



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

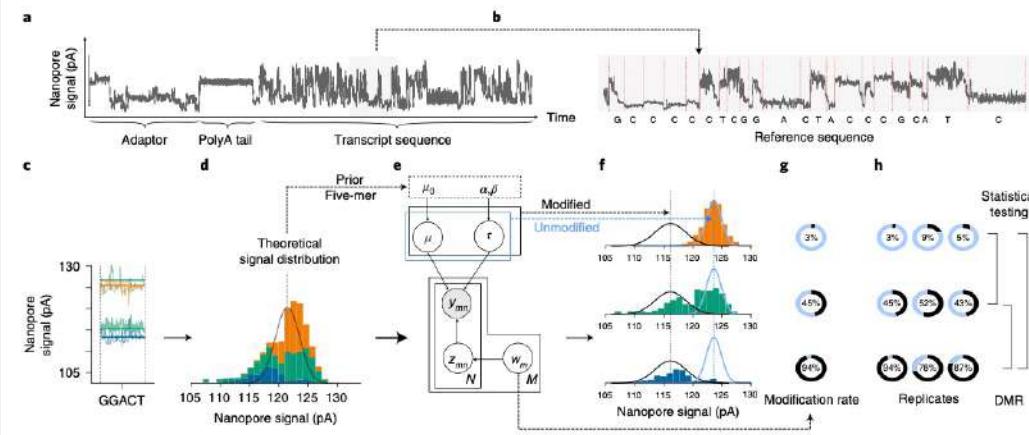
โครงการวิจัยไมโครสโคปนีโอศึกษาเรียนรู้กับฐานการ CODING & AI สำหรับเยาวชน
Model of Learning Ecosystem Platform integrate with Coding & AI for Youth

Add option

3. Bayesian [Multi-Sample] Gaussian Mixture Modelling

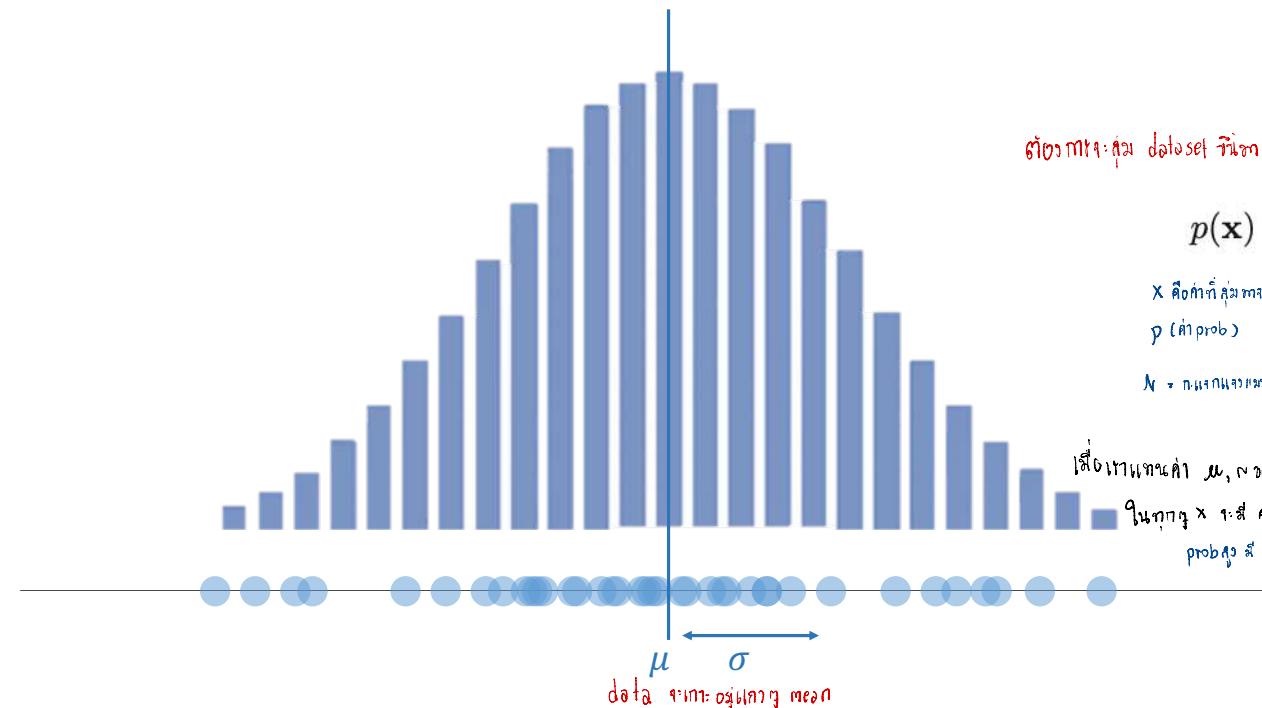
การ model ให้有多個 multiple - sample

L technic



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

Bayesian Multi-Sample Gaussian Mixture Model



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

ສືບກາງ distribution

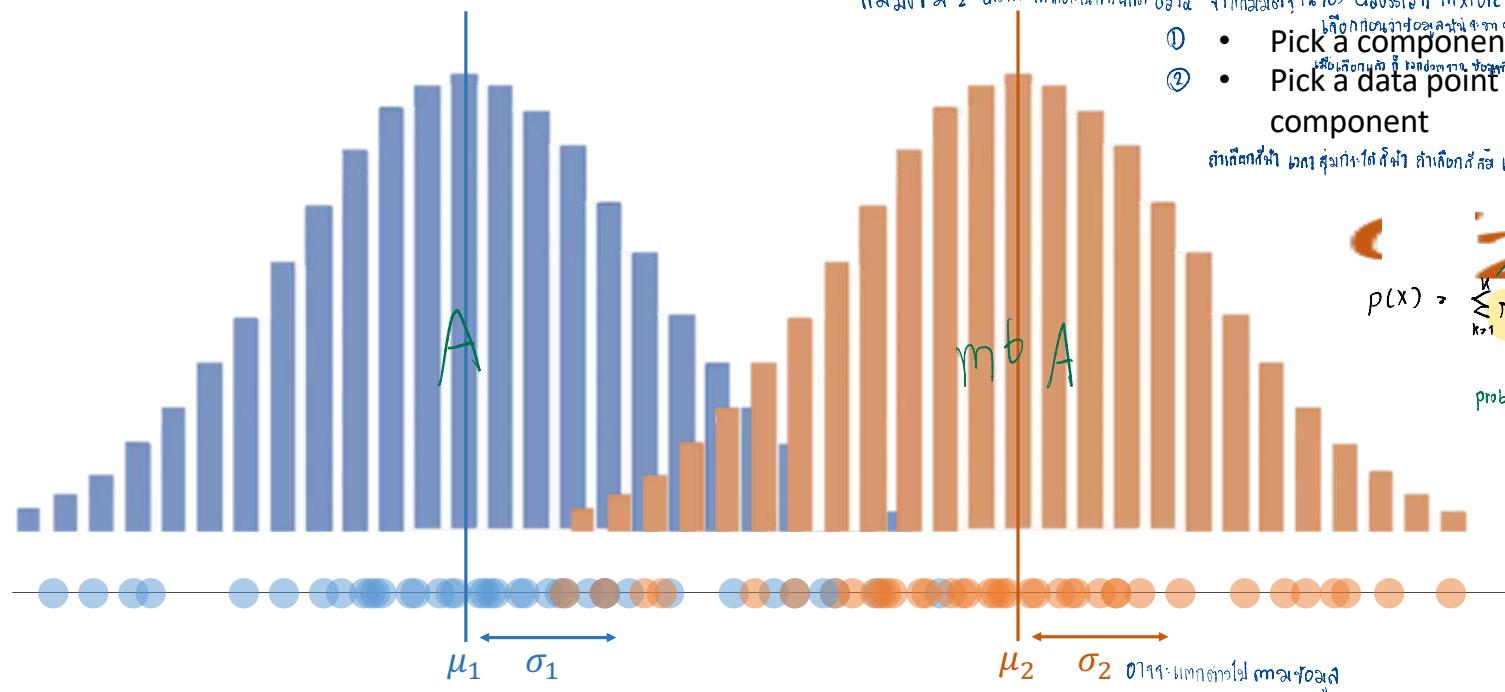
Bayesian Multi-Sample Gaussian Mixture Model

ສ່ວນຕີ ຂີ່ 2 ດິນຕາ ຫຼື ຕີດໃນການອັກ ຢ່າງ ຈາກອ່ອນຕີງໆ ຈາກ ດົບຮູ້ອັກ Mixture

ເນື້ອກຳດູງຈາກອັກຂອງລັບອັກ ທັງອັກທີ່ໄດ້

- ① • Pick a component at random
- ② • Pick a data point from the chosen component

ຄ້າເກີດກໍ່ນີ້ ນັກ ສຸວັນຍະໄສ ກໍ່ນີ້ ສໍາເລັກກໍ່ນີ້ ເລກກໍ່ນີ້ ດັກກໍ່ນີ້ ໄກສໍາ ໃຫ້ ດັກໄສສຸດ ທີ່ກໍ່ນີ້



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

ຄ່າ ດູວ່າວິວ ຖ້າ ດູງເກີດ ເຊັ່ນ component ສໍາດັ່ງ 10% / 10%

What is GMM?

อี.คือ Gaussian Mixture Model

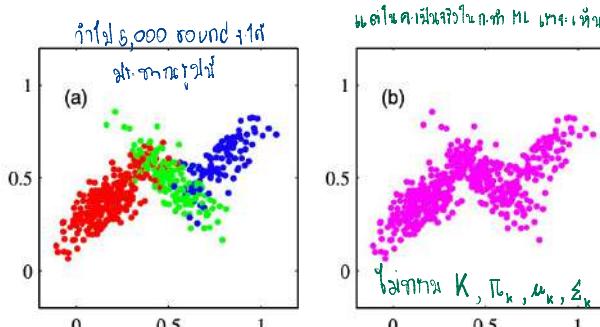
Gaussian Mixture Model ลักษณะตัวฐานในกลุ่ม data คือ

Assumption how data are generated as follows

- There are K components ex. $\frac{data}{3}$ components (r, b, g)
- Each component is defined as a Gaussian distribution
- Pick a component at random After that \rightarrow เที่ยวนคอมเพนท์ที่เลือก ex. สีแดง
- Pick a data point from the chosen component

ก. เลือก component
พิเศษที่ B

* ถ้าเลือก 1 ไม่ต้อง สมมติฐาน
* ทำงานค่าความน่าจะเป็น



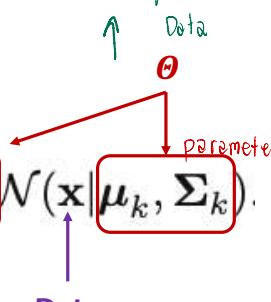
Source: Christopher M. Bishop, "Pattern Recognition and Machine Learning", 2006

โดยมี data ที่จะหา θ แล้ว นำเข้า

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

หมายเหตุ: component ก็คือ μ และ σ คือหน่วยทางกายภาพ
ที่ใช้ normal dist

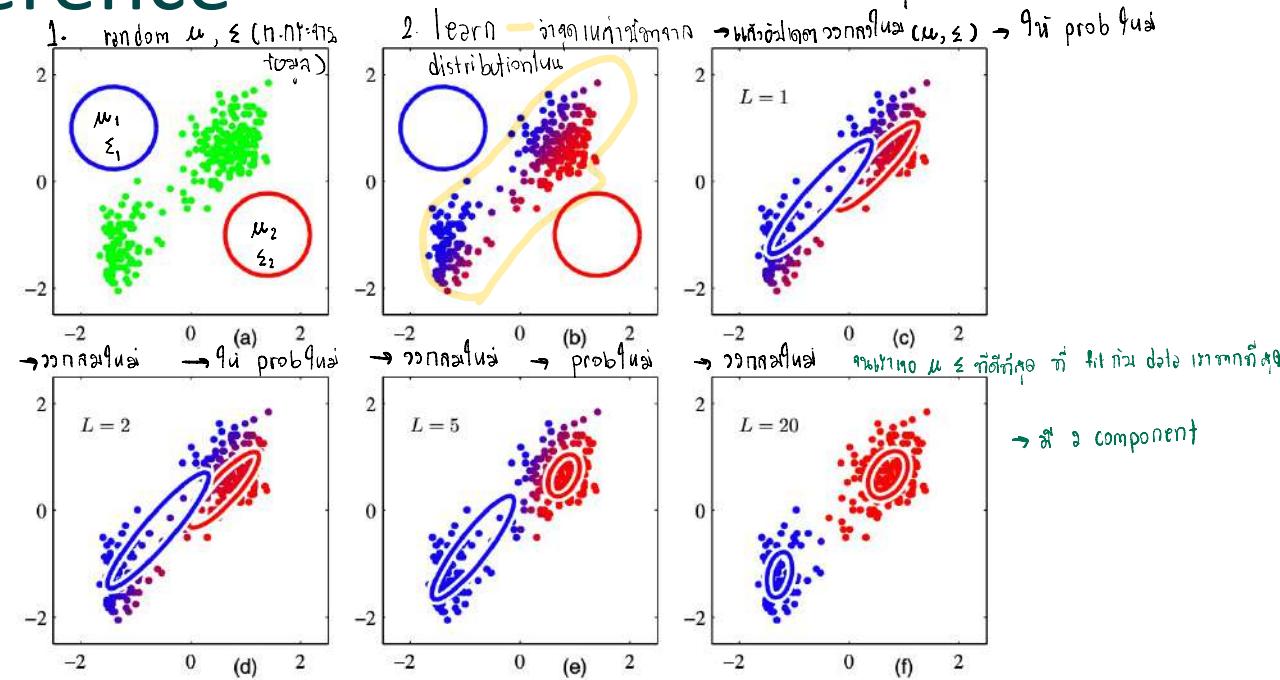
Data



หมายเหตุ: สำหรับ Data ที่ได้จากเรียนรู้
ให้เราสามารถคำนวณค่าความน่าจะเป็นของ concept ของ GMM
สามารถคำนวณค่าความน่าจะเป็นของ component ที่เลือก
มาต้องการค่าความน่าจะเป็นของ component
ก. 1 เลือก component Data ที่ได้จากเรียนรู้
infer / fit learn หาค่า θ ต่างๆ ที่เกี่ยวข้อง กับค่า μ และ σ
ก. 2 คือ หาค่า π_k ของ component ที่เลือก

GMM Inference

Iterative algorithm



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

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

Try Coding

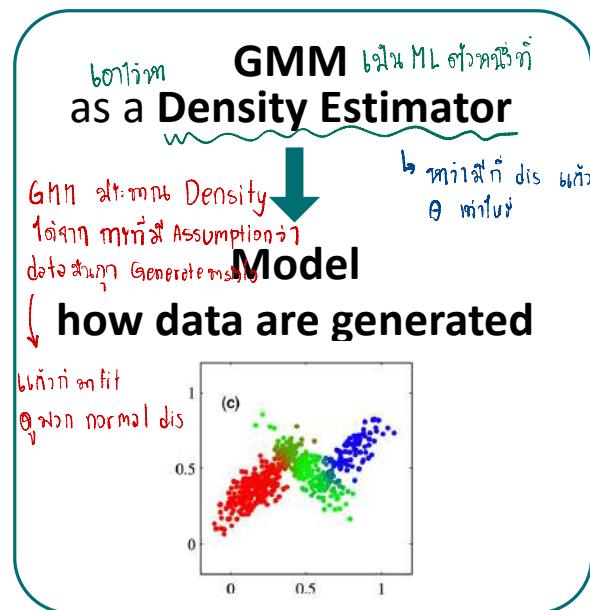


ສ່ວນການ, data ອີ່ໄດ້ bc of Assumption

"Generative AI"

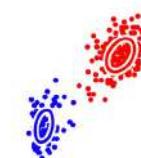
39064

ເບື້ອງການ type π_{ij}

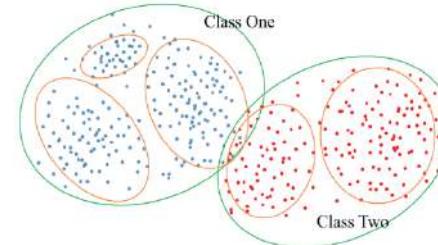


ກໍາ

Clustering for + Gen data 1st



ມີຂະໜາດ (One-Class) Classification



(binary class)

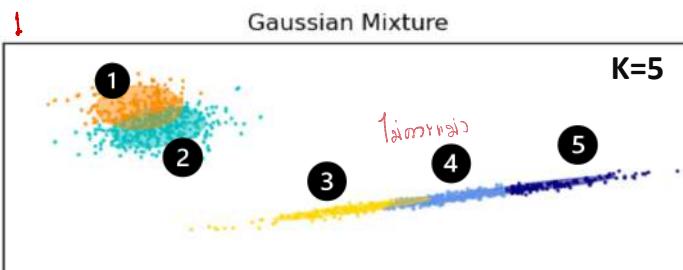
GMM → หัวสร้าง Data



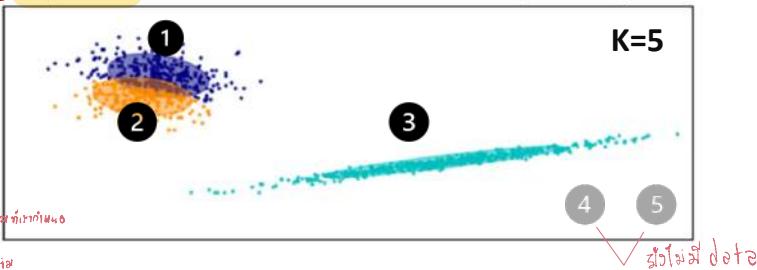
Bayesian Multi-Sample Gaussian Mixture Model

วิธีการหาค่ามาทริซต์ Θ กรณี m, z, π เท่าไหร่ ดีที่สุด

Learning algorithm for making inference on the latent variables



2 Bayesian Gaussian Mixture with a Dirichlet process prior



ก่อนเริ่มการจำลองเราต้องกำหนดค่า parameter
Data ของจุดเดียว Data ทุกจุดเป็นสี่เหลี่ยมจัตุรัส
และต้องตั้งค่า Assumption ให้กับ K จำนวนเท่าไร
1. ต้องกำหนด component ที่ต้องการ คือ K ต้องกำหนด
จำนวน component ที่ต้องการ

2. ต้องตั้งค่า parameter ให้กับ component ที่ต้องการ
คือ ต้องตั้งค่า Prior ให้กับ component ที่ต้องการ ที่ต้องการ

3. ต้องตั้งค่า parameter ให้กับ component ที่ต้องการ
คือ ต้องตั้งค่า Prior ให้กับ component ที่ต้องการ ที่ต้องการ

4. ต้องตั้งค่า parameter ให้กับ component ที่ต้องการ
คือ ต้องตั้งค่า Prior ให้กับ component ที่ต้องการ ที่ต้องการ

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

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

Data

รูปที่ 1

จุดที่ Data อยู่ทางเดียว

Point estimate = Maximum Likelihood

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

↓ Θ
จุด Θ ชุดเดียวจะออก
เหมือนกันทุกจุด

argument
มากกว่า 1

หากมีตัว → หาค่ามาทริซต์ให้ Data ทุกจุด ที่ต้องการ prob ของ Data ทุกจุด ให้คำนึงถึง

Posterior
 $P(\Theta | \text{Data})$

$$= \text{Likelihood} \times \text{Prior}$$

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

Prior
 $P(\Theta)$

คือหมายความว่าค่าเดียวกันทุกจุด
ยกเว้นจุดเดียวที่ต้องการ

↑

Bayesian นี่คือ Prior เท่านั้น

คือไม่ต้องคำนึงถึงจุดเดียวที่ต้องการ distribution ของ Data ทั้งหมด

ใน Bayes นี่คือ Prior (bayesian) bc. ตามที่ distribution ของ Data ทั้งหมด sample

① Frequentist vs Bayesian ②

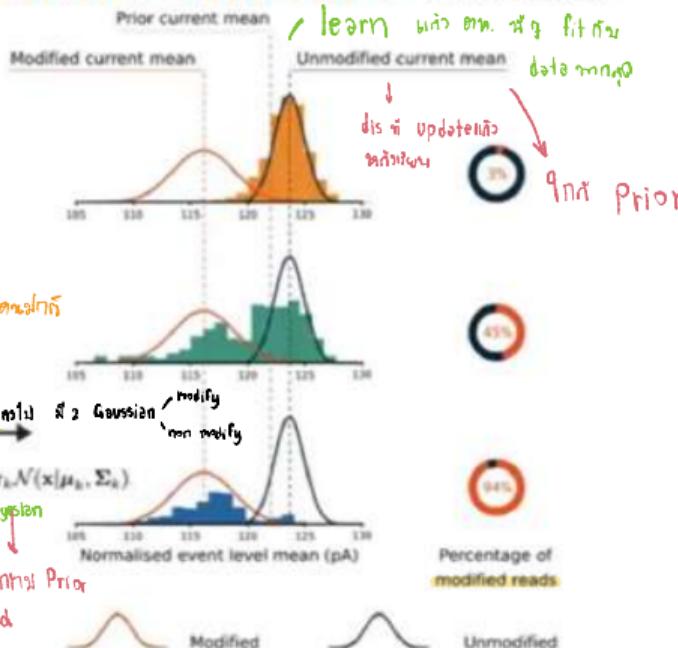
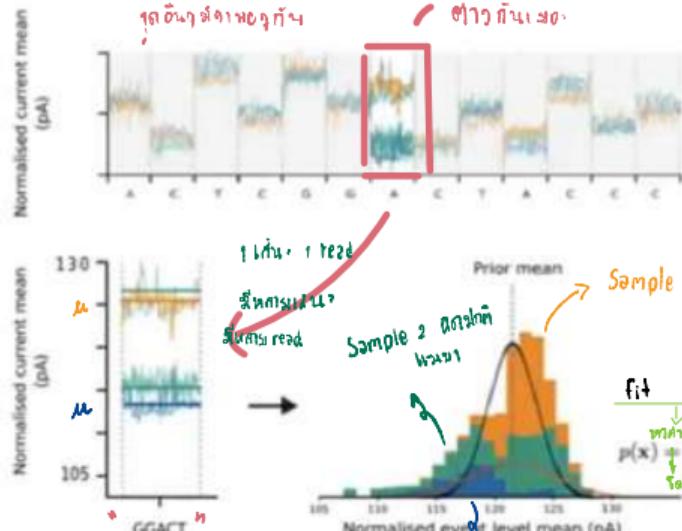
$P(\text{Data} \mid \Theta)$

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

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 <i>b.c. 1 ถึง 10 ต่อ 1 ก้อน กาว</i>	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

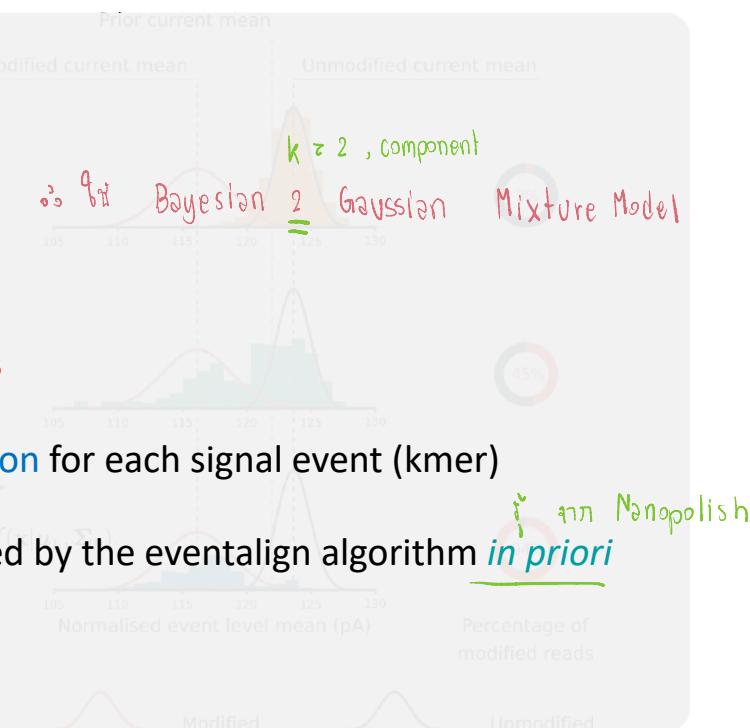
Bayesian Multi-Sample Gaussian Mixture Model

101 RNA 103 913 3 sample
in sequence →
↓
101 RNA 103 913 3



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**



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						0.42	
...
NNGN						-0.01	Least sig
...
NNTNN						-0.01	Least sig

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

Speed-Up ML Experiments

Automated ML models

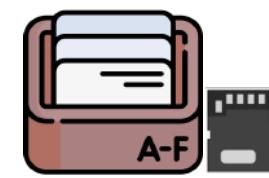
- Hyper-parameter settings
- Multiple datasets
- Different models / methods



- Config file
- Python packaging



- Parallelization



- File indexing

Why config files?

- Automating tasks
- Centralised configuration ຈຸດຕີເລື່ອສັນນາ config 0:ຍິ່ງ YAML, JSON, TOML, and INI are the popular and standardised formats of configuration files
- Documentation ກໍາເທົກການຈຳນວນ
- Portability ຄາມມູນຄົງໃຫຍ່

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



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> 10115114
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-
prefiltering:
    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>
```

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



Python Packaging

ploy-np Merge pull request #115 from

- docs
- figures
- xpore
- .gitattributes
- .gitignore
- LICENSE
- MANIFEST.in
- README.md
- setup.py

```
1  """Setup for the xpore package.***  ຕີກຳ ພັດທະນາ
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/GeekLab/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         'pysensembly>=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?%3Action=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 )
```

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



CODE COMBAT

Google

E-SAN THAILAND
CODING & AI ACADEMY

Parallelization / Multiprocessing

```
import multiprocessing
```

When Data are Too Big to Fit in the Memory

data.index

ຕົວຢ່າເຄີຍນາງກໍາຈາກໆໄຟສັກ

gene_id	start_idx	stop_idx
ENGxx1	0	16856
ENGxx2	16857	29435
...

data.json

```
{'ENGxx1': [123,110,...]}, {'ENGxx1':  
[123,110,...]}, {...}
```

ໂຄຮງການວັດຍໂນເດລຮະບບນິເວສົກເຮົາຮຽນນັ້ກໍບູນການ CODYING & AI ສໍາຫຼັບເຍາວຊັນ
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH



หัวข้อ ที่สนใจ

The slide is titled "4. Evaluation" and features a collage of scientific and technical figures. At the top, there is a banner for "E-SAN THAILAND CODING & AI ACADEMY" and logos for various sponsors including Thailand Coding & AI Academy, Thailand Science Park, Google, The Asia Foundation, DMAP, and THAI PROGRAMMER.

The main content includes:

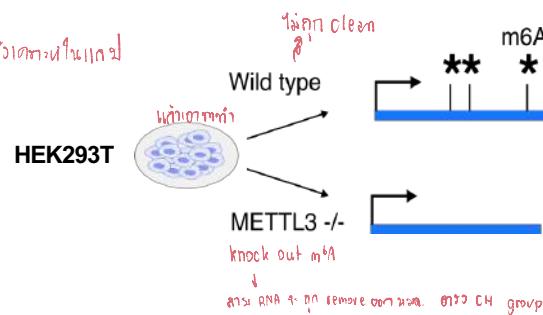
- A diagram comparing Wild type and METTL3 $-/-$ genotypes, showing m6A sites (marked with ** and *) on the Wild type strand.
- A "direct RNA-Seq" plot showing normalized coverage over time.
- A "Gene" section with a heatmap of m6ACE-Seq validation across various genes.
- A scatter plot of True positive rate vs False positive rate with AUC = 0.86, and estimated modification rates (0.27, 0.43, 0.67).
- A density plot comparing XPORE and mSAGE-Seq methods across 5' UTR, CDS, and 3' UTR regions.
- A genomic coordinate plot showing m6A distribution across exons and introns.
- A heatmap of m6A modification rates across different genomic coordinates.
- A scatter plot of accuracy versus top positions for m6ACE-Seq and DNACH methods.
- A heatmap of m6A enrichment for RNF7 across different genotypes (WT, AT, AD).
- A sequence logo for the motifs GGACC and AGACA.
- A schematic of the RNF7 gene structure.

On the right side of the slide, there is a vertical column of text bullet points:

- Experiment setup
- Validation
- Applicability
- Discovery

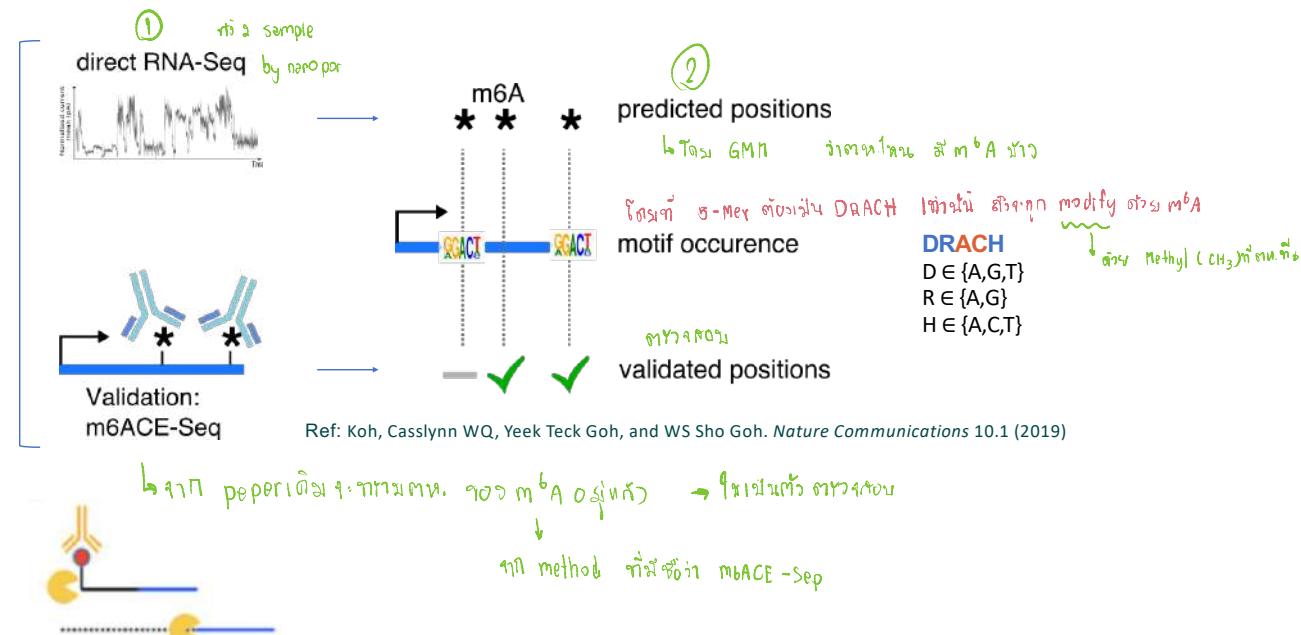
Experiment Setup

รีสอร์ท 101 cell HEK293T ทำตัวอย่างที่ไม่ต้อง
ใช้เคมี

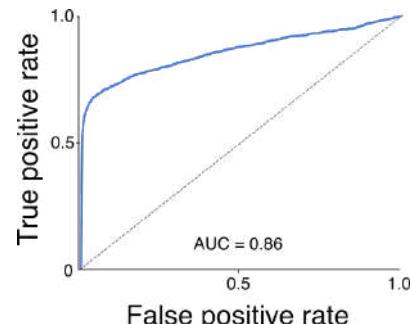


00 กว่าเนื่องจาก ทดลองว่า ชั้นของ RNA ที่ร่วงตกหล่น ซึ่งคืองานวิจัย หรือไม่

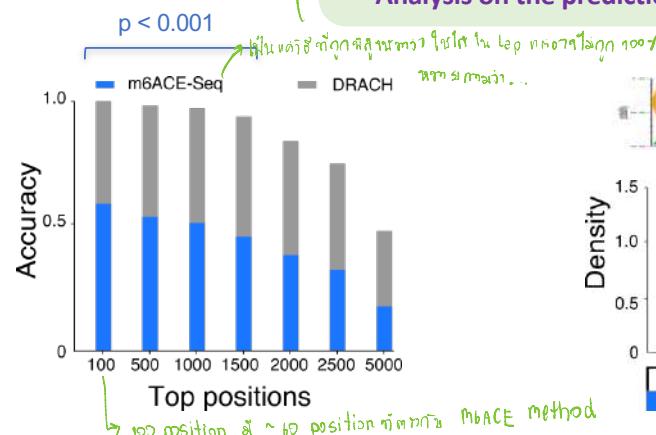
model's



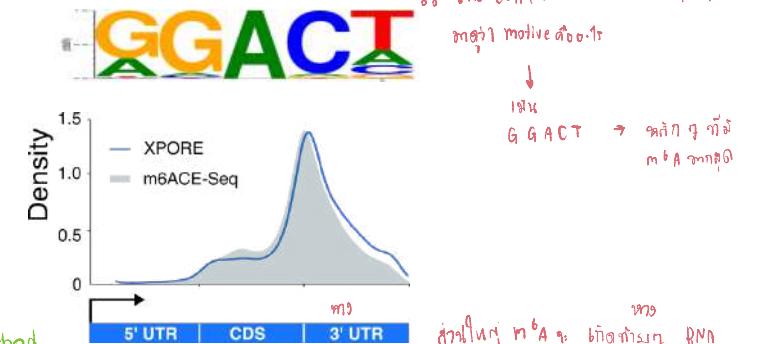
Validation: m6A calling



- ~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. → ຖາຍັນ ດັກ່າ
 - xPore can facilitate motif discovery in any other pairwise comparison.
 - The differentially modified sites are also enriched at stop codons.

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

→ Pore identify mult ✓ → Goal(1)

Goal ② จําหนີສ່ວາຕົວ, ທີ່ modify 1% ສ່ວນ modification rate ມີໃຫຍ່ read 1%.



CODE COMBAT

Google

The Asia Foundation

DMAP
DIGITAL MAKERS
ASIA PACIFIC

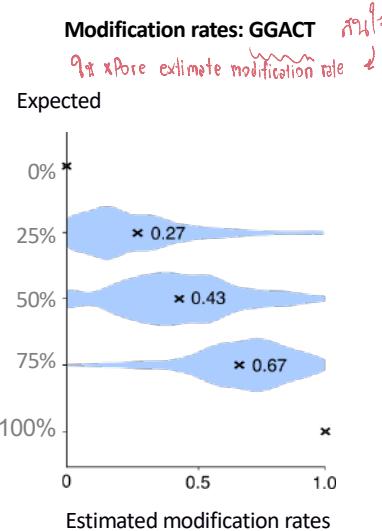
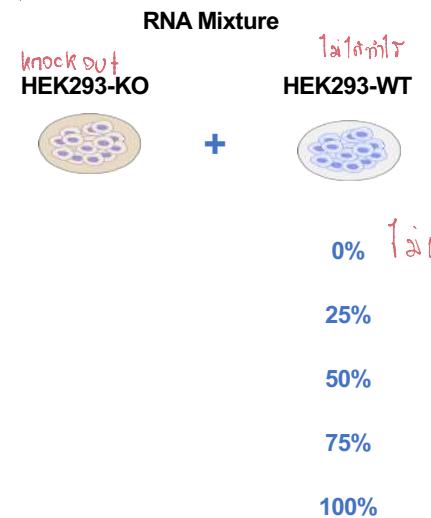
THAI
PROGRAMMER



E-SAN THAILAND
CODING & AI ACADEMY

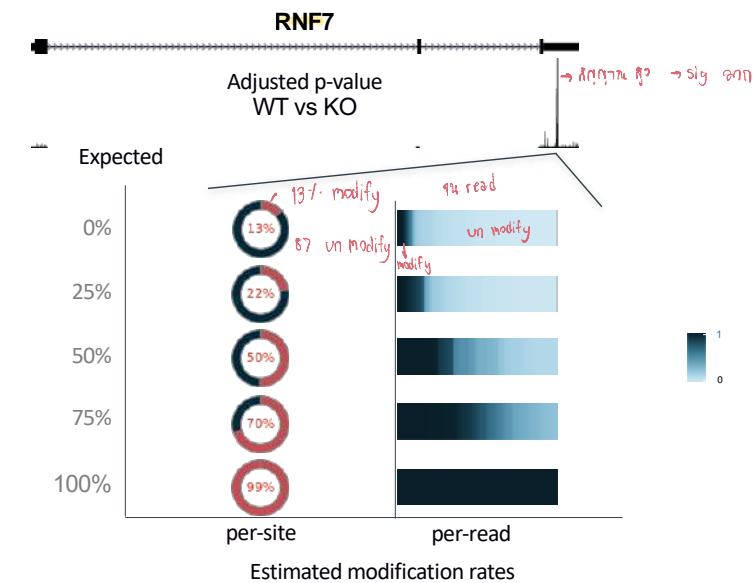
Validation: m6A stoichiometry quantification

ເຮືອມສັນ



- 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.

DN P ດາວກາ

m6A Gene → RNA

m6A Gene

Gene → RNA → transcript

Gene → RNA

to

Gene → RNA → transcript

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

Validation: ML Metrics & Result Analysis

ผลการ validation

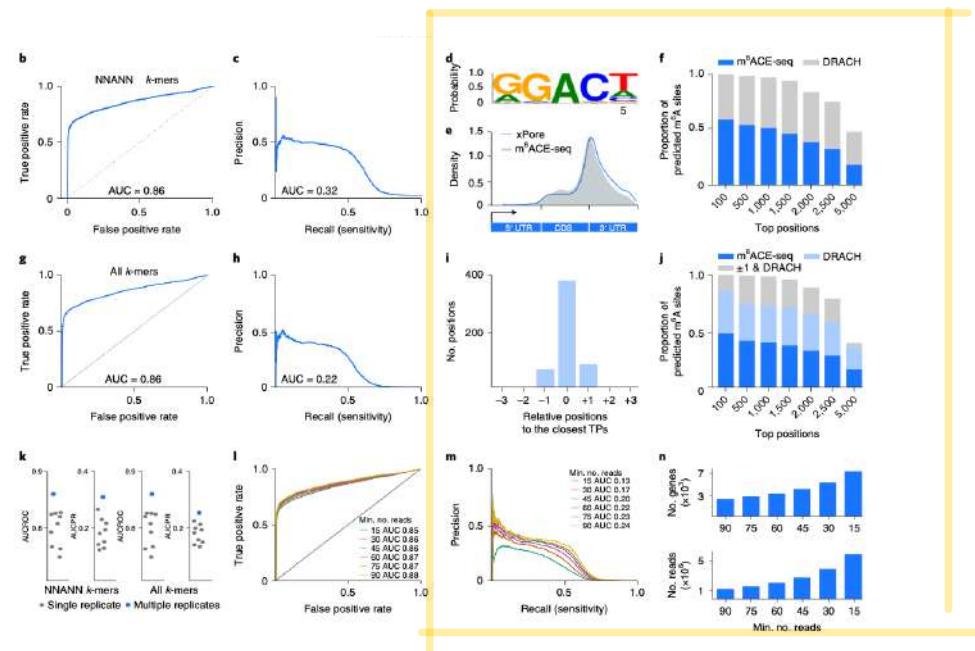
Fig. 2 | Detection of m6A sites in the human transcriptome.

ML Metrics

- ROC Curve
- Precision-Recall Curve
- Accuracy

Analysis

- Domain-specific evaluation
- Effects of the data size



Validation: ML Metrics & Result Analysis

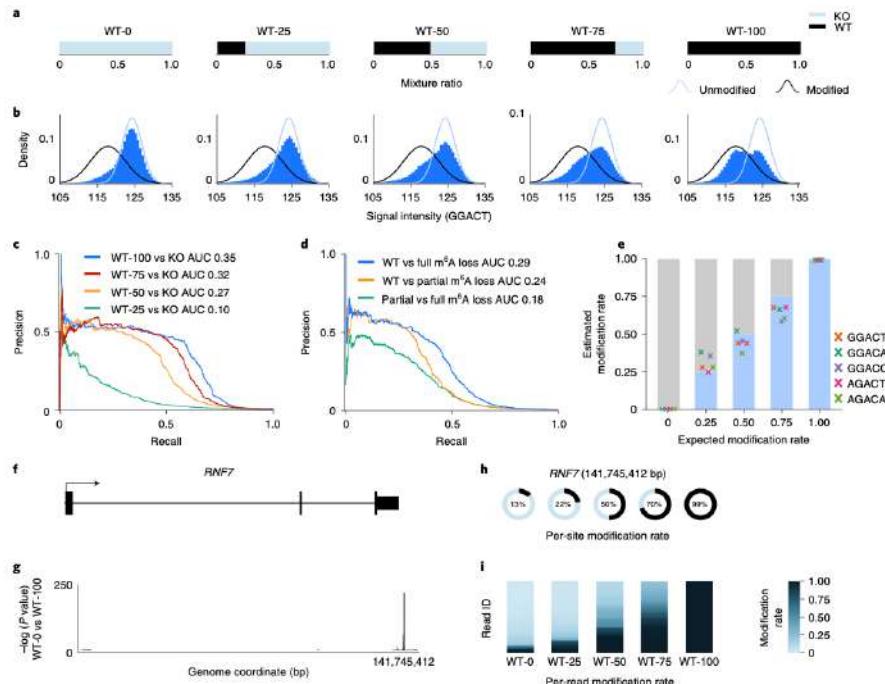
Fig. 3 | xPore modification-rate estimates correspond to the fraction of modified RNA species in the cell

ML Metrics

- ROC Curve
- Precision-Recall Curve
- Accuracy

Analysis

- Domain-specific evaluation
- Effects of the data size

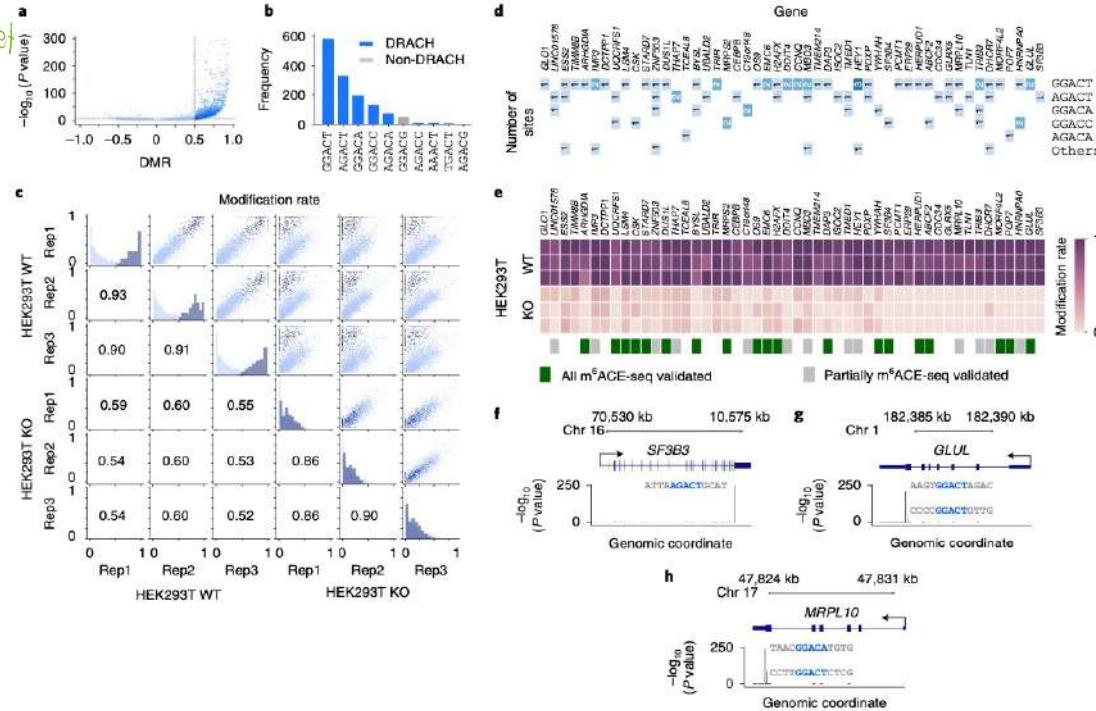


oscillating method in impact test 180 mm

Applicability: Full Dataset

↳ 2019 წლის full data set

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

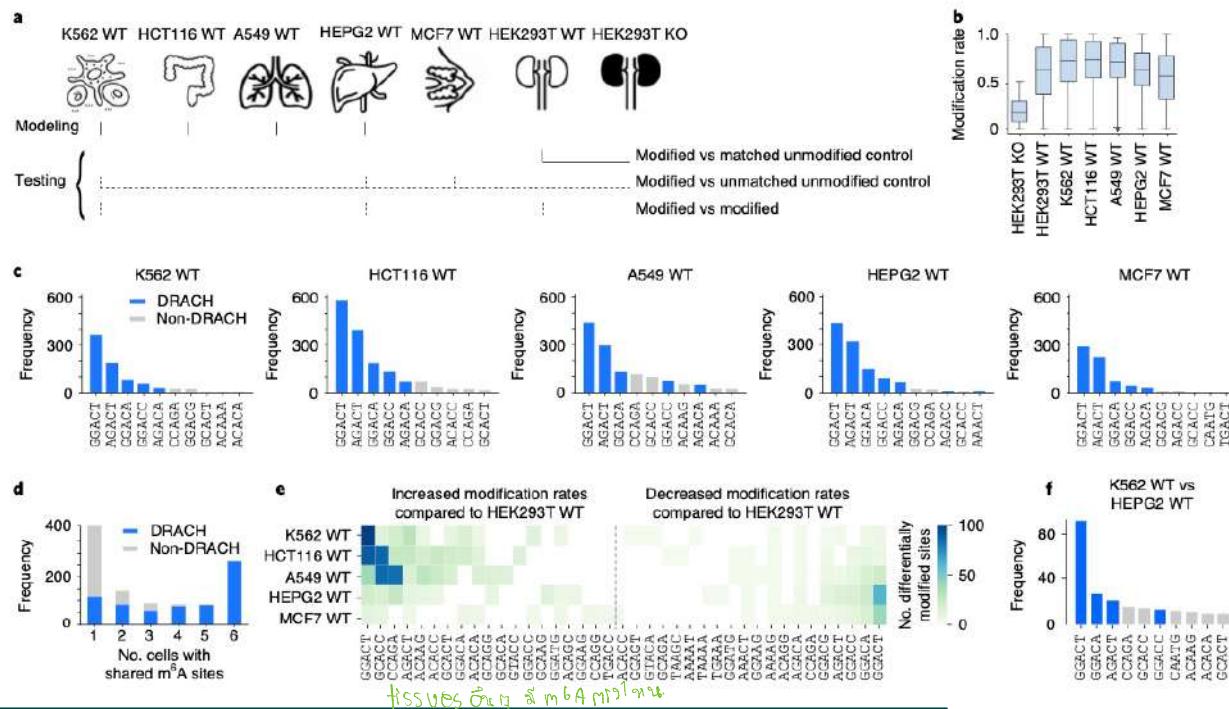


โครงการวิจัยโมเดลระบบปั๊วิศการเรียนรู้ที่บูรณาการ CODING & AI สำหรับเยาวชน
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

Applicability: Other Datasets

ເພີ້ມຂາດໜາຍ cell line ແລະ ຂອງ tissues.

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

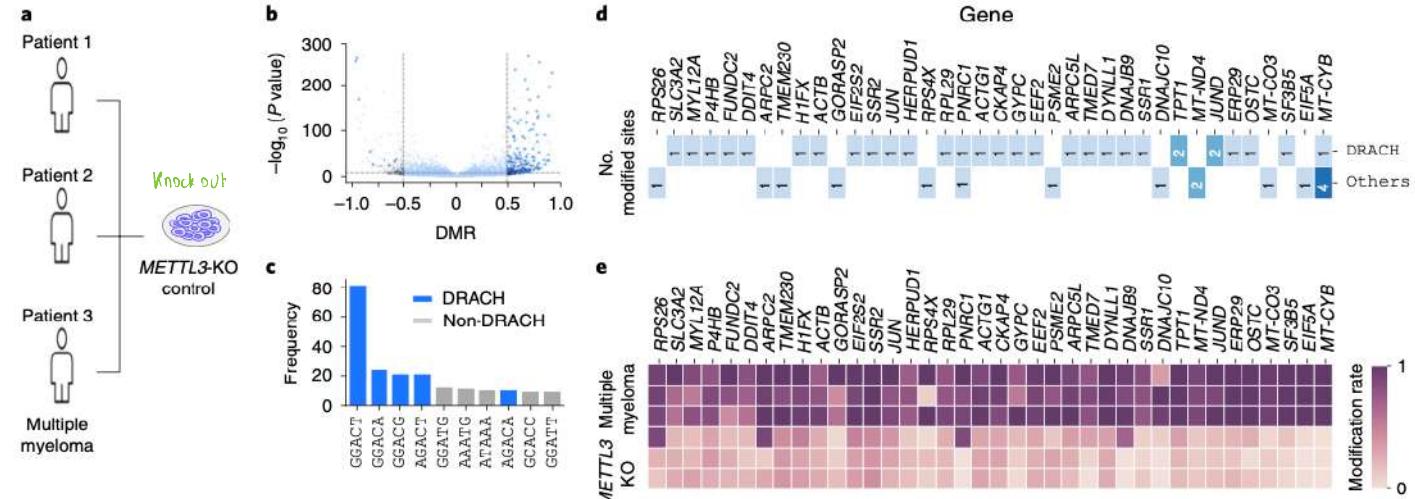


ໂຄງການວັດໄຍໂນໂດຣະບບນິເວສຳການເຮັດວຽກ m6A ສໍາຫຼັບເຢາວ່ານ
MODEL OF LEARNING ECOSYSTEM PLATFORM INTEGRATE WITH CODING & AI FOR YOUTH

Applicability: Clinical Data

→ Date ၁၇၂၇

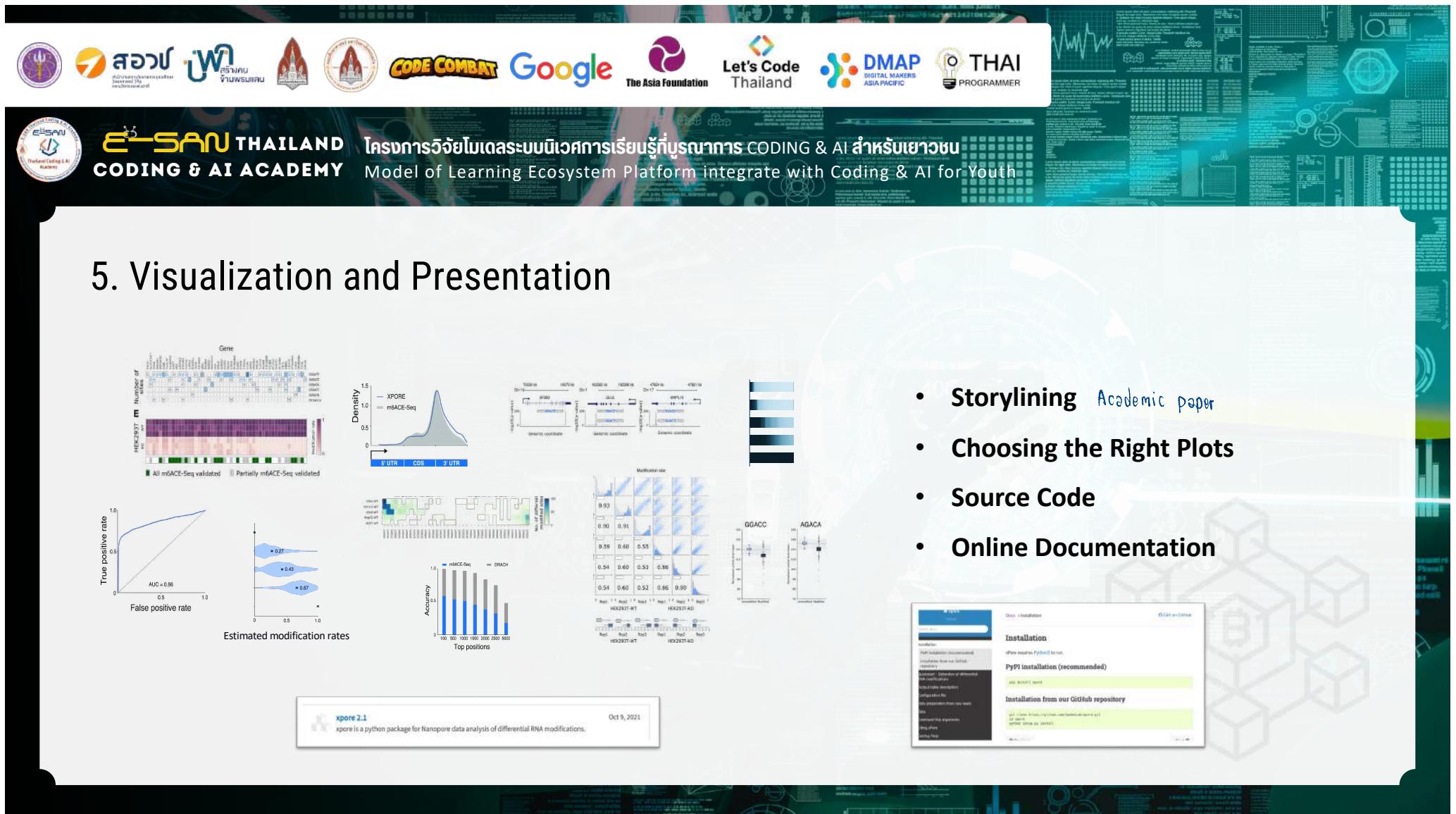
Fig. 6 | Identification of m6A in clinical samples using direct RNA-seq.



Evaluation: Keys Takeaway

- Validation
 - Using appropriate ML metrics
 - Analyzing the results to get more insights
- Comparison with other state-of-the-art methods / software ბეჭდი
- Applicability გთავაზონ
 - External / Other data → Full data set
 - Human evaluation
 - Discovery





ເຊື້ອກຕາມເຈັ້ນ paper ໂດຍຕ່າງໆ ແລະ ດຳວັດທະນາກິດການ ເນັ້ນໃຈກົດ



ໄປເລີນ outline ຫຼື ນຳກົດ section

Storylining

Method overview

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

ເພີ້ນ → ເຫັນ ຕອນ ໂດຍ

Sub topic

ໃນ m6A ອັດຕາກິດ

xPore: identification of differential RNA modifications.
xPore identifies m6A sites at single-base resolution.

Validation

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

ແຜ່ນ ຕອນ ໂດຍ

Replicates increase precision. 1 sample ສັນກະຊາດ

Pooling data increases sensitivity.

read modify goes in identifier

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

xPore identifies modified positions with low stoichiometry.
Quantitative estimation of RNA-modification rates.

ເຫັນໃຈກົດ Applicability & Discovery

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

DMRs as estimates of effect size.

ໄປເລີນຕ້ອງ ນໍາກ່າງ cell line

Identification of m6A across genetically diverse cell lines.

Variation of m6A across different cell lines.

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

Identification of m6A in clinical cancer samples.

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

Choosing the Right Plots

ການអໜີບຕົວ ຈະກັງ

ບມນຂ ເກົ່າໃຫຍ່ກົດ paper

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

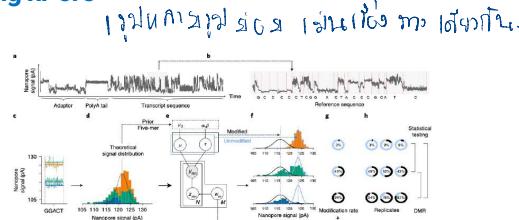


Fig. 4 | Transcriptome-wide identification of differentially modified positions.

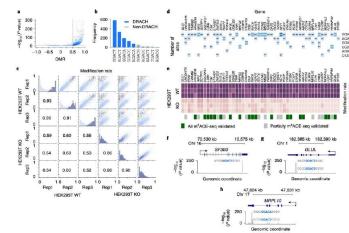


Fig. 2 | Detection of m6A sites in the human transcriptome.

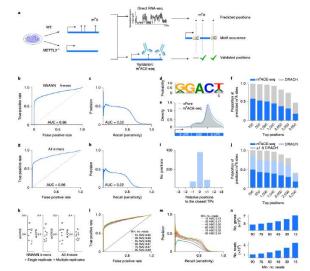


Fig. 3 | xPore modification-rate estimates correspond to the fraction of modified RNA species in the cell

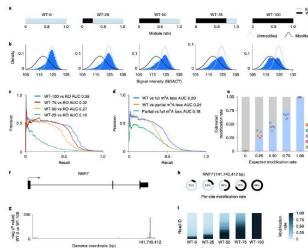


Fig. 5 | Identification of m6A sites across different tissues and cell lines.

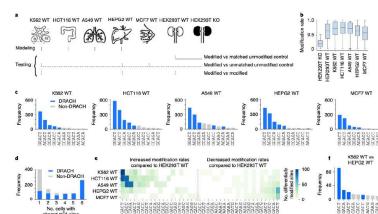
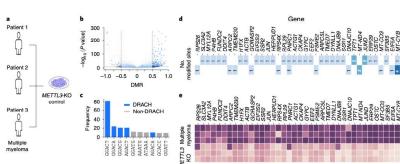


Fig. 6 | Identification of m6A in clinical samples using direct RNA-seq.



Choosing the Right Plots

Set up experiment

นักศึกษาชำนาญ 1. เรื่อง ก 2. เรื่อง จ.

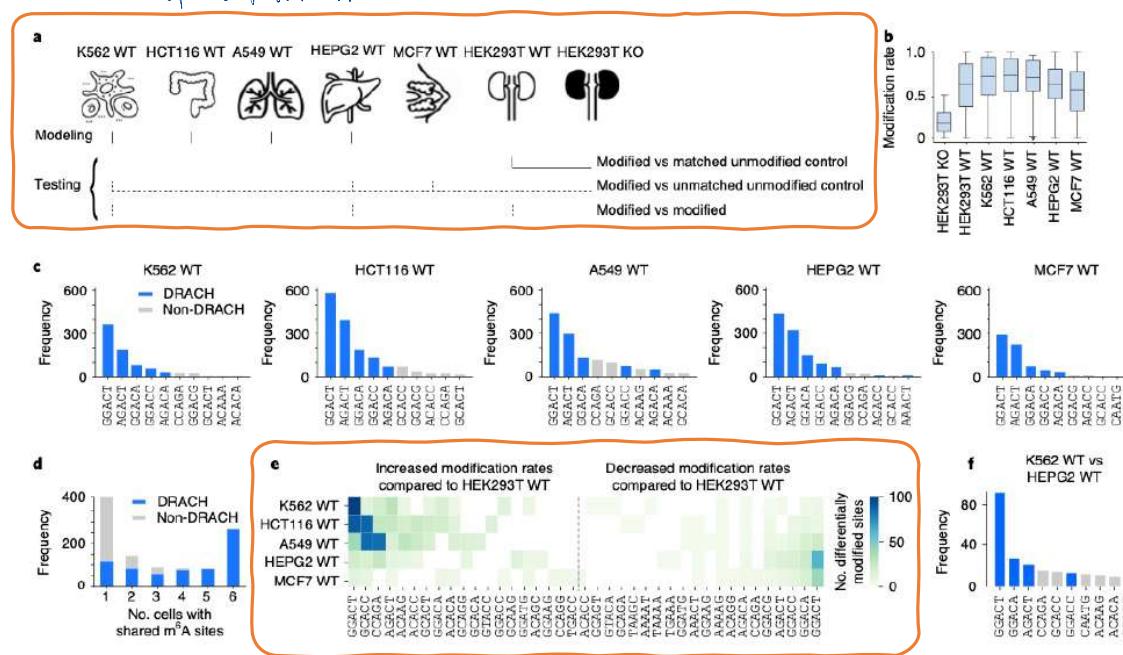


Fig. 5 | Identification of m6A sites across different tissues and cell lines.

3 Key Success to Develop AI-Powered Apps

ก. ก. ร. จ. ป. น. software ที่ 1 ต้องมีรีสอร์ส + ความยั่งยืน

1. Alignment with the actual needs ต้องอธิบายได้

2. Sufficient generalization and evaluation

3. Simple deployment and serving 便利な構造を用意する

- Online documentation
- Easy installation
- Source code
- Data availability
- Lightweight
- Fast 早い ジャンクな開発が望まない

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

ARTICLES

<https://doi.org/10.1038/nbt3872>

nature
biotechnology

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

Ploy N. Pratanwanich^{1,2,3,4*}, Fei Yao^{1,1†}, Ying Chen^{1,1}, Casslyn 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 Choo¹, Wee Joo Chng^{5,6,7}, Sarah B. Ng¹, Alexandre Thierry¹, W. S. Sho Goh^{1,9,2,6} and Jonathan Göke^{1,10,11}

Scopus metrics

78 99th percentile

Citations in Scopus

9.61

Field-Weighted citation impact

downloads 27k

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

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





CODE COMBAT

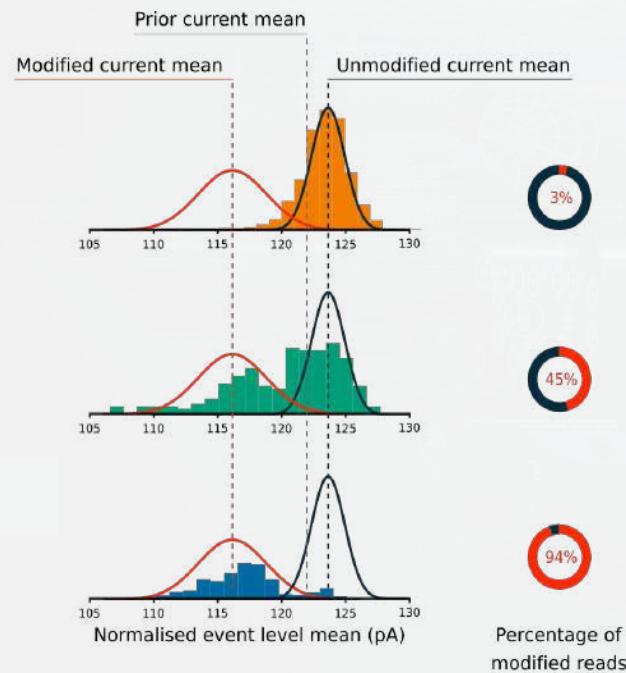
Google



E-SAN THAILAND
CODING & AI ACADEMY

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

6. Future Work ต่อไป



- Identifying the Limitations งานนี้

- Considering Changes in the Future 0.1% ลดลง software

ตัวอย่าง 1 ถ้า 0.1% ลดลง



ทาง 1 ลับเข้าไปในน้ำ

Future Work

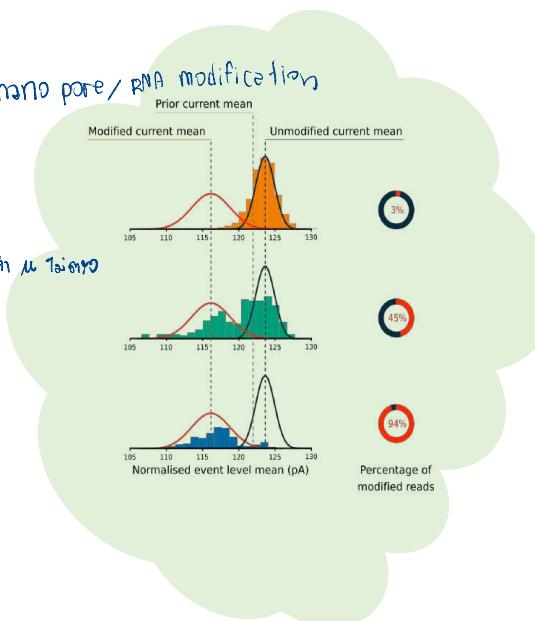
ปี 2024

สิ่งที่ต้องการ

Domain-Oriented

- m6anet • • • nature methods
↑ ตากต้องหัวกับ ร่อง กับ หัว → long tail → ค่า μ ไม่เท่ากัน
- <Gaussian> mixture model
↳ ตากต้องหัว t-dis
- Interpretability
↳ รูปแบบ distribution เกิดจาก...
- Modification or basecalled errors
- End-to-end
 - Why?
→ Nanopolish eventalign / Guppy basecaller are subject to change

อัปเดตในปีนี้ nano pore / RNA modification



Method-Oriented

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

