# IDASH PRIVACY & SECURITY WORKSHOP 2025

*supported by NHGRI R13HG012902

## Organizers:

Track 1: Arif Harmanci (UTHEALTH), Miran Kim (Hanyang), Xiaoqian Jiang (UTHEALTH)
Track 2: Tsung-Ting Kuo (Yale), Lucila Ohno-Machado (Yale)
Workshop: Lucila Ohno-Machado (Yale), Tsung-Ting Kuo (Yale)
Co-organizer: Haixu Tang (IUB)

# Keynote Talk

by Professor Judy Wawira Gichoya, MD, MS, FSIIM
Associate Professor
Interventional Radiology & Informatics
Department of Radiology & Imaging Sciences
Emory University

# Setting the Stage

# Summarized Agenda (US EST)

- 8:30 – 8:35: Registration & Safety
- 8:35 – 8:40: Welcome - Dr. Ohno-Machado
- 8:40 – 9:20: Keynote - Dr. Judy Wawira Gichoya
- 9:20 – 9:35: Setting the Stage
- 9:35 – 10:20: Track I talks
- 10:20 –10:35: Break & Group Photo
- 10:35 – 11:35: Track II talks
- 11:35 – 11:45: Closing & Box Lunch
- 11:45: Adjourn (Lunch Provided)

# Safety Plan

**1. Statement of commitment to provide a safe environment.** The iDASH Privacy & Security Workshop - Secure Genome Analysis Competition ("the iDASH workshop") is committed to provide a safe and harassment-free environment to all workshop co-organizers and participants.

**2. Expectations of behavior**. The following behaviors are considered harassing during the iDASH workshop: body or language, sexual, racial, and ethnic harassments; discrimination, insults or quid-pro-quo; violence, assaults, battery or stalking, and/or threats.

**3. Confidentially report**. For any alleged violations of the expectations of behavior, please contact the iDASH workshop organizer.

**4. Organizer actions**. The iDASH workshop organizer will assess allegations through an internal discussion, and if the allegations are valid, the organizer will intervene to stop the violations, report to the Title XI Office of the organizing university, and provide the university supporting resources to the victims.

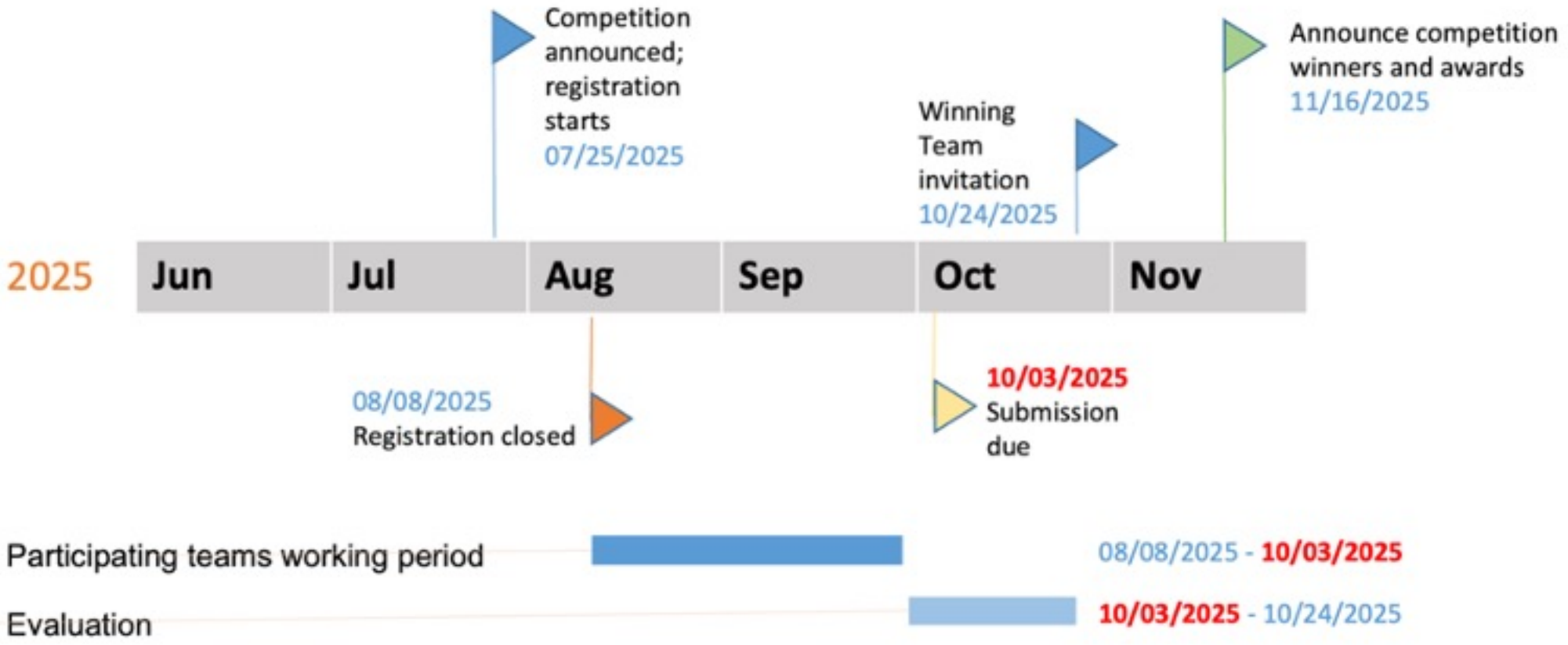
**5. Questions and related information**. Individuals who have questions, concerns or complaints related to harassment are also encouraged to contact the iDASH workshop organizer, or the Department of Health and Human Services (HHS) Office for Civil Rights (OCR).

**6. HHS OCR complaint filing**. Please refer to the following HHS OCR website for more information about how to file a civil right complaint: https://www.hhs.gov/hipaa/filing-a-complaint/index.html.

**7. Process of filing a complaint to the HHS OCR**. Filing a complaint with the iDASH workshop organizer is not required before filing a complaint of discrimination with HHS OCR. Also, seeking assistance from the iDASH workshop organizer in no way prohibits filing complaints with HHS OCR.

**8. Finding help with NIH**. Individuals can notify NIH about concerns of harassment, including sexual harassment, discrimination, and other forms of inappropriate conduct at the iDASH workshop, which is support by NIH: https://grants.nih.gov/grants/policy/harassment/find-help.htm.

# TIMELINE


Competition announced; registration starts 07/25/2025
Winning Team invitation 10/24/2025
Announce competition winners and awards 11/16/2025
2025
Jun
Jul
Aug
Sep
Oct
Nov
08/08/2025 Registration closed
10/03/2025 Submission due
Participating teams working period
08/08/2025 - 10/03/2025
Evaluation
10/03/2025 - 10/24/2025

# TWO TRACKS OF COMPETITION TASKS

- TRACK 1: Secure Evaluation of DNA binding Classification Convolutional Neural Network

- TRACK 2: Access Request Recording and Querying for Biomedical Datasets

# Acknowledgements


- **Sponsors**
  - NIH/NHGRI
- **Venue (Emory)**
  - Vaidy Sunderam, Li Xiong, Eva Stotz, Imani Jones
- **Track 1 (UTH)**
  - Luyao Chen
- **Track 2 (Yale)**
  - Lana Kareem, Winnie Ng, Juliet Lam
- **Workshop (Yale)**
  - Jennifer Koabel, Lana Kareem, Gabriella DiGioia, Frank Morale, Samantha Amidon, Robin Einbinder

National Institute of General Medical Sciences

iDASH

National Human Genome Research Institute

# Track I: Secure Evaluation of DNA binding Classification Convolutional Neural Network

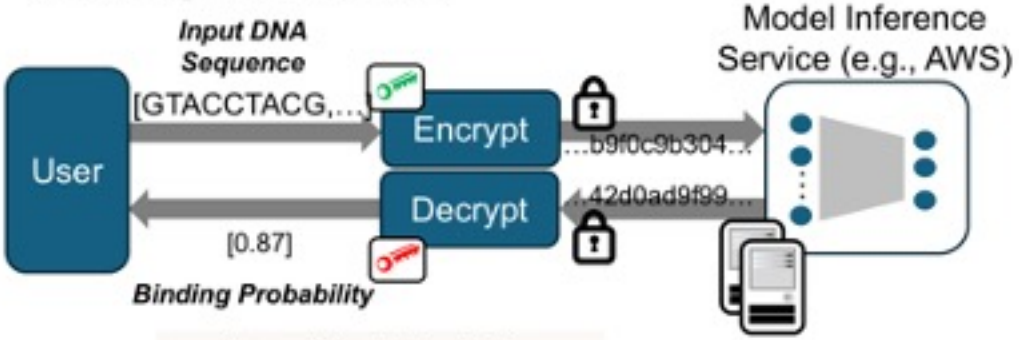
*supported by NHGRI R13HG012902

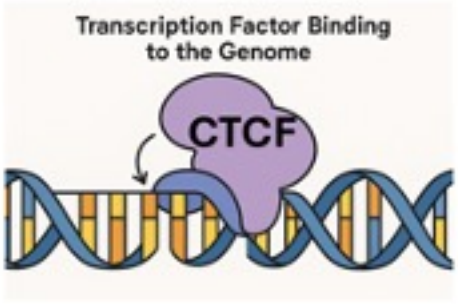
Arif Harmanci[1], Luyao Chen[1], Miran Kim[2], Xiaoqian Jiang[1]

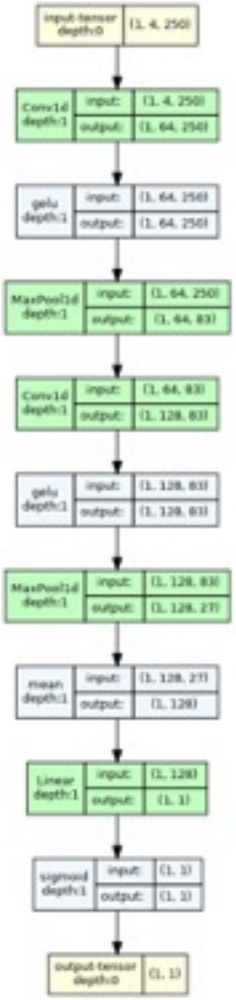
1. University of Texas Health Science Center, Houston
2. Hanyang University, Seoul, Korea

# TRACK I: Challenge Scenario

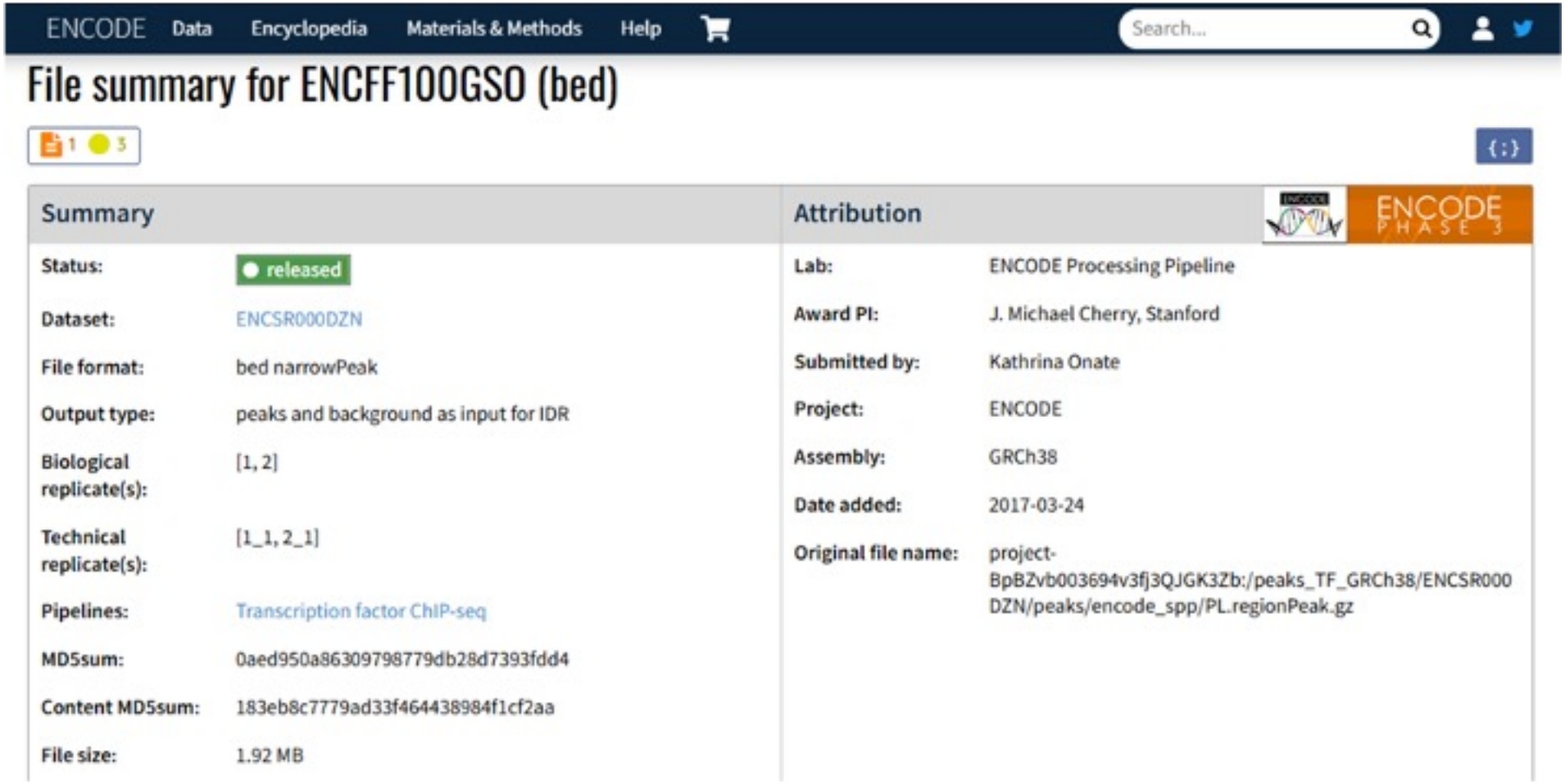
- Emergence of large models often requires model inference to be outsourced (ChatGPT, AlphaFold)
- Secure Evaluation of a Convolutional Neural Network architecture for binding prediction



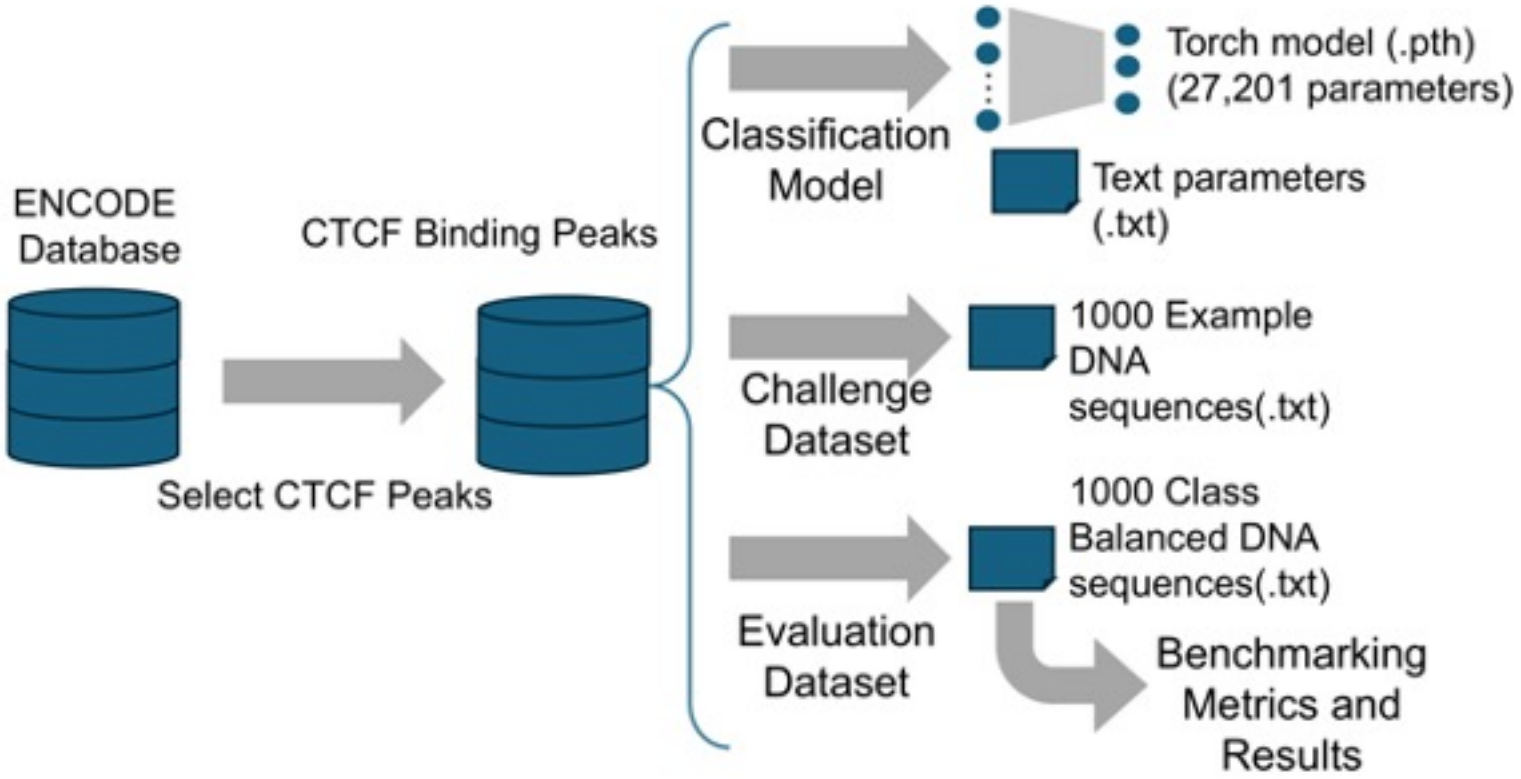
# TRACK I: Challenge Scenario



# TRACK I: Challenge Scenario

- CTCF datasets were downloaded from ENCODE portal (98,805 binding sites)
- 200-bp around the binding peak positions of top 5,000 peaks were extracted from hg38 assembly and used as positives
- To generate negatives, all peaks were shifted by 2,000 base pairs and sequences were extracted
- Classifier CNN was trained on 10,000 total positive/negative sequences with their labels (0/1)
- 10,000 pos/neg testing sequences were extracted for testing model accuracy with no overlap with training data.
- DASHCNN was trained:
  - N_epochs=50
  - Learning rate=0.001
  - Batchsize=64
- Tested max and average pooling but used average pooling to make the task easier for secure inference.

# TRACK I: Challenge Scenario



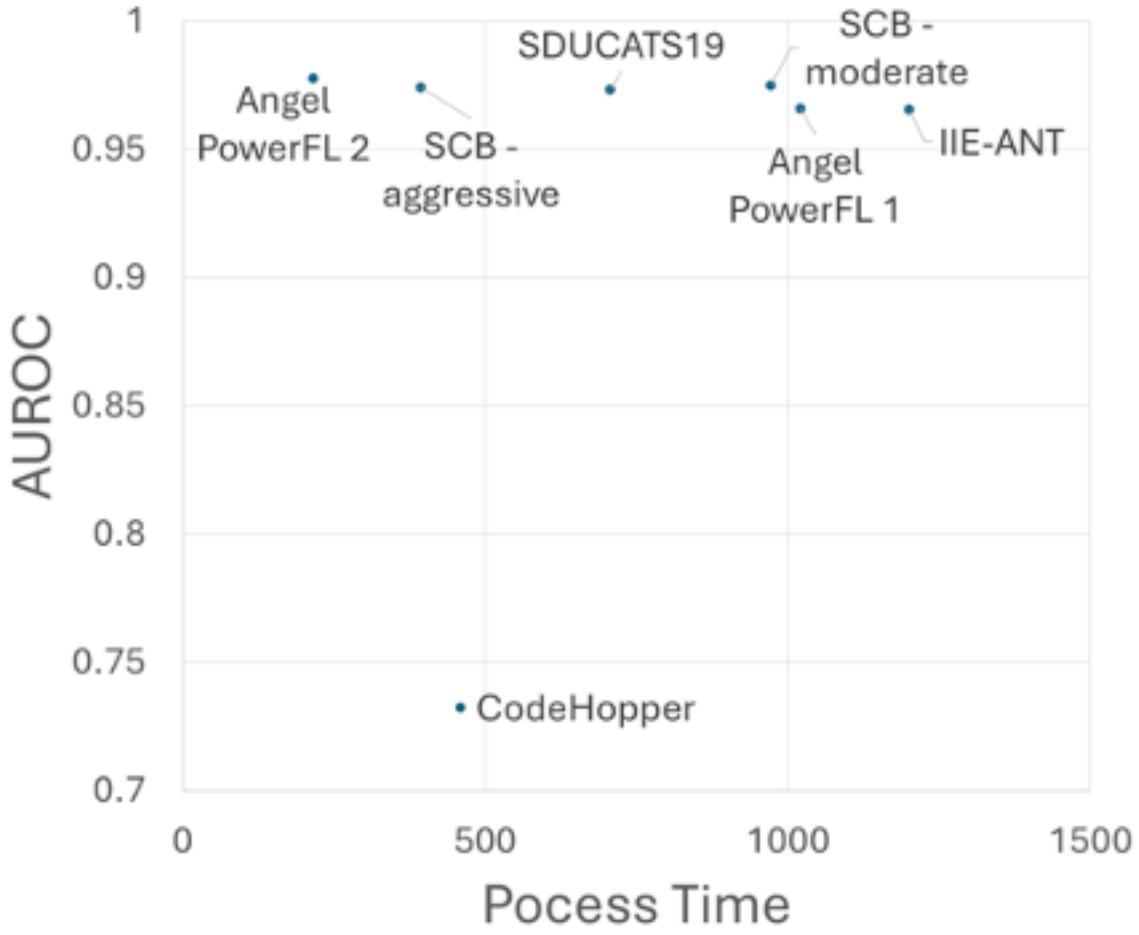
# TRACK I: Evaluation

- **Evaluation Stage:**
  - We received 8 solutions from 6 teams
  - 2000 class balanced DNA sequences from the main database were selected (1000/1000 pos/neg)
  - Each submission was run on the evaluation dataset
  - Submissions that required more than 60 minutes were disqualified
    - When we encounter errors in submissions (happens frequently), we contact the teams and resolve the errors within limits of fairness
  - You can find more information from FAQ: https://docs.google.com/document/d/121BWBgqN3jlSBjNU9994iR8fz2JWJsBf-jOEBCjRv8g/edit?tab=t.0
  - MicroAUC and Time/memory usage were recorded
- **Final Ranking Metrics:**
  - "auROC per exp(wall time per 20 min)" = (auROC / exp(total time in minutes / 20))

# TRACK I: Top Performing Teams

| TEAM | Time (Secs) | AUROC |
|---|---|---|
| Angel PowerFL 2 | 215.716 | 0.977819 |
| SCB - moderate | 972.099 | 0.974979 |
| SCB - aggressive | 393.611 | 0.97425 |
| SDUCATS19 | 706.442 | 0.973337 |
| Angel PowerFL 1 | 1020.861 | 0.966119 |
| IIE-ANT | 1200.645 | 0.96557 |
| CodeHopper | 459.309 | 0.732452 |
| Nouroudine-Curiosity | 3694.072 | 0.49815 |

# TRACK I: Top Performing Teams



# TRACK I: Top Performing Teams

| Rank | Team Name | Affiliation | Team Members |
|---|---|---|---|
| 1st | **Angel PowerFL** (Solution-2) | Tencent Inc. | **Fan Zhang[1], Jiaxuan Fu[1], Chenyu Huang[1], Huiqiang Zhao[1], Peixuan Chen[1], Danqing Huang[1]**<br>[1] Tencent |
| 2nd | **SCB@Chongqing** | Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Sansure Biotech Inc. | **Jingwei Chen[1], Wenyuan Wu[1], Li Yang[2], Yong Feng[1], Kang Wu[2], Lizhong Dai[2], Qiang He[1], Yi Yang[1], Rui Li[1], Weijie Miao[1], Xiaokang Dai[1]**<br>[1] Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences<br>[2] Sansure Biotech Inc. |
| 3rd | **SDUCATS** | Shandong University | **Honghui You[1], Kanyu Ye[1], Yusi Chen[1], Haoping Yang[1], Zhuoran Ji[1], Lei Ju[1], Meiqin Wang[1]**<br>[1]Shandong University |

# Track II: Access Request Recording and Querying for Biomedical Datasets

*supported by NHGRI R13HG012902

Tsung-Ting Kuo[1] and Lucila Ohno-Machado[1]

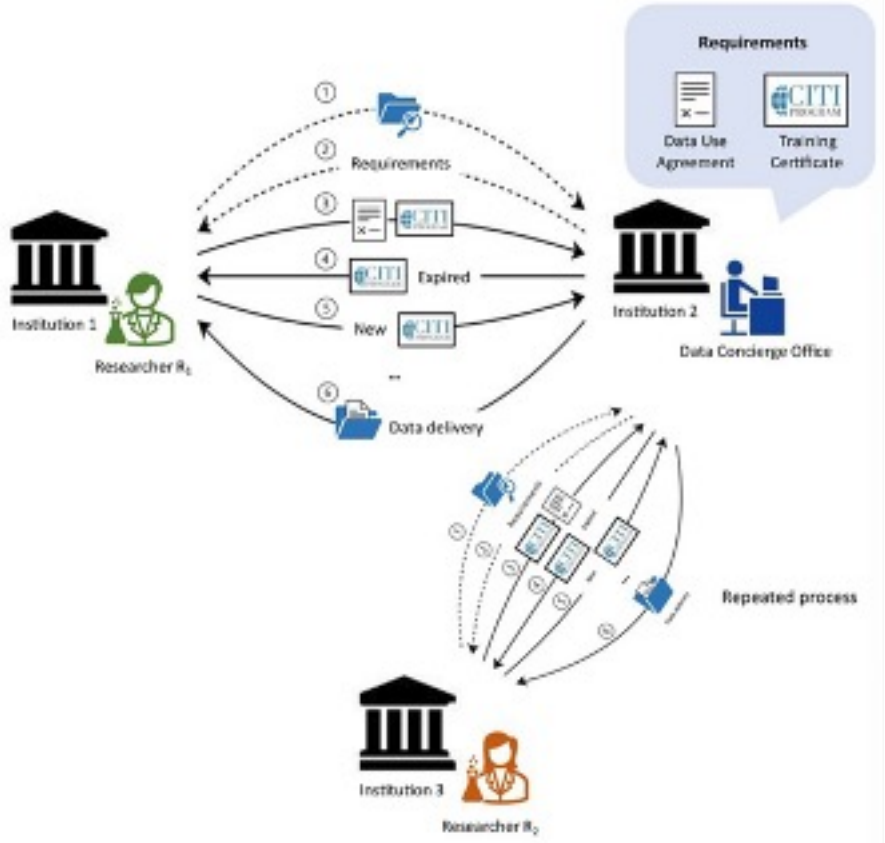
[1] Yale University

# TRACK II: GOAL

- Traditional email process for data request
  - Manual procedures
  - Time-consuming
  - Repetitive for multiple datasets

- Issue of centralized system
  - Single-point-of-failure
  - Data alternation
  - Non-transparency

- Track purpose
  - To develop blockchain-based smart contracts
  - For managing biomedical data requests
  - To facilitate the continuity/efficiency of biomedical, clinical, and genomic research




Reference: Yu Y, Edelson M, Pham A, Pekar JE, Johnson B, Post K, Kuo T-T. Distributed, immutable, and transparent biomedical limited data set request management on multi-capacity network. Journal of the American Medical Informatics Association. 2024.

# TRACK II: CHALLENGE

- Given a set of “header” info
  - Data Usage Agreement (DUA) — #, signed PIs
  - Dataset — Name, Link, DUA #, owner email, required certificates
  - Principal Investigator (PI) — Name, email, certificates and their expiration dates
- And a set of data requests
  - PI Email — e.g., erichmond16@yale.edu
  - Dataset — e.g., Data Set 200
  - Institution Name — e.g., YALE
  - Expected Answer — e.g., N
- Design a time/space efficient data structure and mechanisms
  - To manage (i.e., store and retrieve) these requests
  - Based on Ethereum Solidity

# TRACK II: CHALLENGE (CONT.)

- The input data: 250 request x 4 sites = 1,000 requests
  - Include header info and request records

- Must enforce specific requirements around dataset ownership, PI credentials, and DUA
  - Expected answers (e.g., DUA signed, valid certificate, etc.) are provided to for participants to check

- All data and intermediary data must be saved on-chain via smart contracts
  - Using the skeleton of 3 smart contracts and scripts provided by the co-organizers
  - Participants can implement any algorithm to store, retrieve and present the data correctly/efficiently

- One storage function
  - Required to allow the storage of one record at a time
  - Each participant can determine how each record is represented and saved in the smart contract

- One query function
  - Users should be able to query the data from any of the blockchain nodes

# TRACK II: EVALUATION CRITERIA

- Test data
  - Split into 4 parts to simulate 4 sites
- Storage time evaluation
  - Stored 250 records simultaneously from the 4 nodes (1,000 records total)
  - Measured the average time per-records inserted at each node
- Query time evaluation
  - The solution needs to demonstrate 100% accurate query results
  - Measuring average time per-query for both functions
- Final score
  - 50% for the insertion and 50% for query
  - Averaged from 5 trials

# TRACK II: Teams Completing the Task

**Among 5 participating teams, 4 completed the task**

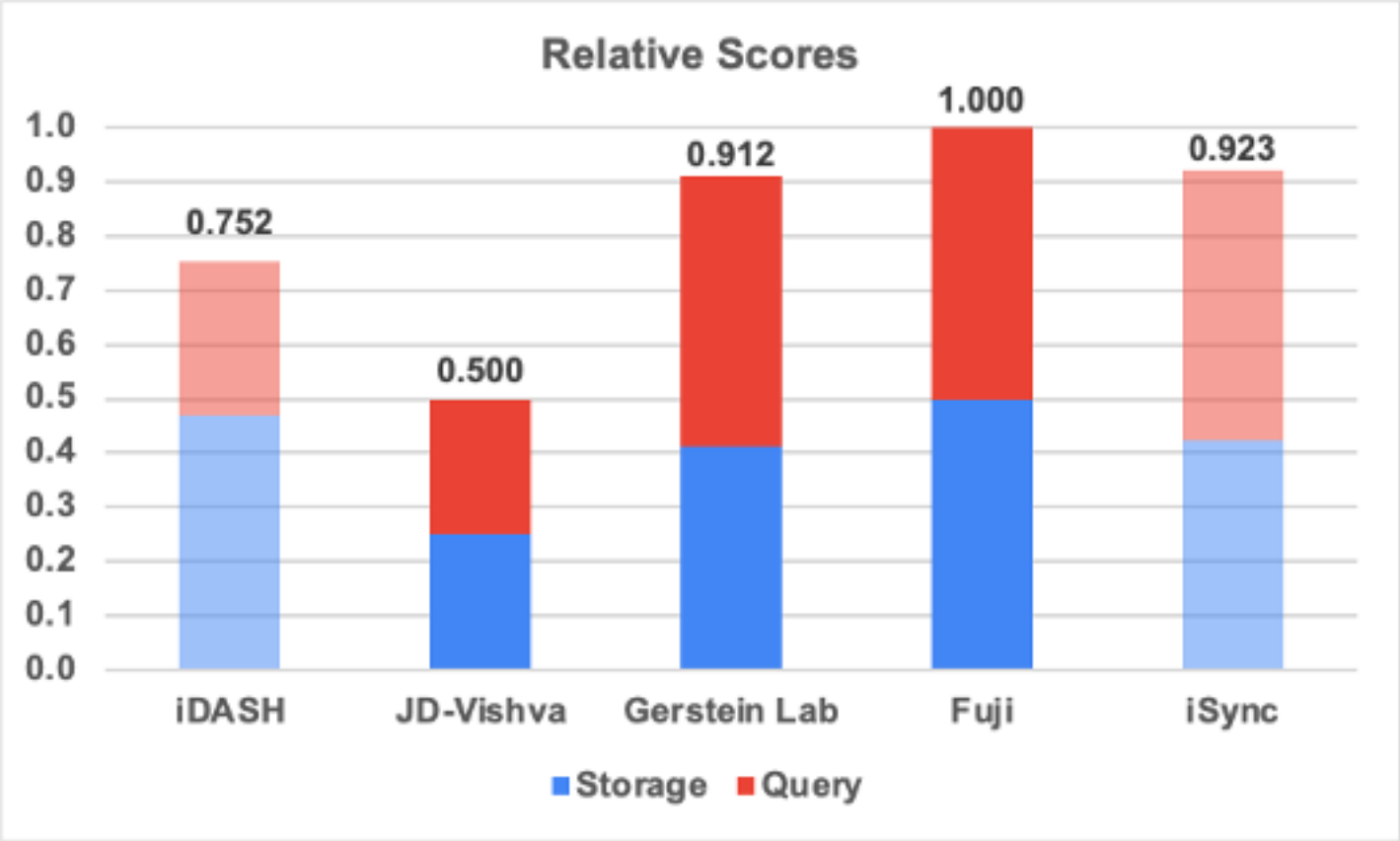
| Team | Affiliation |
|---|---|
| **iDASH** | **Yale University** |
| **JD-Vishva** | **JD Technology Information Technology Co., Ltd** |
| **Gerstein Lab** | **Yale University** |
| **Fuji** | **University of Massachusetts Amherst** |
| **iSync** | **Yale University** |

Co-organized with Lana Kareem, Winnie Ng, and Juliet Lam

# TRACK II: EVALUATION PERFORMANCE (1/2)

| Team | Storage Time (Seconds) | Query Time (Milliseconds) |
|---|---|---|
| iDASH | 2.088 | 61.305 |
| JD-Vishva | 2.111 | 70.635 |
| Gerstein Lab | 2.094 | 1.245 |
| Fuji | 2.084 | 1.310 |
| iSync | 2.092 | 1.690 |

# TRACK II: EVALUATION PERFORMANCE (2/2)



# TRACK II: WINNING TEAMS

| Rank | Team | Member(s) |
|---|---|---|
| **1st Place** | **Fuji** | Arisa Tajima (University of Massachusetts Amherst) |
| **2nd Place** | **Gerstein Lab** | Eric Ni (Yale University)<br>Mark Gerstein (Yale University) |
| **3rd Place** | **JD-Vishva** | Mengqing Yang (JD Technology Information Technology Co., Ltd)<br>Zhongwei Sun (JD Technology Information Technology Co., Ltd)<br>Donglai Jiang (JD Technology Information Technology Co., Ltd) |

# Track I Presentations

Angel_PowerFL

SCB@Chongqing

SDUCATS

# Track II Presentations

JD-Vishva

Gerstein Lab

Fuji

iSync

# Next Steps

- Confidential computing  track - planning to come back in 2026!

# Workshop evaluation