

# Lost in Translation: Harnessing the Power of Nanopore Electrical Signals in Genomics

Arch4Health Workshop (in conj. with HPCA 2026), Sydney, Australia  
January 31, 2026

**Can Firtina** | [canfirtina@gmail.com](mailto:canfirtina@gmail.com) | [cs.umd.edu/~firtina](https://cs.umd.edu/~firtina)

Assistant Professor of Computer Science at UMD



DEPARTMENT OF  
COMPUTER SCIENCE

# Analyzing Genomes Reveals Key Insights



Personalized Medicine



Altering genomes to solve fundamental challenges of life

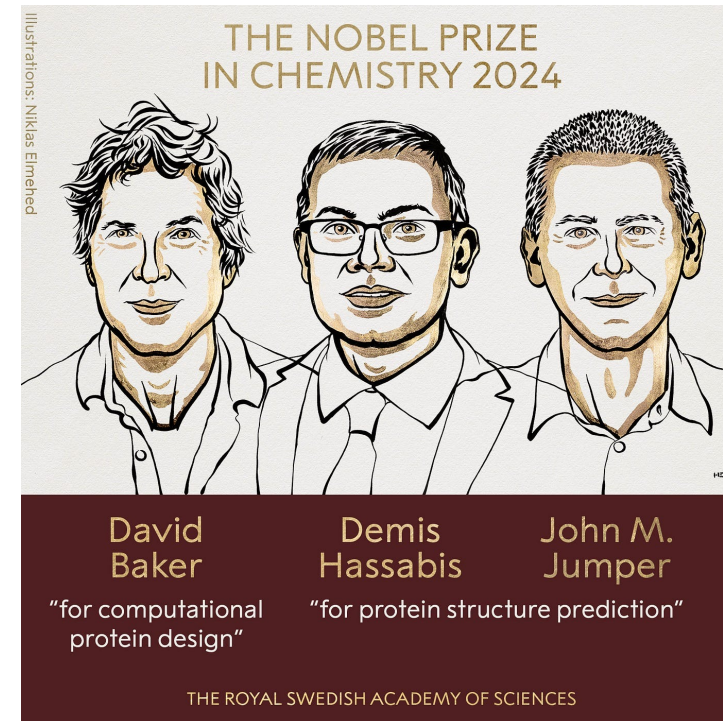
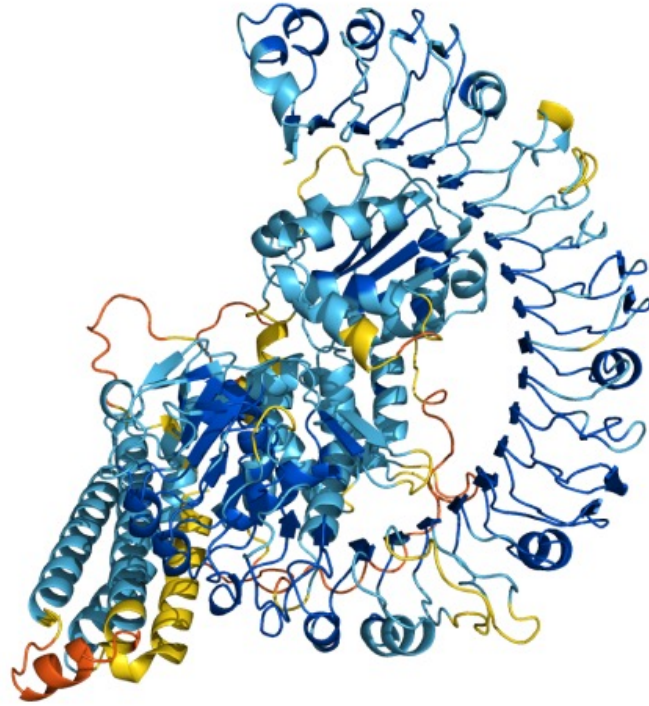


Detecting **pathogens** in the environment



Rapid surveillance of **disease outbreaks**

# Protein Structure Prediction with AI



<https://www.nobelprize.org/prizes/chemistry/2024/press-release/>

# Genome Editing



**The Nobel Prize in Chemistry 2020**  
awarded "for the development of a  
method of genome editing"



<https://www.nobelprize.org/prizes/chemistry/2020/press-release/>

# Population-Scale Analysis



According to the results provided by Aldy, in the Vietnamese population, \*36 + \*10 and \*10 were the two most common *CYP2D6* haplotypes, both known for their decreased function in drug metabolism. In addition to this, there were two non-functional haplotypes, \*36 and \*5 appeared at lower frequencies, that were identified in this dataset. Among the 55 haplotypes detected in this genomic dataset, 26 were identified as SVs, with 22 of these SVs being tandem arrangements, highlighting the significant complexity and diversity of the *CYP2D6* genotype within the Vietnamese population. Furthermore, 40 participants carrying haplotypes recorded an activity score (AS) that had been marked as not available (N/A) in the CPIC guidelines.

## An efficient genotyper and star-allele caller for pharmacogenomics

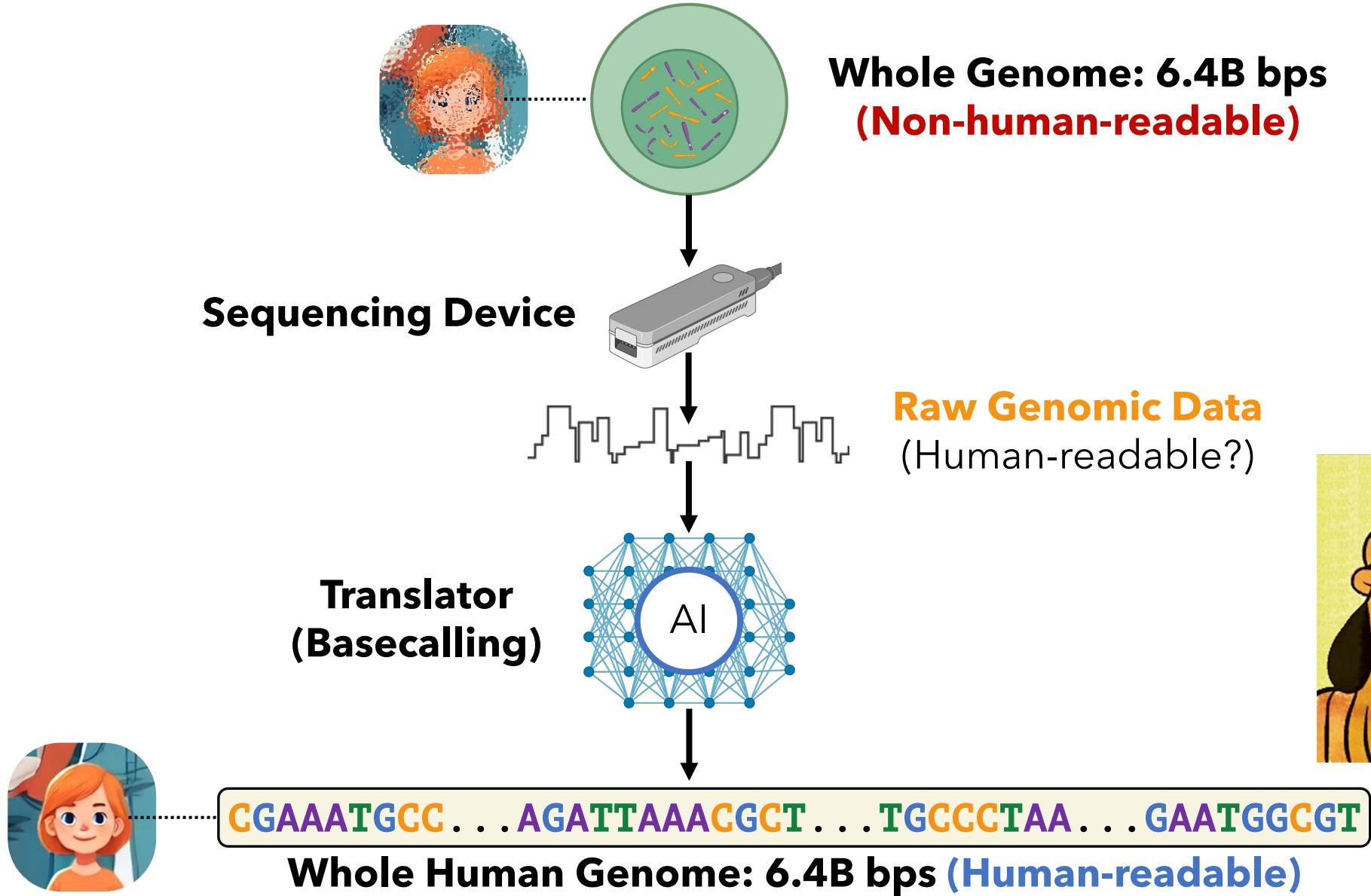
Ananth Hari,<sup>1,2</sup> Qinghui Zhou,<sup>3</sup> Nina Gonzaludo,<sup>4</sup> John Harting,<sup>4</sup> Stuart A. Scott,<sup>5</sup> Xiang Qin,<sup>6</sup> Steve Scherer,<sup>6</sup> S. Cenk Sahinalp,<sup>2</sup> and Ibrahim Numanagić<sup>3</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742, USA; <sup>2</sup>Cancer Data Science Laboratory, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; <sup>3</sup>Department of Computer Science, University of Victoria, Victoria, British Columbia V8P 5C2, Canada; <sup>4</sup>Pacific Biosciences, Menlo Park, California 94025, USA; <sup>5</sup>Department of Pathology, Stanford University, Palo Alto, California 94304, USA; <sup>6</sup>Baylor College of Medicine Human Genome Sequencing Center, Houston, Texas 77030, USA

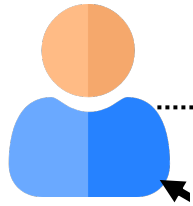
Nguyen+, "A comprehensive assessment of pharmacogenomic annotation tools for next-generation sequencing data: an emphasis on *cyp2d6* and vietnamese genomic data." *Journal of Human Genetics* 2025

Hari+, "An efficient genotyper and star-allele caller for pharmacogenomics", *Genome Research* 2023.

# Generating the Entire Genomic Sequence



# Genomic Data: Giant Jigsaw Puzzle to Solve



**Human (Reference) Genome:** 3.2 Billion bps (haploid)

CGAAATGCC...AGATTAAACGGCT...TGCCCTAA...GAATGGCGT



CGAAATGCC...AGATTAAATGGT...TGCCCTAA...GAATGGCGT

**Entire Genomic Sequence of an Organism**



**Small pieces (reads)**



**Unknown origins**



**Non-identical reference genome**

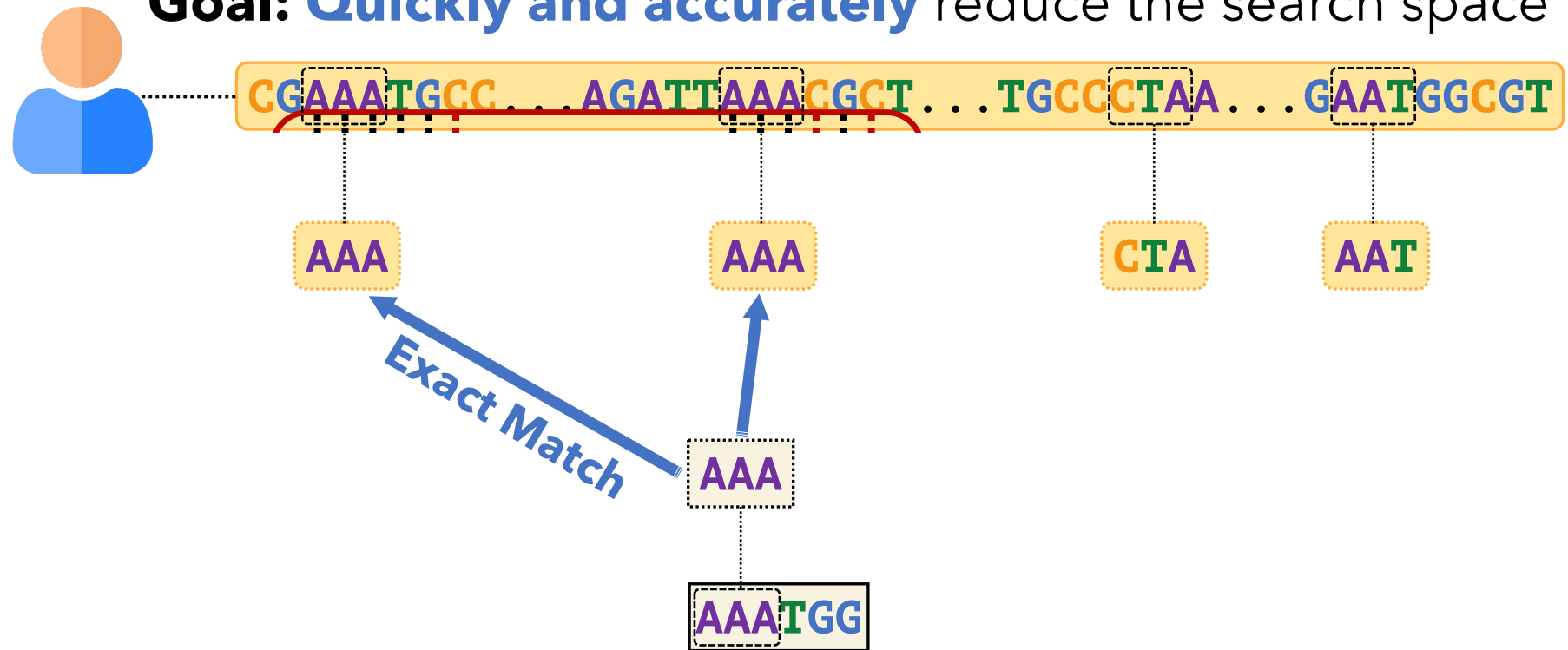


**Large volume of data to analyze**

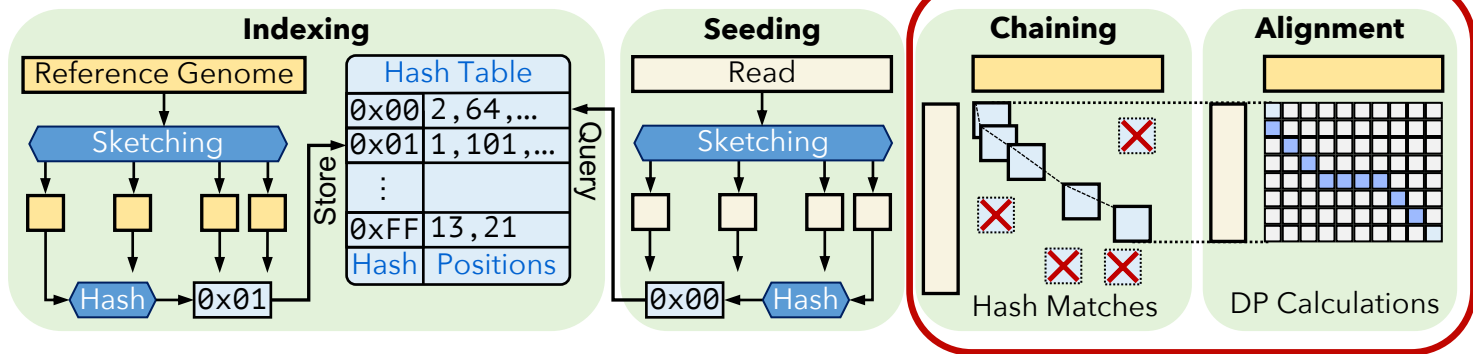


# Solving the Puzzle: The Practical Way

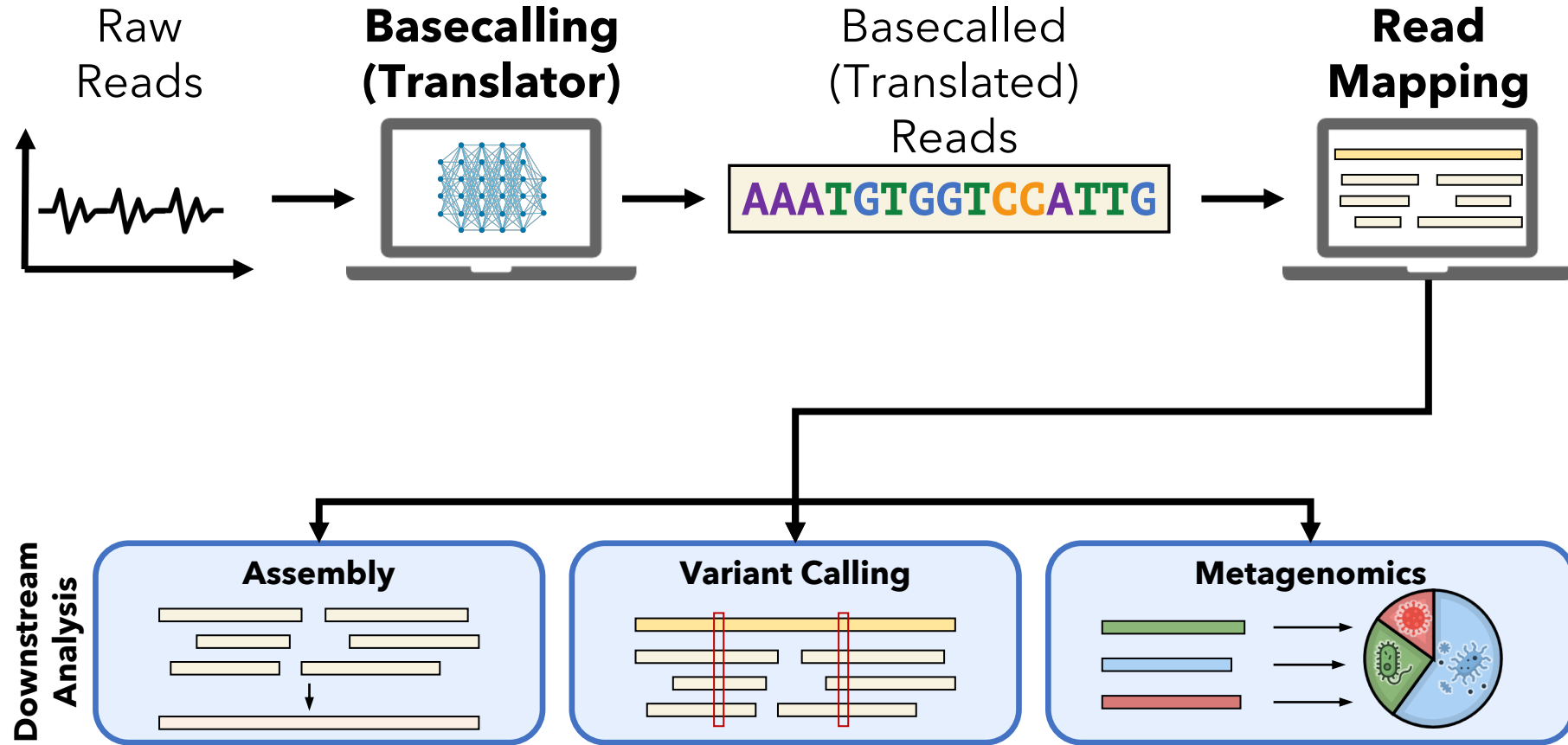
**Goal:** Quickly and accurately reduce the search space



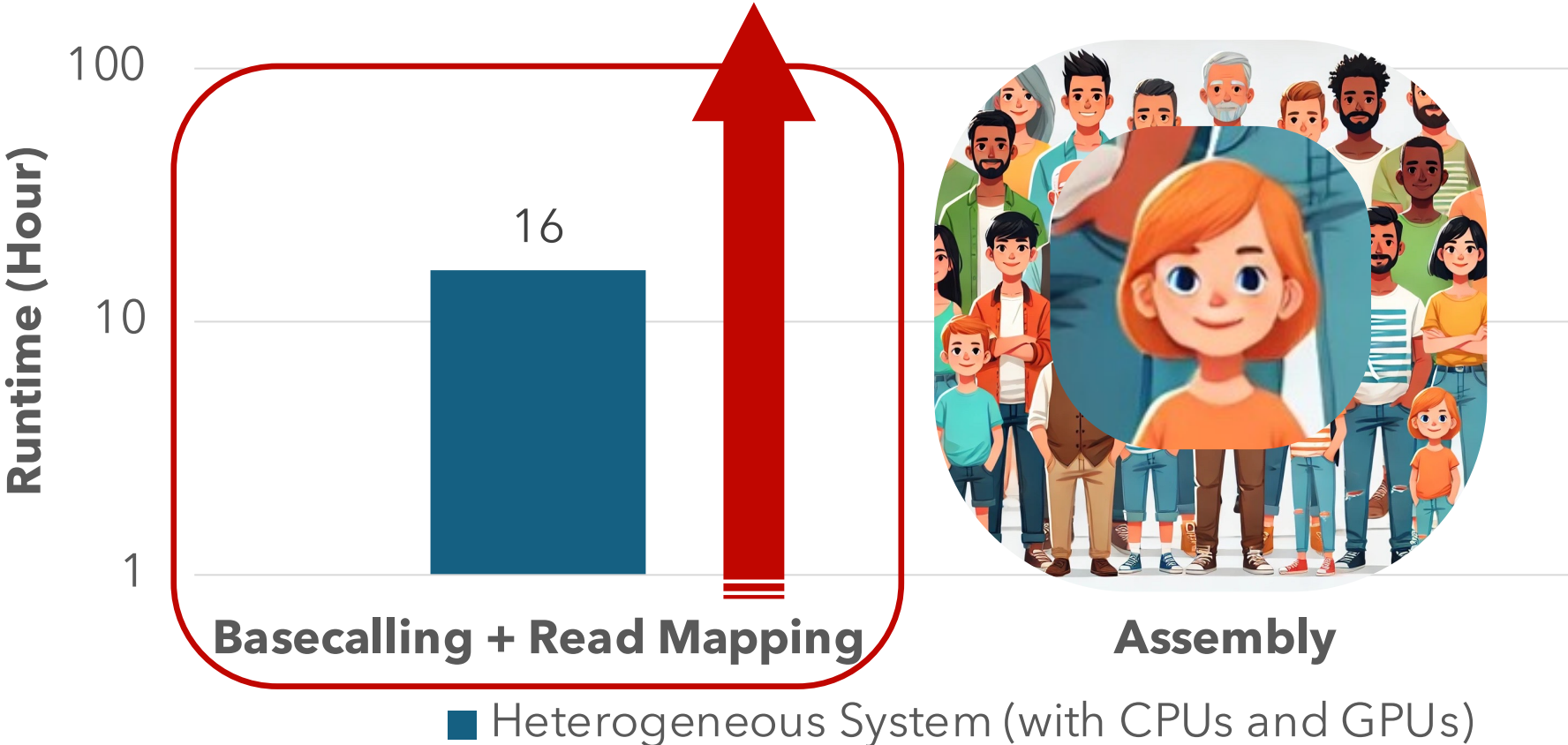
## Read Mapping



# A Common Genome Analysis Pipeline



# Genome Analysis is Computationally Costly



 **Significant computation and energy overhead**

# Genome Analysis Pipeline is Unbalanced

## Short Reads (Illumina)

<b>1 Sequencing</b> Library preparation: 6.5 hours Sequencing: 68.2 Gb/hour	<b>2 Basecalling</b> 104.4 Gb/hour	<b>3 Quality Control</b> 1339.2 Gb/hour	<b>4 Read Mapping</b> 0.2 Gb/hour	<b>5 Variant Calling</b> 1.2 Gb/hour
---	---------------------------------------	--	--------------------------------------	---

## Ultra-long Reads (ONT)

<b>1 Sequencing</b> Library preparation: 24 hours Sequencing: 4.1 Gb/hour	<b>2 Basecalling</b> 0.833 Gb/hour	<b>3 Quality Control</b> 3420 Gb/hour	<b>4 Read Mapping</b> 1.7 Gb/hour	<b>5 Variant Calling</b> 0.044 Gb/hour
---	---------------------------------------	--	--------------------------------------	---

## Accurate Long Reads (PacBio)

<b>1 Sequencing</b> Library preparation: 24 hours Sequencing: 5.3 Gb/hour	<b>2 Basecalling</b> 8.3 Gb/hour	<b>3 Quality Control</b> 1081 Gb/hour	<b>4 Read Mapping</b> 1.4 Gb/hour	<b>5 Variant Calling</b> 1.1 Gb/hour
---	-------------------------------------	--	--------------------------------------	---

# Limited Application Scope and Accuracy



**Energy-efficient analysis for  
Resource-constrained devices**

**Urgent analysis  
within minutes**



**Large-scale analysis  
without sacrificing accuracy  
and performance**



# STORM Group

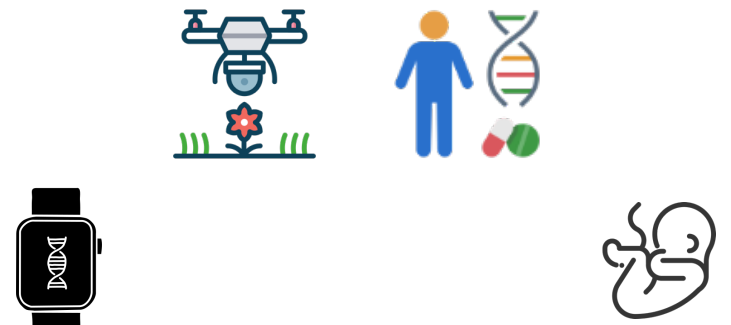
## Research Directions

 **Real-Time and Portable**  
Genomic Signal Analysis

 **Data-centric and AI-Driven** Solutions to Analyze Genomic Data

 Accurate and Fast Algorithms Scalable to **Large Volumes of Genomic Data**

## Today's Talk



## Goal

Enable fundamentally better integration of biological data analysis in our **everyday lives**



# New Frontiers: Raw Signal Analysis [ISMB '23]

- [Can Firtina](#), Nika Mansouri Ghiasi, Joel Lindegger, Gagandeep Singh, Meryem Banu Cavlak, Haiyu Mao, and Onur Mutlu,

## "RawHash: enabling fast and accurate real-time analysis of raw nanopore signals for large genomes"

Proceedings of the [31st Annual Conference on Intelligent Systems for Molecular Biology \(ISMB 2023\)](#) and the [22nd European Conference on Computational Biology \(ECCB 2023\)](#), Lyon, France, July 2023.

Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

Conference Talk [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)

Preprint (bioRxiv) [\[online\]](#) [\[pdf\]](#)

Social Media Thread [\[Twitter \(X\)\]](#)

*Bioinformatics*, 2023, **39**, i297–i307  
<https://doi.org/10.1093/bioinformatics/btad272>  
ISMB/ECCB 2023



---

## RawHash: enabling fast and accurate real-time analysis of raw nanopore signals for large genomes

Can Firtina <sup>1,\*</sup>, Nika Mansouri Ghiasi <sup>1</sup>, Joel Lindegger <sup>1</sup>, Gagandeep Singh <sup>1</sup>,  
Meryem Banu Cavlak <sup>1</sup>, Haiyu Mao <sup>1</sup>, Onur Mutlu <sup>1,\*</sup>

<sup>1</sup>Department of Information Technology and Electrical Engineering, ETH Zurich, 8092 Zurich, Switzerland

\*Corresponding author. Department of Information Technology and Electrical Engineering, ETH Zurich, Gloriastrasse 35, 8092 Zurich, Switzerland.  
E-mail: [firtinac@ethz.ch](mailto:firtinac@ethz.ch) (C.F.), [omutlu@ethz.ch](mailto:omutlu@ethz.ch) (O.M.)

# Real-Time Raw Signal Analysis [Bioinformatics '24]

- [Can Firtina](#), Melina Soysal, Joël Lindegger, and Onur Mutlu, **"RawHash2: Mapping Raw Nanopore Signals Using Hash-Based Seeding and Adaptive Quantization"** [Bioinformatics](#), July 2024.  
Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

*Bioinformatics*, 2024, **40(8)**, btae478  
<https://doi.org/10.1093/bioinformatics/btae478>  
Advance Access Publication Date: 30 July 2024  
**Applications Note**



---

Sequence analysis

## RawHash2: mapping raw nanopore signals using hash-based seeding and adaptive quantization

Can Firtina <sup>1,\*</sup>, Melina Soysal <sup>1</sup>, Joël Lindegger <sup>1</sup>, Onur Mutlu <sup>1,\*</sup>

<sup>1</sup>Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich 8092, Switzerland

\*Corresponding authors. Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich 8092, Switzerland.  
E-mail: [firtinac@ethz.ch](mailto:firtinac@ethz.ch) (C.F.); [omutlu@ethz.ch](mailto:omutlu@ethz.ch) (O.M.)

# New Directions: Assembling Raw Signals

- [Can Firtina](#), Maximilian Mordig, Harun Mustafa, Sayan Goswami, Nika Mansouri Ghiasi, Stefano Mercogliano, Furkan Eris, Joel Lindegger, Andre Kahles, and Onur Mutlu, **"Rawsamble: Overlapping and Assembling Raw Nanopore Signals using a Hash-based Seeding Mechanism"**  
[arXiv](#), October 2024.  
Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)  
[ISMB 2024](#) Talk [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)  
[CSHL 2024 \(Biological Data Science\)](#) Talk [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)  
Social Media Thread [\[Twitter \(X\)\]](#) [\[LinkedIn\]](#)

## **Rawsamble: Overlapping and Assembling Raw Nanopore Signals using a Hash-based Seeding Mechanism**

Can Firtina<sup>1</sup> Maximilian Mordig<sup>1,2</sup> Harun Mustafa<sup>1,3,4</sup> Sayan Goswami<sup>1</sup> Nika Mansouri Ghiasi<sup>1</sup>  
Stefano Mercogliano<sup>1</sup> Furkan Eris<sup>1</sup> Joël Lindegger<sup>1</sup> Andre Kahles<sup>1,3,4</sup> Onur Mutlu<sup>1</sup>  
<sup>1</sup>*ETH Zurich*      <sup>2</sup>*Max Planck Institute for Intelligent Systems*  
<sup>3</sup>*University Hospital Zurich*      <sup>4</sup>*Swiss Institute of Bioinformatics*

# Quickly Aligning Raw Signals [IEEE Access '24]

- Joël Lindegger, Can Firtina, Nika Mansouri Ghiasi, Mohammad Sadrosadati, Mohammed Alser, and Onur Mutlu,

## "RawAlign: Accurate, Fast, and Scalable Raw Nanopore Signal Mapping via Combining Seeding and Alignment"

[IEEE Access](#), December 2024.

Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

## RawAlign: Accurate, Fast, and Scalable Raw Nanopore Signal Mapping via Combining Seeding and Alignment

**JOËL LINDEGGER<sup>ID</sup>, CAN FIRTINA<sup>ID</sup>, NIKA MANSOURI GHIASI, MOHAMMAD SADROSADATI<sup>ID</sup>,  
MOHAMMED ALSER<sup>ID</sup>, AND ONUR MUTLU<sup>ID</sup>, (Fellow, IEEE)**

Department of Information Technology and Electrical Engineering, ETH Zürich, 8092 Zürich, Switzerland

Corresponding authors: Joël Lindegger (jmlindegger@gmail.com), Can Firtina (firtinac@ethz.ch), and Onur Mutlu (omutlu@ethz.ch)

This work was supported in part by the Semiconductor Research Corporation (SRC), in part by the ETH Future Computing Laboratory, in part by European Union's Horizon Programme for Research and Innovation under Grant 101047160-BioPIM, and in part by the Swiss National Science Foundation (SNSF) under Grant 200021\_213084.

# Benchmarking Raw Signal Analysis [ACM-BCB '25]

- Furkan Eris, Ulysse McConnell, Can Firtina, and Onur Mutlu,  
**"RawBench: A Comprehensive Benchmarking Framework  
for Raw Nanopore Signal Analysis Techniques"**

Proceedings of the [16th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics \(ACM-BCB 2025\)](#), Philadelphia, PA, USA, October 2025.

Paper [\[online\]](#) [\[pdf\]](#)

Conference Talk [\[pptx\]](#) [\[pdf\]](#)

## RawBench: A Comprehensive Benchmarking Framework for Raw Nanopore Signal Analysis Techniques

Furkan Eris  
ETH Zurich  
Zurich, Switzerland  
furkan6olm@gmail.com

Ulysse McConnell  
ETH Zurich  
Zurich, Switzerland  
umcconnell@student.ethz.ch

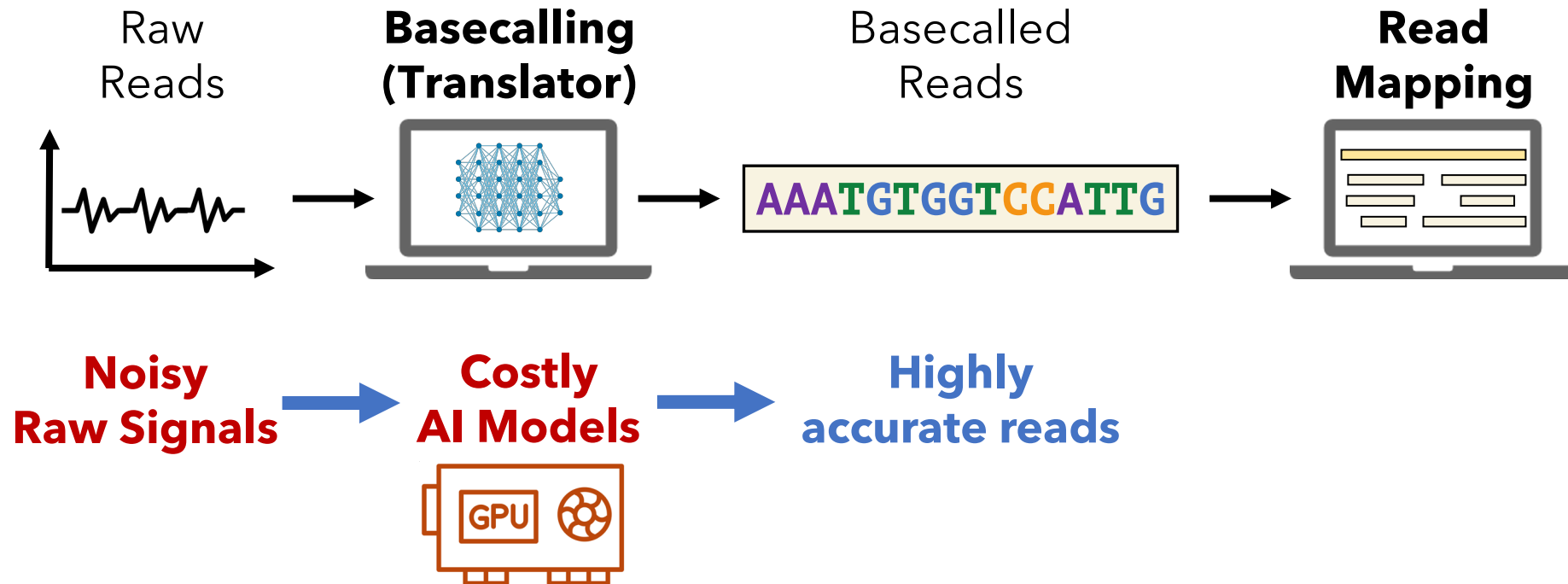
Can Firtina\*  
University of Maryland, College Park  
MD, USA  
firtina@umd.edu

Onur Mutlu\*  
ETH Zurich  
Zurich, Switzerland  
omutlu@gmail.com

# Basecalling is Accurate yet Costly

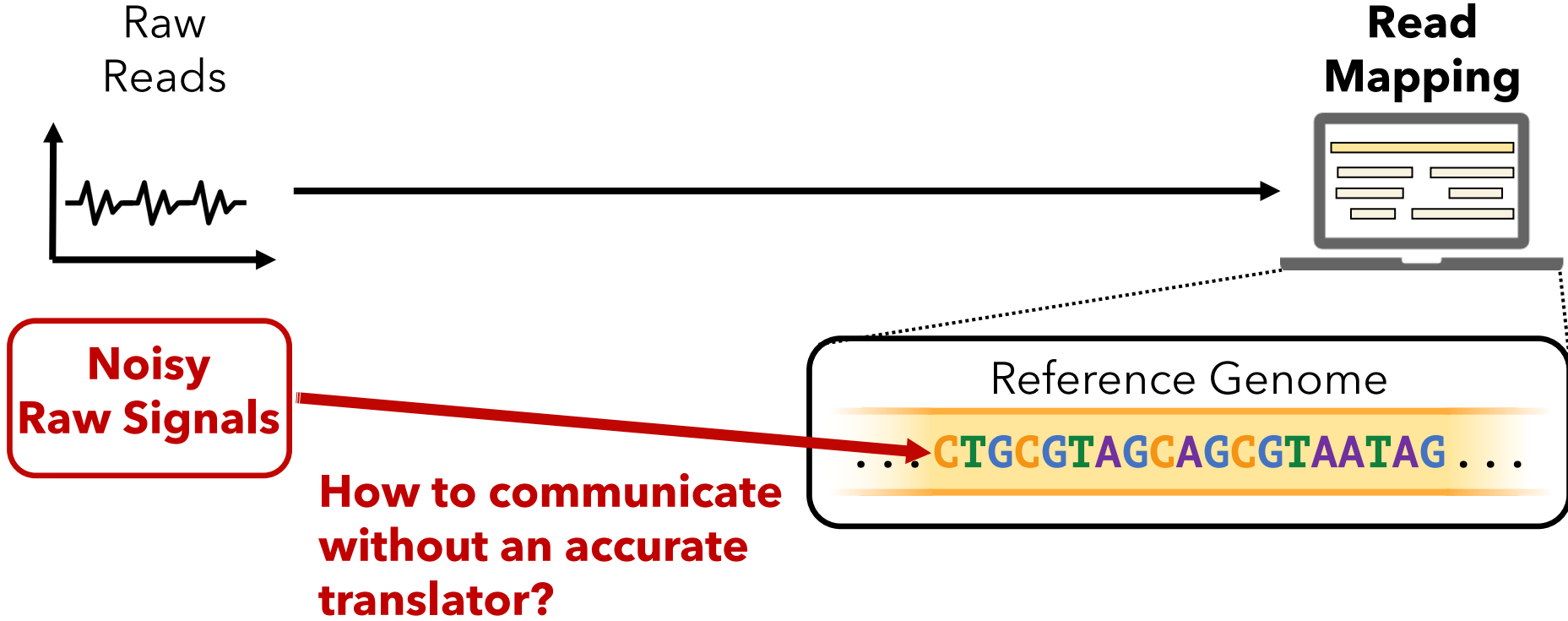


Significant computation and energy overhead

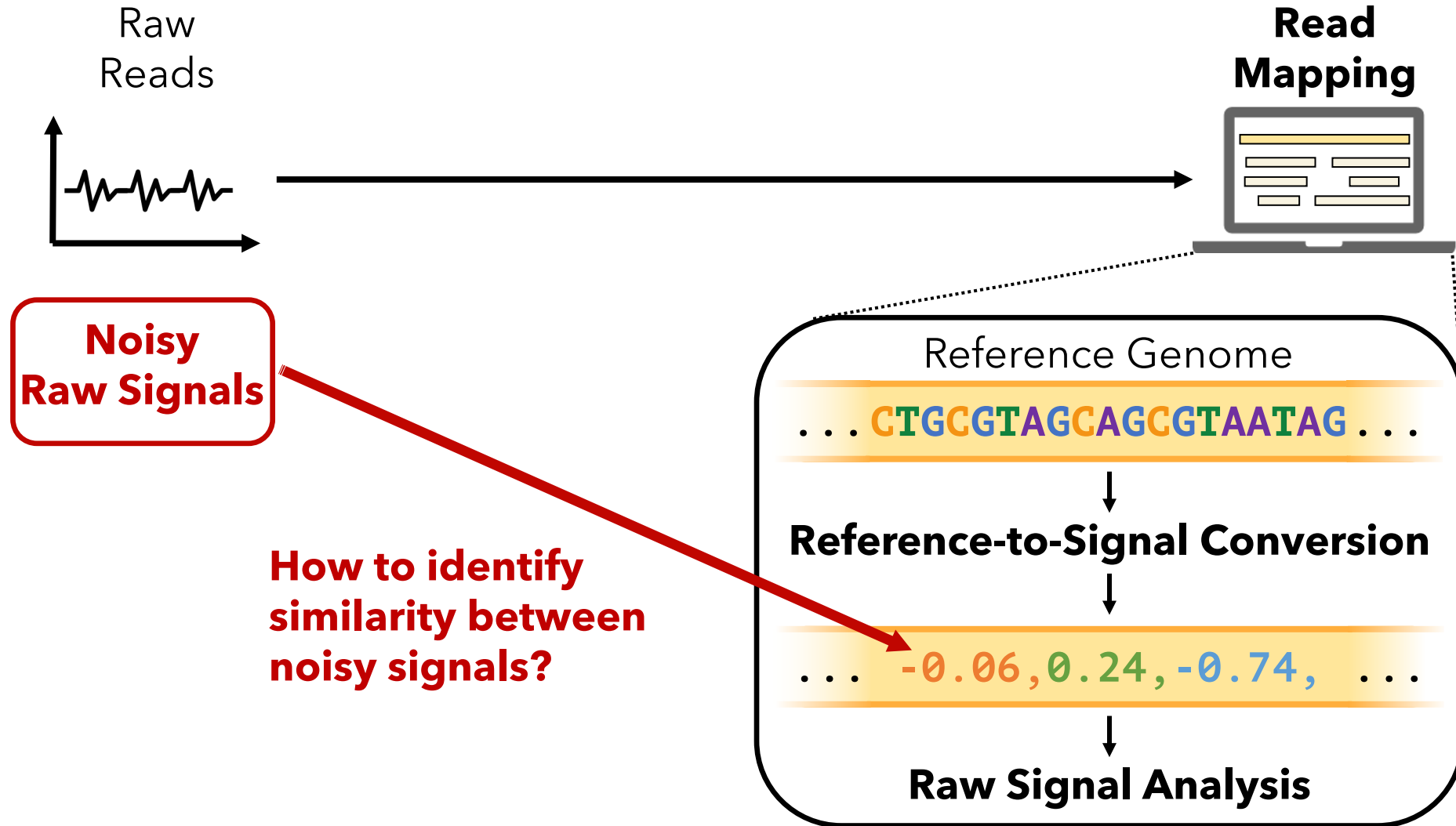


How to make this pipeline cheaper?

# Can We Survive Without Translation?



# Can We Survive With Noisy Signal Analysis?



# Noise in Raw Electrical Signals

**Observation:** Identical molecules generate similar raw signals

Sequencing **CTGCGTAGCA**  
Raw Signals:  $[-0.06, -0.22, -0.75, 0.96, -0.49]$

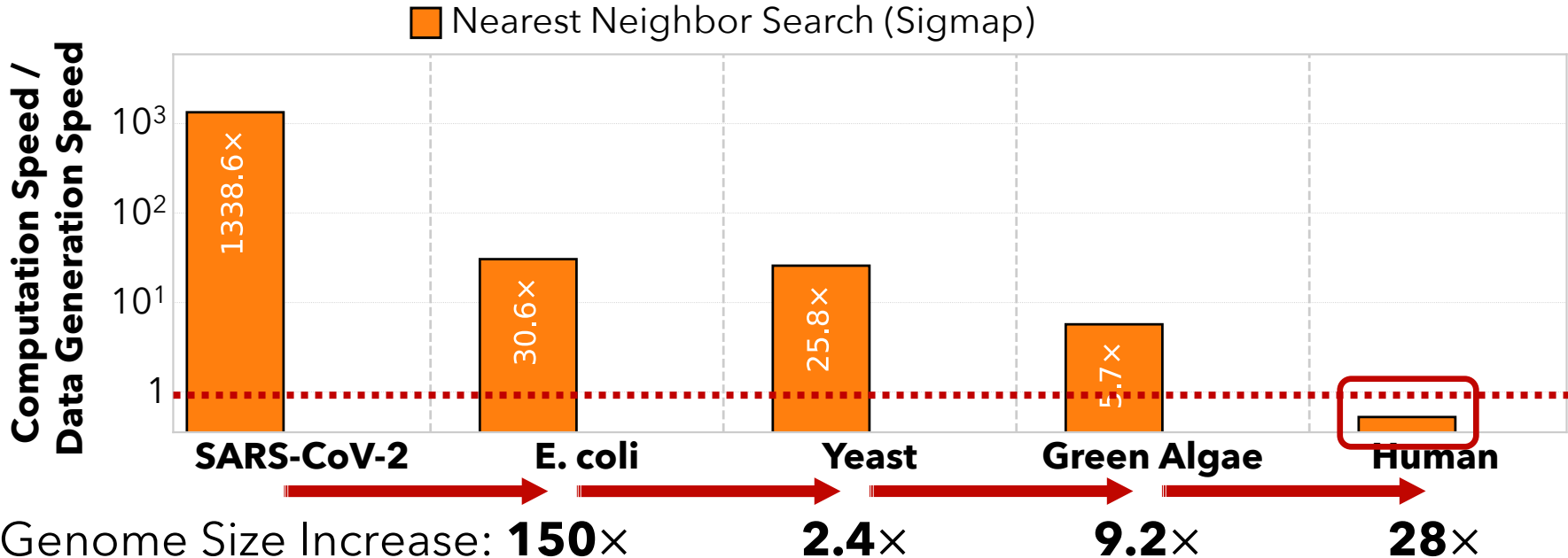
**Key Idea:** Perform similarity estimation based on **vector similarity search**  
**Nearest Neighbor Search**

Raw Signals:  $[-0.05, -0.21, -0.76, 1.01, -0.51]$   
Sequencing **CTGCGTAGCA**



**Problem: Too Costly for Large Genomes**

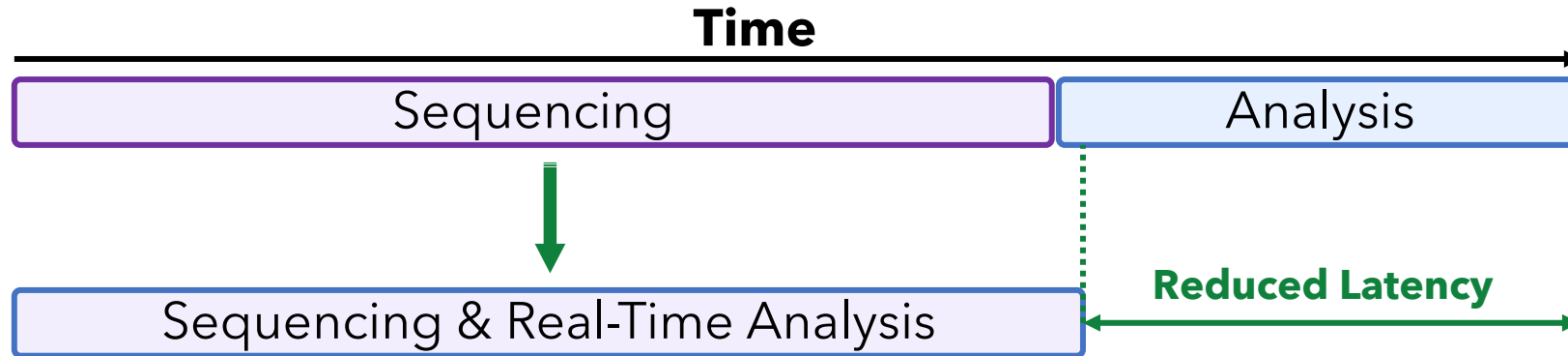
# Scalability Issues Limit Real-Time Analysis



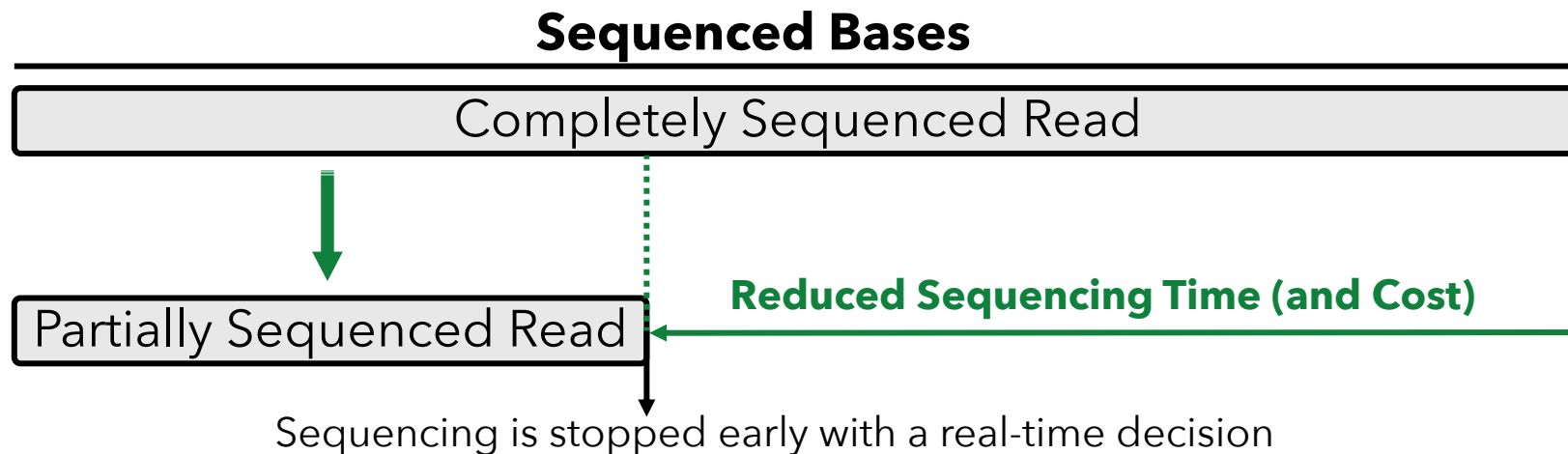
**No Real-Time Analysis for Larger Genomes:**  
Searching mechanism does not scale well

# Benefits of Real-Time Analysis

- ✓ **Reducing latency** by overlapping analysis with sequencing



- ✓ **Reducing sequencing time and cost** by stopping sequencing early

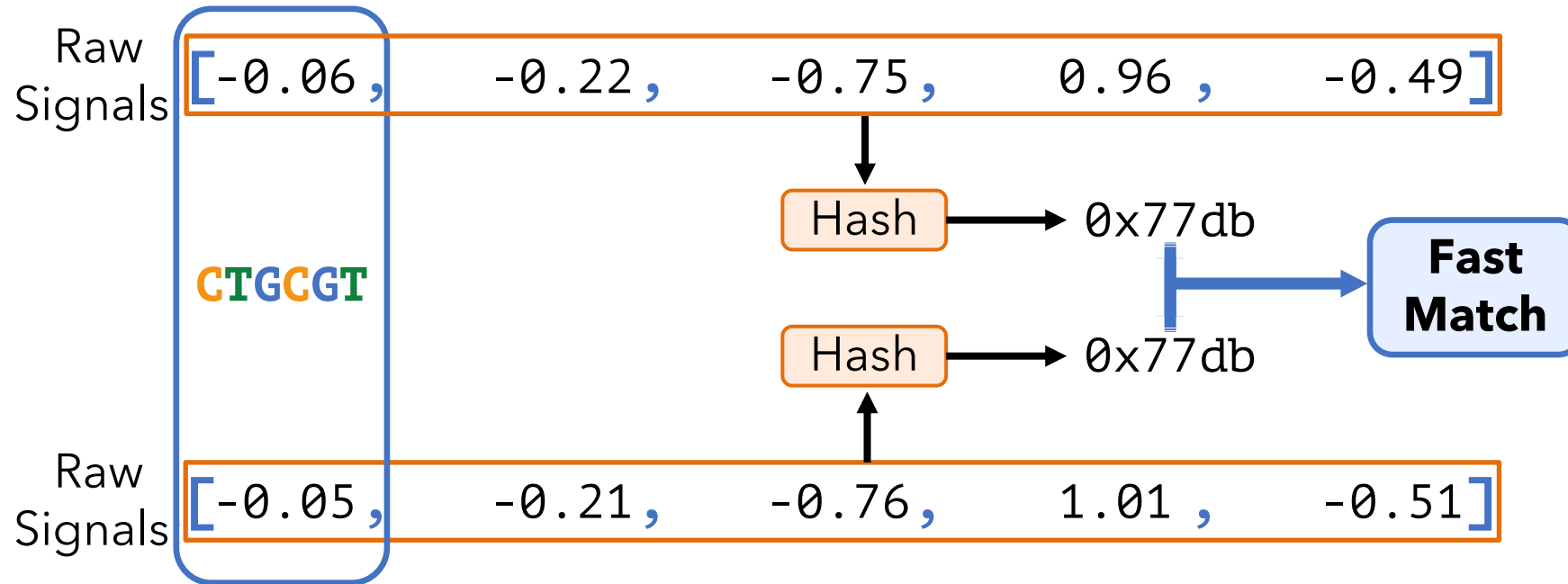


# RawHash's Goal

Enable **Fast, Accurate, and Scalable Real-Time Analysis** for Larger Genomes

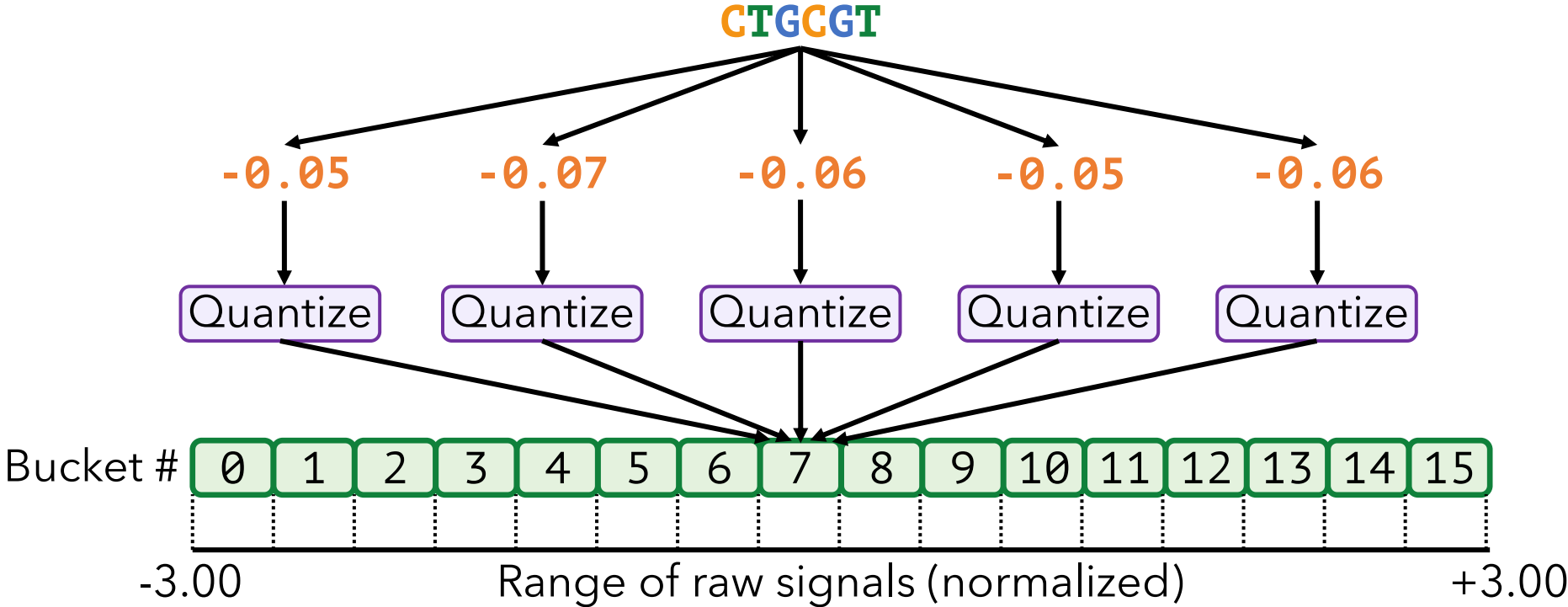
# RawHash's Key Idea

**Key Idea:** Perform similarity estimation based on **quick exact matches**



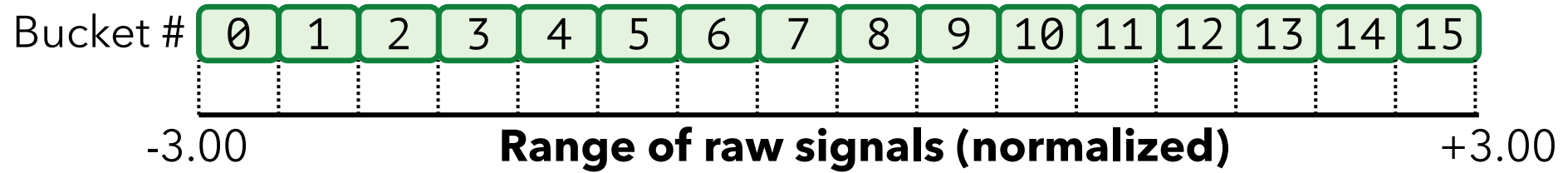
**Challenge:** How to generate **the same hash value** for **similar enough signals with various noise types?**

# RawHash's Key Idea – Quantization

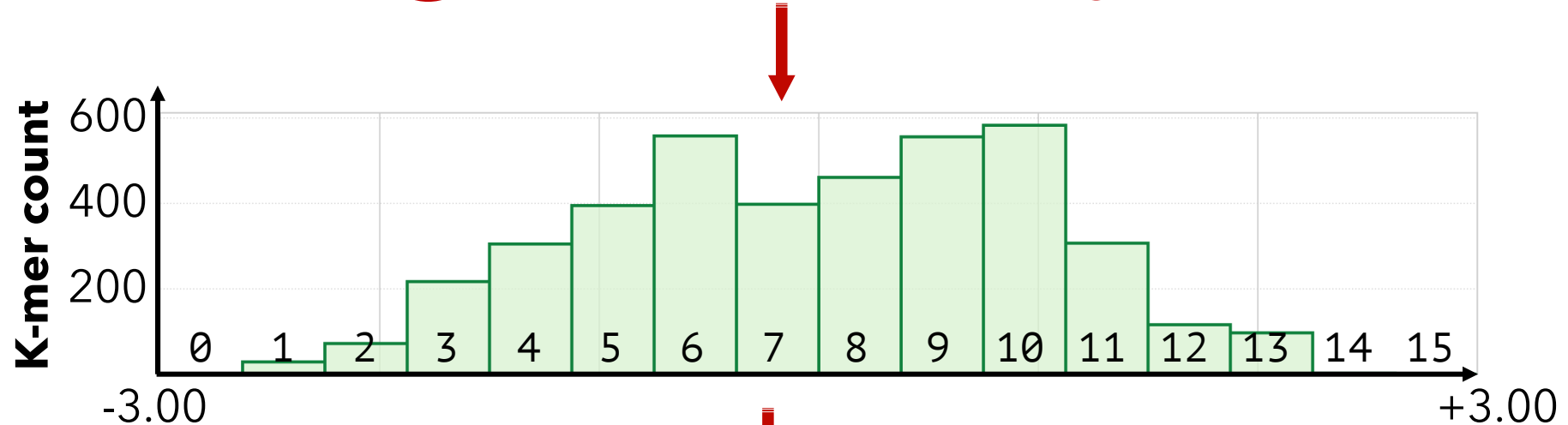


✔ **Enables matching raw signals** by eliminating slight differences

# How to Better Quantize Signals



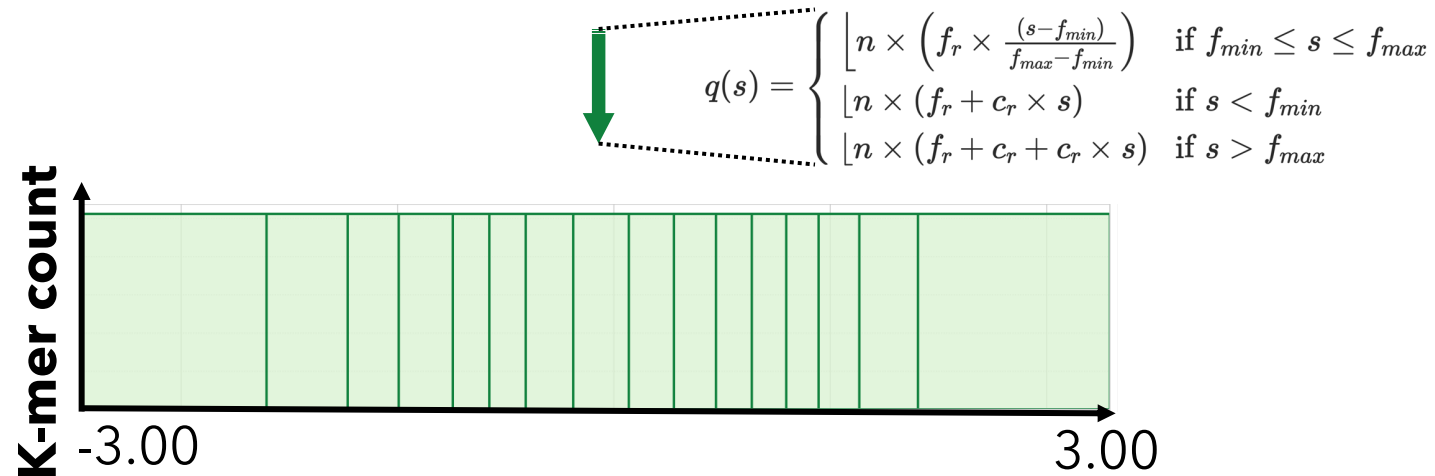
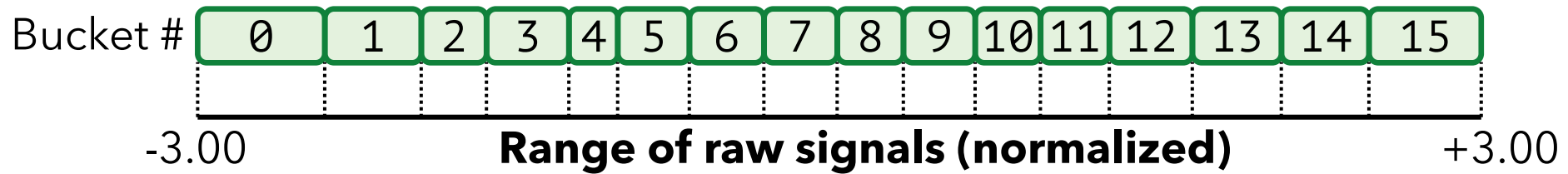
Equal-width buckets leads to **unbalanced bucket loading**



Reduced uniqueness and poor resource utilization

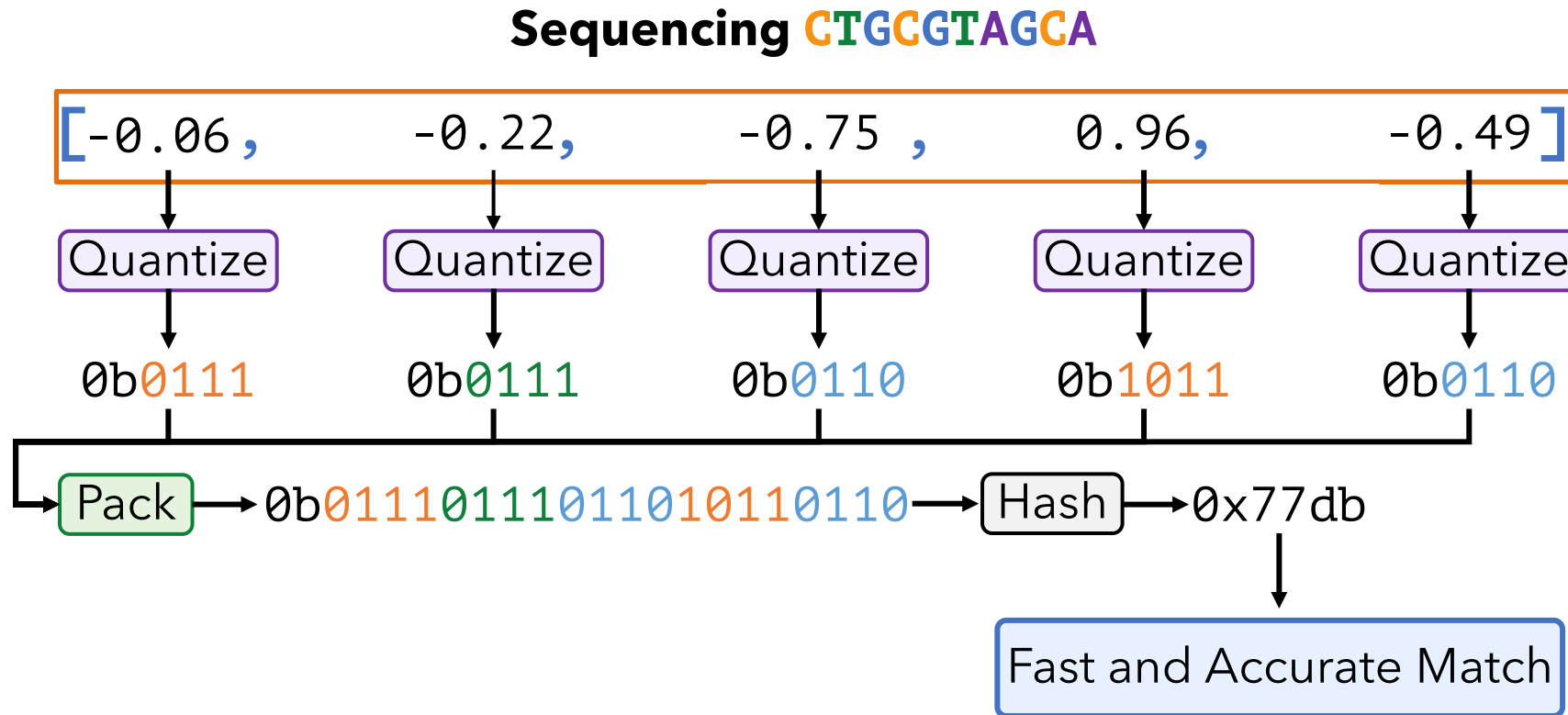
# Adaptive Quantization

- **Key Idea:** Quantizing raw signals with **non-equal bucket widths** by leveraging raw signal distribution



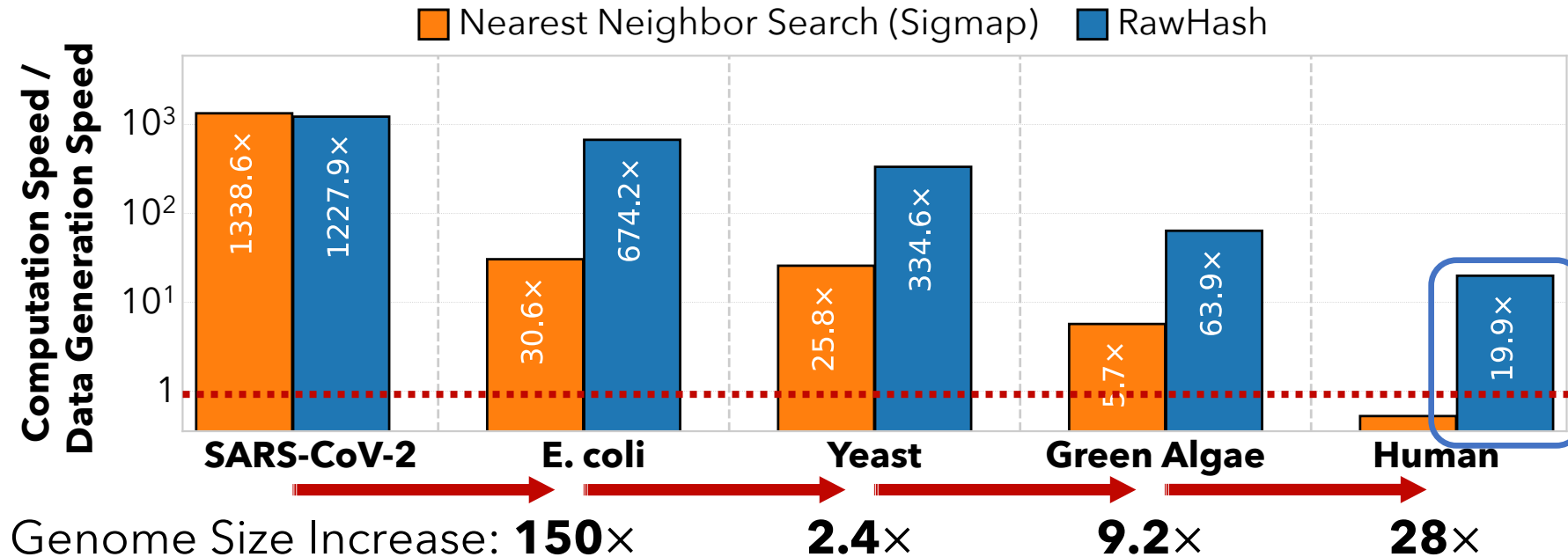
✓ **Better** bucket utilization

# RawHash Key Idea – Hash-based Matching



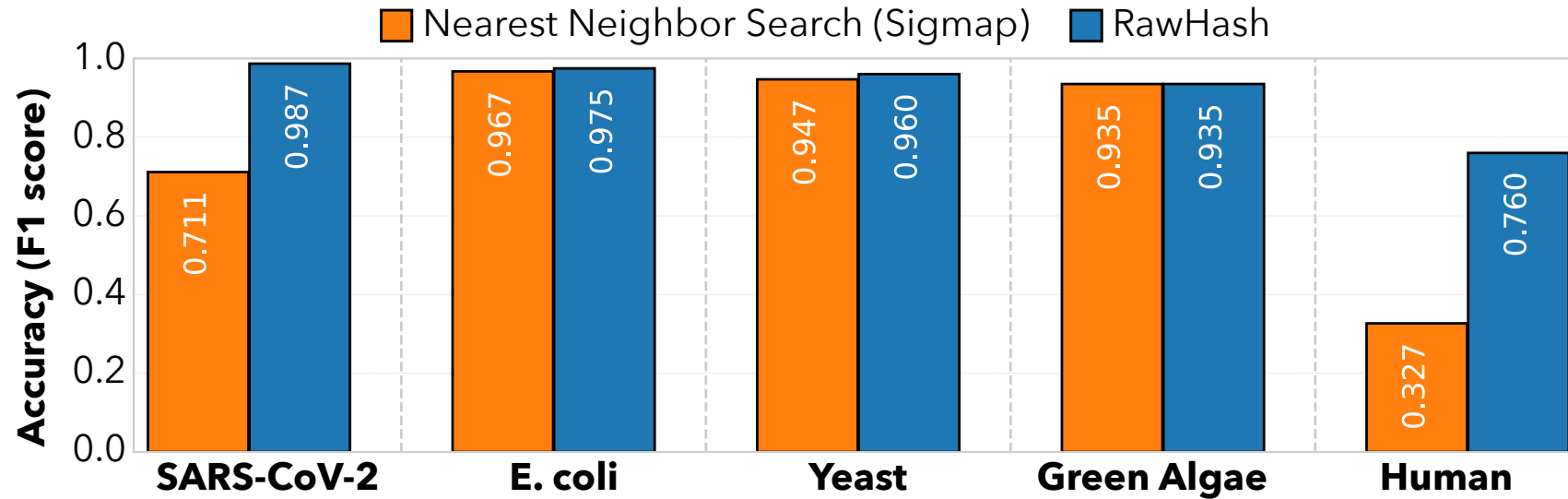
**Key Contribution:** **The first** hash-based search (and indexing) mechanism for raw signals

# Real-Time Analysis Benefits from Fast Search



**Real-time Analysis for Human Genomes:**  
Fast search instead of costly neighbor search

# Fast Search Leads to Accurate Search



## Best accuracy for all genomes:

Hash values span longer sequences than vectors ->  
More unique and informative matches

## (Almost) Identical accuracy to basecalled analysis:

Is the small difference because RawHash is even better?

# RawHash is Open Source

SAFARI RawHash Public

Edit Pins Unwatch 7 Fork 10 Starred 64


main 9 Branches 3 Tags

Go to file Add file Code

canfirtina Rawsamble evaluation scripts cc9f2b9 · 29 days ago 49 Commits

extern	new R10 parameters	6 months ago
gitfigures	rawsamble info in README	2 years ago
src	new R10 parameters	6 months ago
test	Rawsamble evaluation scripts	29 days ago
.gitignore	Rawsamble evaluation scripts	29 days ago
.gitmodules	ZSTD submodule for POD5	3 years ago
LICENSE	R10 k-mer models can be parsed now as well.	3 years ago
Makefile	A minor fix in Makefile	3 years ago
README.md	new R10 parameters	6 months ago
code_of_conduct.md	Moving to multiple headers than a single one to improve ...	3 years ago

README Code of conduct GPL-3.0 license



# RawHash

<https://github.com/CMU-SAFARI/RawHash>

About

RawHash can accurately and efficiently map raw nanopore signals to reference genomes of varying sizes (e.g., from viral to a human genomes) in real-time without basecalling. Described by Firtina et al. (published at [https://academic.oup.com/bioinformatics/article/39/Supplement\\_1/i297/7210440](https://academic.oup.com/bioinformatics/article/39/Supplement_1/i297/7210440)).

[academic.oup.com/bioinformatics/art...](https://academic.oup.com/bioinformatics/art...)

- bioinformatics
- nanopore
- seeding
- segmentation
- event-detection
- genome-analysis
- hash-tables
- contamination
- read-mapping
- relative-abundances
- nanopore-sequencing
- nanopore-analysis-pipeline
- nanopore-reads
- nanopore-data
- nanopore-minion
- raw-signal
- rawhash
- raw-nanopore-signal-analysis

- Readme
- GPL-3.0 license
- Code of conduct
- Activity
- Custom properties
- 64 stars
- 7 watching

So What?

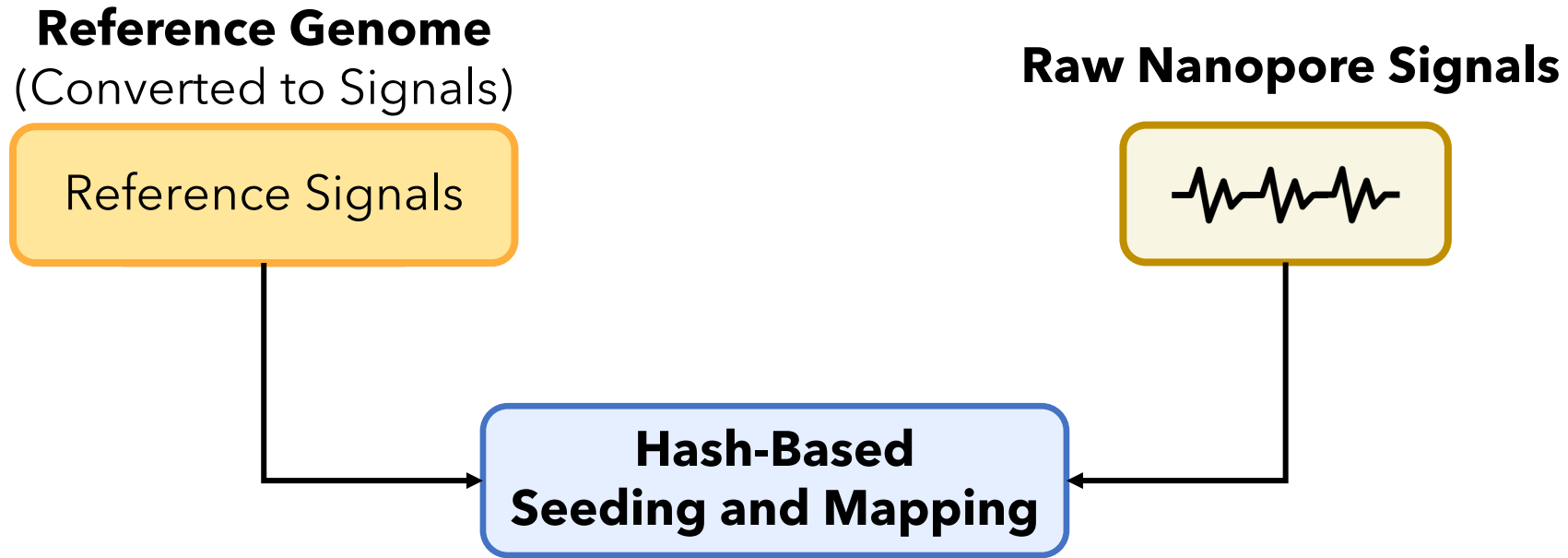
# New Directions: Assembling Raw Signals

- [Can Firtina](#), Maximilian Mordig, Harun Mustafa, Sayan Goswami, Nika Mansouri Ghiasi, Stefano Mercogliano, Furkan Eris, Joel Lindegger, Andre Kahles, and Onur Mutlu, **"Rawsamble: Overlapping and Assembling Raw Nanopore Signals using a Hash-based Seeding Mechanism"**  
[arXiv](#), October 2024.  
Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)  
[ISMB 2024 Talk](#) [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)  
[CSHL 2024 \(Biological Data Science\) Talk](#) [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)  
Social Media Thread [\[Twitter \(X\)\]](#) [\[LinkedIn\]](#)

## **Rawsamble: Overlapping and Assembling Raw Nanopore Signals using a Hash-based Seeding Mechanism**

Can Firtina<sup>1</sup> Maximilian Mordig<sup>1,2</sup> Harun Mustafa<sup>1,3,4</sup> Sayan Goswami<sup>1</sup> Nika Mansouri Ghiasi<sup>1</sup>  
Stefano Mercogliano<sup>1</sup> Furkan Eris<sup>1</sup> Joël Lindegger<sup>1</sup> Andre Kahles<sup>1,3,4</sup> Onur Mutlu<sup>1</sup>  
<sup>1</sup>*ETH Zurich*      <sup>2</sup>*Max Planck Institute for Intelligent Systems*  
<sup>3</sup>*University Hospital Zurich*      <sup>4</sup>*Swiss Institute of Bioinformatics*

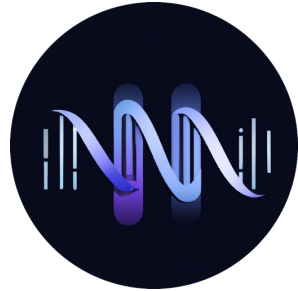
# Beyond Reference Mapping: Overlapping



 **Challenge:** Reference genome is not always available

 **Assembly:** Constructing genome from **overlapping reads**

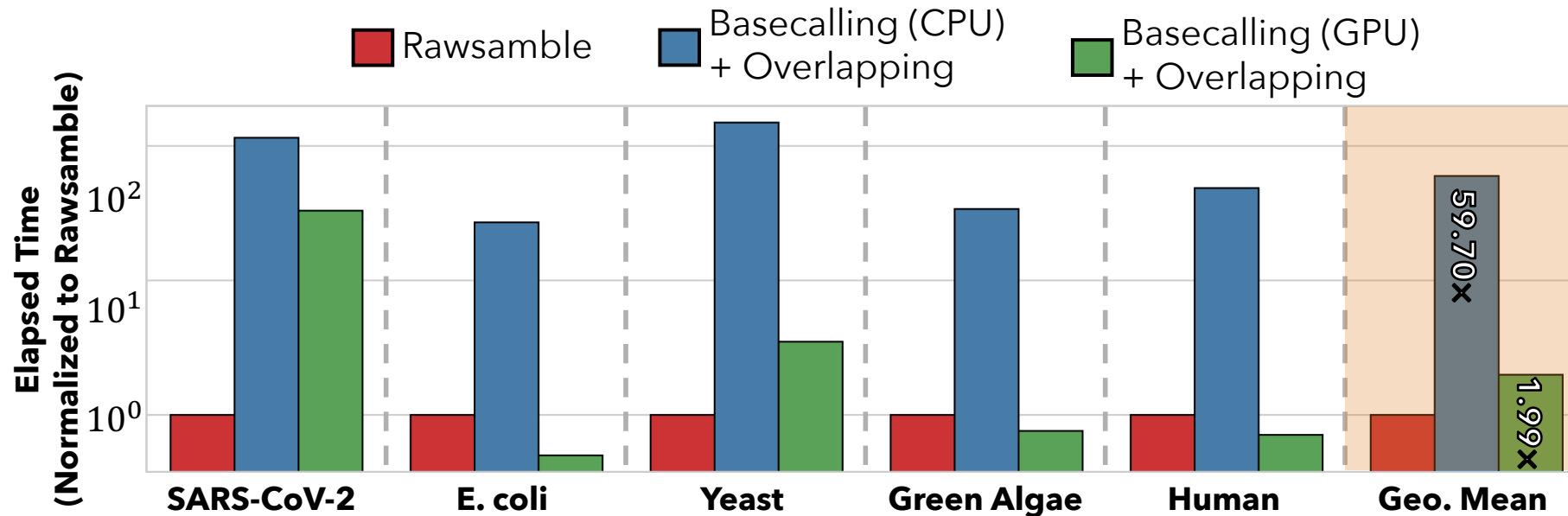
 Existing solutions cannot find overlapping reads **without basecalling**



# Rawsamble

**The first mechanism that can perform  
all-vs-all overlapping from raw signals**

# Eliminating Basecalling from the Pipeline

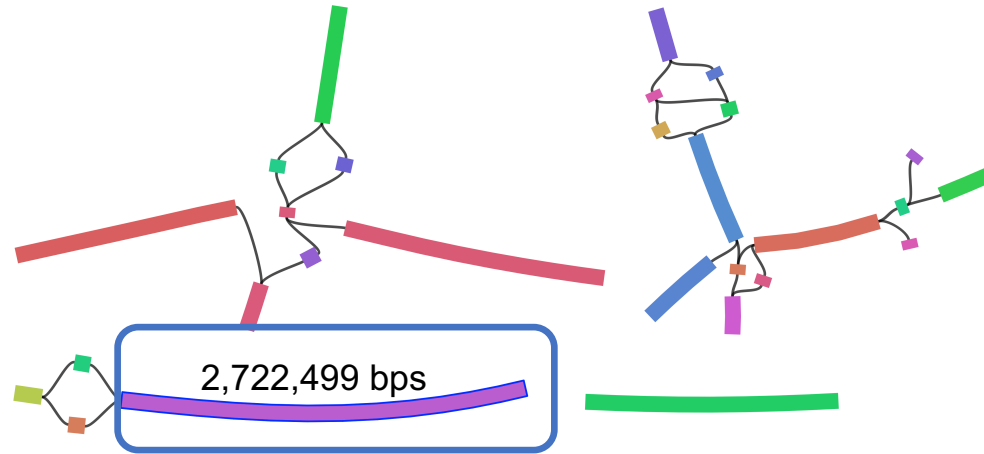


**Average speedup of 60x** than CPU-based basecalling:  
**Eliminating basecalling** leaves a significant room  
for performance improvements

**Average speedup of 2x** than the **GPU-based basecalling**:  
GPU vs. GPU comparison is likely to provide even better results

# Key Results – First Ever Assemblies

**E. coli Assembly  
(From the Rawsample Overlaps)**



**First assemblies  
ever constructed**  
without basecalling

**~400× longer assembled pieces**  
than the average read length:  
Informative overlaps lead to  
longer paths

**It is time to rethink if we want to translate  
individual reads separately**

# Quickly Aligning Raw Signals [IEEE Access '24]

- Joël Lindegger, Can Firtina, Nika Mansouri Ghiasi, Mohammad Sadrosadati, Mohammed Alser, and Onur Mutlu,

## "RawAlign: Accurate, Fast, and Scalable Raw Nanopore Signal Mapping via Combining Seeding and Alignment"

[IEEE Access](#), December 2024.

Paper [\[online\]](#) [\[pdf\]](#) [\[code\]](#)

## RawAlign: Accurate, Fast, and Scalable Raw Nanopore Signal Mapping via Combining Seeding and Alignment

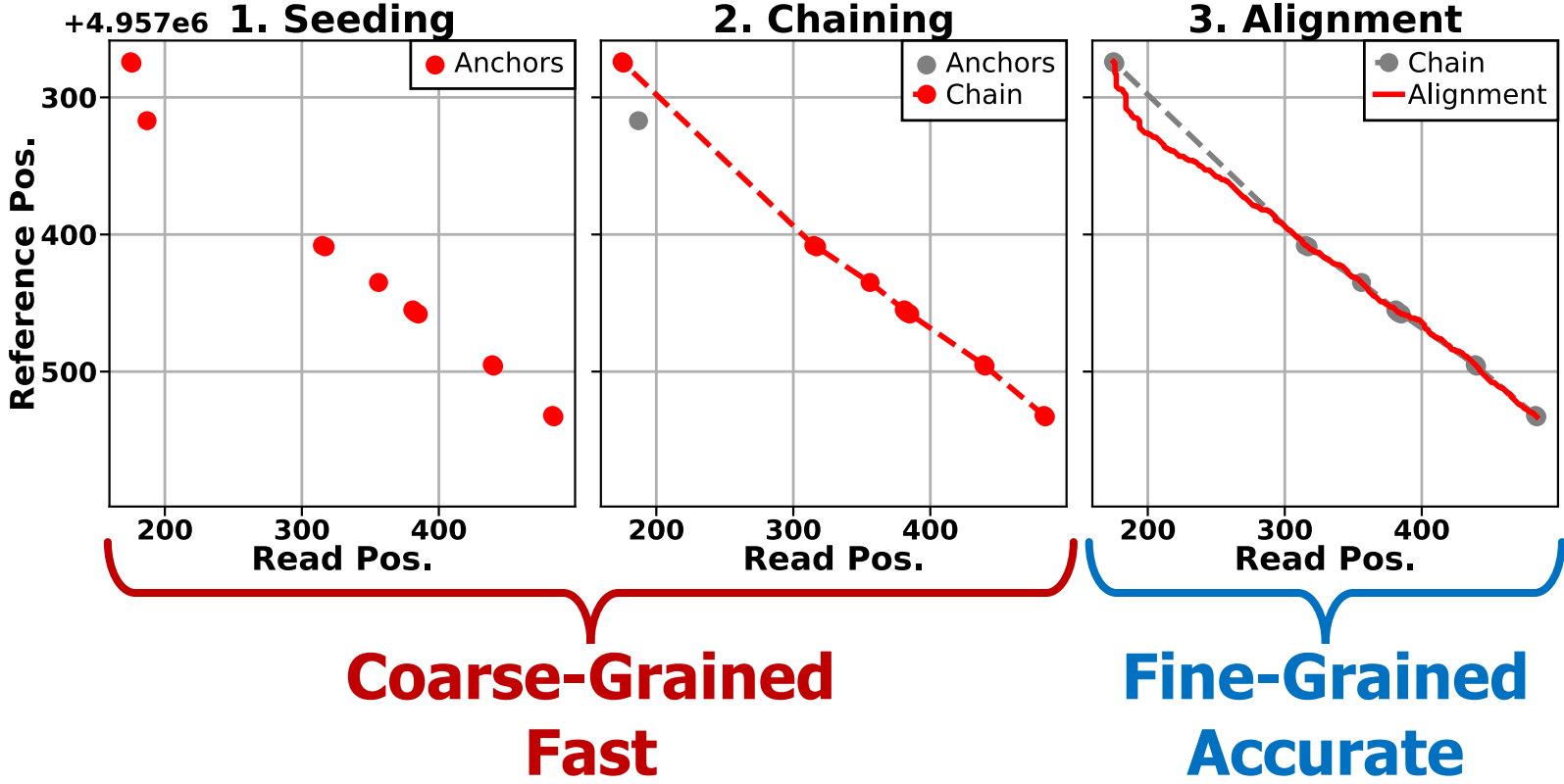
**JOËL LINDEGGER<sup>ID</sup>, CAN FIRTINA<sup>ID</sup>, NIKA MANSOURI GHIASI, MOHAMMAD SADROSADATI<sup>ID</sup>,  
MOHAMMED ALSER<sup>ID</sup>, AND ONUR MUTLU<sup>ID</sup>, (Fellow, IEEE)**

Department of Information Technology and Electrical Engineering, ETH Zürich, 8092 Zürich, Switzerland

Corresponding authors: Joël Lindegger (jmlindegger@gmail.com), Can Firtina (firtinac@ethz.ch), and Onur Mutlu (omutlu@ethz.ch)

This work was supported in part by the Semiconductor Research Corporation (SRC), in part by the ETH Future Computing Laboratory, in part by European Union's Horizon Programme for Research and Innovation under Grant 101047160-BioPIM, and in part by the Swiss National Science Foundation (SNSF) under Grant 200021\_213084.

# RawAlign Overview



# Benchmarking Raw Signal Analysis [ACM-BCB '25]

- Furkan Eris, Ulysse McConnell, Can Firtina, and Onur Mutlu,  
**"RawBench: A Comprehensive Benchmarking Framework  
for Raw Nanopore Signal Analysis Techniques"**

Proceedings of the [16th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics \(ACM-BCB 2025\)](#), Philadelphia, PA, USA, October 2025.

Paper [\[online\]](#) [\[pdf\]](#)

Conference Talk [\[pptx\]](#) [\[pdf\]](#)

## RawBench: A Comprehensive Benchmarking Framework for Raw Nanopore Signal Analysis Techniques

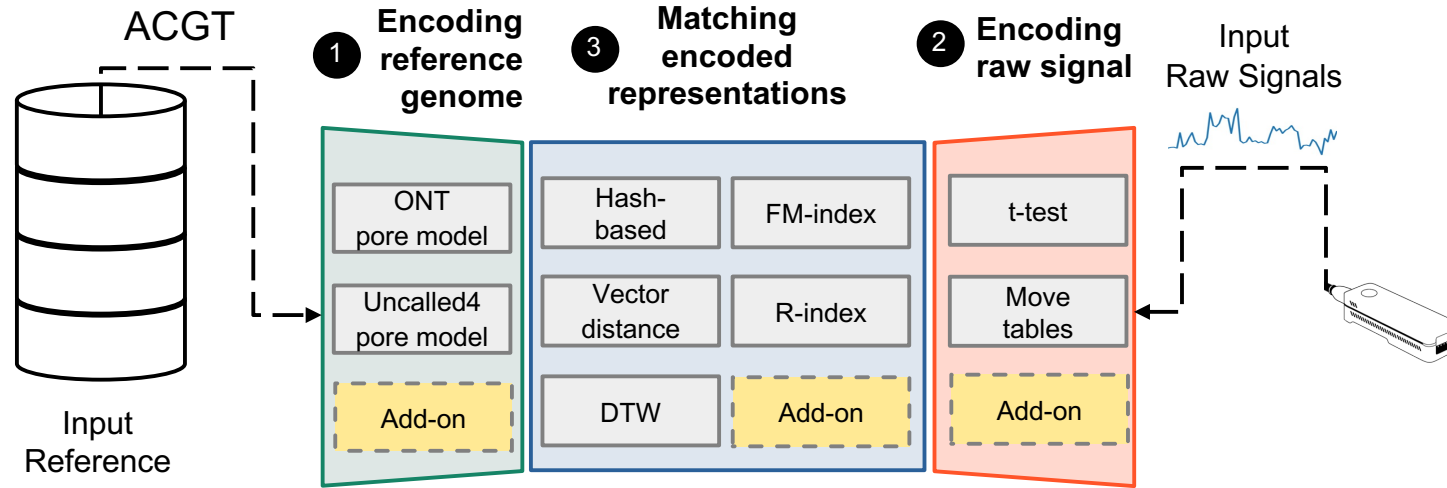
Furkan Eris  
ETH Zurich  
Zurich, Switzerland  
furkan6olm@gmail.com

Ulysse McConnell  
ETH Zurich  
Zurich, Switzerland  
umcconnell@student.ethz.ch

Can Firtina\*  
University of Maryland, College Park  
MD, USA  
firtina@umd.edu

Onur Mutlu\*  
ETH Zurich  
Zurich, Switzerland  
omutlu@gmail.com

# RawBench Benchmarking Framework



## 1 Encoding Reference Genome

- Uses learned pore models by ONT and community

## 2 Encoding Raw Signal

- Utilizes statistical and ML-based encoding techniques

## 3 Matching Encoded Representations

- Compares representations that are now encoded into the same space

# In-Storage Raw Signal Analysis [ICS '25]

- Melina Soysal, Konstantina Koliogeorgi, Can Firtina, Nika Mansouri Ghiasi, Rakesh Nadig, Haiyu Mao, Geraldo Francisco Oliveira Junior, Yu Liang, Klea Zambaku, Mohammad Sadrosadati, and Onur Mutlu,

## **"MARS: Processing-In-Memory Acceleration of Raw Signal Genome Analysis Inside the Storage Subsystem"**

Proceedings of the [39th International Conference on Supercomputing \(ICS 2025\)](#), Salt Lake City, Utah, USA, June 2025.

Paper [\[online\]](#) [\[pdf\]](#)

Conference Talk [\[video\]](#) [\[pptx\]](#) [\[pdf\]](#)

## **MARS: Processing-In-Memory Acceleration of Raw Signal Genome Analysis Inside the Storage Subsystem**

Melina Soysal<sup>†</sup>   Konstantina Koliogeorgi<sup>†</sup>   Can Firtina<sup>†</sup>   Nika Mansouri Ghiasi<sup>†</sup>  
Rakesh Nadig<sup>†</sup>   Haiyu Mayo<sup>\*</sup>   Geraldo F. Oliveira<sup>†</sup>  
Yu Liang<sup>†</sup>   Klea Zambaku<sup>†</sup>   Mohammad Sadrosadati<sup>†</sup>   Onur Mutlu<sup>†</sup>

<sup>†</sup> *ETH Zürich*

<sup>\*</sup> *King's College London*

So What?

# Some Interesting Challenges in Signal Analysis

- Figuring out how to better utilize raw signal analysis
  - Should we combine with basecalled analysis?
  - Should (and can) we perform the full analysis pipelines using only raw signals
- Distinguishing noise from useful information
- Redesigning effective solutions for storage
- Enabling end-to-end and energy-efficient analysis with (portable) sequencing devices

# STORM Group

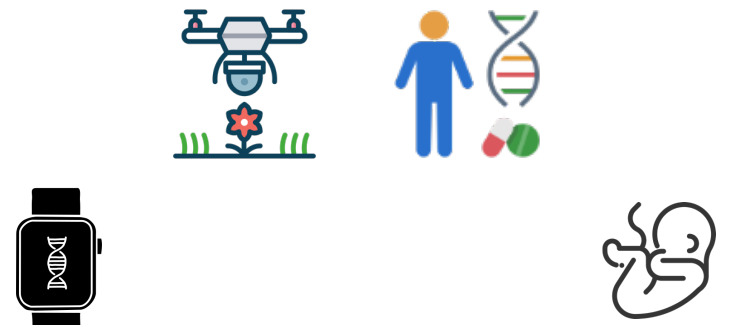
## Research Directions

 **Real-Time and Portable**  
Genomic Signal Analysis

 **Data-centric and AI-Driven** Solutions to Analyze Genomic Data

 Accurate and Fast Algorithms Scalable to **Large Volumes of Genomic Data**

### Today's Talk



### Goal

Enable fundamentally better integration of biological data analysis in our **everyday lives**



# Lost in Translation: Harnessing the Power of Nanopore Electrical Signals in Genomics

Arch4Health Workshop (in conj. with HPCA 2026), Sydney, Australia  
January 31, 2026

**Can Firtina** | [canfirtina@gmail.com](mailto:canfirtina@gmail.com) | [cs.umd.edu/~firtina](https://cs.umd.edu/~firtina)

Assistant Professor of Computer Science at UMD



DEPARTMENT OF  
COMPUTER SCIENCE

# Backup Slides

# Generating Electrical Signals from DNA

