

# P&S Mobile Genomics

## Introduction & Project Proposals

Dr. Mohammed Alser



ETH Zurich

Fall 2021

5 October 2021

# The Role of This Course

# Projects & Seminars: Mobile Genomics

---

- We will cover the **basics** of **genome analysis** to understand the **speed-accuracy tradeoff** in using computationally-lightweight heuristics versus accurate computationally-expensive algorithms.
- Students will **experimentally** evaluate different heuristic **algorithms** and observe their effect on **the end results**.
- This evaluation will give the students the chance to carry out a **hands-on project** to implement one or more of these heuristic algorithms in **their smartphones** and **help the society by enabling on-site analysis of genomic data**.

# Key Objectives

---

- Multiple components that are aimed at improving students'
  - Basic knowledge in genome analysis (**dry lab**)
  - Technical skills in genome analysis and computer architecture
  - Critical thinking and analysis
  - Familiarity with key research directions
  - Technical presentation of your project

# Key Goal

---

(Learn how to)

efficiently implement

one of the key steps in genome

analysis on portable devices

# Prerequisites of the Course

---

- No prior knowledge in bioinformatics or genome analysis is required.
- A good knowledge in C programming language and programming is required.
- Interest in making things efficient and solving problems

# Course Info: Who Are We?

---



## ■ Onur Mutlu

- Full Professor @ ETH Zurich ITET (INFK), since September 2015
- Strecker Professor @ Carnegie Mellon University ECE/CS, 2009-2016, 2016-...
- PhD from UT-Austin, worked at Google, VMware, Microsoft Research, Intel, AMD
- <https://people.inf.ethz.ch/omutlu/>
- [omutlu@gmail.com](mailto:omutlu@gmail.com) (Best way to reach me)
- <https://people.inf.ethz.ch/omutlu/projects.htm>

## ■ Research and Teaching in:

- Computer architecture, computer systems, hardware security, bioinformatics
- Memory and storage systems
- Hardware security, safety, predictability
- Fault tolerance
- Hardware/software cooperation
- Architectures for bioinformatics, health, medicine
- ...

# Course Info: Who Are We?

---



## Mohammed Alser

- Lecturer and Senior Researcher, [SAFARI Research Group, ETH Zürich](#), since Sept. 2018.
- PhD from Bilkent University (Turkey) 2018, worked at UCLA, TU Dresden, and PETRONAS.
- [Received the IEEE Turkey Doctoral Dissertation Award](#) and a number of international prestigious awards.
-  <https://twitter.com/mealser>
- My main research is in **bioinformatics, computational genomics, metagenomics**, and computer architecture.
- I am especially excited about **building** new data structures, algorithms, and architectures that **make intelligent genome analysis a reality**.

# Course Info: Who Are We?



**Juan Gómez Luna**  
Senior Researcher and  
Lecturer

Processing-In-Memory |  
Heterogeneous computing |  
Memory Systems | Bioinformatics |  
Medical imaging



**Nour Almadhoun Alserr**  
Senior Researcher

Data privacy | Bioinformatics |  
Computational Genomics



**Can Firtina**  
PhD Student

Genome Assembly |  
Sequence Analysis &  
Alignment | Biologically-  
Inspired Computing  
Paradigms | Brain-  
Computer Interfaces |  
Phase-change memory

- Get to know them and their research: <https://safari.ethz.ch/safari-group/> <sup>9</sup>

# Course Requirements and Expectations

---

- Attendance required for all meetings
- Study the learning materials
- Each student will carry out a hands-on project
  - Build, implement, code, and design with close engagement from the supervisors
- Participation
  - Ask questions, contribute thoughts/ideas
  - Read relevant papers
- Presentation & GitHub repository

We will help the projects with good progress to get published in good venues!

# Course Website

---

- [https://safari.ethz.ch/projects\\_and\\_seminars/fall2021/doku.php?id=genome\\_seq\\_mobile](https://safari.ethz.ch/projects_and_seminars/fall2021/doku.php?id=genome_seq_mobile)
- Useful information for the course
- Check your email and Moodle frequently for announcements
- We will also have Moodle for Q&A, announcements, ..

# Next Meetings

---

- We will give you a chance to select a project,
- Then, we will have **1-1 meetings** to match your interests, skills, and background with a suitable project.
- It is important that you **study the learning materials** before our next meeting!
- We will start the projects **next week**.

---

# **WHAT IS GENOME ANALYSIS?**

# What is Genome Analysis?

---



## Genomic analysis

 Atom

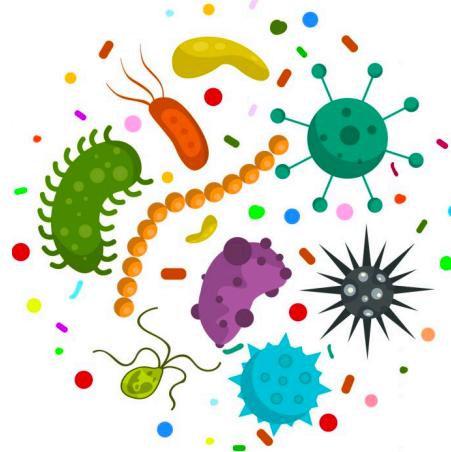
 RSS Feed

Genomic analysis is the identification, measurement or comparison of genomic features such as DNA sequence, structural variation, gene expression, or regulatory and functional element annotation at a genomic scale. Methods for genomic analysis typically require high-throughput sequencing or microarray hybridization and bioinformatics.

# Applications of Genome Analysis



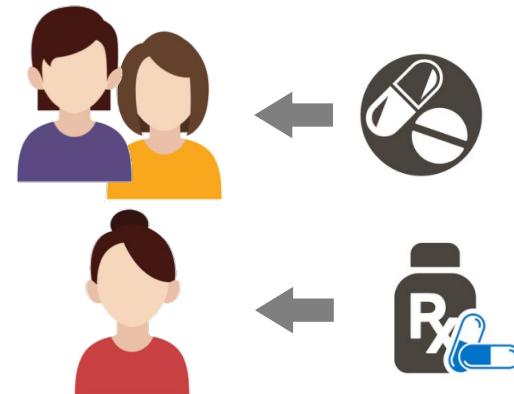
Understanding **genetic variations**



Predicting the presence and relative abundances of **microbes** in a sample



Rapid surveillance of **disease outbreaks**



Developing **personalized medicine**

And many other applications ...

# How to Analyze a Genome?



**NO**

machine gives the **complete**  
**sequence** of genome as output



```
>CCTCCTCAGTGCCACCCAGCCCAGTGGCAGCTCCAAACAGGGCTCTTATTAAAACACCCCTGTCCTGCCCTGGAGTGAGGTGTCAAG  
GACCTAAACTAAAAAAAAAAAAAGAAAAAGAAAAAGAATTAAAATTAAGTAATTCTTGAAAAAAAACTAATTCTAAGCTTCTT  
CATGTCAAGGACCTAACATGTGCTAACAGCACTTTTGACCATTATTTGGATCTGAAAGAAATCAAGAATAATGAAGGACTGATACATTG  
GAAGAGGAGAGTCAGGACCTACAGAAAAAAAAAAAAAGAAAAAGAAAAAGAATTAAAATTAAGTAATTCTTGAAAAAAA  
ACTAATTCTAAGCTTCTTCATGTCAAGGACCTAACATGTCTGTGCAGGTCTTGCATTCCCTGTCAAAAGAAAAAGAATTAAAATTT  
AAGTAATTCTTGAAAAAAAACTAATTCTAAGCTTCTTCAAGGTCAAGGACCTAACATGTCAAGGACCTAACAGAGTTGCAAAAAAAAAGAAAAAA  
GAAAAGAAAAAGAATTAAAATTAAGTAATTCTTGAAAAAAAACTAATTCTAAGCTTCTTCAAGGTCAAGGACCTAACATGTAGCCAGAATGG  
TTGTGGGATGGGAGCCTCTGGACCGACCAGGTAGCTCTTTCCACACTGTAGTCTCAAAGCTTCTTCAAGGTCAAGGACCTAACACACCACCTCCAGAAAGCTTCTCA.....
```

# DNA Testing



Fall DNA special  
Just 55 CHF ~~89 CHF~~

Order now

The promotion ends today in 12 more hours!



<https://www.myheritage.ch/dna>

# DNA Testing



Fall DNA special  
Just 55 CHF ~~89 CHF~~



<https://www.myheritage.ch/dna>  
<https://www.23andme.com/>



Health + Ancestry Service

\$199

- Includes everything in Ancestry + Traits Service

*PLUS*

- 10+ Health Predisposition reports\*
- 5+ Wellness reports
- 40+ Carrier Status reports\*



# High-Throughput Sequencers



Illumina MiSeq



Illumina NovaSeq 6000



Pacific Biosciences RS II

Oxford  
Nanopore  
PromethION



Pacific  
Biosciences  
Sequel II



Oxford Nanopore MinION



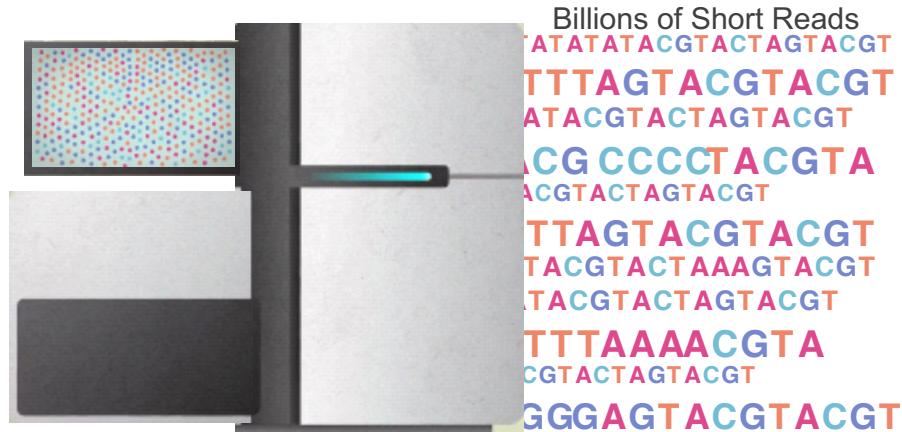
Oxford  
Nanopore  
SmidgION

**... and more! All produce data with different properties.**

# Genome Sequencer is a Chopper

---

Regardless the sequencing machine,  
reads still lack information about their order and location  
(which part of genome they are originated from)



# Solving the Puzzle

---

.FASTA file



Reference genome

.FASTQ file



Reads

<https://www.pacb.com/smrt-science/smrt-sequencing/hifi-reads-for-highly-accurate-long-read-sequencing/>

# Reference Genome

---

.FASTA file:

```
>NG_008679.1:5001-38170 Homo sapiens paired box 6 (PAX6)
ACCCTTTTCTTATCATTGACATTAAACTCTGGGGCAGGTCTCGCGTAGAACGCGGCTGTCAGATCT
GCCACTTCCCCTGCCGAGCGCGGTGAGAAGTGTGGAACCGCGCTGCCAGGCTCACCTGCCTCCCCGC
CCTCCGCTCCCAGGTAAACGCCCGGGCTCCGGCCCCGGCTCGGGGCCCGGGCCTCTCCGCTG
CCAGCGACTGCTGTCCCCAAATCAAAGCCGCCCAAGTGGCCCCGGGCTTGATTTGCTTTAAAAG
GAGGCATAAAAGATGGAAGCGAGTTACTGAGGGAGGGATAGGAAGGGGGTGGAGGAGGGACTTGTCTT
TGCCGAGTGTGCTCTTGCAAAAGTAGCAAAATGTTCACTCCTAAGAGTGGACTTCCAGTCCGGCCCT
GAGCTGGAGTAGGGGGCGGGAGTCTGCTGCTGTCTGCTAAAGCCACTCGCGACCGCGAAAAATGCA
GGAGGTGGGGACGCACTTGCATCCAGACCTCCTGCATCGCAGTTCACGACATCCACGCTGGAAAG
TCCGTACCCGCGCCTGGAGCGCTAAAGACACCCCTGCCCGGGTCGGCGAGGTGCAGCAGAAGTTCCC
GCGGTTGCAAAGTGCAGATGGCTGGACCGCAACAAAGTCTAGAGATGGGTTCTCAGAAAGACGC
```

# Obtaining the Human Reference Genome

---

- **GRCh38.p13**
- Description: Genome Reference Consortium Human Build 38 patch release 13 (GRCh38.p13)
- Organism name: [Homo sapiens \(human\)](#)
- Date: 2019/02/28
- 3,099,706,404 bases
- Compressed .fna file (964.9 MB)
- [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.39](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39)

```
>NC_00001.11 Homo sapiens chromosome 1, GRCh38.p13 Primary Assembly  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  
....
```

---

# Genomic Reads

---

.FASTQ file:

Identifier ————— @HWI-EAS209\_0006\_FC706VJ:5:58:5894:21141#ATCACG/1  
Sequence ————— TTAATTGGTAAATAAATCTCCTAATAGCTTAGATNTTACCTNNNNNNNNNTAGTTCTTGAGA  
'+' sign ————— +  
Quality scores ————— efcfffffffcfеffffcfffffdff`feed] ` ] \_Ba\_ ^ \_\_ [ YBBBBBBBBBBRTT\ ] ] [ ] dddd`  
  
Base T  
phred Quality ] = 29

# Obtaining .FASTQ Files

- <https://www.ncbi.nlm.nih.gov/sra/ERR240727>

The screenshot shows the NCBI SRA search interface. At the top, there's a navigation bar with the NCBI logo, 'Resources' (with a dropdown arrow), and 'How To'. Below the bar, the text 'SRA' is displayed, followed by a dropdown menu set to 'SRA' and an empty search input field. A link to 'Advanced' search is also present. A prominent orange banner at the top of the main content area contains a black exclamation mark icon and the text 'COVID-19 is an emerging, rapidly evolving situation.' Below the banner are links to 'Public health information (CDC)', 'Research information (NIH)', 'SARS-CoV-2 data (NCBI)', and 'Prevention and treatment information (WHO)'. The main content area displays study details for 'ERX215261: Whole Genome Sequencing of human TSI NA20754'. It includes a summary, design information, submitter details, study information, sample details, library information, and run statistics.

Full ▾

Send to: ▾

**ERX215261: Whole Genome Sequencing of human TSI NA20754**

1 ILLUMINA (Illumina HiSeq 2000) run: 4.1M spots, 818.7M bases, 387.2Mb downloads

**Design:** Illumina sequencing of library 6511095, constructed from sample accession SRS001721 for study accession SRP000540. This is part of an Illumina multiplexed sequencing run (9340\_1). This submission includes reads tagged with the sequence TTAGGCAT.

**Submitted by:** The Wellcome Trust Sanger Institute (SC)

**Study:** Whole genome sequencing of (TSI) Toscani in Italia HapMap population

[PRJNA33847](#) • [SRP000540](#) • [All experiments](#) • [All runs](#)

**Sample:** Coriell GM20754

[SAMN00001273](#) • [SRS001721](#) • [All experiments](#) • [All runs](#)

**Organism:** *Homo sapiens*

**Library:**

**Name:** 6511095

**Instrument:** Illumina HiSeq 2000

**Strategy:** WGS

**Source:** GENOMIC

**Selection:** RANDOM

**Layout:** PAIRED

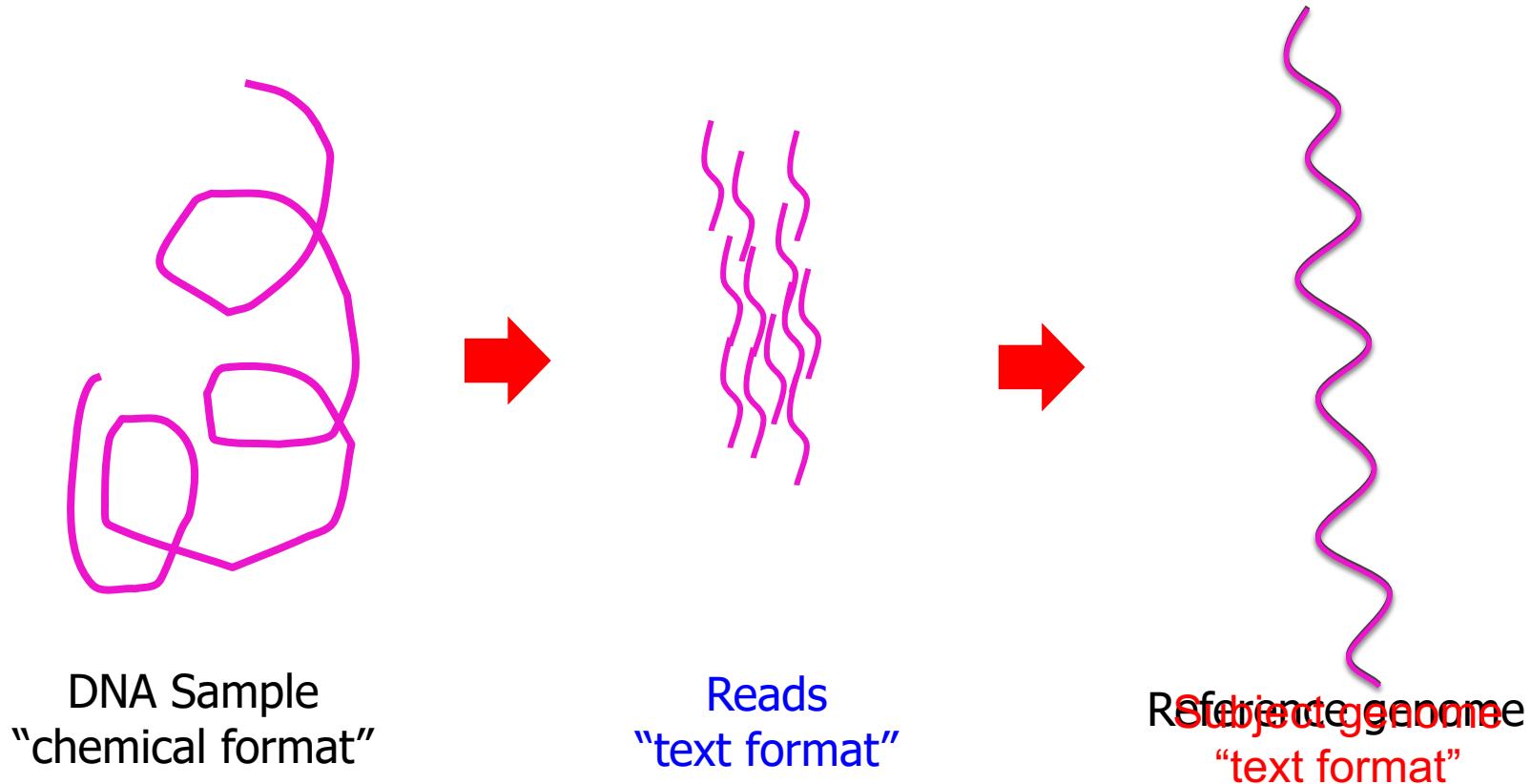
**Construction protocol:** Standard

**Runs:** 1 run, 4.1M spots, 818.7M bases, 387.2Mb

| Run                       | # of Spots | # of Bases | Size    | Published  |
|---------------------------|------------|------------|---------|------------|
| <a href="#">ERR240727</a> | 4,093,747  | 818.7M     | 387.2Mb | 2013-03-22 |

# Read Mapping

Map **reads** to a known reference genome with some minor differences allowed



# Matching Each Read with Reference Genome

---

## .FASTA file:

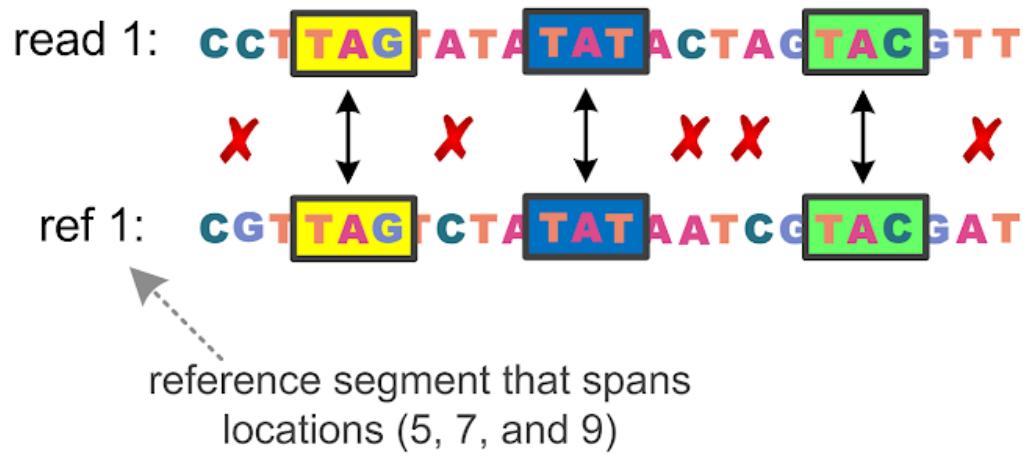
```
>NG_008679.1:5001-38170 Homo sapiens paired box 6 (PAX6)
ACCCCT[REDACTED]TCATTGACATTAAACTCTGGGGCAGG[REDACTED]GAACGC GGCTGT CAGATCT
GCCACTTCCCCTGCCGAGCGGCGGTGAGAAGTGTGGGAACC CGCGCTGCCAGGCTCACCTGCCTCCCCGC
CCTCCGCTCCCAGGTAACCGCC[REDACTED]CCCCGGCCGGCTCGGGGCCCGCGGGCCTCTCCGCTG
CCAGCGACTGCTGTCCCCAAATCAAAGCCCGCCCAAGTGGCCCCGGGGCTTGATTTGCTTTAAAAG
GAGGCATACAAAGATGGAAGCGAGTTACTGAGGGAGGGATAGGAAGGGGGGTGGAGGAGGGACTTGTCTT
TGCCGAGGTG[REDACTED]CAAAAGTAGCA[REDACTED]CTCCTA[REDACTED]TCCAGTCCGGCCCT
GAGCTGGGAGTAGGGGGCGGGAGTCTGCTGCTGTCTGCTAAAGCCACTCGCGACCGCGAAAAATGCA
GGAGGTGGGGACGCACTTGCATCCAGACCTCCTCTGCATCGCAGTTC[REDACTED]CGCTTGGGAAAG
TCCGTACCCGCGCCT[REDACTED]AAAGACACCCCTGCCGCGGGTCGGCGAGGTGCAGCAGAAGTTCCC
GCGGTTGCAAAGTGCAGATGGCTGGACCGCAACAAAGTCTAGAGATGGGGTTCGTTCTCAGAAAGACGC
```

## .FASTQ file:

```
@HWI-EAS209_0006_FC706VJ:5:58:5894:21141#ATCACG/1
T[REDACTED]AATAAAATCT[REDACTED]TTAGATN[REDACTED]NNNNNNNNNTAG
+HWI-EAS209_0006_FC706VJ:5:58:5894:21141#ATCACG/1
efcfffffcfeffffcfffffffdf`feed]`]_Ba_`__[YBBBBBBBBBBRTT
```

# Base-by-Base Comparison

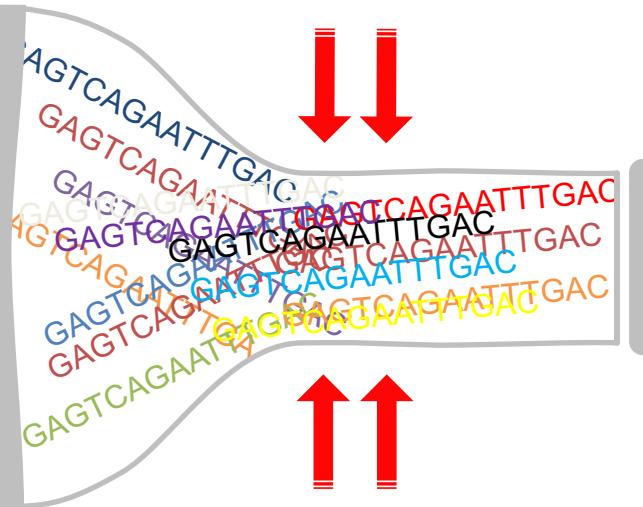
---



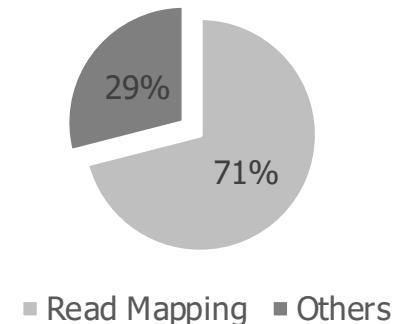
# Analysis is Bottlenecked in Read Mapping!!

**48** Human whole genomes  
at 30 × coverage  
in about 2 days

Illumina NovaSeq 6000



**1** Human genome  
32 CPU hours  
on a 48-core processor



# What is Intelligent Genome Analysis?

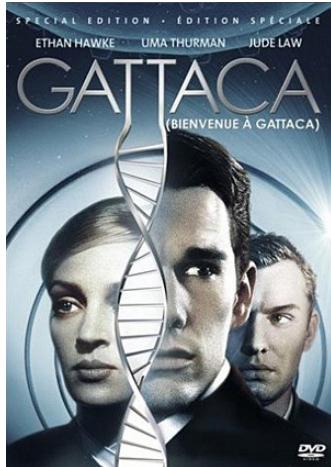
---

- Fast genome analysis Bandwidth
  - *Real-time analysis*
- Using intelligent architectures Energy-efficiency & Latency
  - *Specialized HW with less data movement*
- DNA is a valuable asset Privacy
  - *Controlled-access analysis*
- Population-scale genome analysis Scalability
  - *Sequence anywhere at large scale!*
- Avoiding erroneous analysis Accuracy
  - *E.g., your father is not your father*

# Fast Genome Analysis?

- **Fast** genome analysis in mere seconds using **limited computational resources** (i.e., personal computer or small hardware).

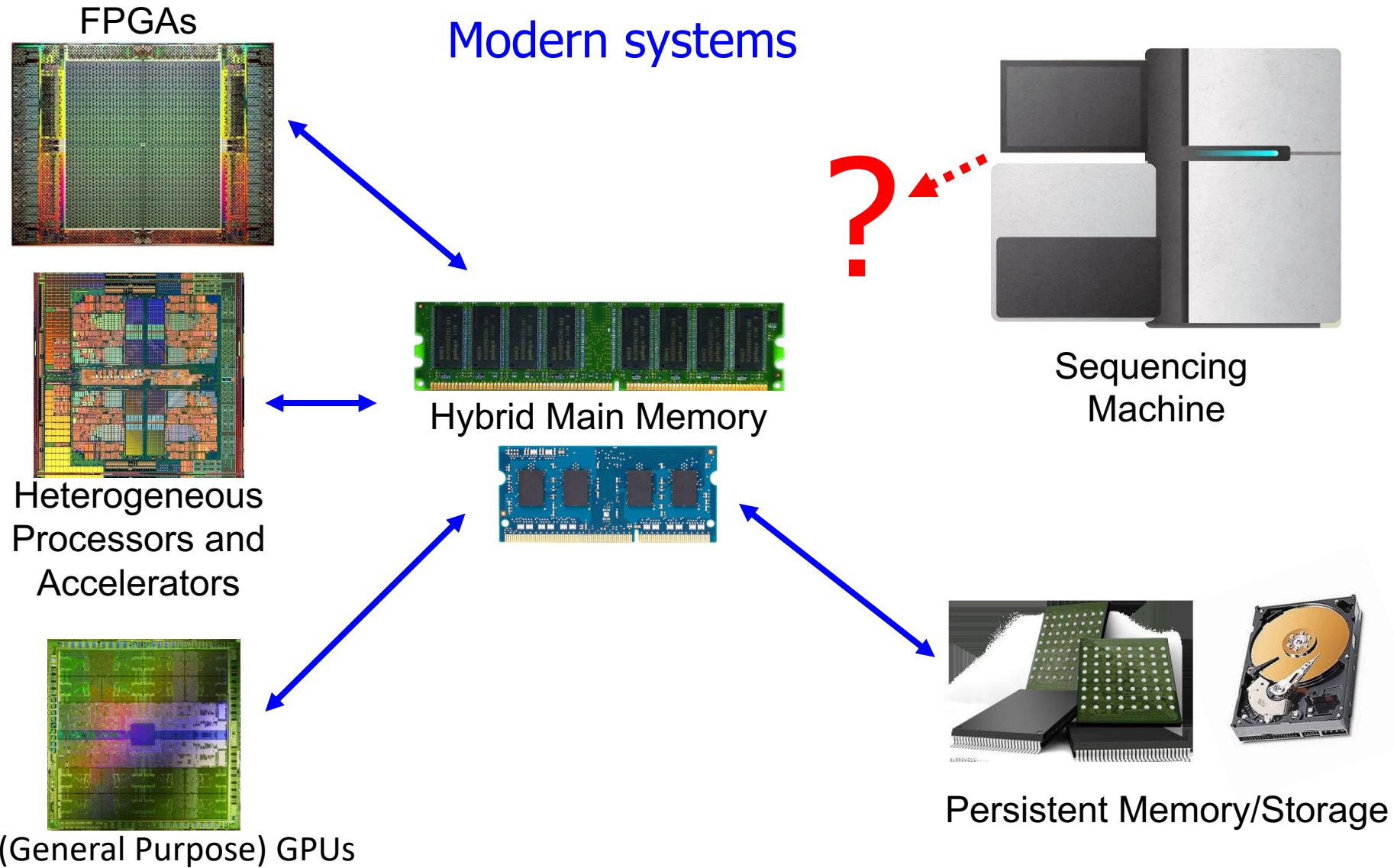
1997



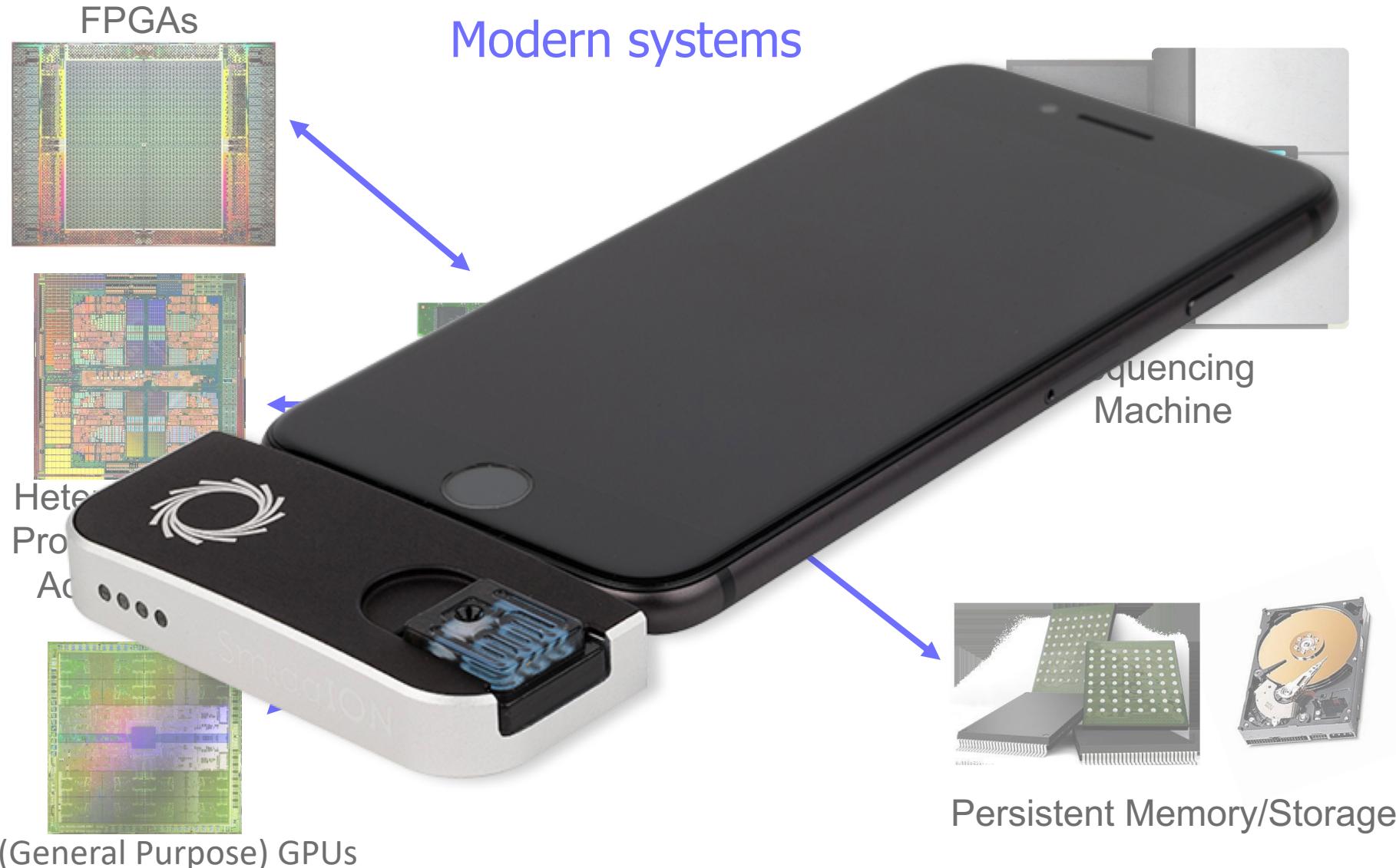
2015



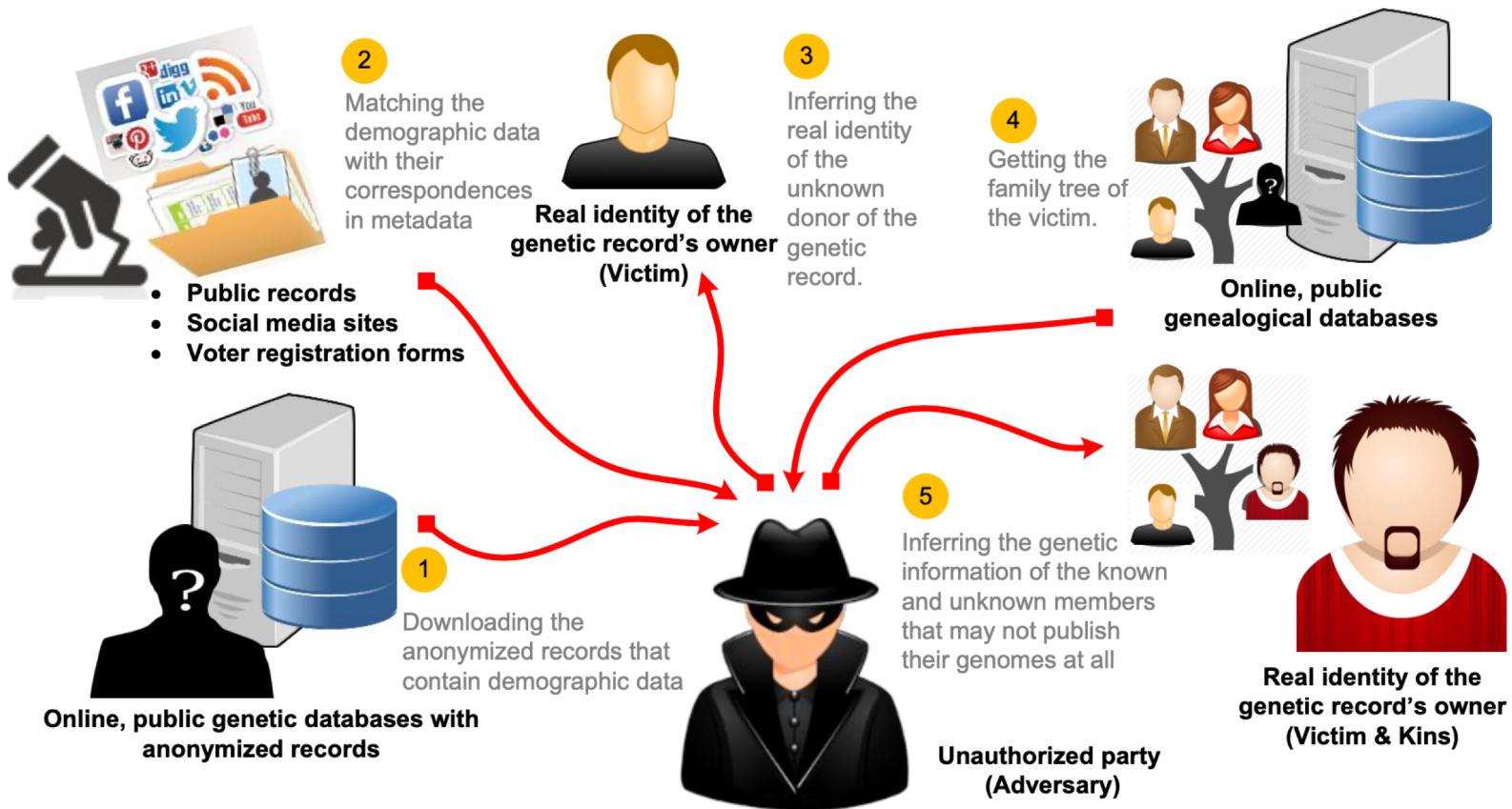
# Intelligent Architecture?



# Intelligent Architecture?



# Privacy-Preserving Genome Analysis?



**Fig. 5.** A completion attack.

Alser+, "[Can you really anonymize the donors of genomic data in today's digital world?](#)" 10th International Workshop on Data Privacy Management (DPM), 2015.

# Can you Really Anonymize the Donors?

---

## (Position Paper) Can You Really Anonymize the Donors of Genomic Data in Today's Digital World?

Mohammed Alser, Nour Almadhoun, Azita Nouri, Can Alkan, and Erman Ayday

Computer Engineering Department, Bilkent University, 06800 Bilkent, Ankara, Turkey

**Abstract.** The rapid progress in genome sequencing technologies leads to availability of high amounts of genomic data. Accelerating the pace of biomedical breakthroughs and discoveries necessitates not only collecting millions of genetic samples but also granting open access to genetic databases. However, one growing concern is the ability to protect the privacy of sensitive information and its owner. In this work, we survey a wide spectrum of cross-layer privacy breaching strategies to human genomic data (using both public genomic databases and other public non-genomic data). We outline the principles and outcomes of each technique, and assess its technological complexity and maturation. We then review potential privacy-preserving countermeasure mechanisms for each threat.

**Keywords:** Genomics, Privacy, Bioinformatics

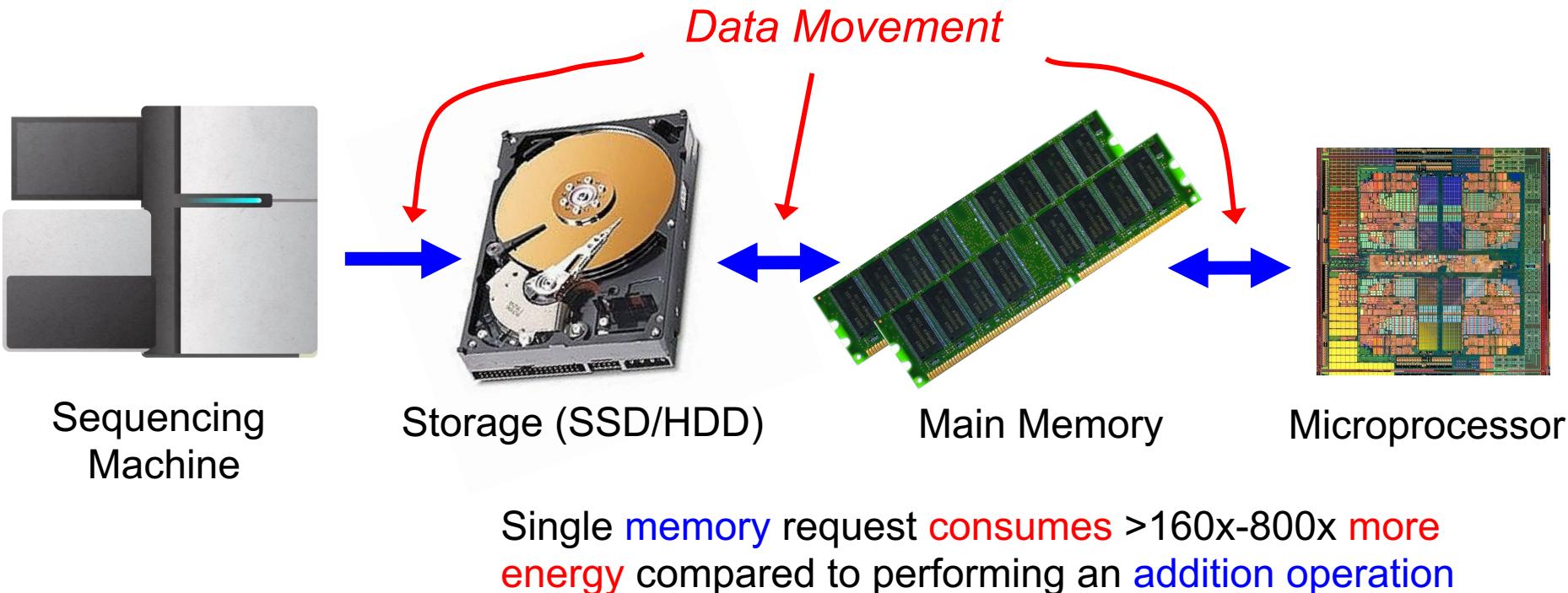


Alser+, "[Can you really anonymize the donors of genomic data in today's digital world?](#)" *10th International Workshop on Data Privacy Management (DPM)*, 2015.

---

# Pushing Towards New Architectures

- **Data movement** dominates performance and is a **major system energy bottleneck** (accounting for 40%-62%)

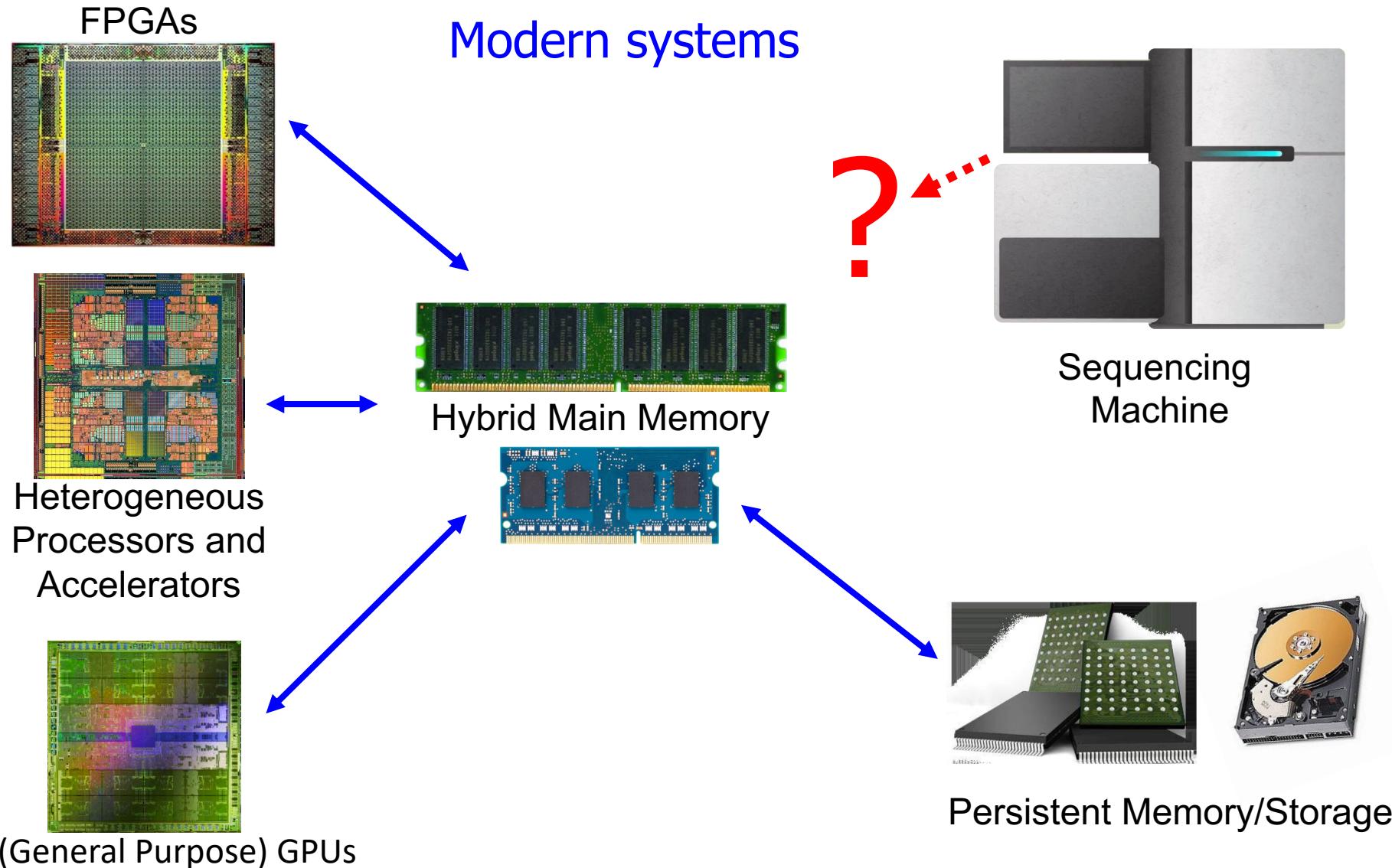


\* Boroumand et al., “Google Workloads for Consumer Devices: Mitigating Data Movement Bottlenecks,” ASPLOS 2018

★ Kestor et al., “Quantifying the Energy Cost of Data Movement in Scientific Applications,” IISWC 2013

★ Pandiyan and Wu, “Quantifying the energy cost of data movement for emerging smart phone workloads on mobile platforms,” IISWC 2014

# Processing Genomic Data Where it Makes Sense



# Achieving Intelligent Genome Analysis?

---

How and where to enable  
fast, accurate, cheap,  
privacy-preserving, and exabyte scale  
analysis of genomic data?

---

Most speedup comes from **parallelism** enabled  
by **novel architectures** and **algorithms**

# Read Mapping in 111 pages!

In-depth analysis of 107 read mappers (1988-2020)

**Mohammed Alser**, Jeremy Rotman, Dhrithi Deshpande, Kodi Taraszka, Huwenbo Shi, Pelin Icer Baykal, Harry Taegyun Yang, Victor Xue, Sergey Knyazev, Benjamin D. Singer, Brunilda Balliu, David Koslicki, Pavel Skums, Alex Zelikovsky, Can Alkan, Onur Mutlu, Serghei Mangul

"[Technology dictates algorithms: Recent developments in read alignment](#)"

Genome Biology, 2021

[[Source code](#)]

Alser *et al.* *Genome Biology* (2021) 22:249  
<https://doi.org/10.1186/s13059-021-02443-7>

Genome Biology

REVIEW

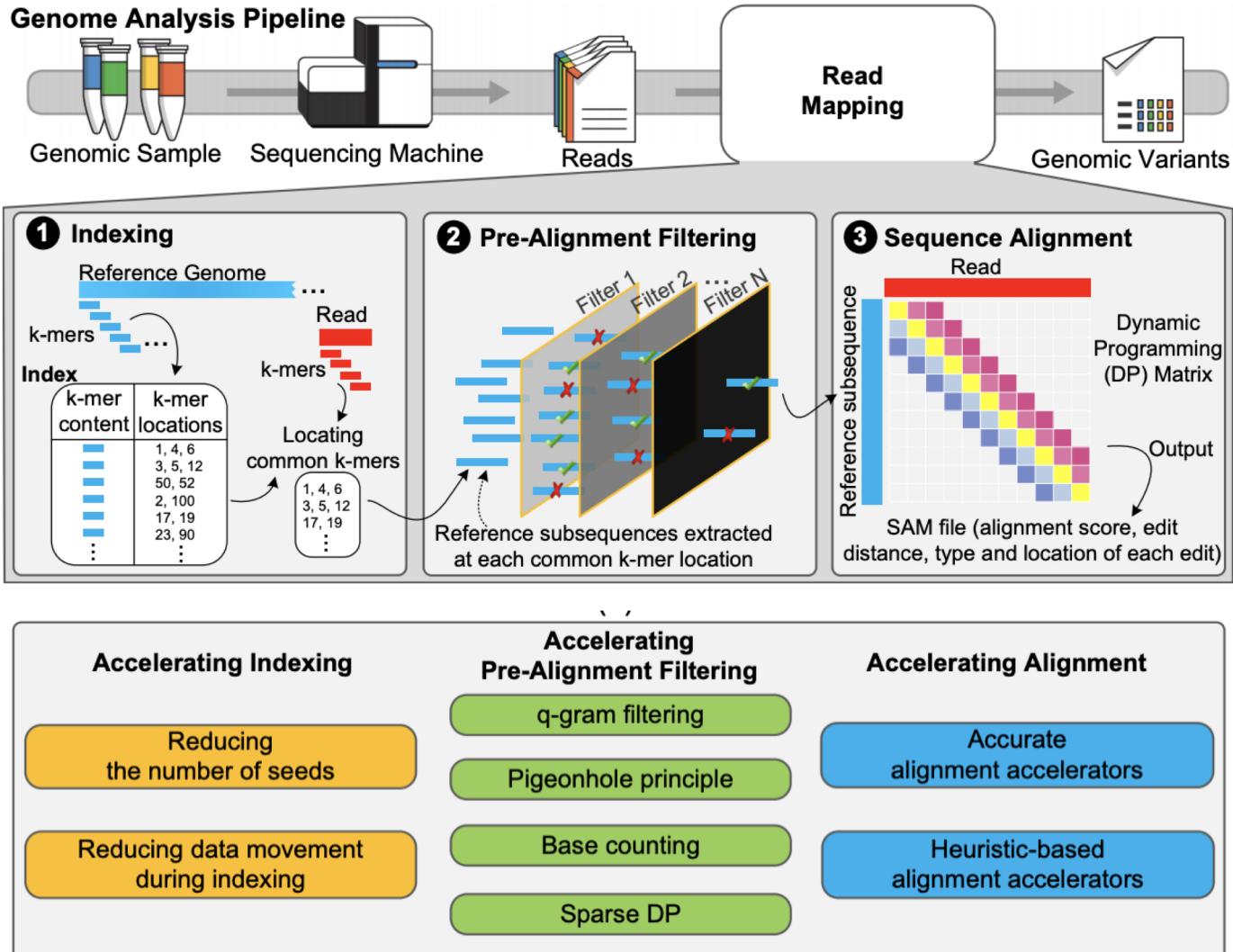
Open Access



## Technology dictates algorithms: recent developments in read alignment

Mohammed Alser<sup>1,2,3†</sup>, Jeremy Rotman<sup>4†</sup>, Dhrithi Deshpande<sup>5</sup>, Kodi Taraszka<sup>4</sup>, Huwenbo Shi<sup>6,7</sup>, Pelin Icer Baykal<sup>8</sup>, Harry Taegyun Yang<sup>4,9</sup>, Victor Xue<sup>4</sup>, Sergey Knyazev<sup>8</sup>, Benjamin D. Singer<sup>10,11,12</sup>, Brunilda Balliu<sup>13</sup>, David Koslicki<sup>14,15,16</sup>, Pavel Skums<sup>8</sup>, Alex Zelikovsky<sup>8,17</sup>, Can Alkan<sup>2,18</sup>, Onur Mutlu<sup>1,2,3†</sup> and Serghei Mangul<sup>5\*†</sup>

# Accelerating Read Mapping



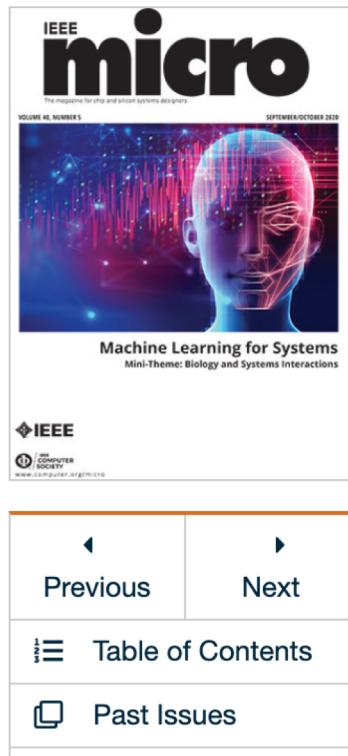
Alser+, “[Accelerating Genome Analysis: A Primer on an Ongoing Journey](#)”, IEEE Micro, 2020.

# Detailed Analysis of Tackling the Bottleneck

Mohammed Alser, Zülal Bingöl, Damla Senol Cali, Jeremie Kim, Saugata Ghose, Can Alkan, Onur Mutlu

[“Accelerating Genome Analysis: A Primer on an Ongoing Journey”](#)

IEEE Micro, August 2020.



[Home](#) / [Magazines](#) / [IEEE Micro](#) / [2020.05](#)

*IEEE Micro*

## Accelerating Genome Analysis: A Primer on an Ongoing Journey

Sept.-Oct. 2020, pp. 65-75, vol. 40

DOI Bookmark: [10.1109/MM.2020.3013728](https://doi.org/10.1109/MM.2020.3013728)

### Authors

[Mohammed Alser](#), ETH Zürich

[Zulal Bingol](#), Bilkent University

[Damla Senol Cali](#), Carnegie Mellon University

[Jeremie Kim](#), ETH Zurich and Carnegie Mellon University

[Saugata Ghose](#), University of Illinois at Urbana–Champaign and Carnegie Mellon University

[Can Alkan](#), Bilkent University

[Onur Mutlu](#), ETH Zurich, Carnegie Mellon University, and Bilkent University

# More on Accelerating Genome Analysis ...

■ Mohammed Alser,

**"Accelerating Genome Analysis: A Primer on an Ongoing Journey"**

Talk at [RECOMB 2021](#), Virtual, August 30, 2021.

[[Slides \(pptx\)](#) ([pdf](#))]

[[Talk Video](#) (27 minutes)]

[[Related Invited Paper \(at IEEE Micro, 2020\)](#)]

The slide is titled "Our Contributions" and is divided into two main sections:

- Near-memory/In-memory Pre-alignment Filtering** (Green box):
  - GRIM-Filter [[BMC Genomics'18](#)]
  - SneakySnake [[IEEE Micro'21](#)]
  - GenASM [[MICRO 2020](#)]
- Specialized Pre-alignment Filtering Accelerators (GPU, FPGA)** (Blue box):
  - GateKeeper [[Bioinformatics'17](#)]
  - MAGNET [[AACBB'18](#)]
  - Shouji [[Bioinformatics'19](#)]
  - GateKeeper-GPU [[arXiv'21](#)]
  - SneakySnake [[Bioinformatics'20](#)]

Below the sections, there is a diagram illustrating the flow of data between storage, memory, and the microprocessor. It shows a hard drive connected to Main Memory (RAM) and a Microprocessor (CPU). Red dashed arrows indicate the movement of data from storage to memory and from memory to the microprocessor. Blue arrows indicate the movement of data between storage and memory, and between memory and the microprocessor.

Promotional elements at the bottom left include a reminder to "Premieres in 23 hours" on "October 5, 4:30 PM" and a "Set reminder" button. The word "SAFARI" is also present.

Accelerating Genome Analysis: A Primer on an Ongoing Journey - RECOMB 2021 talk by Mohammed Alser

# More on Intelligent Genome Analysis ...

- Mohammed Alser,  
[\*\*"Computer Architecture - Lecture 8: Intelligent Genome Analysis"\*\*](#)  
ETH Zurich, Computer Architecture Course, Lecture 8, Virtual, 15 October 2021.  
[\[Slides \(pptx\) \(pdf\)\]](#)  
[\[Talk Video \(2 hour 54 minutes, including Q&A\)\]](#)  
[\[Related Invited Paper \(at IEEE Micro, 2020\)\]](#)

The screenshot shows a video player interface with a presentation slide on the left and a video feed of the speaker on the right.

**Our Solution: GateKeeper**

**1<sup>st</sup>** FPGA-based Alignment Filter.

**x10<sup>12</sup> mappings**

**x10<sup>3</sup> mappings**

High throughput DNA sequencing (HTS) technologies

Read Pre-Alignment Filtering

Read Alignment

Low Speed & High Accuracy  
Medium Speed, Medium Accuracy  
High Speed, Low Accuracy

Billions of Short Reads

108

SAFARI

2:08:58 / 2:54:18 · GateKeeper >

ETH ZENTRUM

Computer Architecture - Lecture 8: Intelligent Genome Analysis (ETH Zürich, Fall 2020)

44

# More on Fast Genome Analysis ...

- Onur Mutlu,

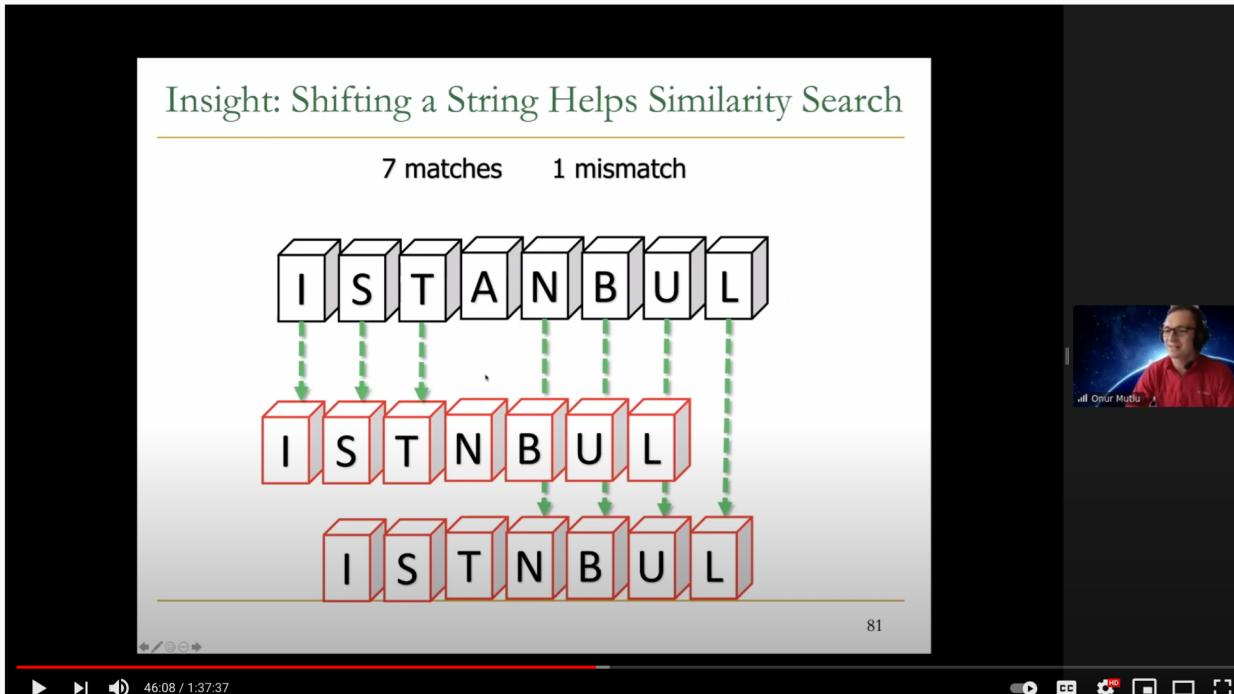
[\*\*"Accelerating Genome Analysis: A Primer on an Ongoing Journey"\*\*](#)

*Invited Lecture at [Technion](#), Virtual, 26 January 2021.*

[[Slides \(pptx\)](#) ([pdf](#))]

[[Talk Video](#) (1 hour 37 minutes, including Q&A)]

[[Related Invited Paper \(at IEEE Micro, 2020\)](#)]



Onur Mutlu - Invited Lecture @Technion: Accelerating Genome Analysis: A Primer on an Ongoing Journey

566 views • Premiered Feb 6, 2021

1 31 0 SHARE SAVE ...



Onur Mutlu Lectures  
13.9K subscribers

ANALYTICS EDIT VIDEO

# Detailed Lectures on Genome Analysis

---

- Computer Architecture, Fall 2020, Lecture 3a
  - **Introduction to Genome Sequence Analysis** (ETH Zürich, Fall 2020)
  - <https://www.youtube.com/watch?v=CrRb32v7SJc&list=PL5Q2soXY2Zi9xidyIgBxUz7xRPS-wisBN&index=5>
- Computer Architecture, Fall 2020, Lecture 8
  - **Intelligent Genome Analysis** (ETH Zürich, Fall 2020)
  - <https://www.youtube.com/watch?v=ygmQpdDTL7o&list=PL5Q2soXY2Zi9xidyIgBxUz7xRPS-wisBN&index=14>
- Computer Architecture, Fall 2020, Lecture 9a
  - **GenASM: Approx. String Matching Accelerator** (ETH Zürich, Fall 2020)
  - <https://www.youtube.com/watch?v=XoLpzmN-Pas&list=PL5Q2soXY2Zi9xidyIgBxUz7xRPS-wisBN&index=15>
- Accelerating Genomics Project Course, Fall 2020, Lecture 1
  - **Accelerating Genomics** (ETH Zürich, Fall 2020)
  - <https://www.youtube.com/watch?v=rgjl8ZyLsAg&list=PL5Q2soXY2Zi9E2bBVAgCqLgwiDRQDTyId>

# Prior Research on Genome Analysis (1 / 2)

---

- Alser+, "[Technology dictates algorithms: Recent developments in read alignment](#)", *Genome Biology*, 2021.
- Alser + "[SneakySnake: A Fast and Accurate Universal Genome Pre-Alignment Filter for CPUs, GPUs, and FPGAs.](#)", *Bioinformatics*, 2020.
- Senol Cali+, "[GenASM: A High-Performance, Low-Power Approximate String Matching Acceleration Framework for Genome Sequence Analysis](#)", *MICRO* 2020.
- Kim+, "[AirLift: A Fast and Comprehensive Technique for Translating Alignments between Reference Genomes](#)", *arXiv*, 2020
- Alser+, "[Accelerating Genome Analysis: A Primer on an Ongoing Journey](#)", *IEEE Micro*, 2020.

# Prior Research on Genome Analysis (2/2)

---

- Firtina+, "[Apollo: a sequencing-technology-independent, scalable and accurate assembly polishing algorithm](#)", *Bioinformatics*, 2019.
- Alser+, "[Shouji: a fast and efficient pre-alignment filter for sequence alignment](#)", *Bioinformatics* 2019.
- Kim+, "[GRIM-Filter: Fast Seed Location Filtering in DNA Read Mapping Using Processing-in-Memory Technologies](#)", *BMC Genomics*, 2018.
- Alser+, "[GateKeeper: A New Hardware Architecture for Accelerating Pre-Alignment in DNA Short Read Mapping](#)", *Bioinformatics*, 2017.
- Alser+, "[MAGNET: understanding and improving the accuracy of genome pre-alignment filtering](#)", *IPSI Transaction*, 2017.

# P&S Mobile Genomics

## Introduction & Project Proposals

Dr. Mohammed Alser



ETH Zurich

Fall 2021

5 October 2021