolish eventalign --reads <PATH/T0/FASTQ_FILE> \  
--bam <PATH/T0/BAM_FILE> \  
--genome <PATH/T0/FASTA_FILE> \  
--signal-index \  
--scale-events \  
--summary <PATH/T0/summary.txt> \  
--threads 32 > <PATH/T0/eventalign.txt>
```





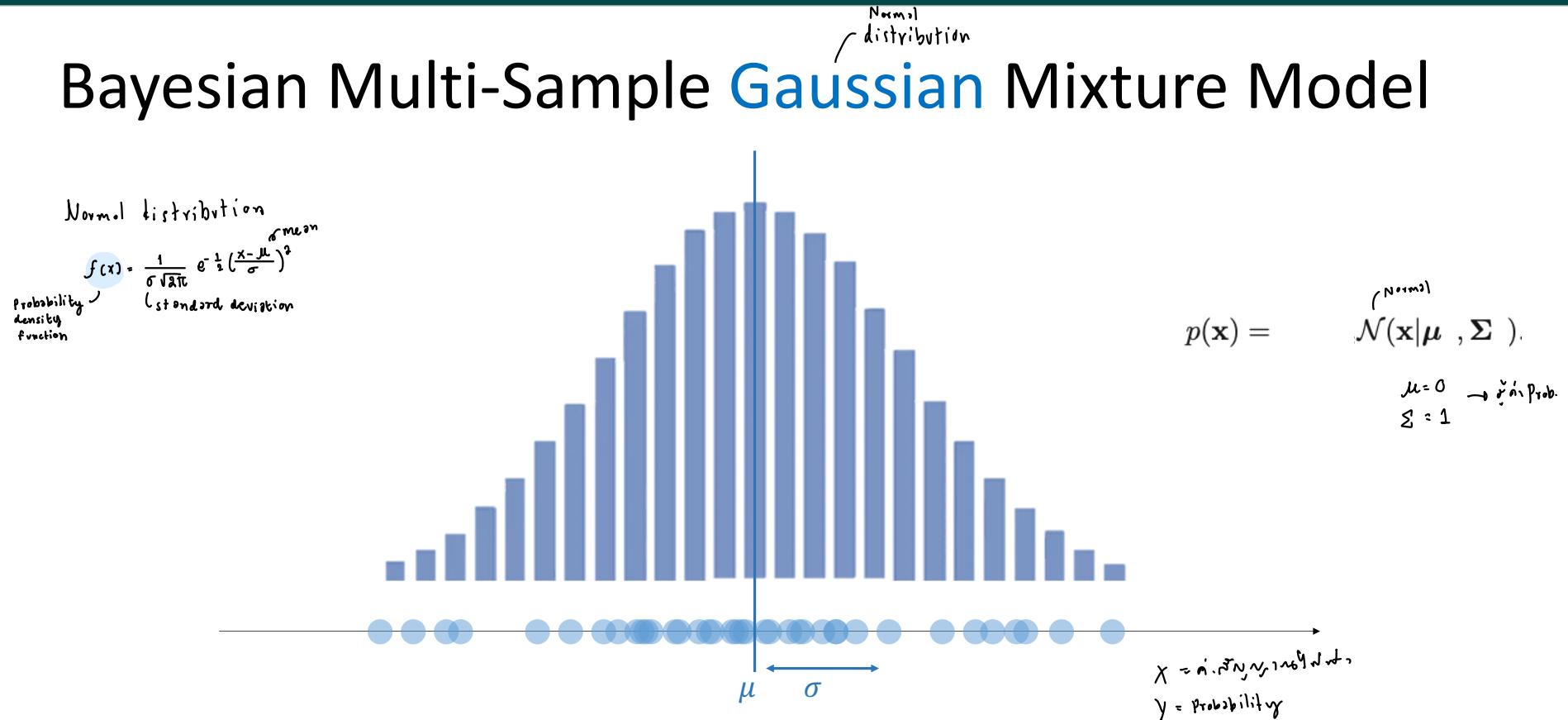
### 3. Bayesian [Multi-Sample] Gaussian Mixture Modelling

↑  
ឧបករណ៍មុខងារ



- [Bayesian] GMM
- Where did the idea come from?
- How Multi-Sample?
- Why Bayesian?
- Speed-Up ML Experiments

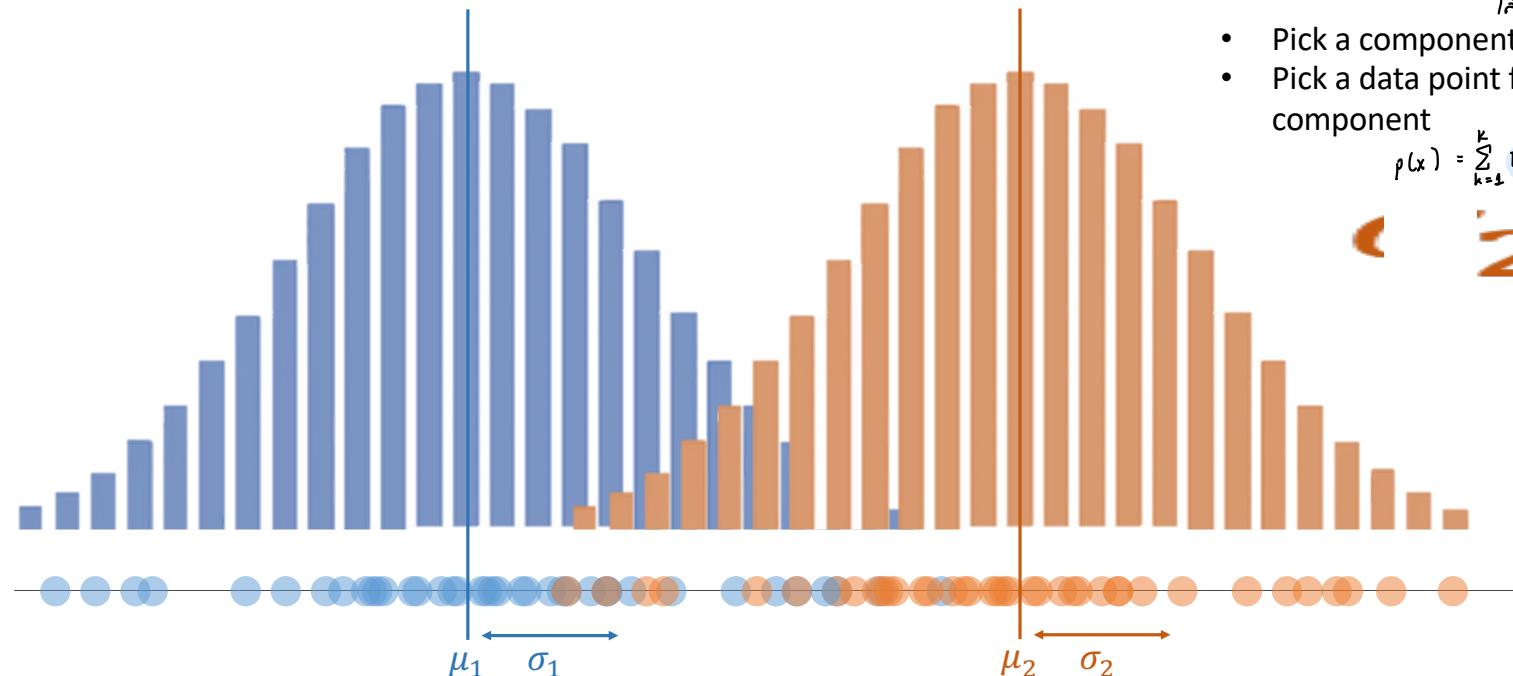
# Bayesian Multi-Sample Gaussian Mixture Model



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

ศูนย์ฯ → สถาบันฯ  
 ศูนย์ฯ

# Bayesian Multi-Sample Gaussian Mixture Model



- Pick a component at random
- Pick a data point from the chosen component

$$p(x) = \sum_{k=1}^K \pi_k N(x|\mu_k, \sigma_k^2)$$

↳ Prob. นิยมที่สุด  
เลือกมาใช้

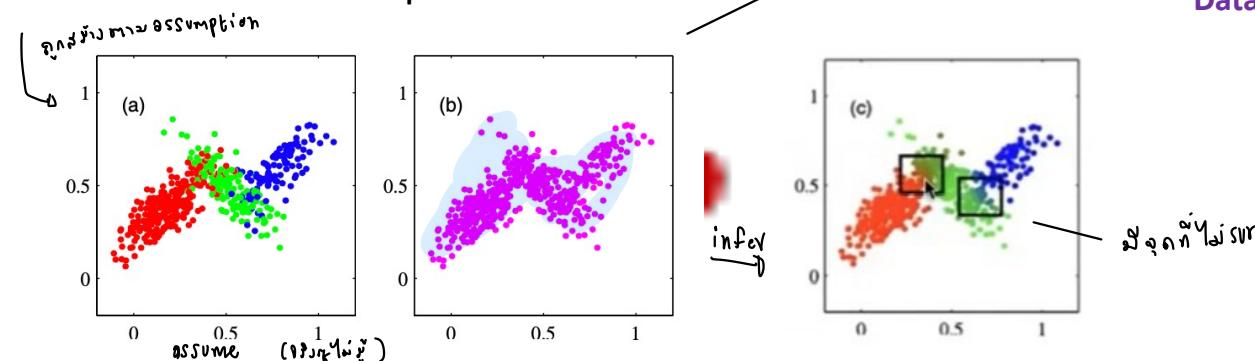
โครงการวิจัยโมเดลระบบโค้ดและการเรียนรู้ที่บูรณาการ CODING & AI สำหรับเยาวชน  
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

จุฬาลงกรณ์มหาวิทยาลัย

# What is GMM?

Assumption how data are generated as follows

- There are K components
- Each component is defined as a Gaussian distribution
- Pick a component at random
- Pick a data point from the chosen component



Source: Christopher M. Bishop, "Pattern Recognition and Machine Learning", 2006

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

fit விவரங்களை

இயல்வதை நம்முடைய விரிவாக என்று கீழே கொண்டு வரவேண்டும்.

$\theta$

Data



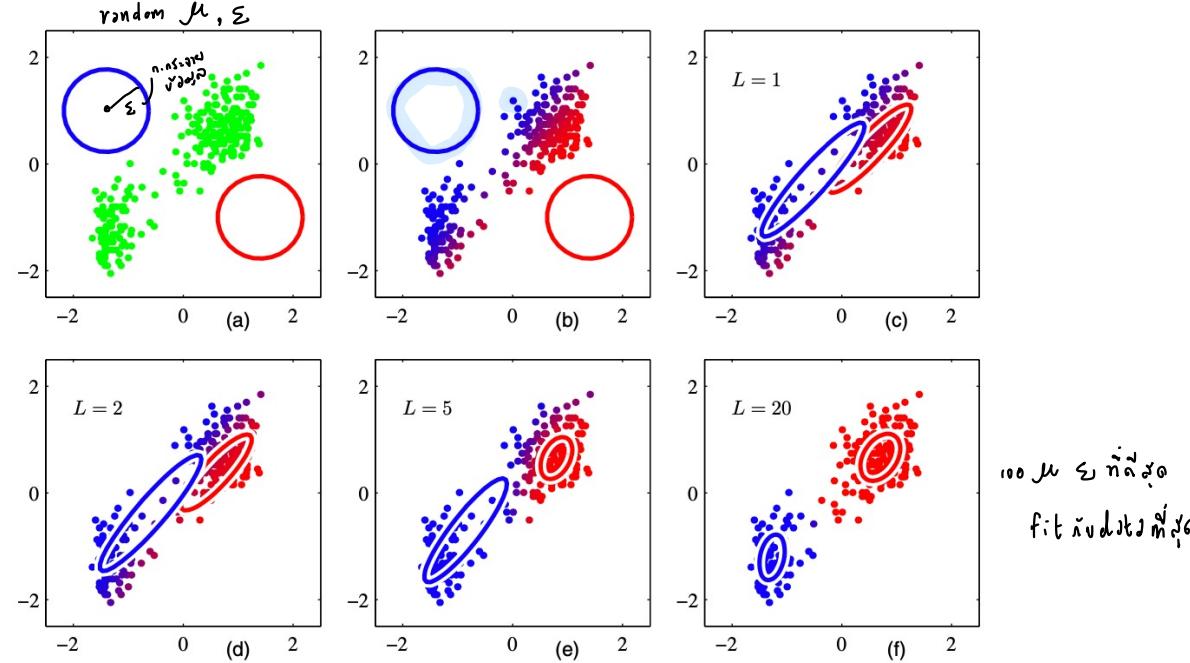
CODE COMBAT

Google

E-SAN THAILAND  
CODING & AI ACADEMY

# GMM Inference

Iterative  
algorithm



Source: Christopher M. Bishop, "Pattern Recognition and Machine Learning", 2006

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

Try Coding



นักศึกษา

## Generative AI

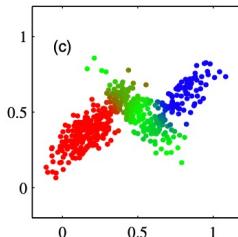
39064

สหัสฯ รุ่นที่ ๖๔



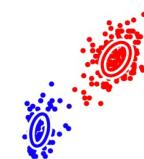
ML

GMM จีэмี  
as a Density Estimator  
จีэмีเป็นเครื่องมือ<sup>ที่</sup>คำนวณความถี่  
Model  
how data are generated

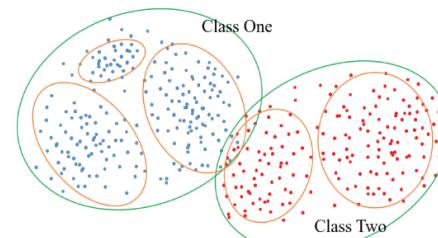


model ใช้กับ generative AI

## Clustering



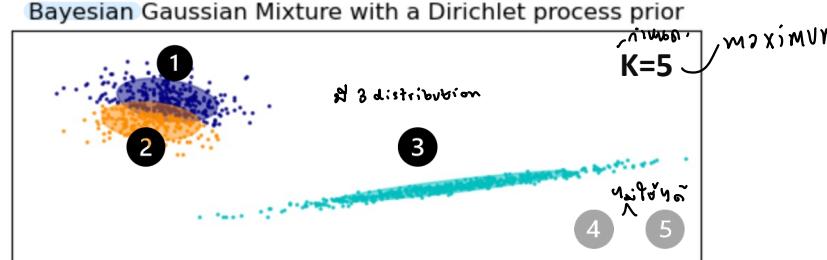
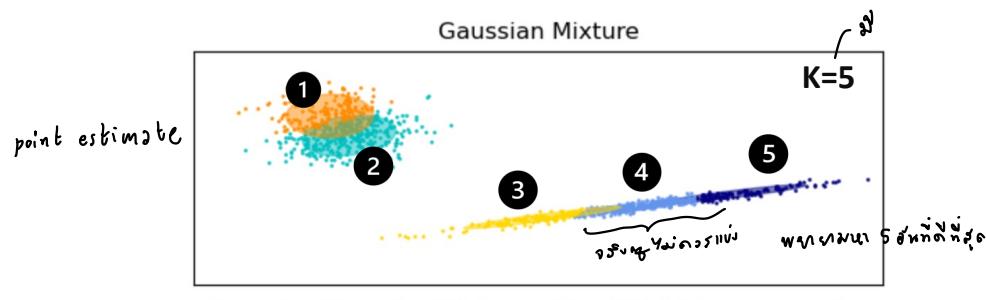
## (One-Class) Classification



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

# Bayesian Multi-Sample Gaussian Mixture Model

Learning algorithm for making inference on the **latent** variables



$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$\theta$

Data

Point estimate = Maximum Likelihood

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} P(\text{Data} | \theta)$$

maximum  
promoter van data 9 u's prob. 5 5 5

Posterior = Likelihood  $\times$  Prior

$$P(\theta | \text{Data}) = P(\text{Data} | \theta) \times P(\theta)$$

believe  
distribution

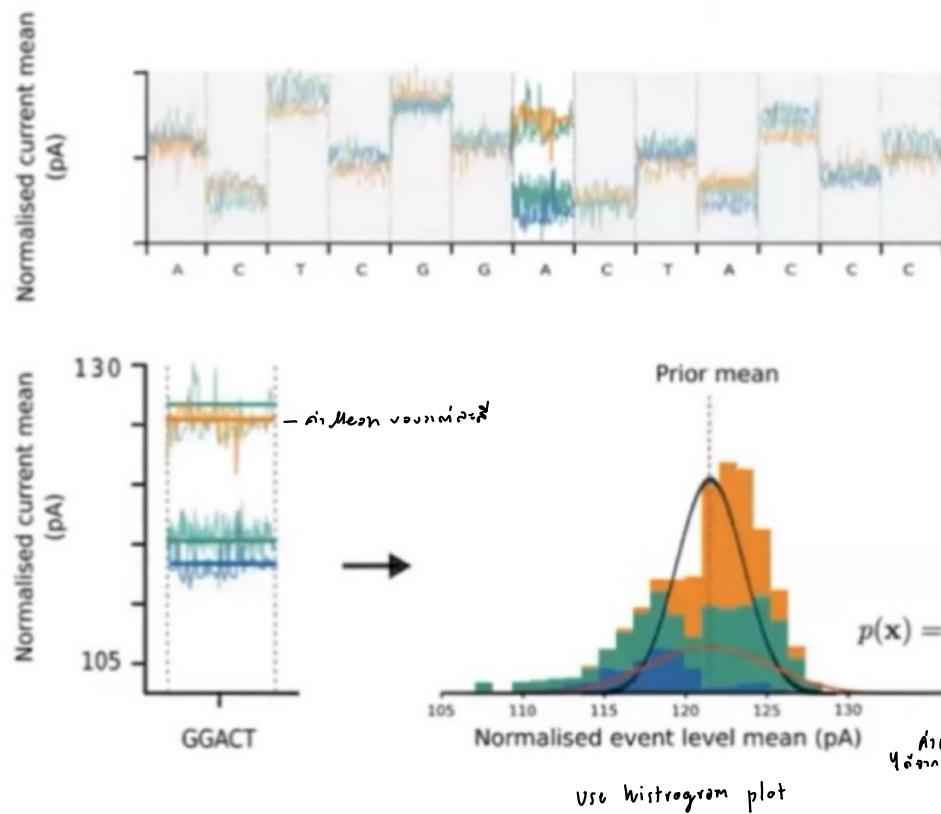
‘ทฤษฎีความน่าจะเป็น’  
**Frequentist vs Bayesian** *— but complex! —*

$$P(\text{Data} \mid \Theta)$$

$$P(\Theta \mid \text{Data}) = P(\text{Data} \mid \Theta) \times P(\Theta)$$

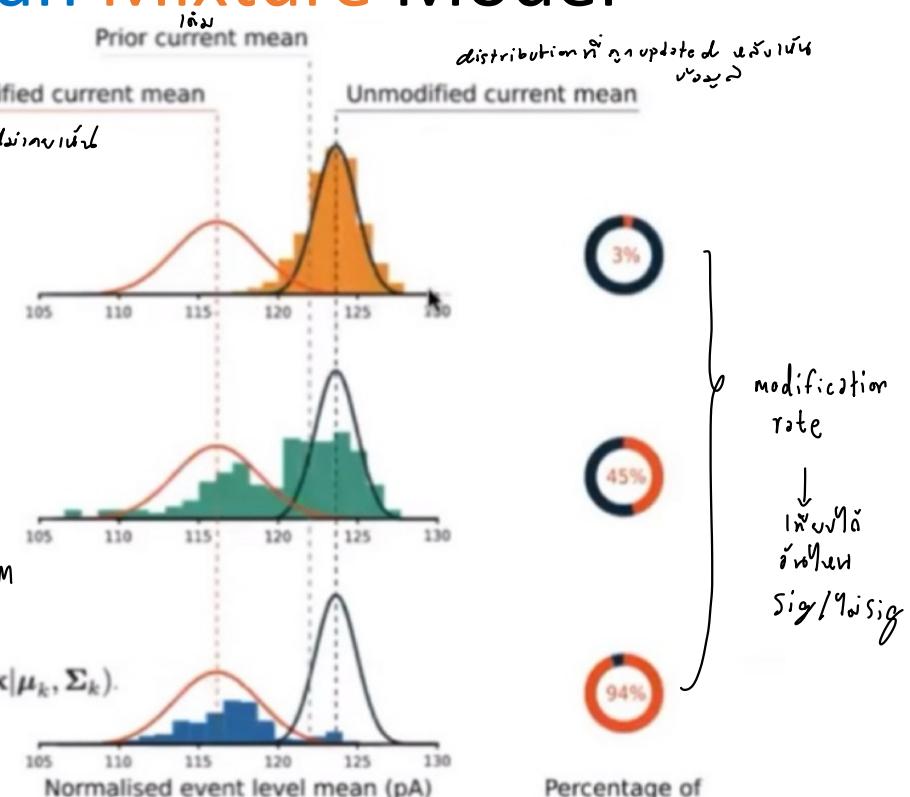
Aspect	Frequentist	Bayesian
Probability interpretation	Long term <u>frequency</u>	<u>Posterior</u>
Treatment of parameters	Fixed / <u>Point estimates</u>	Random / Probability <u>distributions</u>
Prior information	No	Yes
Sample size requirement	Larger	Smaller
Interpretation of results	Focused on the <u>observed</u> data	In the context of <u>prior beliefs</u> and their updates based on the <u>observed</u> data
Computational complexity	Simpler	More complex

# Bayesian Multi-Sample Gaussian Mixture Model



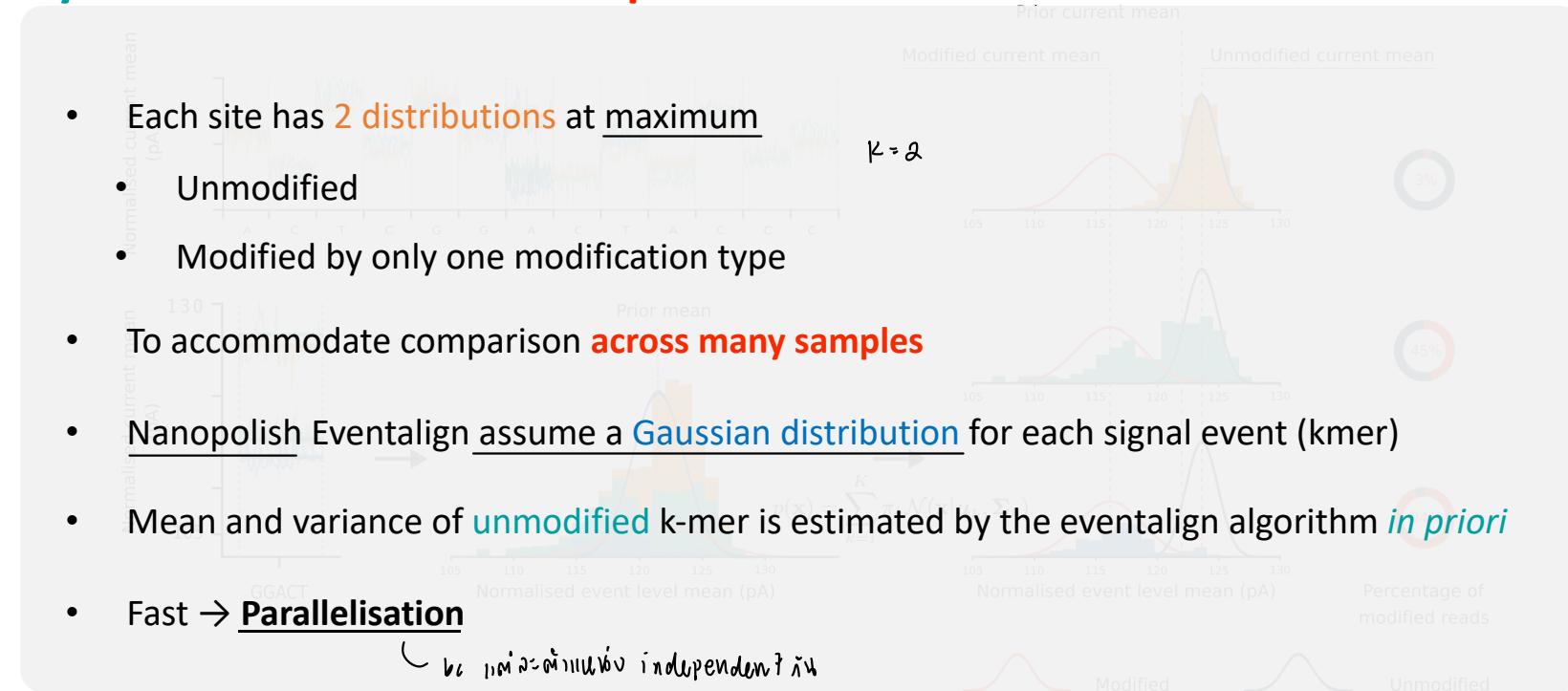
$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

fit GMM



# Bayesian Multi-Sample Gaussian Mixture Model

- Each site has **2 distributions** at maximum
  - Unmodified
  - Modified by only one modification type
- To accommodate comparison **across many samples**
- Nanopolish Eventalign assume a **Gaussian distribution** for each signal event (kmer)
- Mean and variance of **unmodified** k-mer is estimated by the eventalign algorithm *in priori*
- Fast → **Parallelisation**  
*↳ บริการนี้สามารถ independent ทั้งหมด*



# Output Table

Genomic positions	5-mer	Gaussian properties		Modification rates		Differential modification rates	
		Unmod	Mod	KO	WT	$\bar{W}_{WT} - \bar{W}_{KO}$	P-value
NNANN						0.81	Most sig
...	...	...	...	...	...	...	...
NNCNN				3%	94%		
...	...	...	...	...	...	...	...
NGNGN				3%	45%	0.42	
...	...	...	...	...	...	...	...
NNTNN						-0.01	Least sig
...	...	...	...	...	...	...	...

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

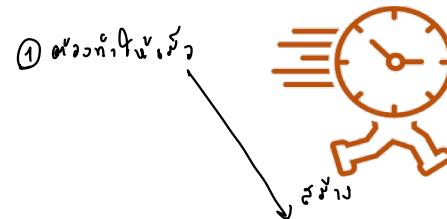
นักวิจัย อาจารย์ ผู้ช่วยศาสตราจารย์ ดร. วงศ์สุข วงศ์สุข

# Speed-Up ML Experiments

## Automated ML models



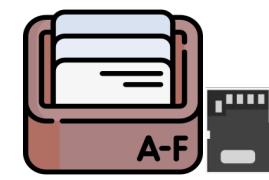
- Hyper-parameter settings
- Multiple datasets
- Different models / methods



- Config file    set up all parameter in file 1 sec
- Python packaging



- Parallelization



- File indexing

# Why config files?

- Automating tasks  
ກະຊວງສ່າງໄວ config ຍັງຍຸ
- Centralised configuration
- Documentation  
ອານຸຈິດຈົດລົງລວມ
- Portability

YAML, JSON, TOML, and INI are the popular and standardised formats of configuration files

```
xpore diffmod --config Hek293T_config.yml
```



CODE COMBAT

Google

ESAN THAILAND  
CODING & AI ACADEMY

# Configuration file

xpore / xpore / diffmod / configurator.py

Code Blame 78 lines (63 loc) · 2.68 KB

```
1 import yaml
2 import os
3 from collections import defaultdict
4
5 from ..utils import misc
6
7 def get_condition_run_name(condition_name, run_name):
8     return '-'.join([condition_name, run_name])
9
10 class Configurator(object):
11     def __init__(self, config_filepath):
12         self.filepath = os.path.abspath(config_filepath)
13         self.filename = self.filepath.split('/')[-1]
14         self.yaml = yaml.safe_load(open(self.filepath, 'r'))
15
16     def get_paths(self):
17         paths = {}
18
19         if 'prior' in self.yaml:
20             paths['model_kmer'] = os.path.abspath(self.yaml['prior'])
21         else:
22             paths['model_kmer'] = os.path.join(os.path.dirname(__file__), 'model_kmer.csv')
23
24         paths['out_dir'] = os.path.join(os.path.abspath(self.yaml['out']))
25         paths.update(misc.makedirs(paths['out_dir'], sub_dirs=['models']))
26         paths['model_filepath'] = os.path.join(paths['out_dir'], 'models', '%s.model')
27
28         return paths
```

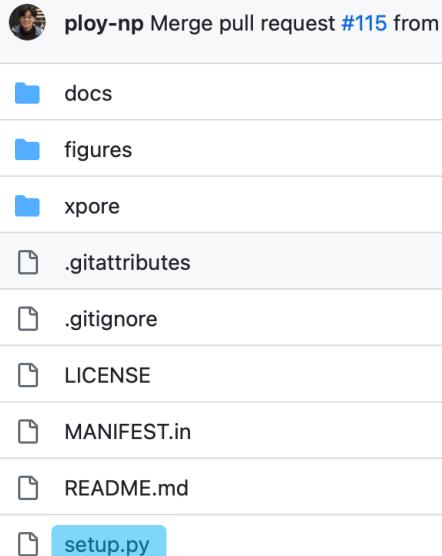
```
config = Configurator(config_filepath)
paths = config.get_paths()
data_info = config.get_data_info()
method = config.get_method()
criteria = config.get_criteria()
prior_params = config.get_priors()
```

data:  
<CONDITION\_NAME\_1>  
  <REP1>: <DIR\_PATH\_TO\_DATA\_JSON>  
  ...  
<CONDITION\_NAME\_2>  
  <REP1>: <DIR\_PATH\_TO\_DATA\_JSON>  
  ...  
out: <DIR\_PATH\_FOR\_OUTPUTS>  
  
criteria:  
  readcount\_min: <15>  
  readcount\_max: <1000>  
  
method:  
  # To speed up xpore-diffmod, you can use a statistical test (currently only t-test is implemented)  
  # to remove positions that are unlikely to be differentially modified. So, xpore-diffmod will  
  # those significant positions by the statistical test -- usually the P\_VALUE\_THRESHOLD very  
  # If you want xpore to test every genomic/transcriptomic position, please remove this pre-filtering:  
    method: t-test  
    threshold: <P\_VALUE\_THRESHOLD>  
  
  # Here are the parameters for Bayesian inference. The default values shown in <> are used,  
  max\_iters: <500>  
  stopping\_criteria: <0.00001>

hyperparameter

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

# Python Packaging



```

1     """Setup for the xpore package."""
2
3     from setuptools import setup,find_packages
4
5     __pkg_name__ = 'xpore'
6
7
8     with open('README.md') as f:
9         README = f.read()
10
11    setup(
12        author="Ploy N. Pratanwanich",
13        maintainer_email="narueemon.p@chula.ac.th",
14        name=__pkg_name__,
15        license="MIT",
16        description='xpore is a python package for Nanopore data analysis of differential RNA modifications.',
17        version='v2.1',
18        long_description=README,
19        long_description_content_type='text/markdown',
20        url='https://github.com/Goekelab/xpore',
21        packages=find_packages(),
22        include_package_data=True,
23        install_requires=[
24            'numpy>=1.18.0',
25            'pandas>=0.25.3',
26            'scipy>=1.4.1',
27            'PyYAML',
28            'h5py>=2.10.0',
29            'pyensembl>=1.8.5',
30            'ujson>=4.0.1'
31        ],
32        python_requires ">=3.8",
33        entry_points={'console_scripts': ["xpore={}.scripts.xpore:main".format(__pkg_name__)]},
34        classifiers=[
35            # Trove classifiers
36            # (https://pypi.python.org/pypi?%3Aaction=list_classifiers)
37            'Development Status :: 1 - Planning',
38            'License :: OSI Approved :: MIT License',
39            'Programming Language :: Python',
40            'Programming Language :: Python :: 3.8',
41            'Topic :: Software Development :: Libraries',
42            'Topic :: Scientific/Engineering :: Bio-Informatics',
43            'Intended Audience :: Science/Research',
44        ],
45    )

```

โครงสร้างแพลตฟอร์ม

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

# Parallelization / Multiprocessing

```
import multiprocessing
```

## When Data are Too Big to Fit in the Memory

	data.index	in index file 10	data.json
gene_id	start_idx	stop_idx	
ENGxx1	0	16856	{'ENGxx1': [123,110,...]}, {'ENGxx1':
ENGxx2	16857	29435	[123,110,...]}, {...}
...	...	...	





CODE COMBAT

Google

The Asia Foundation

Let's Code Thailand

DMAP  
DIGITAL MAKERS  
ASIA PACIFIC

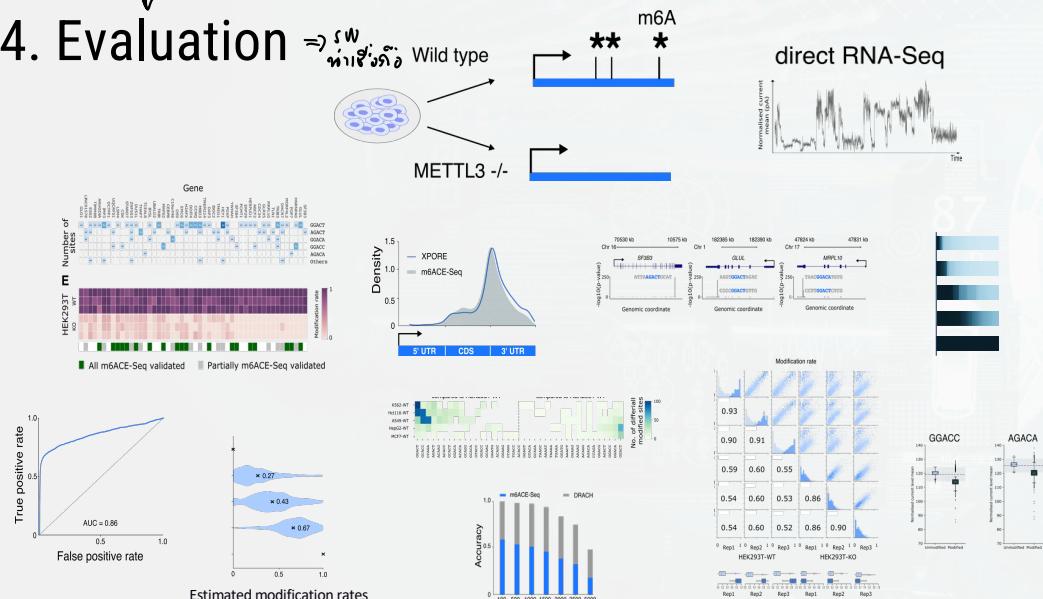
THAI  
PROGRAMMER

E-SAN THAILAND  
CODING & AI ACADEMY

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

คลิกสั่งงาน run model แล้ว

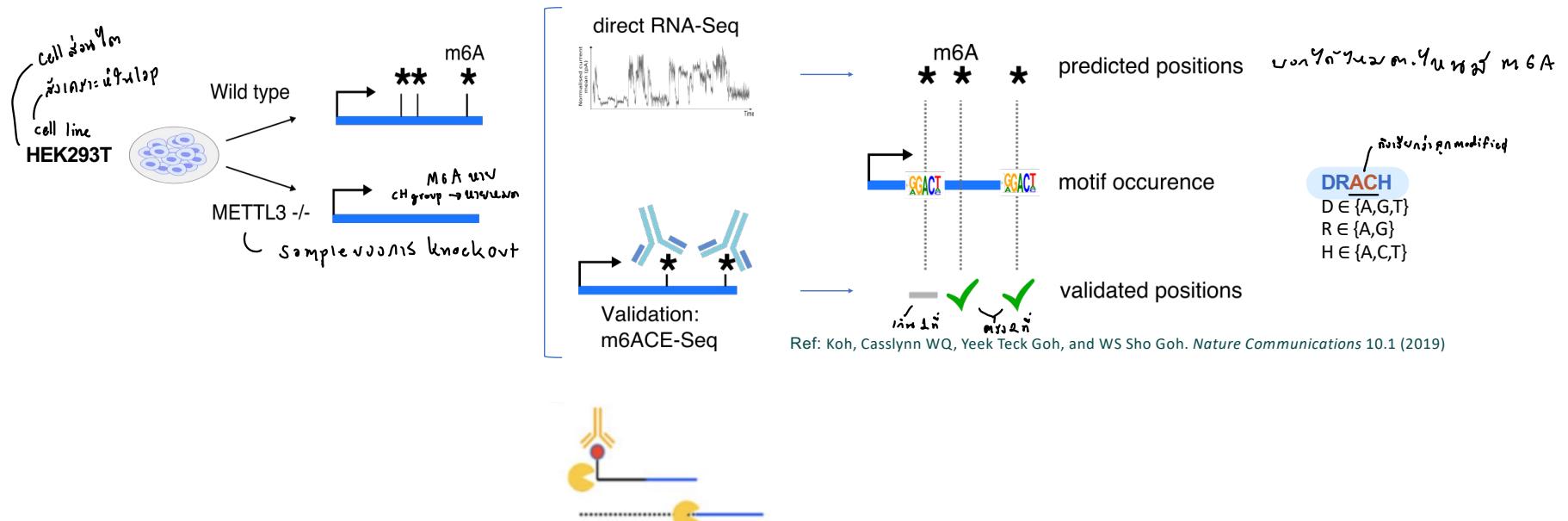
## 4. Evaluation



- Experiment setup
- Validation
- Applicability → 9 ชุดข้อมูลสำหรับการประยุกต์ใช้
- Discovery

ກອງສອນວ່າ ຕອນໂທກ່ຽວຂ້ອງມີການປະສົງປະນຸມ

# Experiment Setup

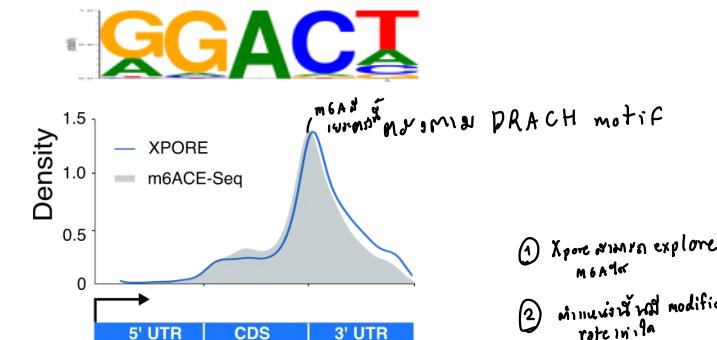
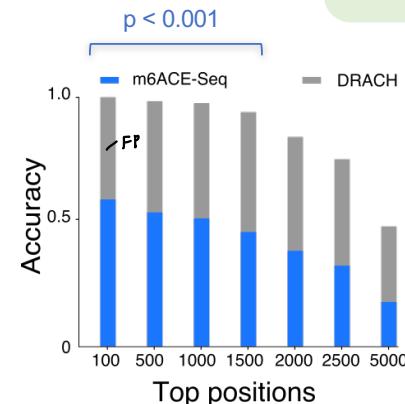
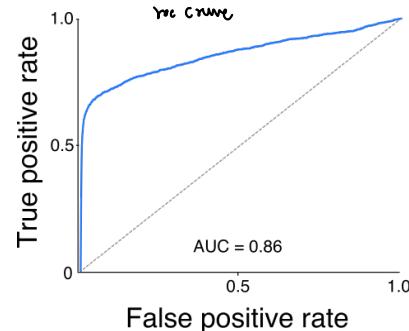


ໂຄງການວັດຍໂນເລຣະບນໄຟເວົຄກາຣເຮັຍນັ້ກືບູນກາຣ Coding & AI ສໍາຫັບເຍາວຊນ  
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

# Validation: m6A calling

In bioinformatics,

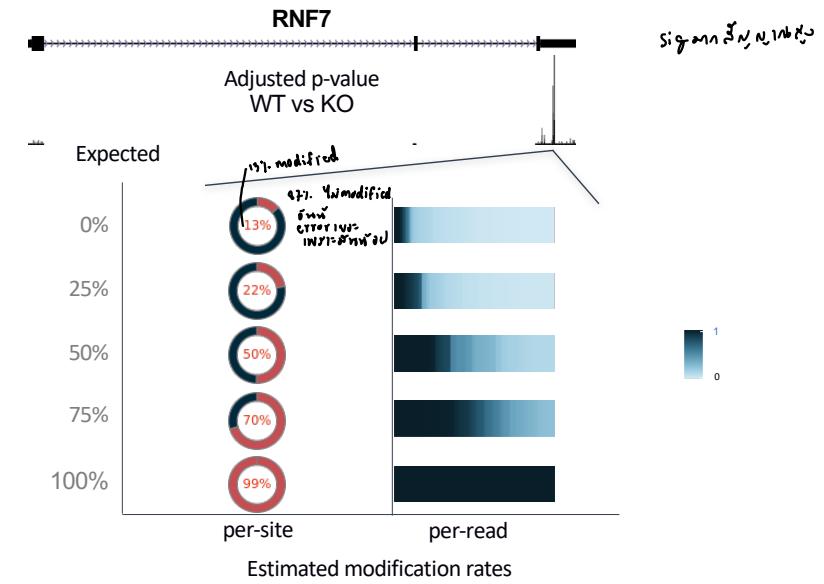
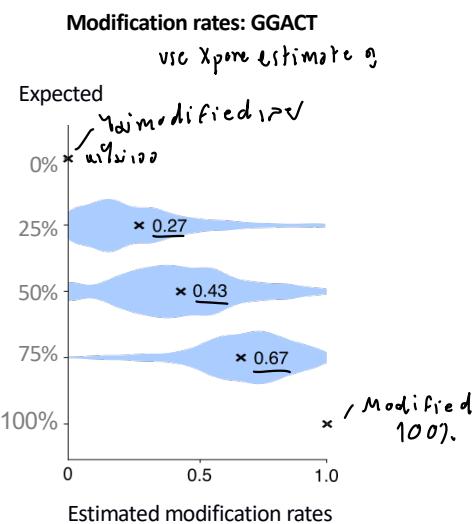
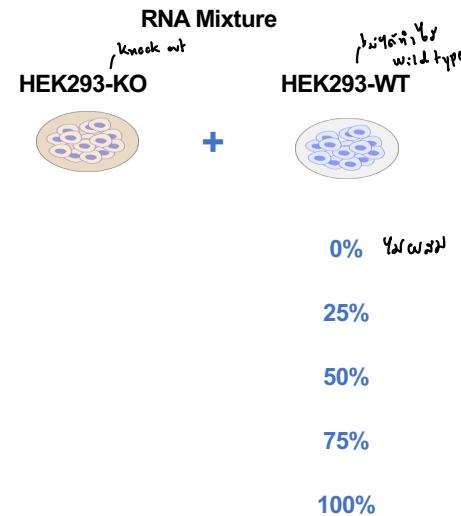
- as the **labels are incomplete**,
- the **false positives may not be wrong**.
- **Analysis on the predictions** is usually required to **get more insights**.



- ~1 million sites were tested.
- xPore achieves AUCROC of 86% to call differentially m6A sites.
- Around half were identified by m6ACE-Seq.
- With m6ACE-Seq + DRACH, the accuracy is up to >95%.
- dRNA-Seq helps identify a different set of modified sites that had been otherwise missed.
- m6A motifs e.g. GGACT, AGACT are confirmed.
- xPore can facilitate motif discovery in any other pairwise comparison.
- The differentially modified sites are also enriched at stop codons.

- ① Xpore สามารถ explore m6A motif
- ② คำแนะนำเพื่อเพิ่ม rate modification rate ใน RNA

# Validation: m6A stoichiometry quantification



- xPore models all RNA mixture samples at once.
- Estimated modification rates closely match to the expected.

- Modification rates estimated by xPore can be interpretable as fractions of modified reads in a cell.
- This allows the analysis of differential modifications.

# Validation: ML Metrics & Result Analysis

ป. ก. ก. validate

**Fig. 2 | Detection of m6A sites in the human transcriptome.**

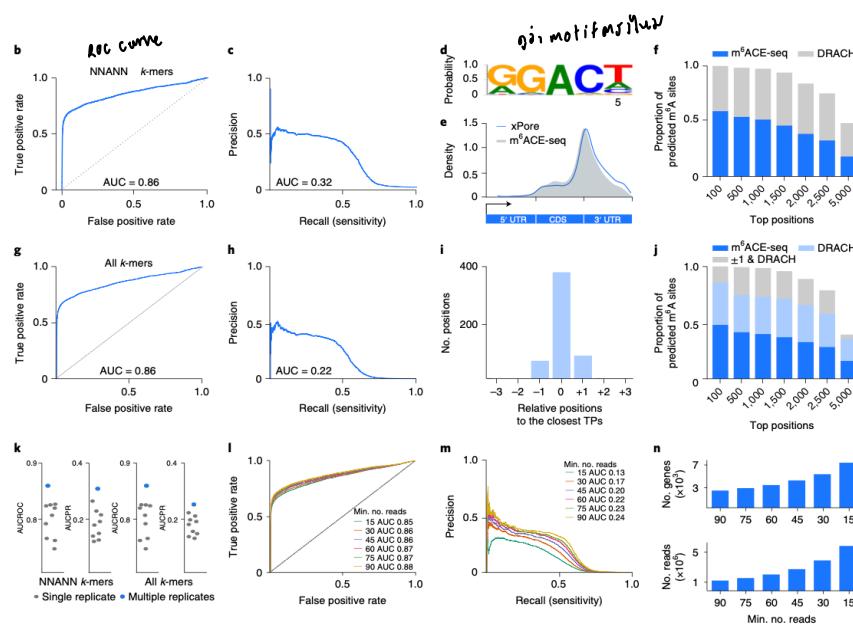
## ML Metrics

- ROC Curve
- Precision-Recall Curve
- Accuracy

## Analysis

- Domain-specific evaluation
- Effects of the data size

↓ data ใหญ่ → Acc. ที่ดีกว่า



# Validation: ML Metrics & Result Analysis

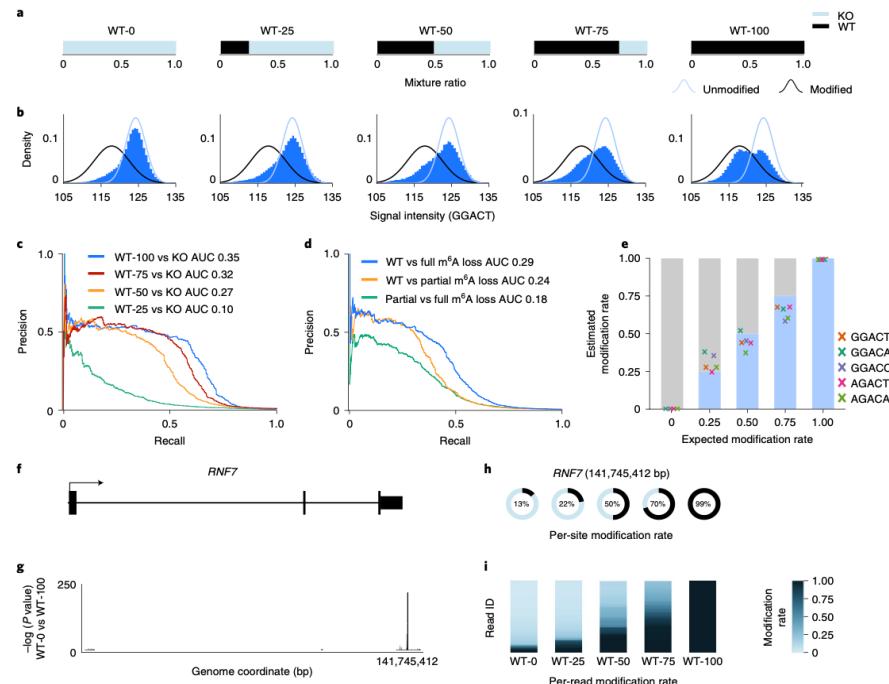
**Fig. 3 | xPore modification-rate estimates correspond to the fraction of modified RNA species in the cell**

## ML Metrics

- ROC Curve
- Precision-Recall Curve
- Accuracy

## Analysis

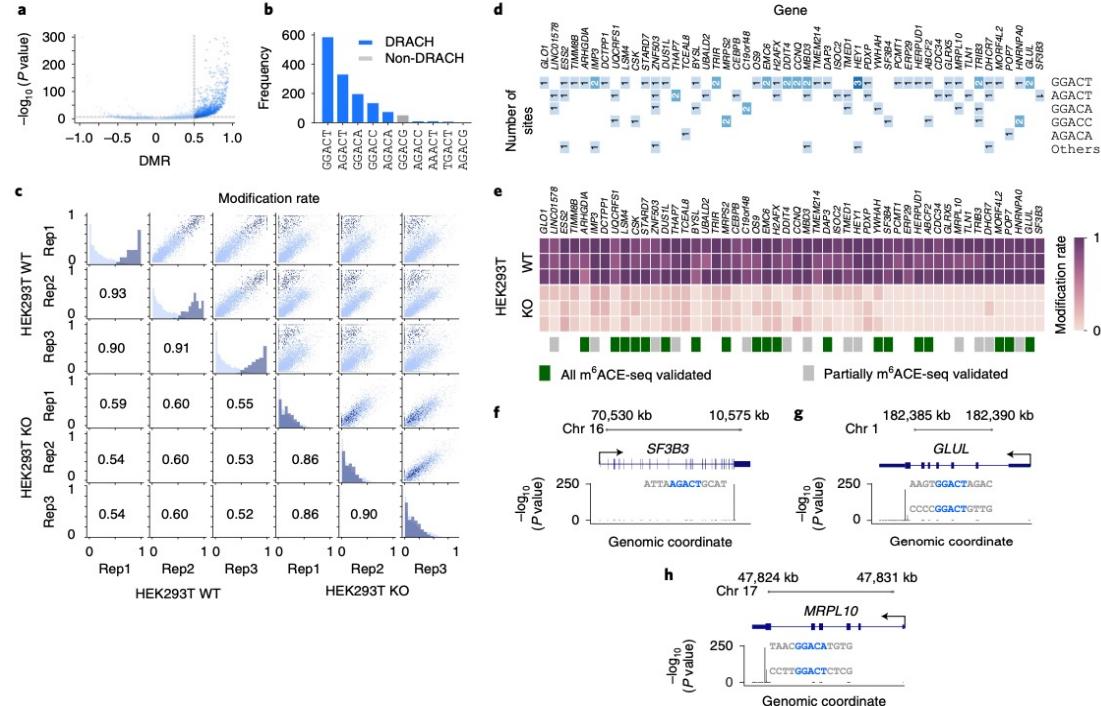
- Domain-specific evaluation
- Effects of the data size



# Applicability: Full Dataset

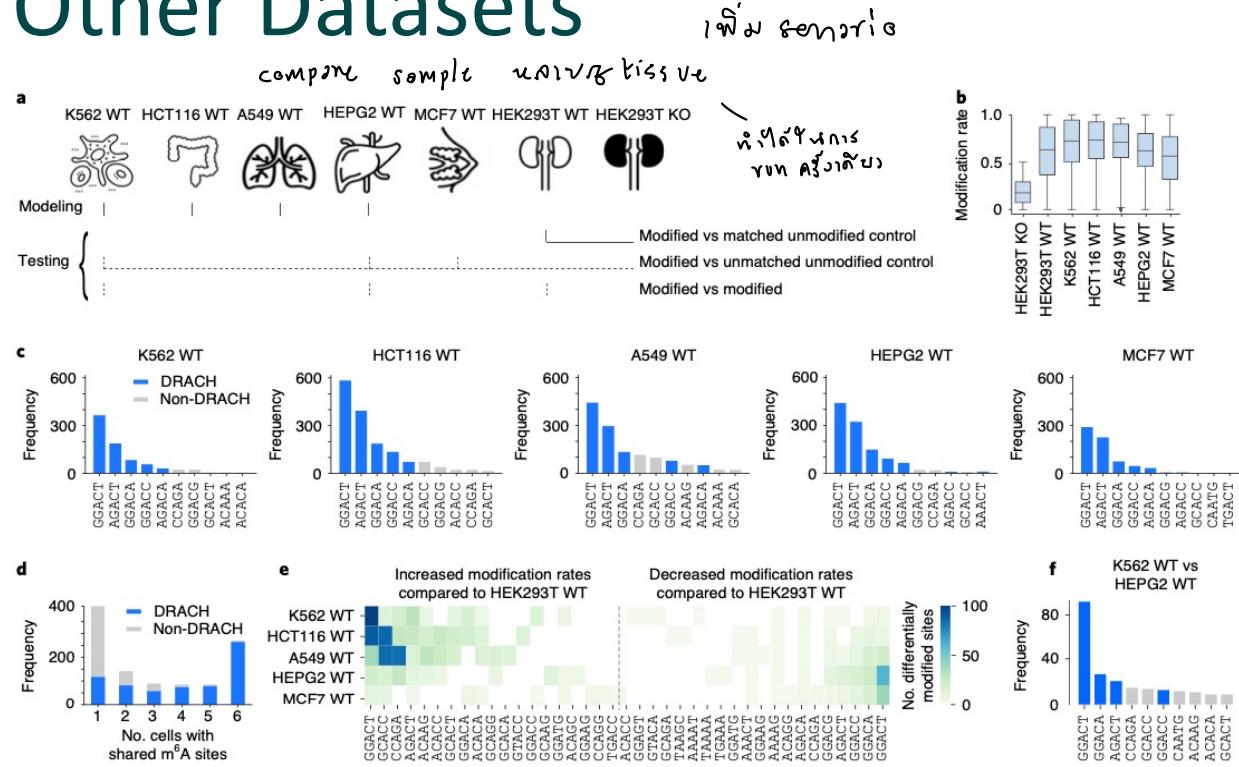
**Fig. 4 | Transcriptome-wide identification of differentially modified positions.**

Xpare qđ gđ Full dataset  
↓  
100% รุ่น



# Applicability: Other Datasets

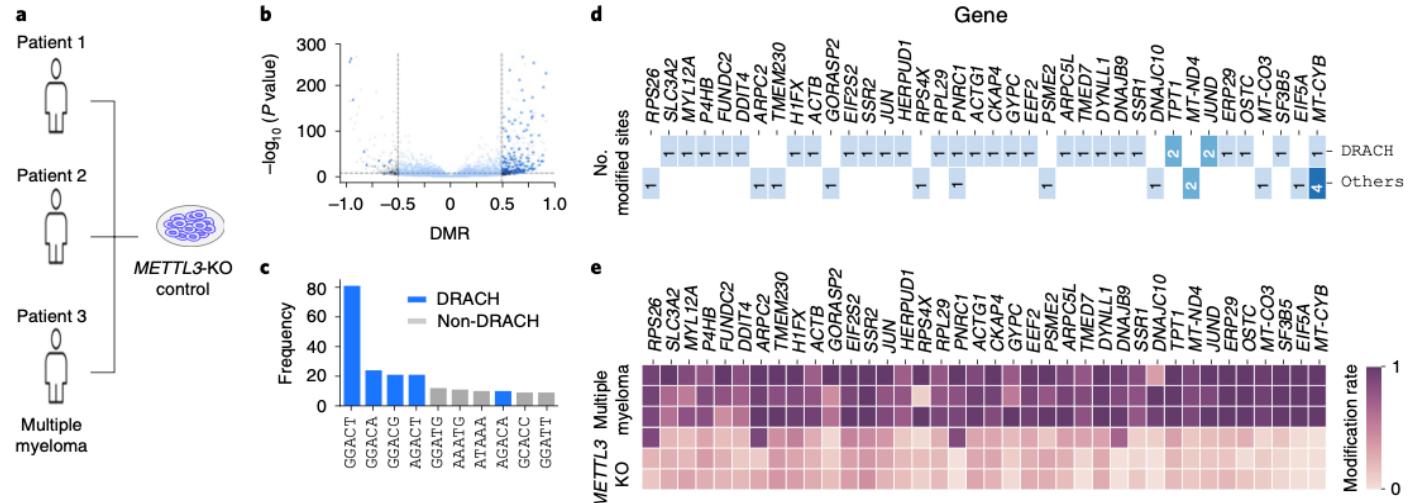
**Fig. 5 | Identification of m6A sites across different tissues and cell lines.**



# Applicability: Clinical Data

ກ່ຽວຂ້ອງ  
ກ່ຽວຂ້ອງ

**Fig. 6 | Identification of m6A in clinical samples using direct RNA-seq.**



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

កិឡាលោមអ្នក — ពីរបៀប

## Evaluation: Keys Takeaway

- Validation នៅវិវាទនៃរបៀប គេត្រូវបង្ហាញថា វាបានសម្រាប់ប្រើប្រាស់នៅលើផ្តល់ទម្រង់
  - Using appropriate ML metrics
  - Analyzing the results to get more insights
- Comparison with other state-of-the-art methods នៅរបៀបណាស់នឹង S.W. ដូចជាដំឡើង
- Applicability នៅរបៀបណាស់នឹងផ្តល់ទម្រង់
  - External / Other data full dataset
  - Human evaluation
  - Discovery



Present paper ที่ทางสถาบันฯ

E-SAN Thailand Coding & AI Academy

**E-SAN THAILAND CODING & AI ACADEMY**

โครงการวิจัยโมเดลระบบปั้นเวศการเรียนรู้กับบูรณาการ CODING & AI สำหรับเยาวชน

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

## 5. Visualization and Presentation

True positive rate vs False positive rate (ROC curve) showing AUC = 0.86.

Estimated modification rates plot showing values for XPORE, m6ACE-Seq, and DRACH.

Number of genes vs Genomic coordinate for various genes (Dyr1, SF3B1, GLUC, TSHZ5, SFRP5, GLP1R, MPV10, 47S110).

Modification rate heatmaps for HEK293T WT and KO cells across different conditions.

Box plots of Normalized enrichment for GGACC and AGACA motifs.

Screenshot of the xpore 2.1 GitHub repository showing installation instructions for Python and from the GitHub repository.

- Storylining @ academic paper
- Choosing the Right Plots
- Source Code
- Online Documentation

# Storylining

Method overview → ดูใน  
Fig. 1 | Schematic workflow: quantification of RNA modifications from direct RNA-seq data using xPore

Validation

Fig. 2 | Detection of m6A sites in the human transcriptome.

Fig. 3 | xPore modification-rate estimates correspond to the fraction of modified RNA species in the cell

Applicability & Discovery

Fig. 4 | Transcriptome-wide identification of differentially modified positions.

Fig. 5 | Identification of m6A sites across different tissues and cell lines.

Fig. 6 | Identification of m6A in clinical samples using direct RNA-seq.

ການອະນຸມາດວິທີນິວ້າ — ຖະໜາຍົກ  
ເປັນຫຼາຍໃຈຂອງການອະນຸມາດ (ການອະນຸມາດ)

xPore: identification of differential RNA modifications.

xPore identifies m6A sites at single-base resolution.

Replicates increase precision. 1 sample ແລະ ນົບຕົວ

Pooling data increases sensitivity.

xPore identifies modified positions with low stoichiometry.

Quantitative estimation of RNA-modification rates.

DMRs as estimates of effect size.

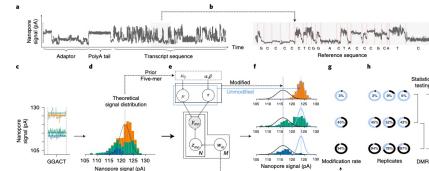
Identification of m6A across genetically diverse cell lines.

Variation of m6A across different cell lines.

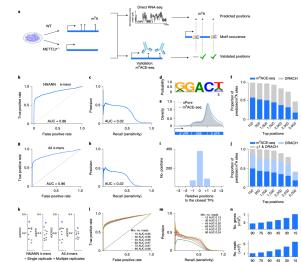
Identification of m6A in clinical cancer samples.

# Choosing the Right Plots

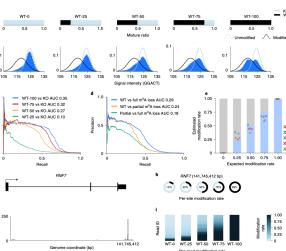
**Fig. 1 | Schematic workflow: quantification of RNA modifications from direct RNA-seq data using xPore**



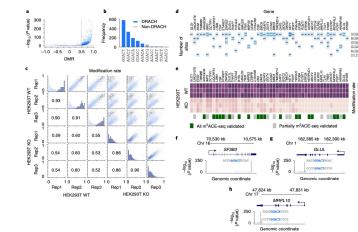
**Fig. 2 | Detection of m6A sites in the human transcriptome.**



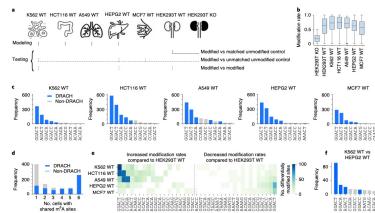
**Fig. 3 | xPore modification-rate estimates correspond to the fraction of modified RNA species in the cell**



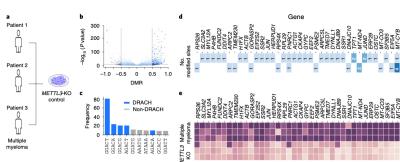
**Fig. 4 | Transcriptome-wide identification of differentially modified positions.**



**Fig. 5 | Identification of m6A sites across different tissues and cell lines.**



**Fig. 6 | Identification of m6A in clinical samples using direct RNA-seq.**



# Choosing the Right Plots

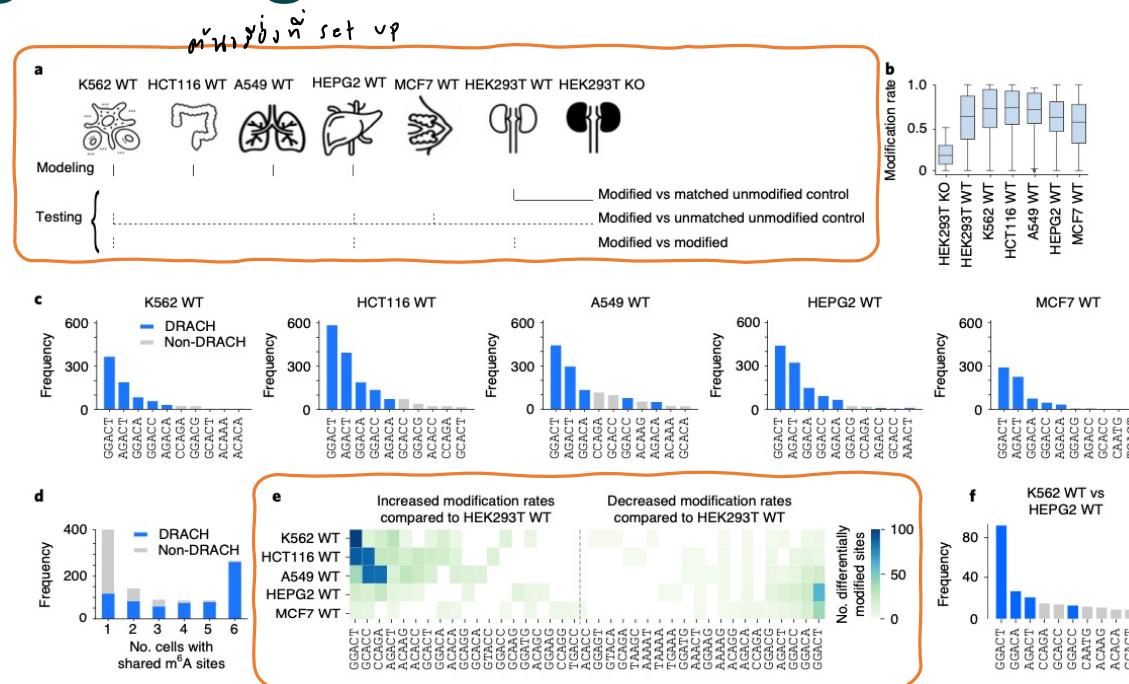


Fig. 5 | Identification of m<sup>6</sup>A sites across different tissues and cell lines.



CODE COMBAT

Google



ESAN THAILAND  
CODING & AI ACADEMY

# 3 Key Success to Develop AI-Powered Apps

1. Alignment with the actual needs
2. Sufficient generalization and evaluation
3. Simple deployment and serving
  - Online documentation
  - Easy installation
  - Source code
  - Data availability
  - Lightweight
  - Fast

<https://github.com/GoekeLab/xPore>

ARTICLES

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

nature  
biotechnology

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

Ploy N. Pratanwanich<sup>1,2,3,5\*</sup>, Fei Yao<sup>1,3</sup>, Ying Chen<sup>1,3</sup>, Casslyn W. Q. Koh<sup>1,3</sup>, Yuk Kei Wan<sup>1,3</sup>, Christopher Hendra<sup>1,4</sup>, Polly Poon<sup>1</sup>, Yeek Teck Goh<sup>1</sup>, Phoebe M. L. Yap<sup>1</sup>, Jing Yuan Chooi<sup>1</sup>, Wee Joo Chng<sup>4,5</sup>, Sarah B. Ng<sup>1</sup>, Alexandre Thierry<sup>6</sup>, W. S. Sho Goh<sup>1,2,3</sup> and Jonathan Göke<sup>1,3,6,7</sup>

Scopus metrics

78  
99th percentile  
Citations in Scopus

9.61  
Field-Weighted citation impact

downloads 27k

[xpore.readthedocs.io/](https://xpore.readthedocs.io/)

python machine-learning rna-seq  
nanopore genomics rna  
transcriptomics modification  
nanopore-sequencing rna-modifications

[Readme](#)

[MIT license](#)

[Activity](#)

[121 stars](#)

[9 watching](#)

[22 forks](#)

[Report repository](#)

[Releases 9](#)

[xPore v2.1](#) Latest  
on Oct 9, 2021

[+ 8 releases](#)

โครงการวิจัยโน้ตเดลร์บบันเวศการเรียนรู้ที่บูรณาการ 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 PACIFIC

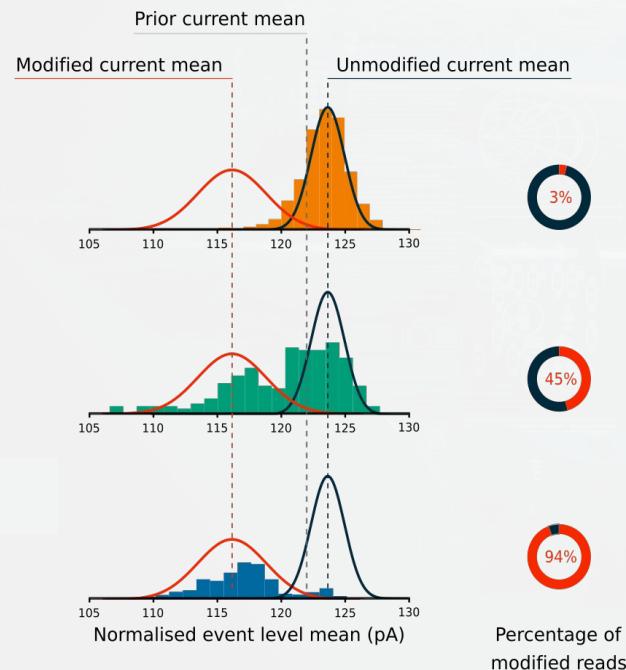
THAI  
PROGRAMMER



E-SAN THAILAND  
CODING & AI ACADEMY

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

## 6. Future Work

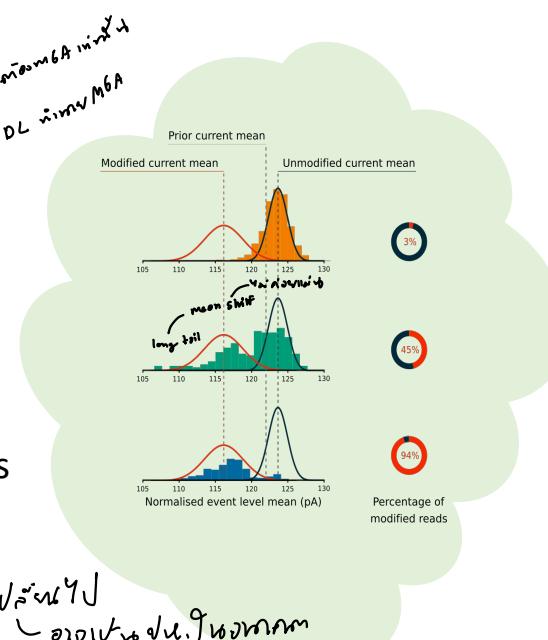


- Identifying the Limitations
- Considering Changes in the Future

# Future Work

## Domain-Oriented

- m6anet
- <Gaussian> mixture model
  - sensitive to long tail
- Interpretability
  - Modification or basecalled errors
- End-to-end
  - Why? / 3rd party SW.  
→ Nanopolish eventalign /  
Guppy basecaller are subject to change



## Method-Oriented

- Deep autoencoder + GMM
- CNN + GMM
- Other models + GMM

