P&S Mobile Genomics

Lecture 1a: Course Introduction

Can Firtina

ETH Zürich
Spring 2023
2 March 2023

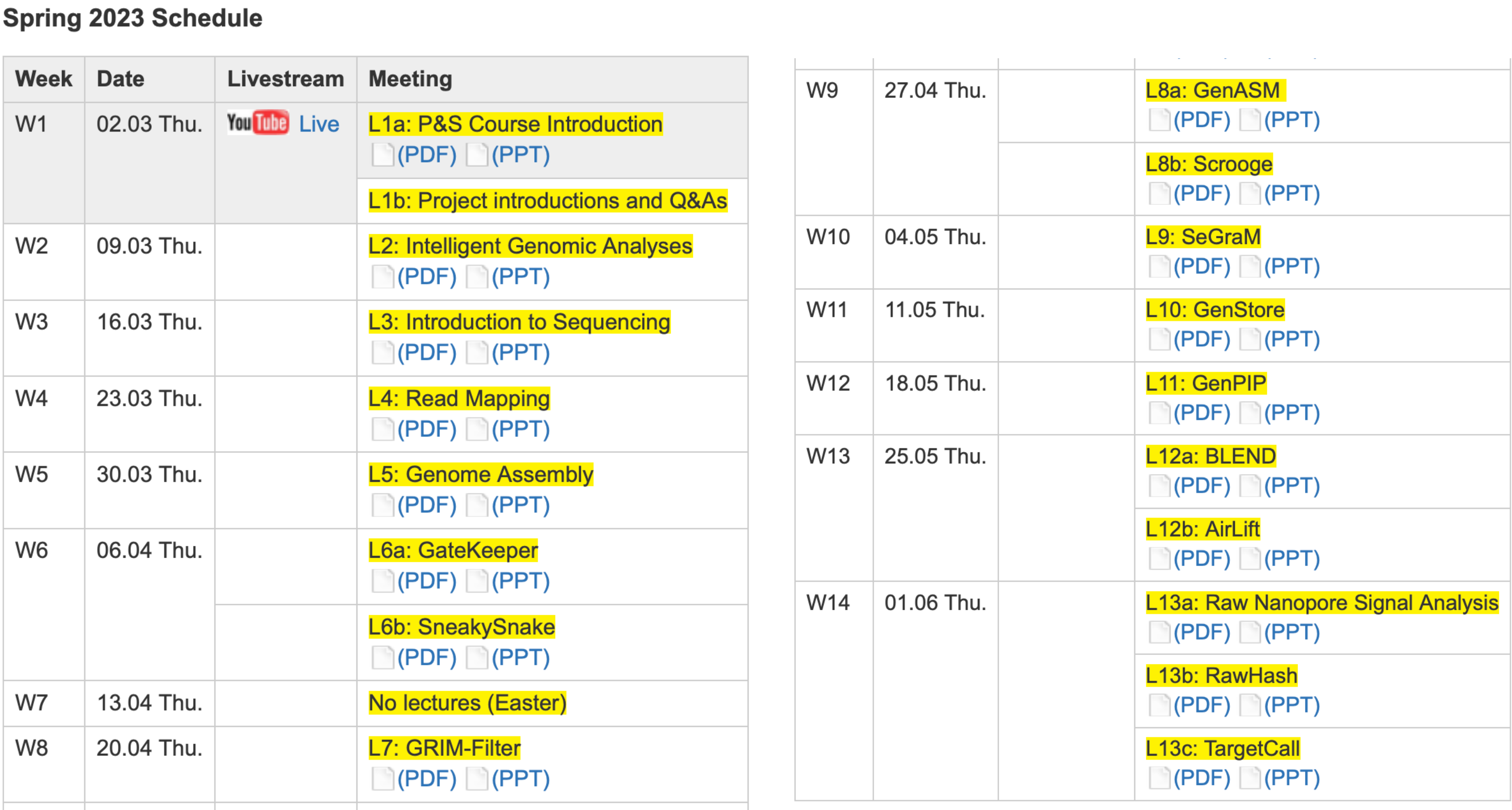
Two P&S Genomics Courses (Spring 2023)

- **P&S Mobile Genomics** (this course)
  - Designing hardware and software mainly for constrained computational resources (e.g., low-power designs)

- **P&S Accelerating Genomics**
  - Focuses on co-designing hardware and software to accelerate genome analysis

- Both are project-based courses
  - Taken by Bachelor’s/Master’s students
  - Weekly lectures on genomics
  - Hands-on research exploration with mentors
  - Improving research and presentation skills

Weekly Lectures on YouTube (Tentative)



Spring 2023 Schedule

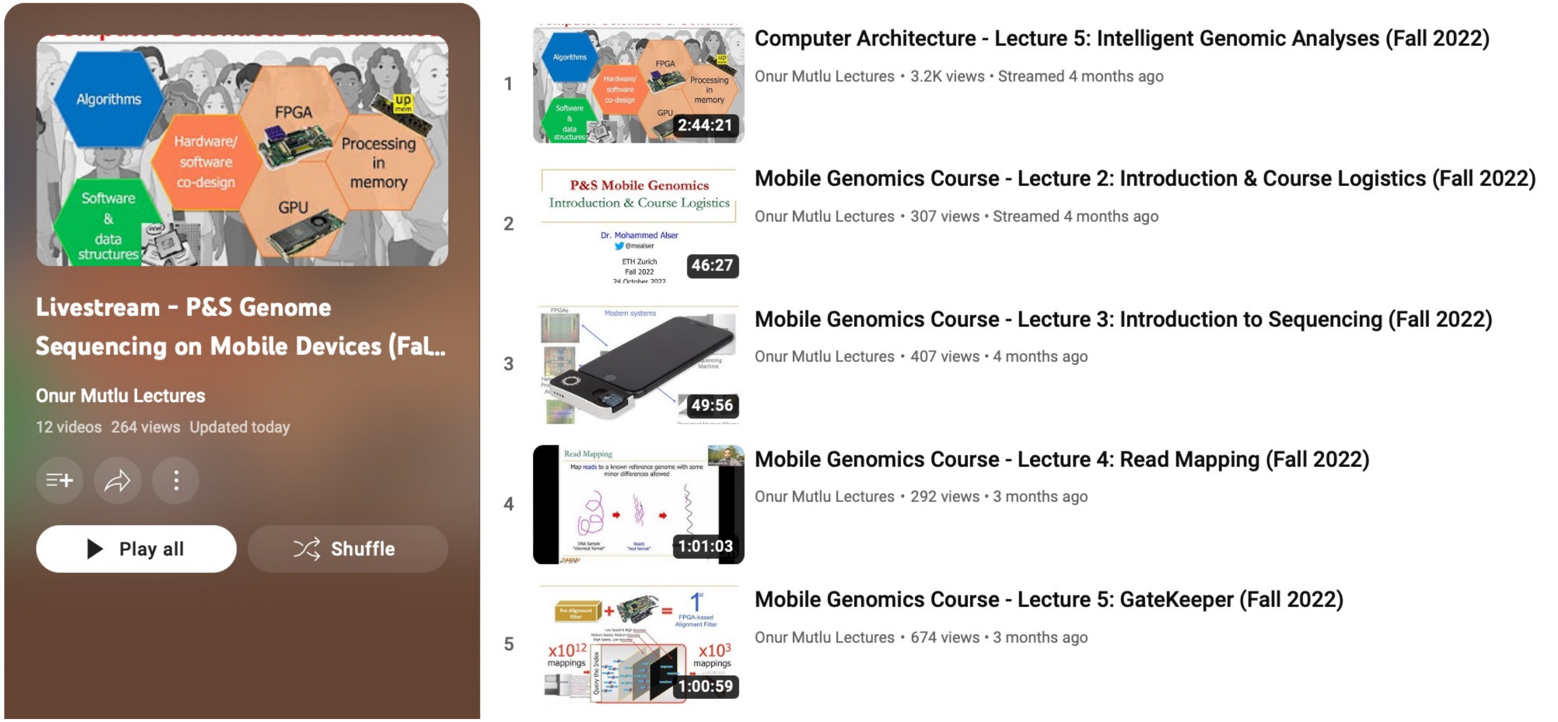
Week	Date	Livestream	Meeting
W1	02.03 Thu.	YouTube Live	L1a: P&S Course Introduction (PDF) (PPT)
			L1b: Project introductions and Q&As
W2	09.03 Thu.		L2: Intelligent Genomic Analyses (PDF) (PPT)
W3	16.03 Thu.		L3: Introduction to Sequencing (PDF) (PPT)
W4	23.03 Thu.		L4: Read Mapping (PDF) (PPT)
W5	30.03 Thu.		L5: Genome Assembly (PDF) (PPT)
W6	06.04 Thu.		L6a: GateKeeper (PDF) (PPT)
			L6b: SneakySnake (PDF) (PPT)
W7	13.04 Thu.		No lectures (Easter)
W8	20.04 Thu.		L7: GRIM-Filter (PDF) (PPT)
W9	27.04 Thu.		L8a: GenASM (PDF) (PPT)
			L8b: Scrooge (PDF) (PPT)
W10	04.05 Thu.		L9: SeGraM (PDF) (PPT)
W11	11.05 Thu.		L10: GenStore (PDF) (PPT)
W12	18.05 Thu.		L11: GenPIP (PDF) (PPT)
W13	25.05 Thu.		L12a: BLEND (PDF) (PPT)
			L12b: AirLift (PDF) (PPT)
W14	01.06 Thu.		L13a: Raw Nanopore Signal Analysis (PDF) (PPT)
			L13b: RawHash (PDF) (PPT)
			L13c: TargetCall (PDF) (PPT)

https://safari.ethz.ch/projects_and_seminars/spring2023/doku.php?id=genome_seq_mobile

P&S Mobile Genomics (Fall 2022)

- **Fall 2022 Edition:**
  - https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile

- **Youtube Livestream (Full list of lectures):**
  - https://www.youtube.com/playlist?list=PL5Q2soXY2Zi-adIrcamBmxfAhMbj5fmc4



Livestream - P&S Genome Sequencing on Mobile Devices (Fal...

Onur Mutlu Lectures

12 videos 264 views Updated today

1 Computer Architecture - Lecture 5: Intelligent Genomic Analyses (Fall 2022)
Onur Mutlu Lectures • 3.2K views • Streamed 4 months ago

2 Mobile Genomics Course - Lecture 2: Introduction & Course Logistics (Fall 2022)
Onur Mutlu Lectures • 307 views • Streamed 4 months ago

3 Mobile Genomics Course - Lecture 3: Introduction to Sequencing (Fall 2022)
Onur Mutlu Lectures • 407 views • 4 months ago

4 Mobile Genomics Course - Lecture 4: Read Mapping (Fall 2022)
Onur Mutlu Lectures • 292 views • 3 months ago

5 Mobile Genomics Course - Lecture 5: GateKeeper (Fall 2022)
Onur Mutlu Lectures • 674 views • 3 months ago

Past Semesters (P&S Mobile Genomics)

- **Fall 2022:** https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile
- **Spring 2022:** https://safari.ethz.ch/projects_and_seminars/spring2022/doku.php?id=genome_seq_mobile
- **Fall 2021:** https://safari.ethz.ch/projects_and_seminars/fall2021/doku.php?id=genome_seq_mobile
- **Spring 2021:** https://safari.ethz.ch/projects_and_seminars/spring2021/doku.php?id=genome_seq_mobile
- **Fall 2020:** https://safari.ethz.ch/projects_and_seminars/fall2020/doku.php?id=genome_seq_mobile

The Role of This Course

Projects & Seminars: Mobile Genomics

- **Weekly lectures** will cover the **basics** of **genome analysis** to understand the **computational steps** of the entire genome analysis pipeline.

- **Hands-on projects with mentors**
  - **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 **computationally constrained resources** (e.g., mobile phones) 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.
- Interest in making things efficient and solving problems
- A good knowledge in C or C++ programming language is required.

Course Requirements and Expectations

- Each student will carry out a hands-on project
  - Build, implement, code, and design with close engagement from the mentors
  - Meet regularly with your mentors
- Participation
  - Follow the lectures to understanding the basics of genome analysis
  - Ask questions, contribute thoughts/ideas
  - Read the relevant papers that your mentor may suggest
- Presentation & GitHub repository

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

Your Responsibilities

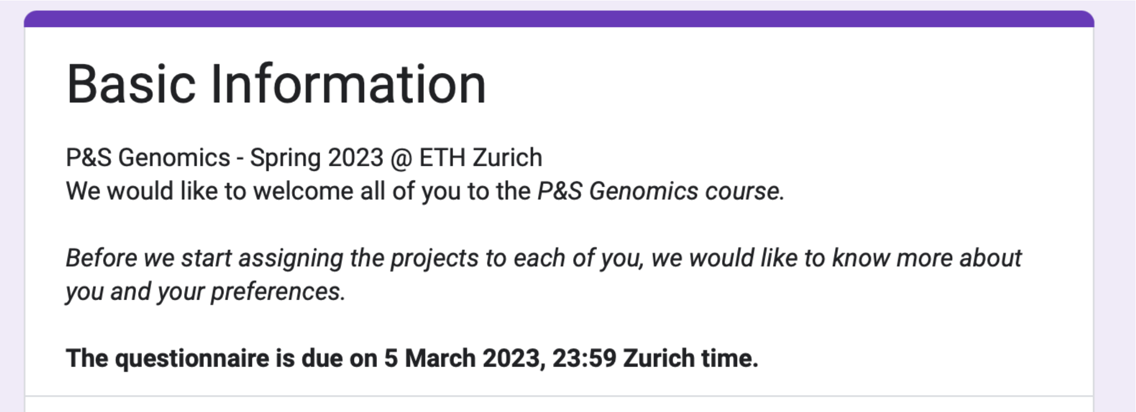
- Weekly lectures
  - Schedule is on the [course website]()
  - Following the lectures is strongly suggested

- Working on your project for ~6 hours per week

- Regularly meeting with your mentors (e.g., weekly) is required

- Regularly follow the course website, Moodle, RingCentral, and your e-mails
  - Respond to your mentors timely.

Assignment 1: Basic Information

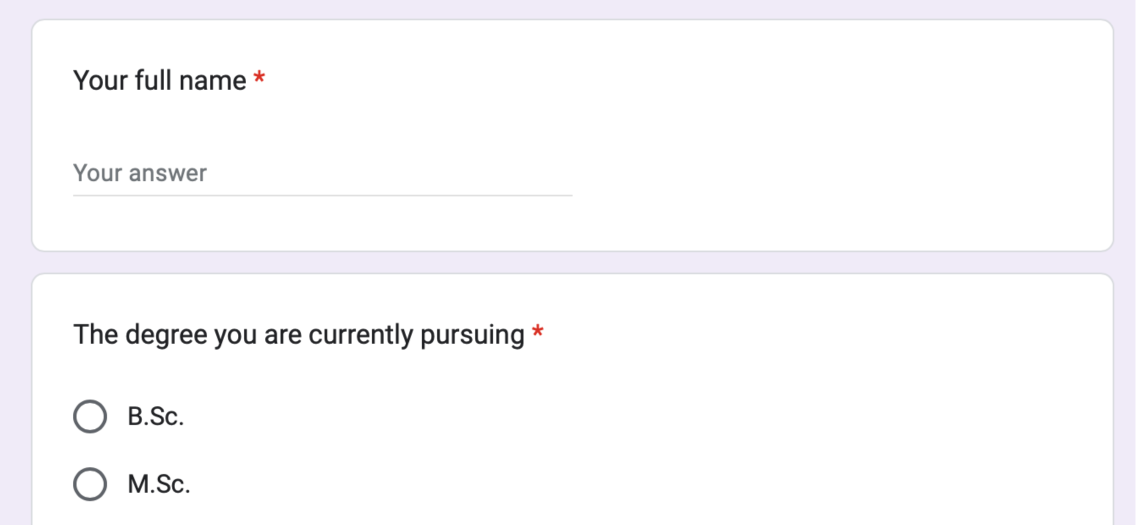


Basic Information

P&S Genomics - Spring 2023 @ ETH Zurich
We would like to welcome all of you to the *P&S Genomics course.*

Before we start assigning the projects to each of you, we would like to know more about you and your preferences.

The questionnaire is due on 5 March 2023, 23:59 Zurich time.



Your full name *

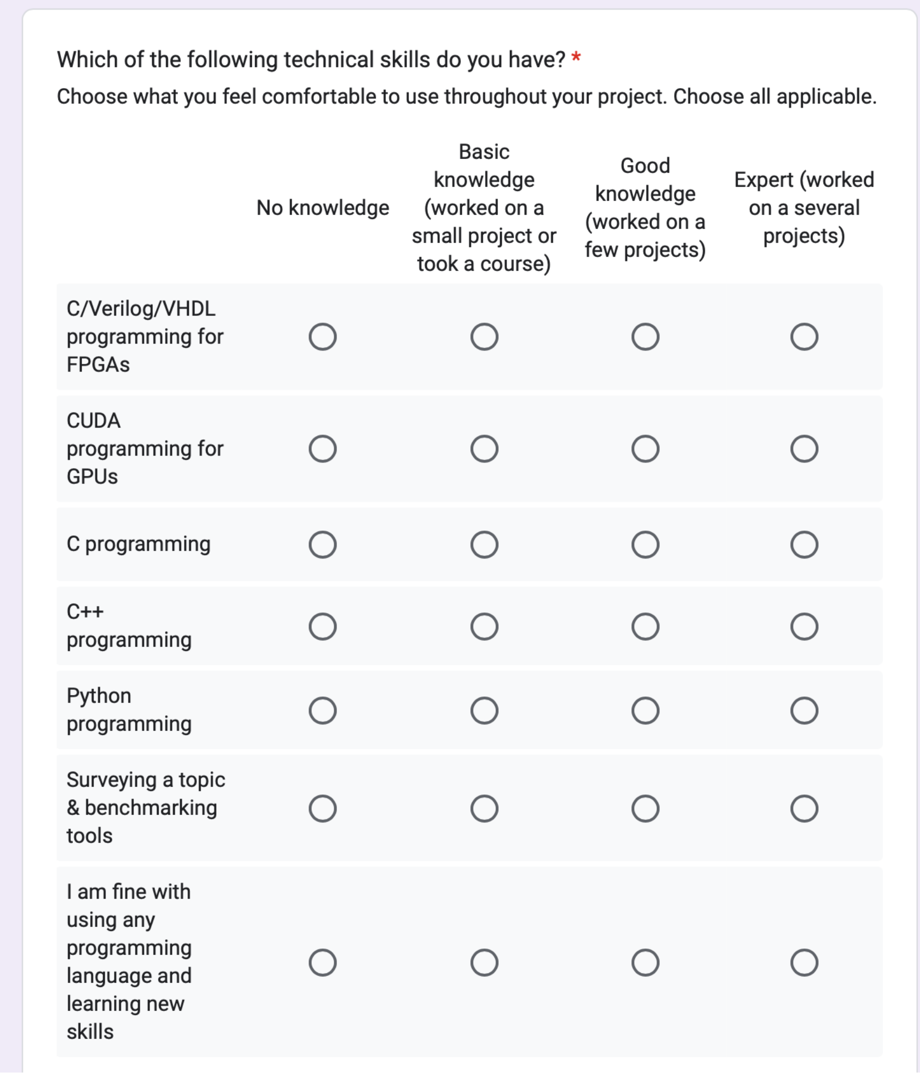
Your answer

The degree you are currently pursuing *

○ B.Sc.

○ M.Sc.

Assignment 1: Project Preferences



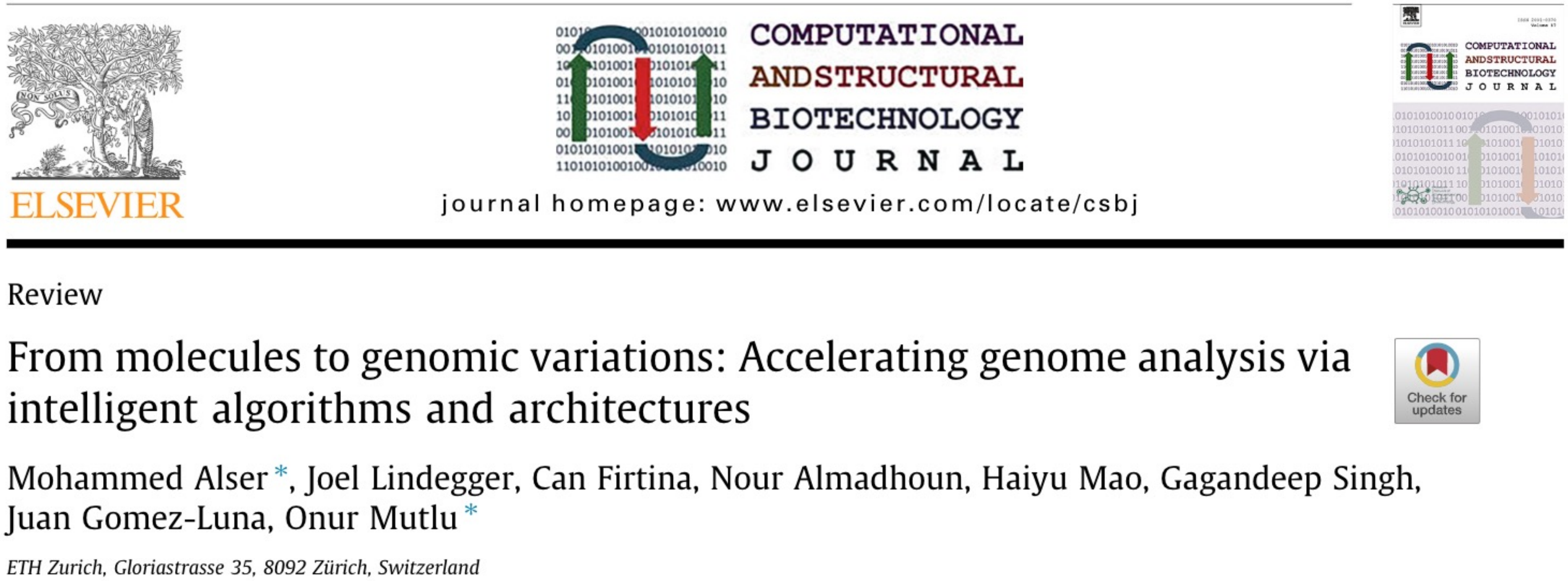
Which of the following technical skills do you have? *

Choose what you feel comfortable to use throughout your project. Choose all applicable.

	No knowledge	Basic knowledge (worked on a small project or took a course)	Good knowledge (worked on a few projects)	Expert (worked on a several projects)
C/Verilog/VHDL programming for FPGAs	○	○	○	○
CUDA programming for GPUs	○	○	○	○
C programming	○	○	○	○
C++ programming	○	○	○	○
Python programming	○	○	○	○
Surveying a topic & benchmarking tools	○	○	○	○
I am fine with using any programming language and learning new skills	○	○	○	○

Assignment 2: Intelligent Genome Analysis

Mohammed Alser, Joel Lindegger, Can Firtina, Nour Almadhoun, Haiyu Mao, Gagandeep Singh, Juan Gomez-Luna, Onur Mutlu
"From Molecules to Genomic Variations: Intelligent Algorithms and Architectures for Intelligent Genome Analysis"
Computational and Structural Biotechnology Journal, 2022
[Source code]



ELSEVIER

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Review

From molecules to genomic variations: Accelerating genome analysis via intelligent algorithms and architectures

Check for updates

Mohammed Alser *, Joel Lindegger, Can Firtina, Nour Almadhoun, Haiyu Mao, Gagandeep Singh, Juan Gomez-Luna, Onur Mutlu *

ETH Zurich, Gloriastrasse 35, 8092 Zürich, Switzerland

Project Assignment

- Mentors will explain the projects in our in-person meeting
  - **March 2 (today) at 14:15**
  - Attend or watch the recording

- Complete assignments #1 and #2

- **Select the top 5 projects** you want to work on
  - You can select them on Moodle starting from **March 3**
  - **Deadline: March 8, 23:59.**
  - We **match** your interests, skills, and background with a suitable project.
  - We assign you a project and put you in touch with a mentor.

Course Mentors: Who Are We?

- **Can Firtina**
  - Ph.D. Student in the SAFARI Research Group

- Research interests:
  - Bioinformatics, real-time genome analysis, genome editing, hardware acceleration of genome analysis

- Get to know us and our research
  - https://safari.ethz.ch/
  - Contact me: canfirtina@gmail.com. Personal academic website: https://cfirtina.com
  - Twitter: https://twitter.com/FirtinaC

Course Mentors: Who Are We? (I)

Mohammed Alser

Senior Researcher and Lecturer

Bioinformatics | Computational Genomics | Hardware/Software Cooperation | Specialized/Heterogeneous Computing Systems | Processing-in-Memory

Juan Gómez Luna

Senior Researcher and Lecturer

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

Haiyu Mao

Senior Researcher

Non-Volatile Memories | Processing-In-Memory | Memory Security | Machine Learning | Heterogeous Memory Systems |

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

Course Mentors: Who Are We? (II)

Can Firtina

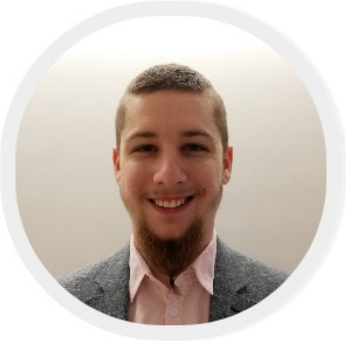
PhD Student

Bioinformatics | Real-time signal analysis | Genome editing | Accelerating Genome Analysis

Nika Mansourighiasi

PhD Student

Processing-In-Memory | Emerging Memory & Processing Technologies



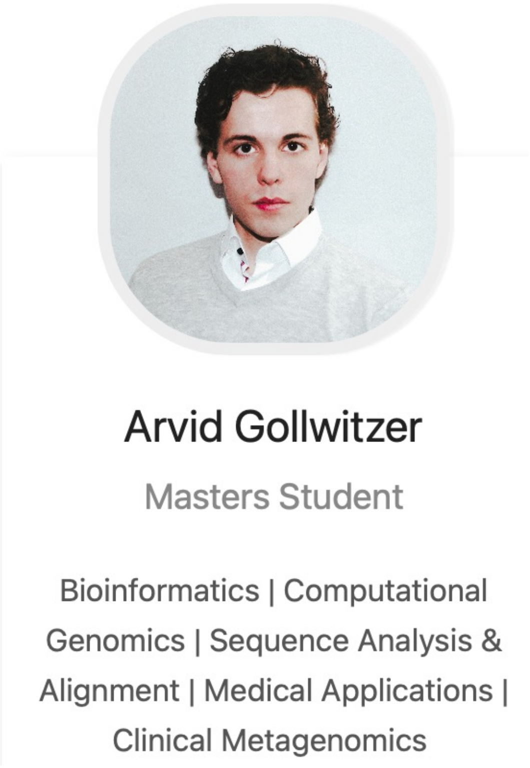
Joël Lindegger

PhD Student

Acceleration of the bioinformatics pipeline | Current and future computer architectures | All kinds of algorithms and data structures

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

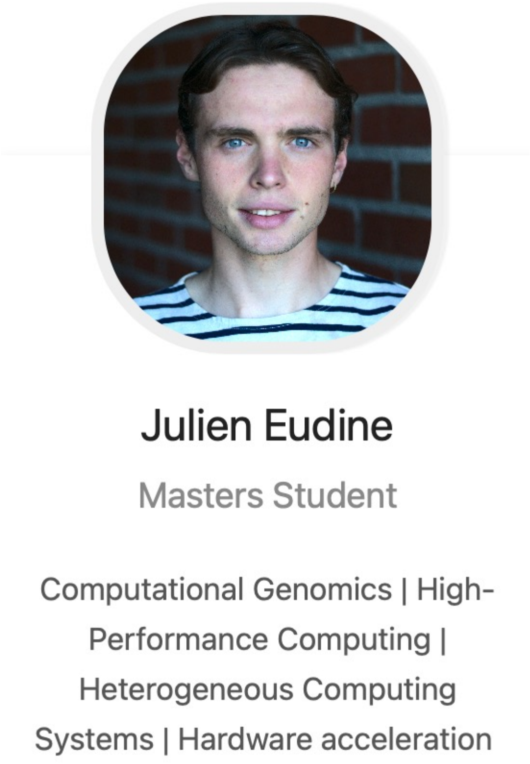
Course Mentors: Who Are We? (III)



Arvid Gollwitzer

Masters Student

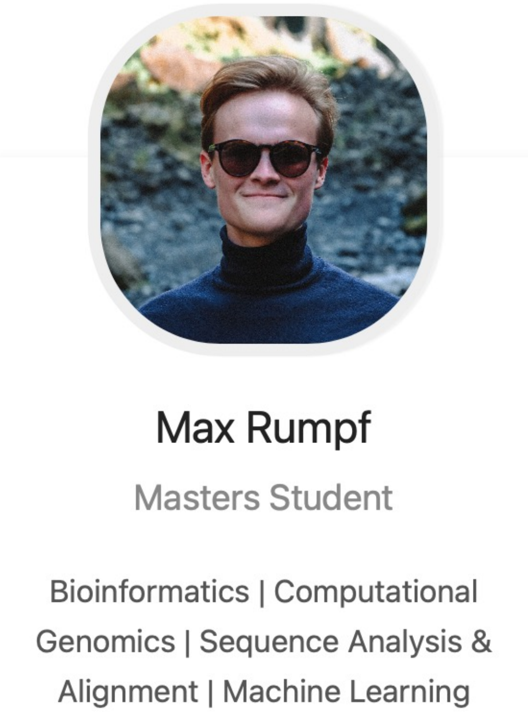
Bioinformatics | Computational Genomics | Sequence Analysis & Alignment | Medical Applications | Clinical Metagenomics



Julien Eudine

Masters Student

Computational Genomics | High-Performance Computing | Heterogeneous Computing Systems | Hardware acceleration



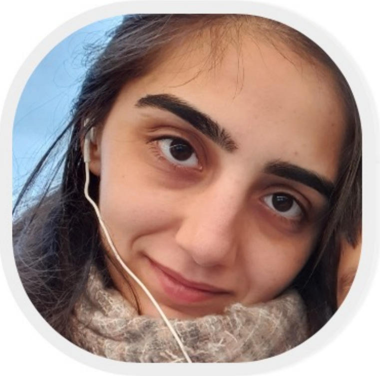
Max Rumpf

Masters Student

Bioinformatics | Computational Genomics | Sequence Analysis & Alignment | Machine Learning

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

Course Mentors: Who Are We? (IV)



Banu Cavlak

Masters Student

Bioinformatics | Computational Genomics | DRAM | Machine Learning

Younjoo Lee

Mobility Student

Computer architecture | FPGA-acceleration | Bioinformatics | Processing in Memory.

Filipe Mulonde

Visiting Research Fellow, Swiss Government Excellence Scholar

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

What is Intelligent Genome Analysis?

- Fast genome analysis
  - *Real-time analysis*

Bandwidth

- Using intelligent architectures
  - *Specialized HW with less data movement*

Energy-efficiency & Latency

- DNA is a valuable asset
  - *Controlled-access analysis*

Privacy

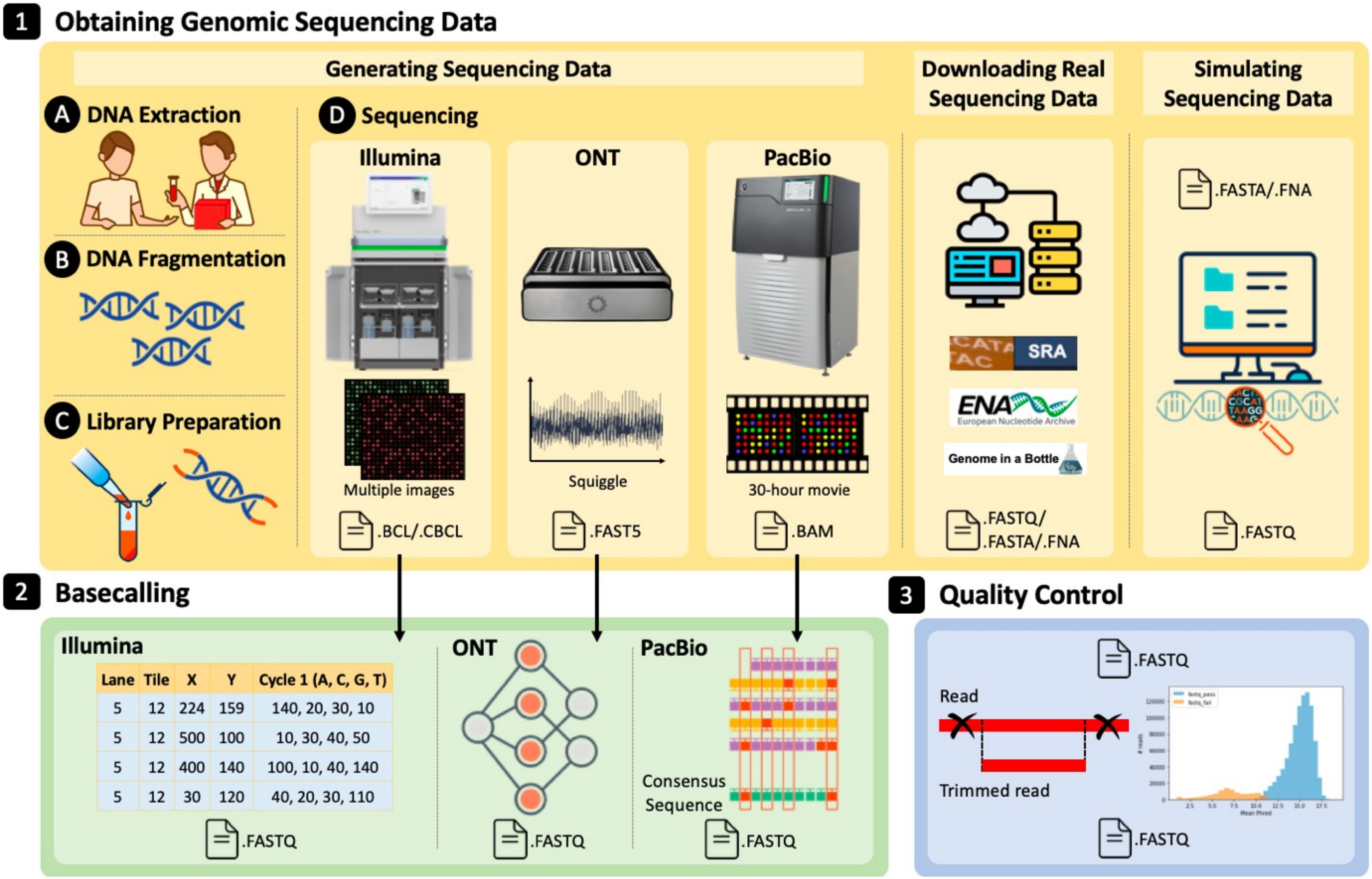
- Population-scale genome analysis
  - *Sequence anywhere at large scale*

Scalability

- Avoiding erroneous analysis
  - *Incorrect diagnosis of disease*

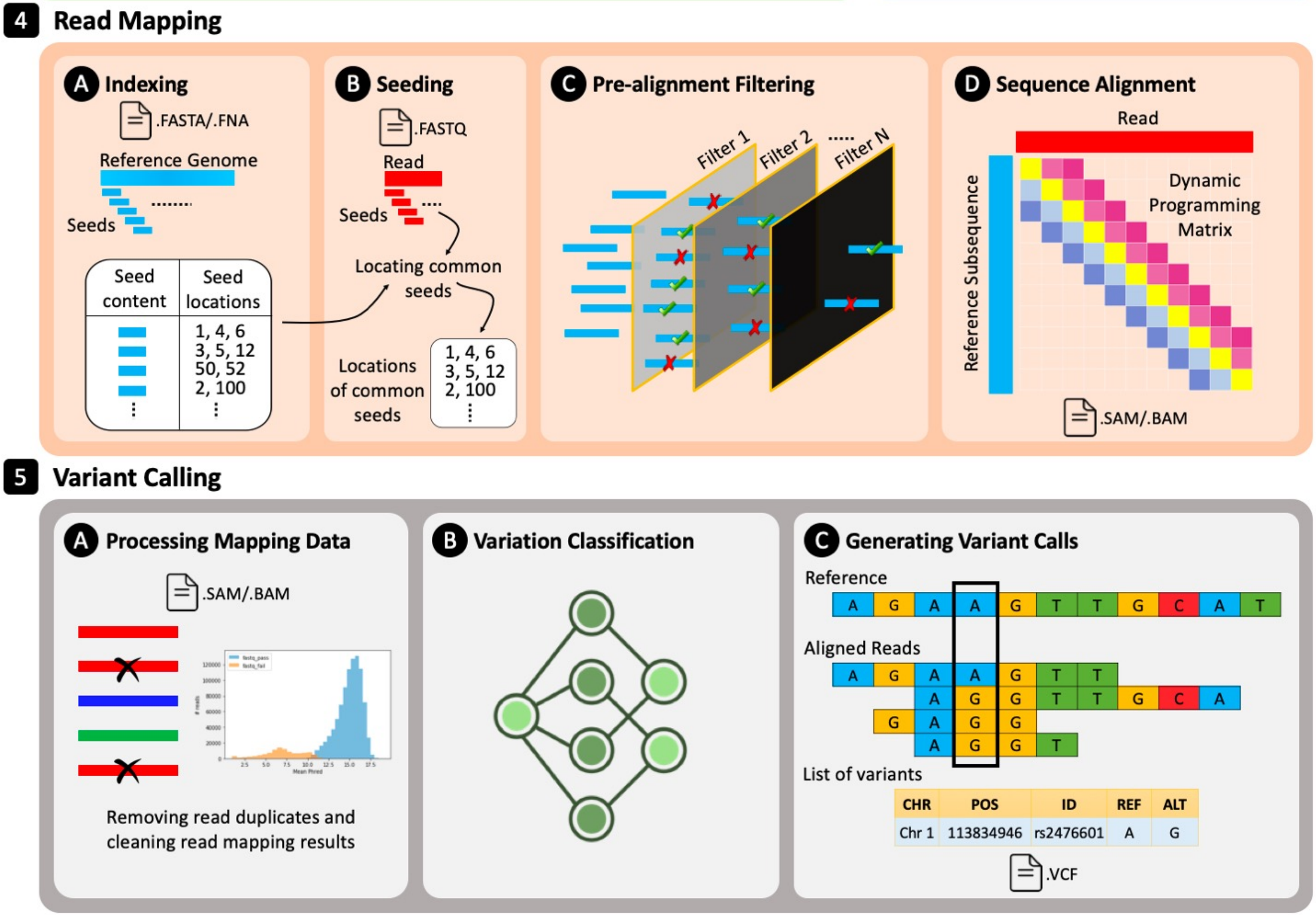
Accuracy

Next Meetings



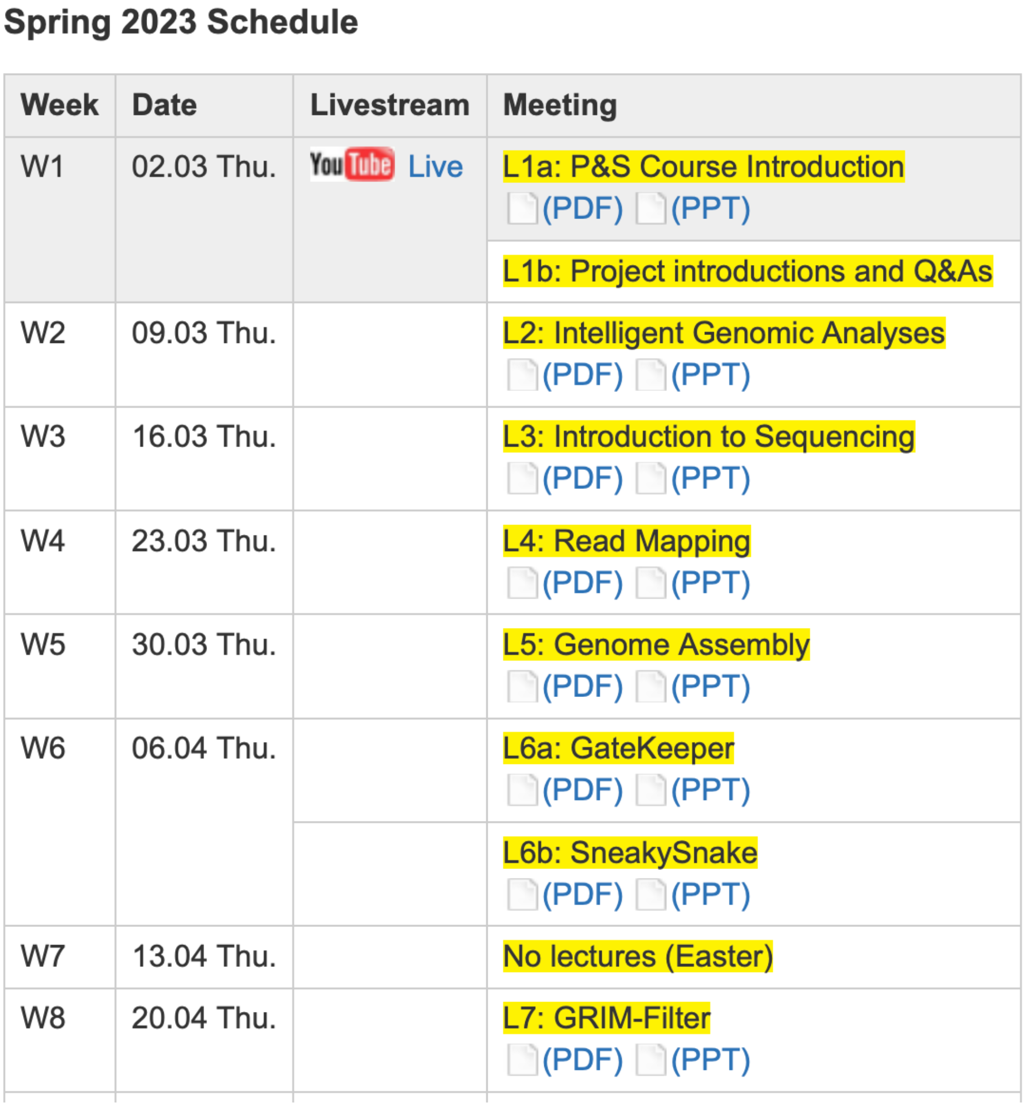
Lane	Tile	X	Y	Cycle 1 (A, C, G, T)
5	12	224	159	140, 20, 30, 10
5	12	500	100	10, 30, 40, 50
5	12	400	140	100, 10, 40, 140
5	12	30	120	40, 20, 30, 110

Next Meetings



CHR	POS	ID	REF	ALT
Chr 1	113834946	rs2476601	A	G

Topics To Be Covered I (Tentative)



Spring 2023 Schedule

Week	Date	Livestream	Meeting
W1	02.03 Thu.	YouTube Live	L1a: P&S Course Introduction (PDF) (PPT)
			L1b: Project introductions and Q&As
W2	09.03 Thu.		L2: Intelligent Genomic Analyses (PDF) (PPT)
W3	16.03 Thu.		L3: Introduction to Sequencing (PDF) (PPT)
W4	23.03 Thu.		L4: Read Mapping (PDF) (PPT)
W5	30.03 Thu.		L5: Genome Assembly (PDF) (PPT)
W6	06.04 Thu.		L6a: GateKeeper (PDF) (PPT)
			L6b: SneakySnake (PDF) (PPT)
W7	13.04 Thu.		No lectures (Easter)
W8	20.04 Thu.		L7: GRIM-Filter (PDF) (PPT)

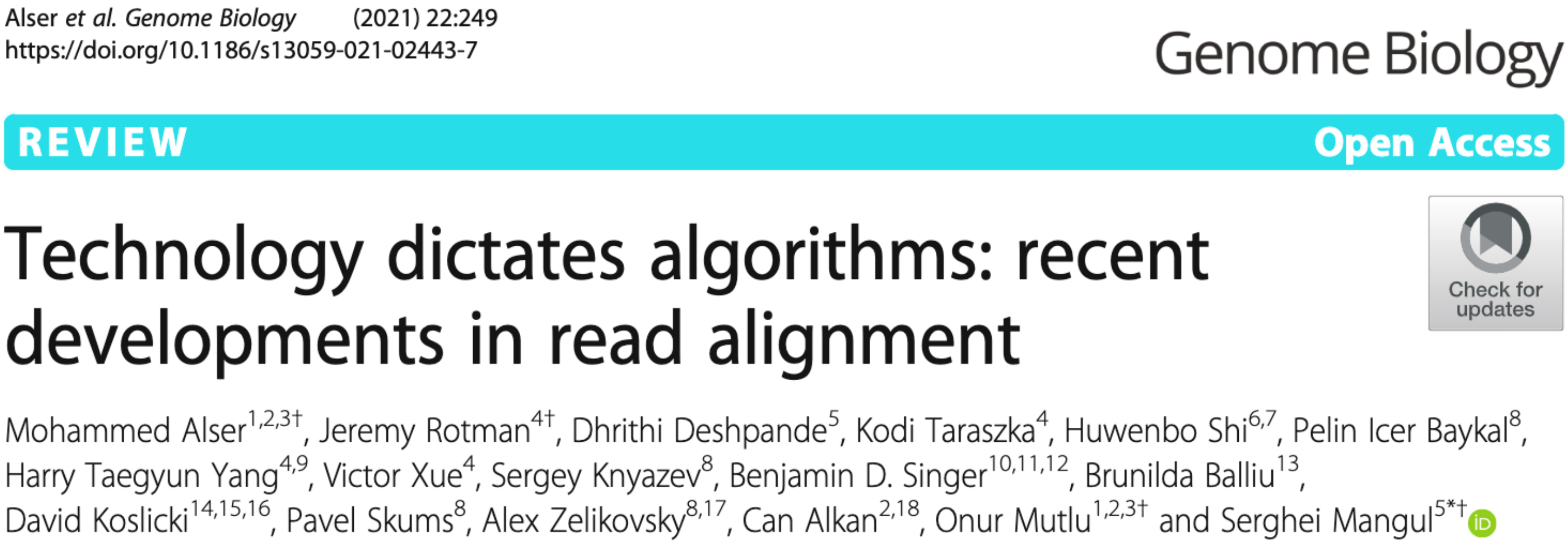
Topics To Be Covered II (Tentative)

W9	27.04 Thu.		L8a: GenASM (PDF) (PPT)
			L8b: Scrooge (PDF) (PPT)
W10	04.05 Thu.		L9: SeGraM (PDF) (PPT)
W11	11.05 Thu.		L10: GenStore (PDF) (PPT)
W12	18.05 Thu.		L11: GenPIP (PDF) (PPT)
W13	25.05 Thu.		L12a: BLEND (PDF) (PPT)
			L12b: AirLift (PDF) (PPT)
W14	01.06 Thu.		L13a: Raw Nanopore Signal Analysis (PDF) (PPT)
			L13b: RawHash (PDF) (PPT)
			L13c: TargetCall (PDF) (PPT)

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

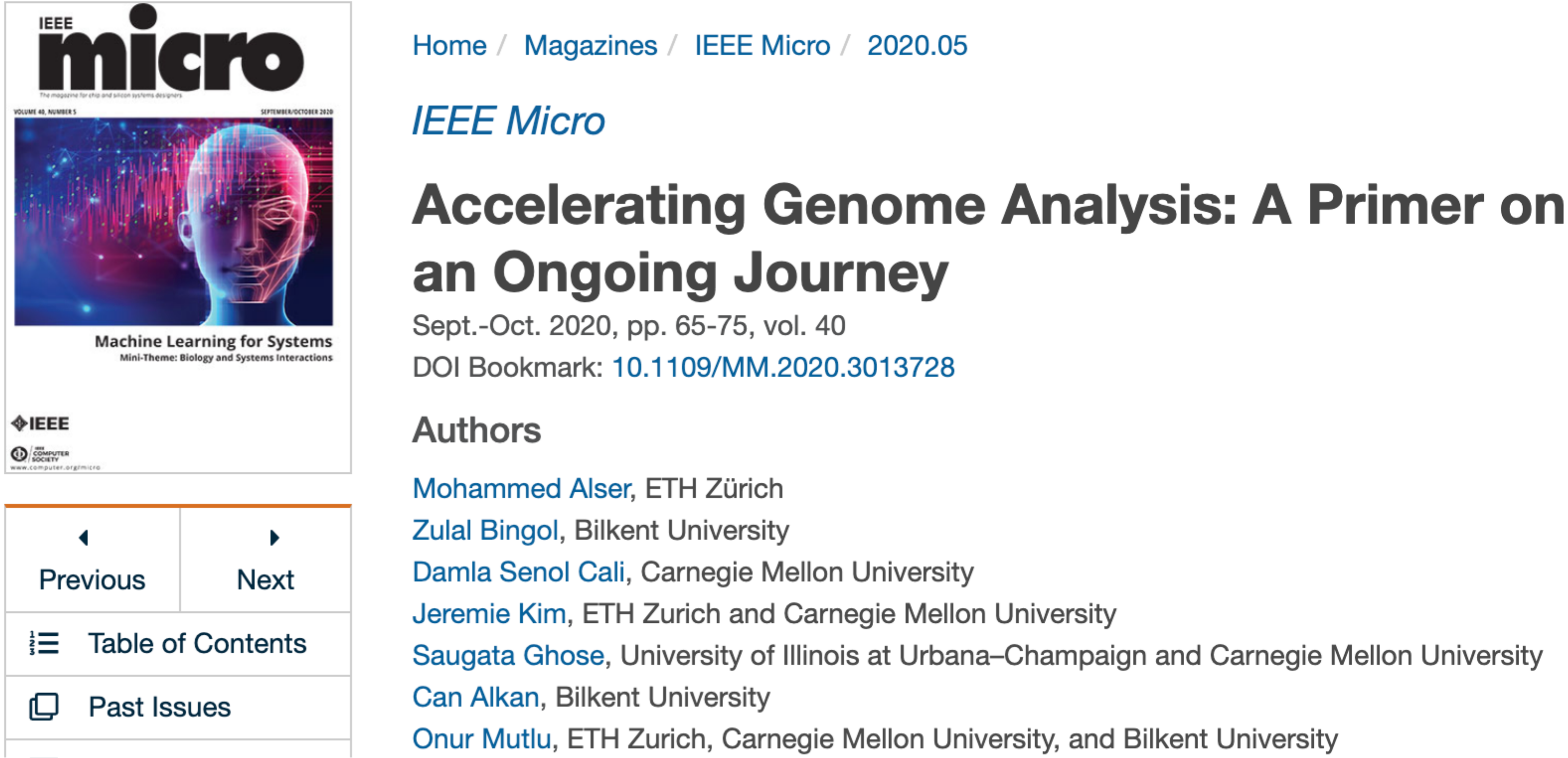
REVIEW **Open Access**

Technology dictates algorithms: recent developments in read alignment

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

Near-memory Pre-alignment Filtering

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

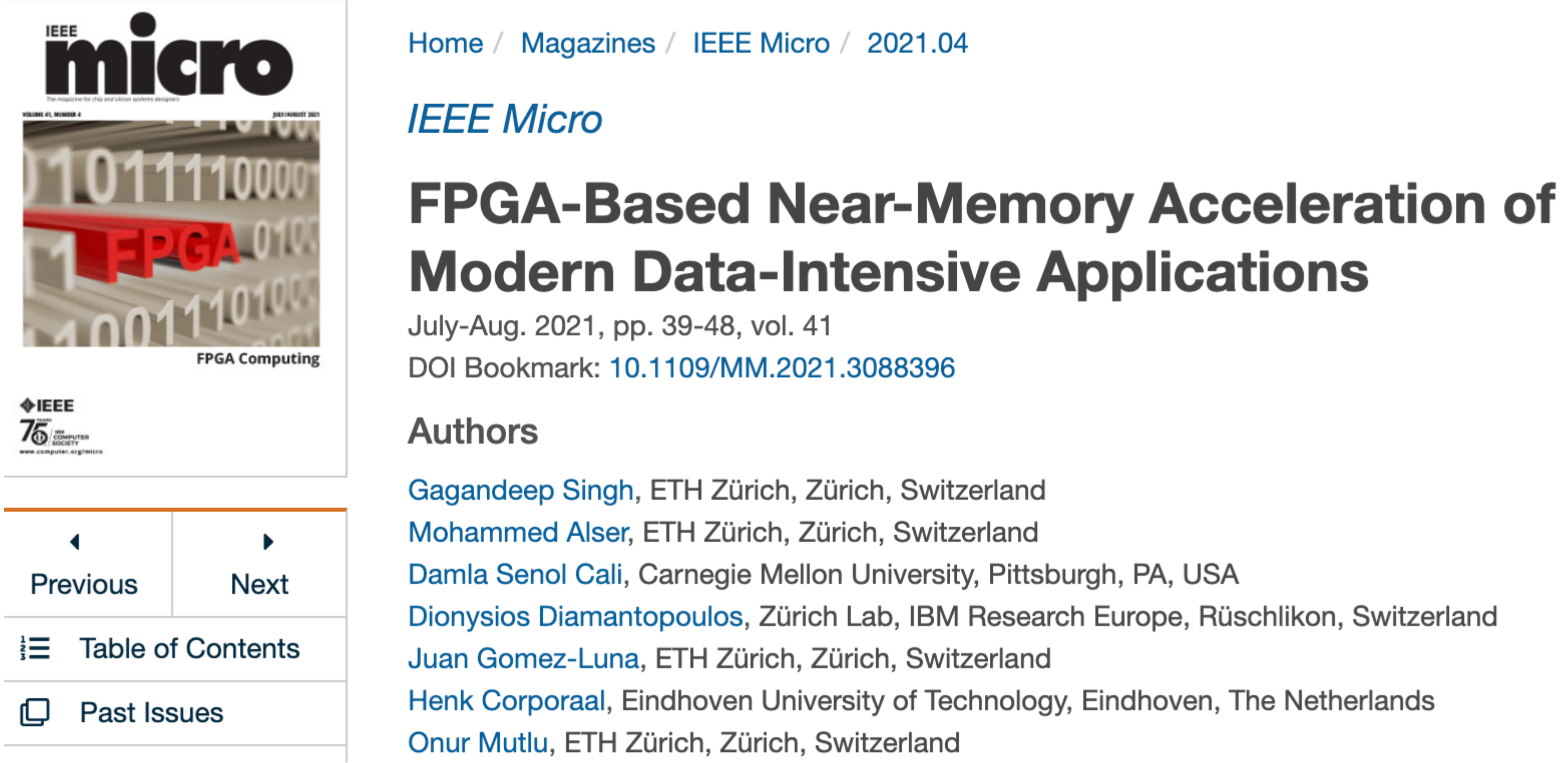
Sept.-Oct. 2020, pp. 65-75, vol. 40
DOI Bookmark: 10.1109/MM.2020.3013728

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

Near-memory Pre-alignment Filtering

Gagandeep Singh, Mohammed Alser, Damla Senol Cali, Dionysios Diamantopoulos, Juan Gomez-Luna, Henk Corporaal, Onur Mutlu,
"FPGA-Based Near-Memory Acceleration of Modern Data-Intensive Applications"
IEEE Micro, 2021.
[Source Code]



Home / Magazines / IEEE Micro / 2021.04

IEEE Micro

FPGA-Based Near-Memory Acceleration of Modern Data-Intensive Applications

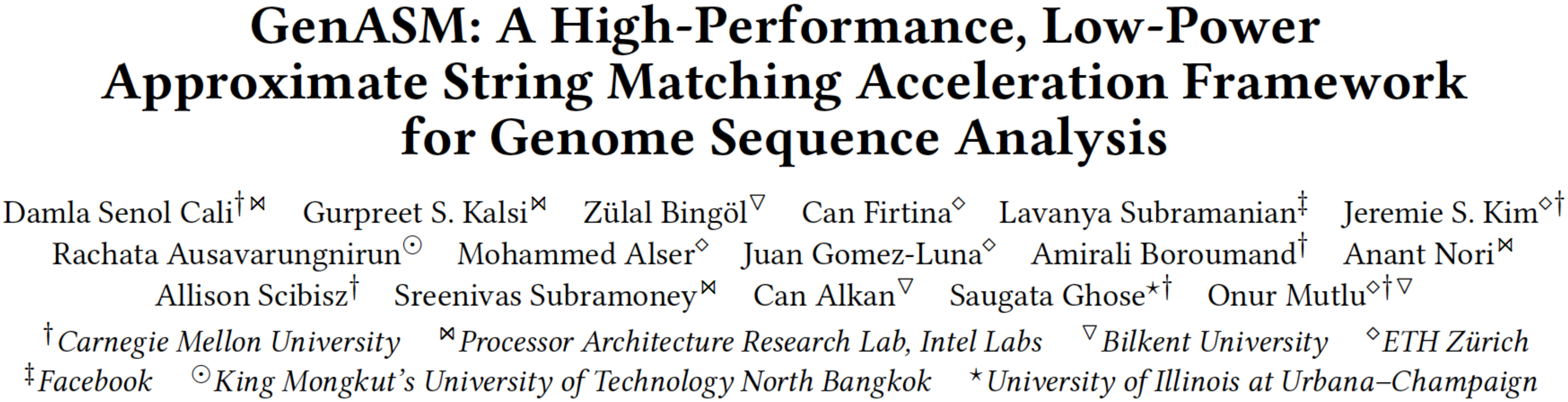
July-Aug. 2021, pp. 39-48, vol. 41
DOI Bookmark: 10.1109/MM.2021.3088396

Authors

Gagandeep Singh, ETH Zürich, Zürich, Switzerland
Mohammed Alser, ETH Zürich, Zürich, Switzerland
Damla Senol Cali, Carnegie Mellon University, Pittsburgh, PA, USA
Dionysios Diamantopoulos, Zürich Lab, IBM Research Europe, Rüschlikon, Switzerland
Juan Gomez-Luna, ETH Zürich, Zürich, Switzerland
Henk Corporaal, Eindhoven University of Technology, Eindhoven, The Netherlands
Onur Mutlu, ETH Zürich, Zürich, Switzerland

GenASM Framework **[MICRO 2020]**

Damla Senol Cali, Gurpreet S. Kalsi, Zulal Bingol, Can Firtina, Lavanya Subramanian, Jeremie S. Kim, Rachata Ausavarungnirun, Mohammed Alser, Juan Gomez-Luna, Amirali Boroumand, Anant Nori, Allison Scibisz, Sreenivas Subramoney, Can Alkan, Saugata Ghose, and Onur Mutlu,
"GenASM: A High-Performance, Low-Power Approximate String Matching Acceleration Framework for Genome Sequence Analysis"
Proceedings of the 53rd International Symposium on Microarchitecture (***MICRO***), Virtual, October 2020.
[Lightning Talk Video (1.5 minutes)]
[Lightning Talk Slides (pptx) (pdf)]
[Talk Video (18 minutes)]
[Slides (pptx) (pdf)]

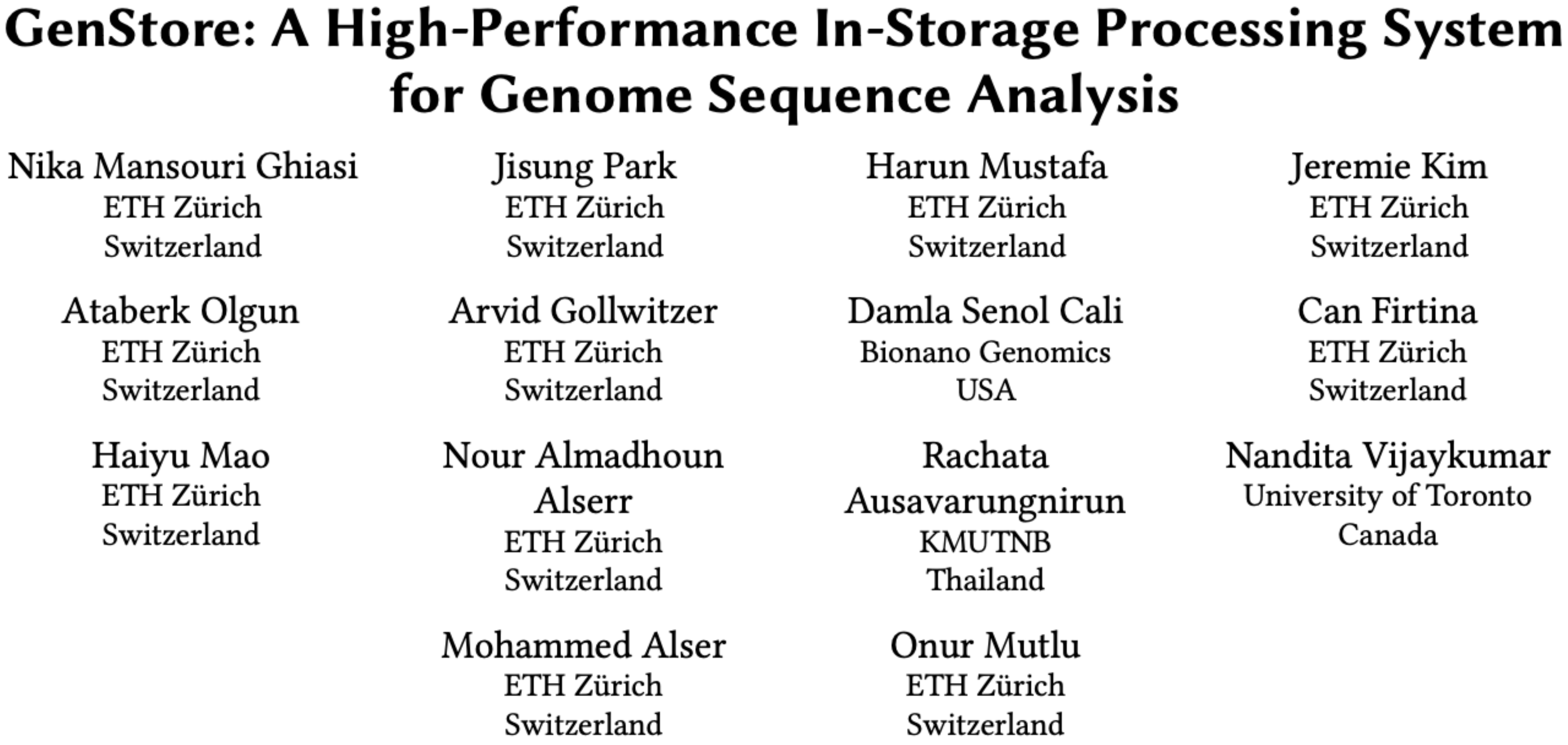


GenASM: A High-Performance, Low-Power Approximate String Matching Acceleration Framework for Genome Sequence Analysis

Damla Senol Cali[†⋈] Gurpreet S. Kalsi[⋈] Zülal Bingöl[▽] Can Firtina[◇] Lavanya Subramanian[‡] Jeremie S. Kim[◇†]
Rachata Ausavarungnirun[⊙] Mohammed Alser[◇] Juan Gomez-Luna[◇] Amirali Boroumand[†] Anant Nori[⋈]
Allison Scibisz[†] Sreenivas Subramoney[⋈] Can Alkan[▽] Saugata Ghose[⋆†] Onur Mutlu[◇†▽]
[†]*Carnegie Mellon University* [⋈]*Processor Architecture Research Lab, Intel Labs* [▽]*Bilkent University* [◇]*ETH Zürich*
[‡]*Facebook* [⊙]*King Mongkut's University of Technology North Bangkok* [⋆]*University of Illinois at Urbana–Champaign*

GenStore (ASPLOS 2022)

Nika Mansouri Ghiasi, Jisung Park, Harun Mustafa, Jeremie Kim, Ataberk Olgun, Arvid Gollwitzer, Damla Senol Cali, Can Firtina, Haiyu Mao, Nour Almadhoun Alserr, Rachata Ausavarungnirun, Nandita Vijaykumar, Mohammed Alser, Onur Mutlu
"GenStore: A High-Performance and Energy-Efficient In-Storage Computing System for Genome Sequence Analysis",
ASPLOS 2022



GenStore: A High-Performance In-Storage Processing System
for Genome Sequence Analysis

Nika Mansouri Ghiasi
ETH Zürich
Switzerland

Jisung Park
ETH Zürich
Switzerland

Harun Mustafa
ETH Zürich
Switzerland

Jeremie Kim
ETH Zürich
Switzerland

Ataberk Olgun
ETH Zürich
Switzerland

Arvid Gollwitzer
ETH Zürich
Switzerland

Damla Senol Cali
Bionano Genomics
USA

Can Firtina
ETH Zürich
Switzerland

Haiyu Mao
ETH Zürich
Switzerland

Nour Almadhoun Alserr
ETH Zürich
Switzerland

Rachata Ausavarungnirun
KMUTNB
Thailand

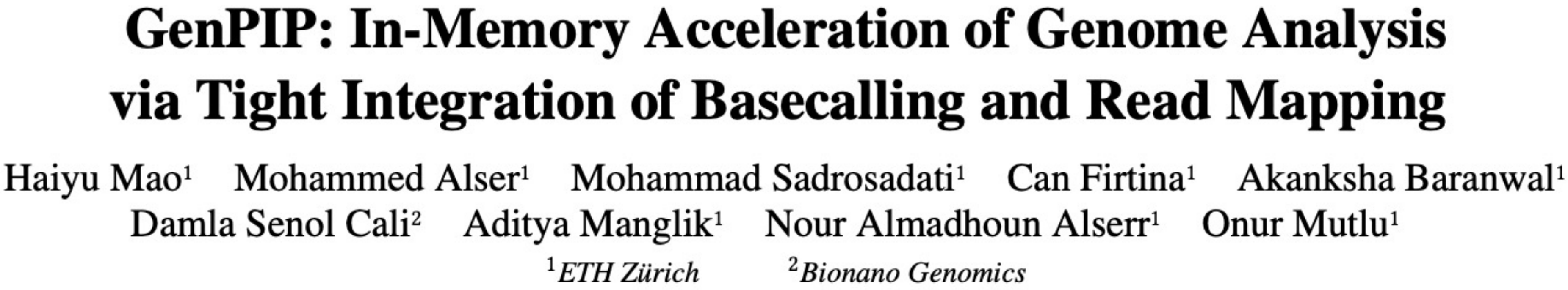
Nandita Vijaykumar
University of Toronto
Canada

Mohammed Alser
ETH Zürich
Switzerland

Onur Mutlu
ETH Zürich
Switzerland

GenPIP (MICRO 2022)

Haiyu Mao, Mohammed Alser, Mohammad Sadrosadati, Can Firtina, Akanksha Baranwal, Damla Senol Cali, Aditya Manglik, Nour Almadhoun Alserr, Onur Mutlu
"GenPIP: In-Memory Acceleration of Genome Analysis via Tight Integration of Basecalling and Read Mapping"
Proceedings of the 55rd International Symposium on Microarchitecture (***MICRO***), 2022.

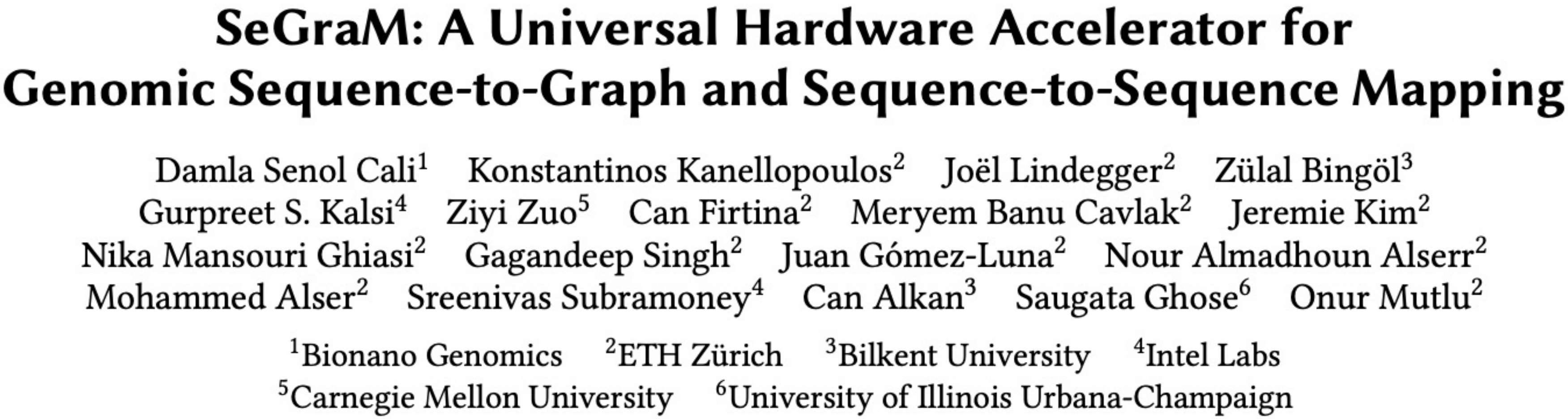


GenPIP: In-Memory Acceleration of Genome Analysis via Tight Integration of Basecalling and Read Mapping

Haiyu Mao[1] Mohammed Alser[1] Mohammad Sadrosadati[1] Can Firtina[1] Akanksha Baranwal[1]
Damla Senol Cali[2] Aditya Manglik[1] Nour Almadhoun Alserr[1] Onur Mutlu[1]
[1]*ETH Zürich* [2]*Bionano Genomics*

SeGraM (ISCA 2022)

Damla Senol Cali, Konstantinos Kanellopoulos, Joel Lindegger, Zülal Bingöl, Gurpreet S. Kalsi, Ziyi Zuo, Can Firtina, Meryem Banu Cavlak, Jeremie Kim, Nika Mansouri Ghiasi, Gagandeep Singh, Juan Gómez-Luna, Nour Almadhoun Alserr, Mohammed Alser, Sreenivas Subramoney, Can Alkan, Saugata Ghose, Onur Mutlu
"SeGraM: A Universal Hardware Accelerator for Genomic Sequence-to-Graph and Sequence-to-Sequence Mapping"
ISCA 2022



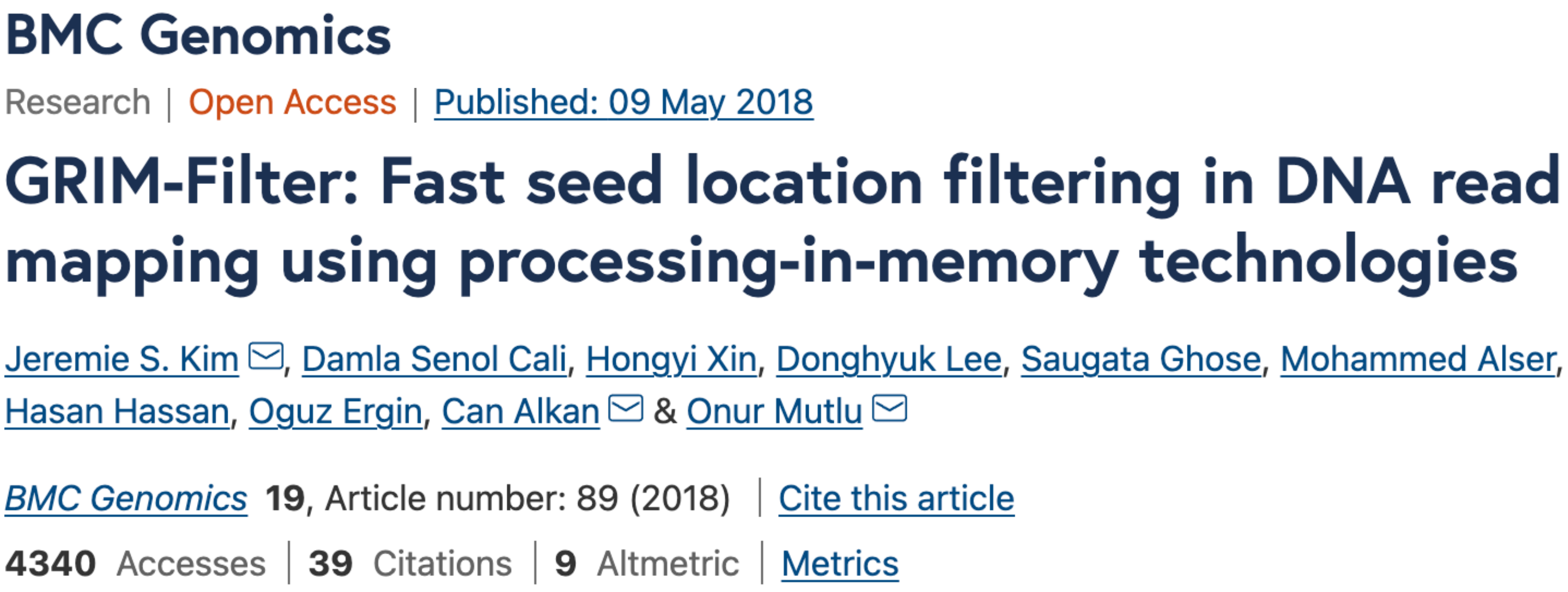
SeGraM: A Universal Hardware Accelerator for Genomic Sequence-to-Graph and Sequence-to-Sequence Mapping

Damla Senol Cali[1] Konstantinos Kanellopoulos[2] Joël Lindegger[2] Zülal Bingöl[3]
Gurpreet S. Kalsi[4] Ziyi Zuo[5] Can Firtina[2] Meryem Banu Cavlak[2] Jeremie Kim[2]
Nika Mansouri Ghiasi[2] Gagandeep Singh[2] Juan Gómez-Luna[2] Nour Almadhoun Alserr[2]
Mohammed Alser[2] Sreenivas Subramoney[4] Can Alkan[3] Saugata Ghose[6] Onur Mutlu[2]

[1]Bionano Genomics [2]ETH Zürich [3]Bilkent University [4]Intel Labs
[5]Carnegie Mellon University [6]University of Illinois Urbana-Champaign

In-Memory Sequence Analysis GRIM-Filter

- Jeremie S. Kim, Damla Senol Cali, Hongyi Xin, Donghyuk Lee, Saugata Ghose, **Mohammed Alser**, Hasan Hassan, Oguz Ergin, Can Alkan, and Onur Mutlu, **"GRIM-Filter: Fast Seed Location Filtering in DNA Read Mapping Using Processing-in-Memory Technologies"**
  to appear in ***BMC Genomics***, 2018.
  *Proceedings of the 16th Asia Pacific Bioinformatics Conference (**APBC**),*
  Yokohama, Japan, January 2018.
  arxiv.org Version (pdf)



BMC Genomics

Research | Open Access | Published: 09 May 2018

GRIM-Filter: Fast seed location filtering in DNA read mapping using processing-in-memory technologies

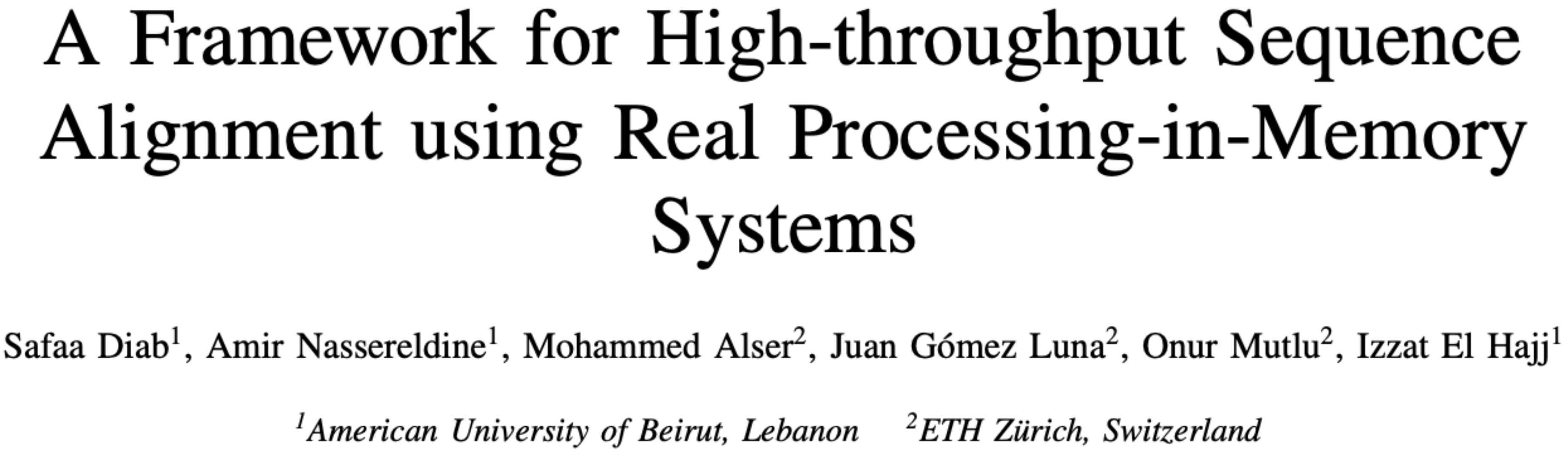
Jeremie S. Kim, Damla Senol Cali, Hongyi Xin, Donghyuk Lee, Saugata Ghose, Mohammed Alser, Hasan Hassan, Oguz Ergin, Can Alkan & Onur Mutlu

BMC Genomics **19**, Article number: 89 (2018) | Cite this article

4340 Accesses | **39** Citations | **9** Altmetric | Metrics

AIM (PIM Sequence Alignment Framework)

Safaa Diab, Amir Nassereldine, Mohammed Alser, Juan Gómez-Luna,
Onur Mutlu, Izzat El Hajj
"A Framework for High-throughput Sequence Alignment using Real Processing-in-Memory Systems"
arXiv, 2022
[Source code]



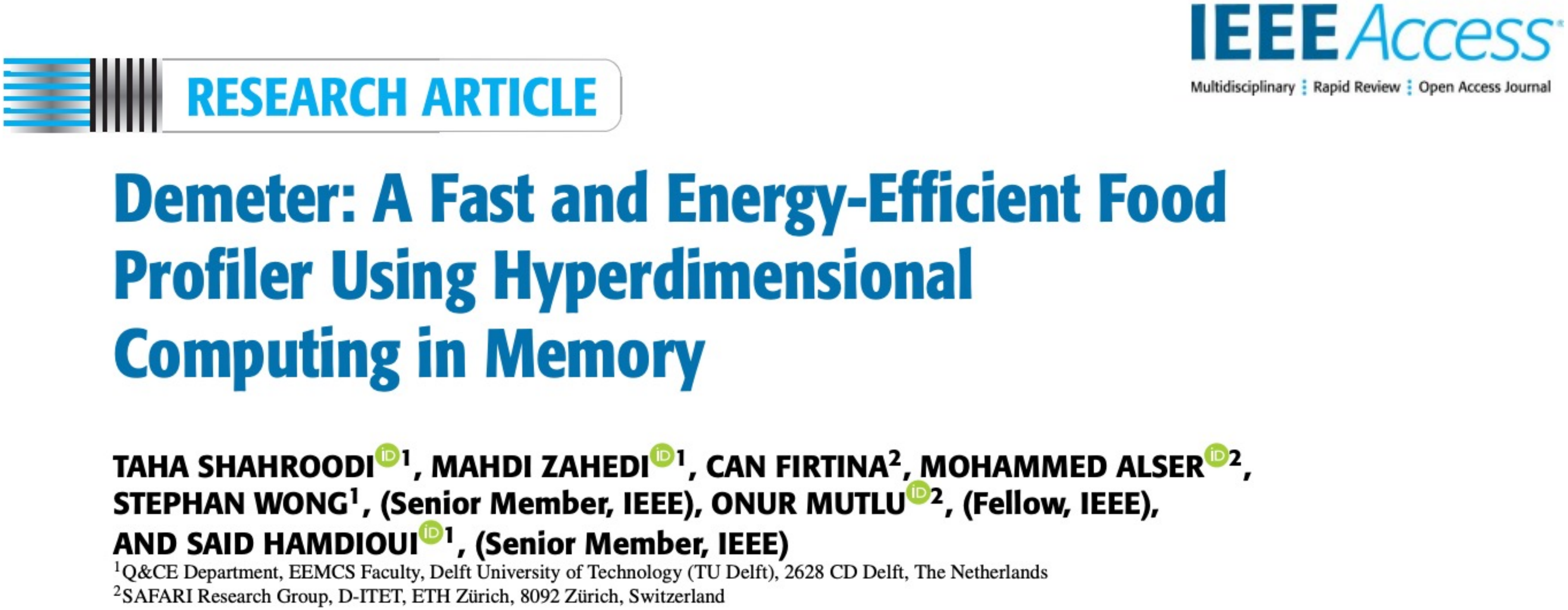
A Framework for High-throughput Sequence Alignment using Real Processing-in-Memory Systems

Safaa Diab[1], Amir Nassereldine[1], Mohammed Alser[2], Juan Gómez Luna[2], Onur Mutlu[2], Izzat El Hajj[1]

[1]*American University of Beirut, Lebanon* [2]*ETH Zürich, Switzerland*

Demeter (HD Food Microbiome Profiling)

Taha Shahroodi, Mahdi Zahedi, Can Firtina, Mohammed Alser, Stephan Wong, Onur Mutlu, Said Hamdioui
"Demeter: A Fast and Energy-Efficient Food Profiler using Hyperdimensional Computing in Memory"
IEEE Access, 2022



IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

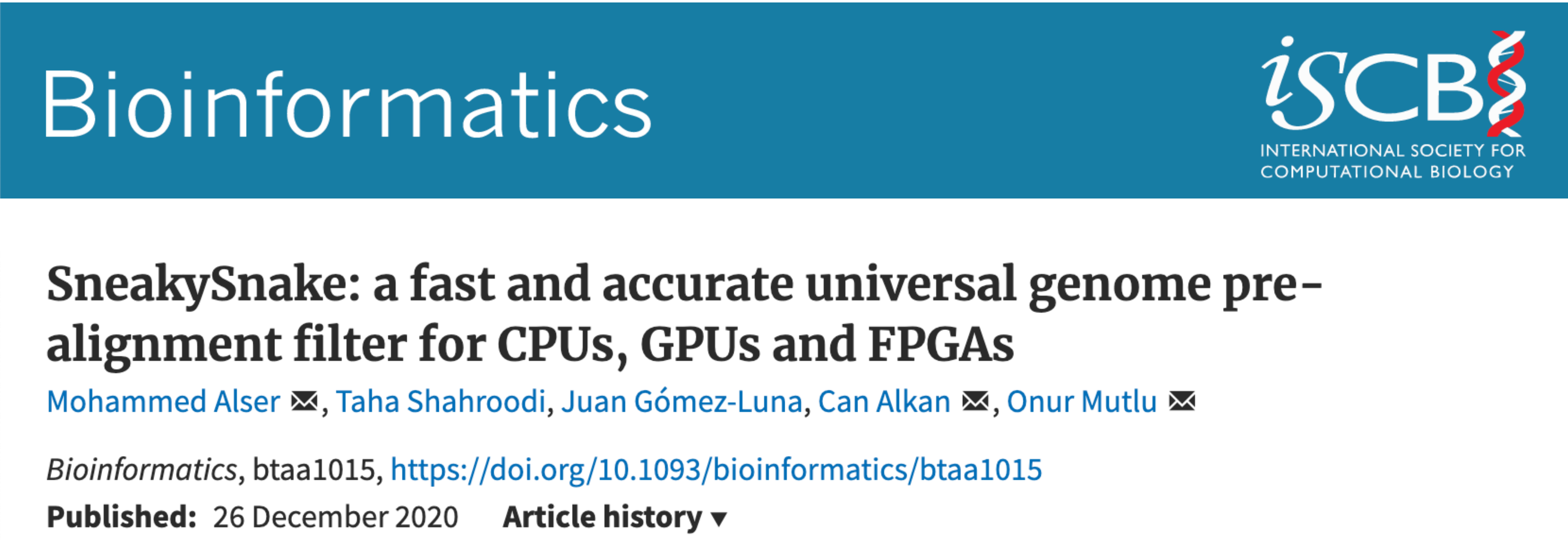
RESEARCH ARTICLE

Demeter: A Fast and Energy-Efficient Food Profiler Using Hyperdimensional Computing in Memory

TAHA SHAHROODI[1], MAHDI ZAHEDI[1], CAN FIRTINA[2], MOHAMMED ALSER[2], STEPHAN WONG[1], (Senior Member, IEEE), ONUR MUTLU[2], (Fellow, IEEE), AND SAID HAMDIOUI[1], (Senior Member, IEEE)
[1]Q&CE Department, EEMCS Faculty, Delft University of Technology (TU Delft), 2628 CD Delft, The Netherlands
[2]SAFARI Research Group, D-ITET, ETH Zürich, 8092 Zürich, Switzerland

SneakySnake [Bioinformatics 2020]

Mohammed Alser, Taha Shahroodi, Juan-Gomez Luna, Can Alkan, and Onur Mutlu,
"SneakySnake: A Fast and Accurate Universal Genome Pre-Alignment Filter for CPUs, GPUs, and FPGAs"
Bioinformatics, 2020.
[Source Code]
[Online link at Bioinformatics Journal]



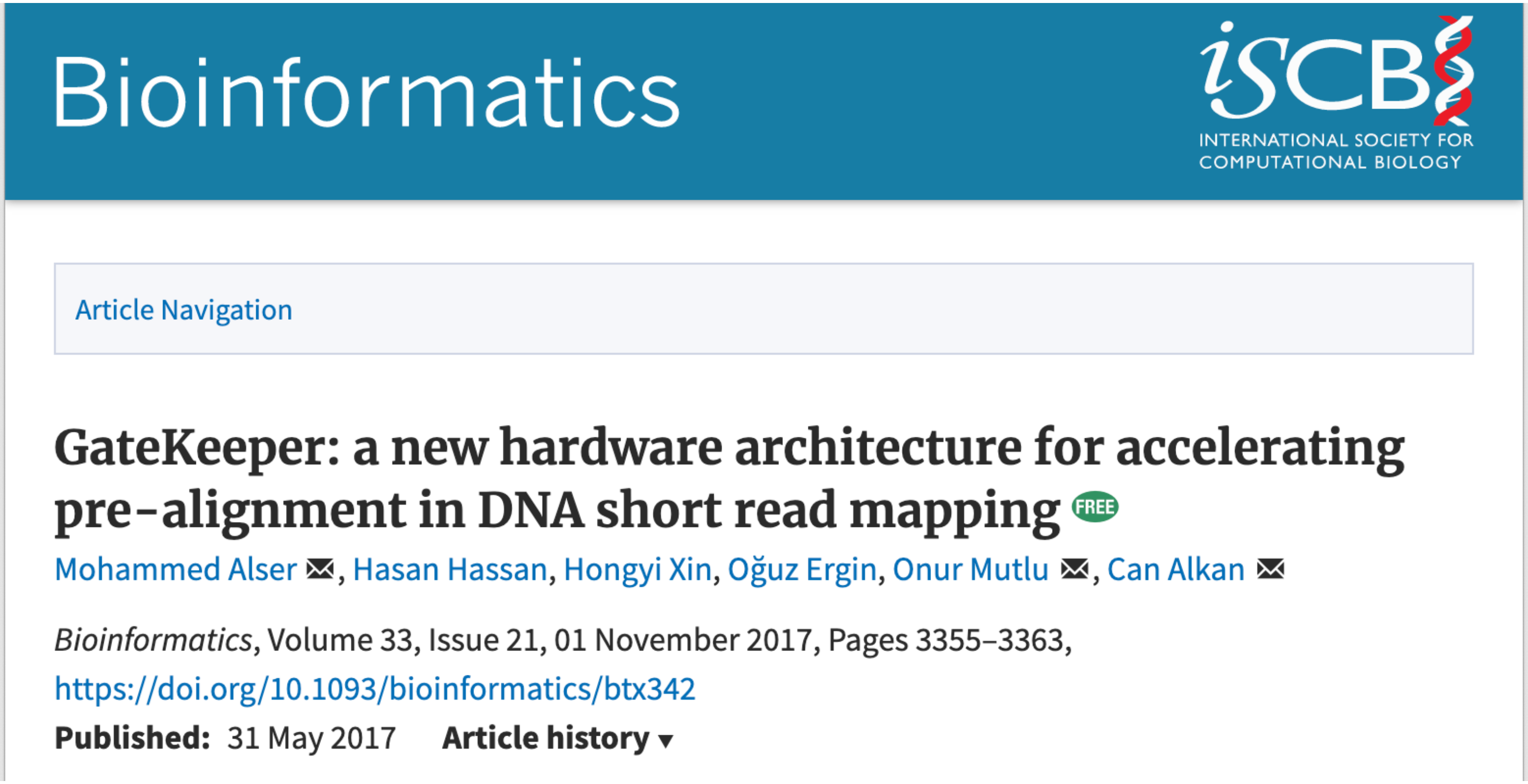
SneakySnake: a fast and accurate universal genome pre-alignment filter for CPUs, GPUs and FPGAs

Mohammed Alser ✉, Taha Shahroodi, Juan Gómez-Luna, Can Alkan ✉, Onur Mutlu ✉

Bioinformatics, btaa1015, https://doi.org/10.1093/bioinformatics/btaa1015
Published: 26 December 2020 **Article history ▾**

GateKeeper [Alser+, Bioinformatics 2017]

Mohammed Alser, Hasan Hassan, Hongyi Xin, Oguz Ergin, Onur Mutlu, and Can Alkan
"GateKeeper: A New Hardware Architecture for Accelerating Pre-Alignment in DNA Short Read Mapping"
Bioinformatics, [published online, May 31], 2017.
[Source Code]
[Online link at Bioinformatics Journal]



GateKeeper: a new hardware architecture for accelerating pre-alignment in DNA short read mapping FREE

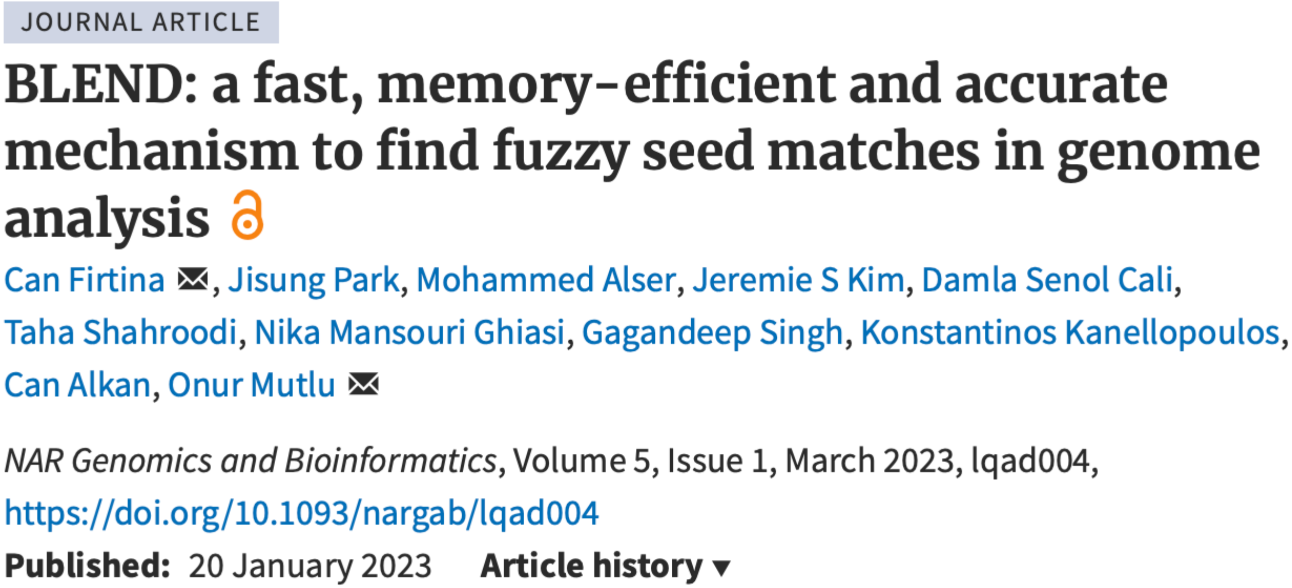
Mohammed Alser, Hasan Hassan, Hongyi Xin, Oğuz Ergin, Onur Mutlu, Can Alkan

Bioinformatics, Volume 33, Issue 21, 01 November 2017, Pages 3355–3363, https://doi.org/10.1093/bioinformatics/btx342

Published: 31 May 2017 **Article history** ▾

Finding Approximate Seed Matches

- Can Firtina, Jisung Park, Mohammed Alser, Jeremie S. Kim, Damla Senol Cali, Taha Shahroodi, Nika Mansouri-Ghiasi, Gagandeep Singh, Konstantinos Kanellopoulos, Can Alkan, and Onur Mutlu,
  "BLEND: A Fast, Memory-Efficient, and Accurate Mechanism to Find Fuzzy Seed Matches"
  NAR Genomics and Bioinformatics, January 2023.
  [Source Code and Data]



JOURNAL ARTICLE

BLEND: a fast, memory-efficient and accurate mechanism to find fuzzy seed matches in genome analysis

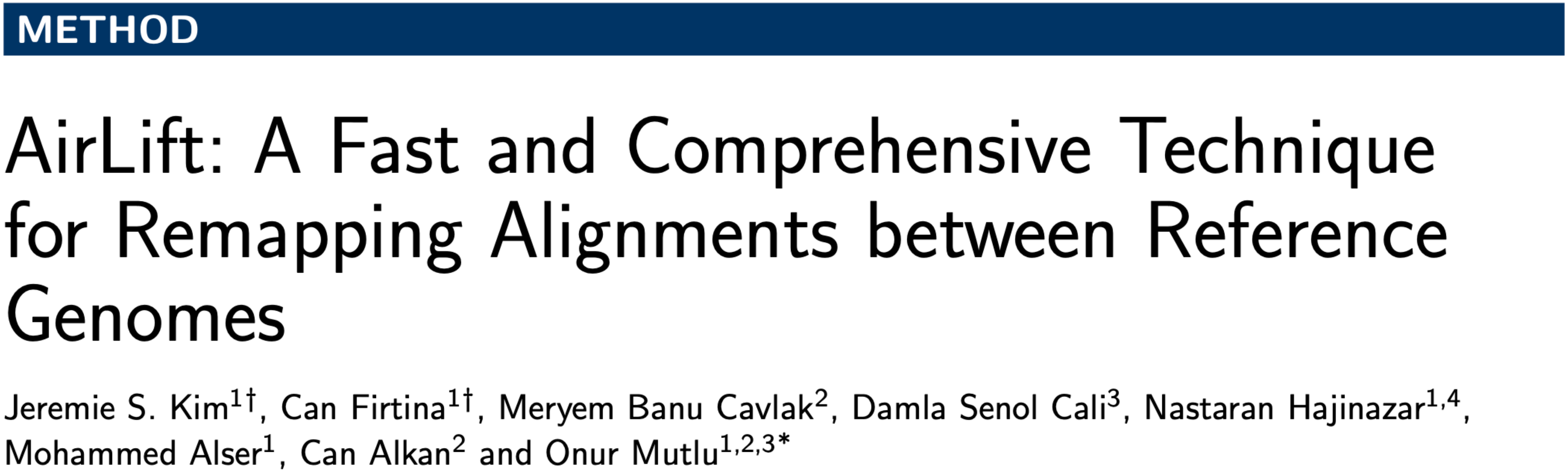
Can Firtina ✉, Jisung Park, Mohammed Alser, Jeremie S Kim, Damla Senol Cali, Taha Shahroodi, Nika Mansouri Ghiasi, Gagandeep Singh, Konstantinos Kanellopoulos, Can Alkan, Onur Mutlu ✉

NAR Genomics and Bioinformatics, Volume 5, Issue 1, March 2023, lqad004, https://doi.org/10.1093/nargab/lqad004

Published: 20 January 2023 **Article history ▾**

Remapping Alignments Between References

- Jeremie S. Kim, Can Firtina, Meryem Banu Cavlak, Damla Senol Cali, Nastaran Hajinazar, Mohammed Alser, Can Alkan, and Onur Mutlu,
  "AirLift: A Fast and Comprehensive Technique for Remapping Alignments between Reference Genomes"
  *Preprint in **arXiv** and **bioRxiv***, 17 February 2021.
  [bioRxiv preprint]
  [arXiv preprint]
  [Source Code and Data]



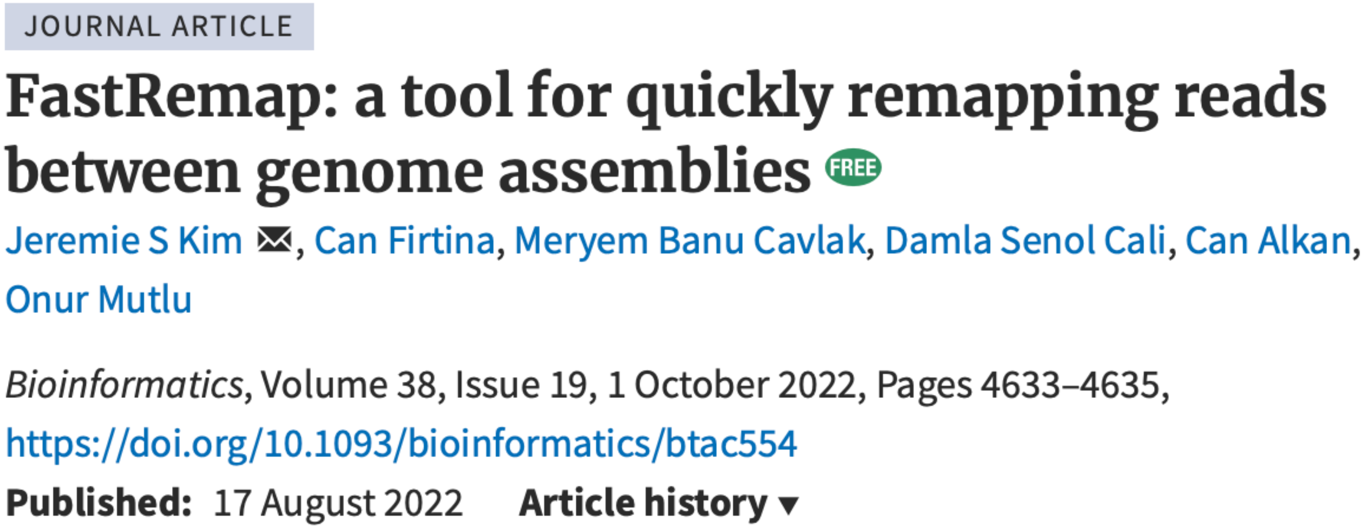
METHOD

AirLift: A Fast and Comprehensive Technique for Remapping Alignments between Reference Genomes

Jeremie S. Kim[1]†, Can Firtina[1]†, Meryem Banu Cavlak[2], Damla Senol Cali[3], Nastaran Hajinazar[1,4], Mohammed Alser[1], Can Alkan[2] and Onur Mutlu[1,2,3]*

Mapping Constant Regions Between References

- Jeremie S. Kim, Can Firtina, Meryem Banu Cavlak, Damla Senol Cali, Can Alkan, and Onur Mutlu,
  "FastRemap: A Tool for Quickly Remapping Reads between Genome Assemblies"
  Bioinformatics, 26 January 2022.
  [Source Code]



JOURNAL ARTICLE

FastRemap: a tool for quickly remapping reads between genome assemblies FREE

Jeremie S Kim ✉, Can Firtina, Meryem Banu Cavlak, Damla Senol Cali, Can Alkan, Onur Mutlu

Bioinformatics, Volume 38, Issue 19, 1 October 2022, Pages 4633–4635, https://doi.org/10.1093/bioinformatics/btac554

Published: 17 August 2022 **Article history** ▾

Mapping Nanopore raw signals in real-time

- 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"
  Preprint in *biorXiv*, January 2023.
  [Source Code and Data]

RawHash: Enabling Fast and Accurate Real-Time Analysis of Raw Nanopore Signals for Large Genomes

Can Firtina[1] Nika Mansouri Ghiasi[1] Joel Lindegger[1] Gagandeep Singh[1]
Meryem Banu Cavlak[1] Haiyu Mao[1] Onur Mutlu[1]
[1] *ETH Zurich*

Accelerating Genome Analysis

How Large is a Genome?

Prime Tower, Zurich

~3.2 billion genomic bases

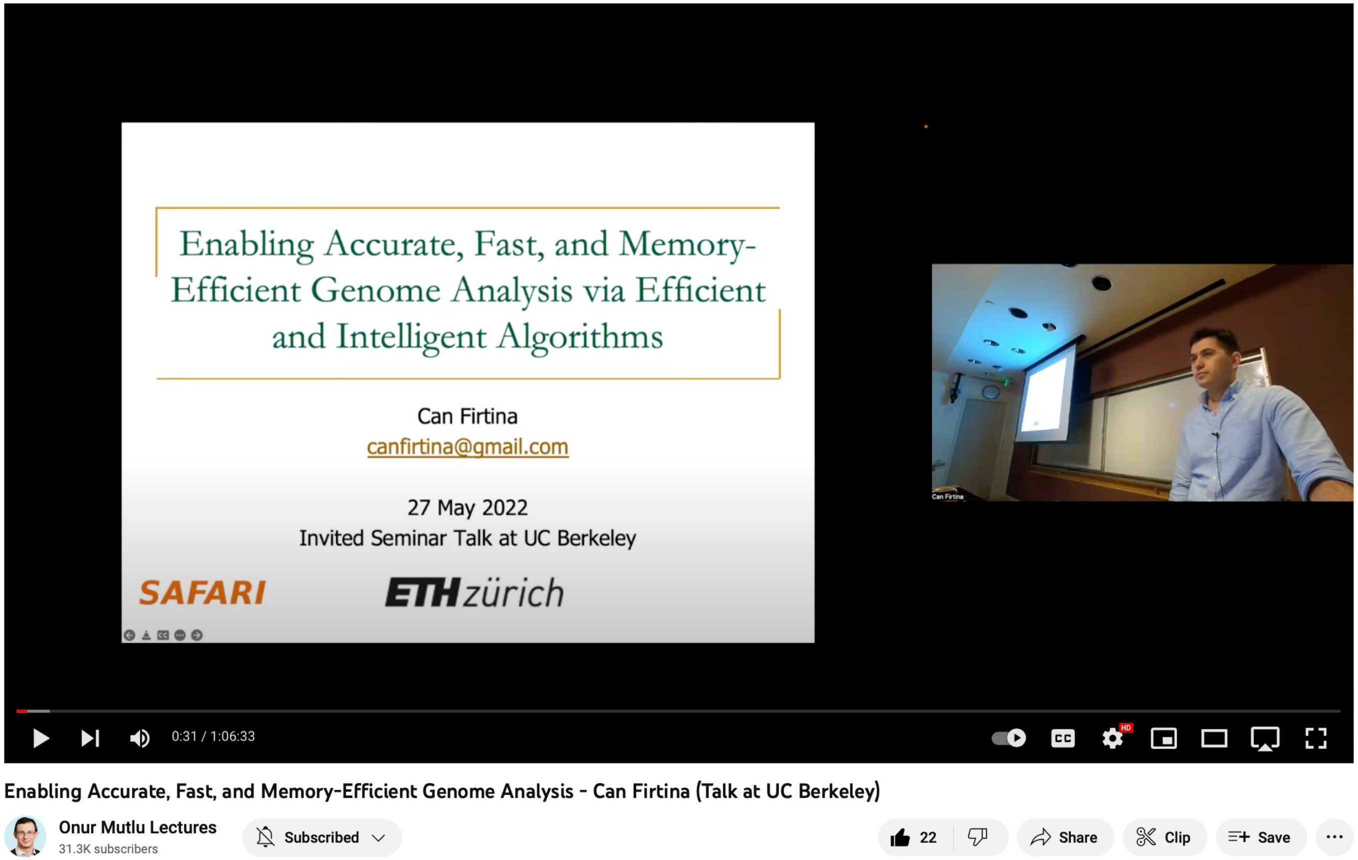
Livestream - Seminar in Computer Architecture - ETH Zürich (Spring 2022)

Seminar in Computer Arch. - Lecture 5: Accelerating Genome Analysis (Spring 2022)

https://www.youtube.com/watch?v=qPIiiwUVFug

More on Accelerating Genome Analysis …

- Can Firtina,
  "Enabling Accurate, Fast, and Memory-Efficient Genome Analysis via Efficient and Intelligent Algorithms"
  Talk at UC Berkeley, Berkeley, CA, United States, May 27, 2022.
  [Slides (pptx) (pdf)]
  [Talk Video (1 hour 6 minutes)]



Enabling Accurate, Fast, and Memory-Efficient Genome Analysis via Efficient and Intelligent Algorithms

Can Firtina
canfirtina@gmail.com

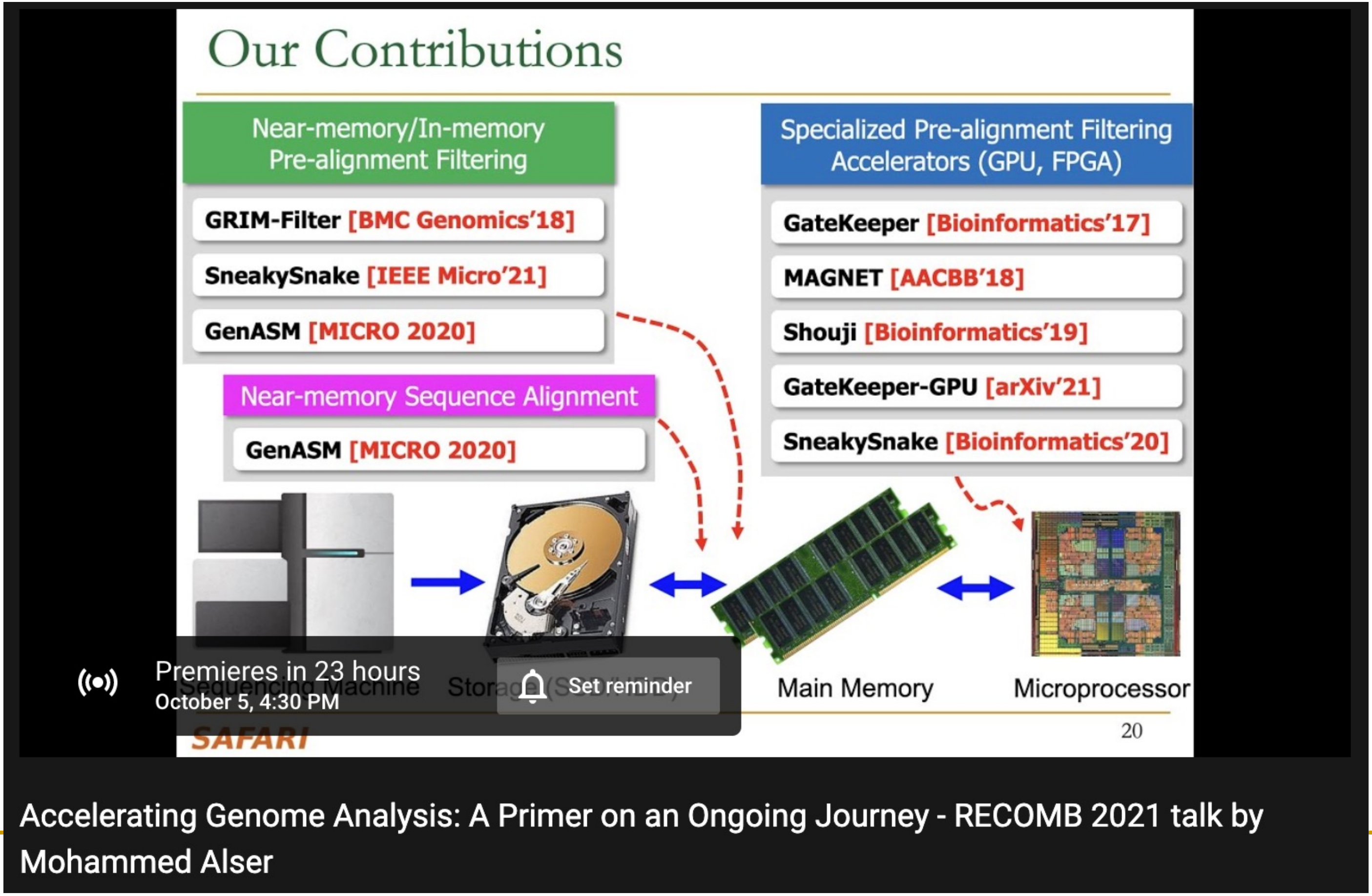
27 May 2022
Invited Seminar Talk at UC Berkeley

Enabling Accurate, Fast, and Memory-Efficient Genome Analysis - Can Firtina (Talk at UC Berkeley)

Onur Mutlu Lectures
31.3K subscribers

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)]



Our Contributions

Near-memory/In-memory Pre-alignment Filtering

- **GRIM-Filter** [BMC Genomics'18]
- **SneakySnake** [IEEE Micro'21]
- **GenASM** [MICRO 2020]

Near-memory Sequence Alignment

- **GenASM** [MICRO 2020]

Specialized Pre-alignment Filtering Accelerators (GPU, FPGA)

- **GateKeeper** [Bioinformatics'17]
- **MAGNET** [AACBB'18]
- **Shouji** [Bioinformatics'19]
- **GateKeeper-GPU** [arXiv'21]
- **SneakySnake** [Bioinformatics'20]

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

More on Intelligent Genome Analysis …

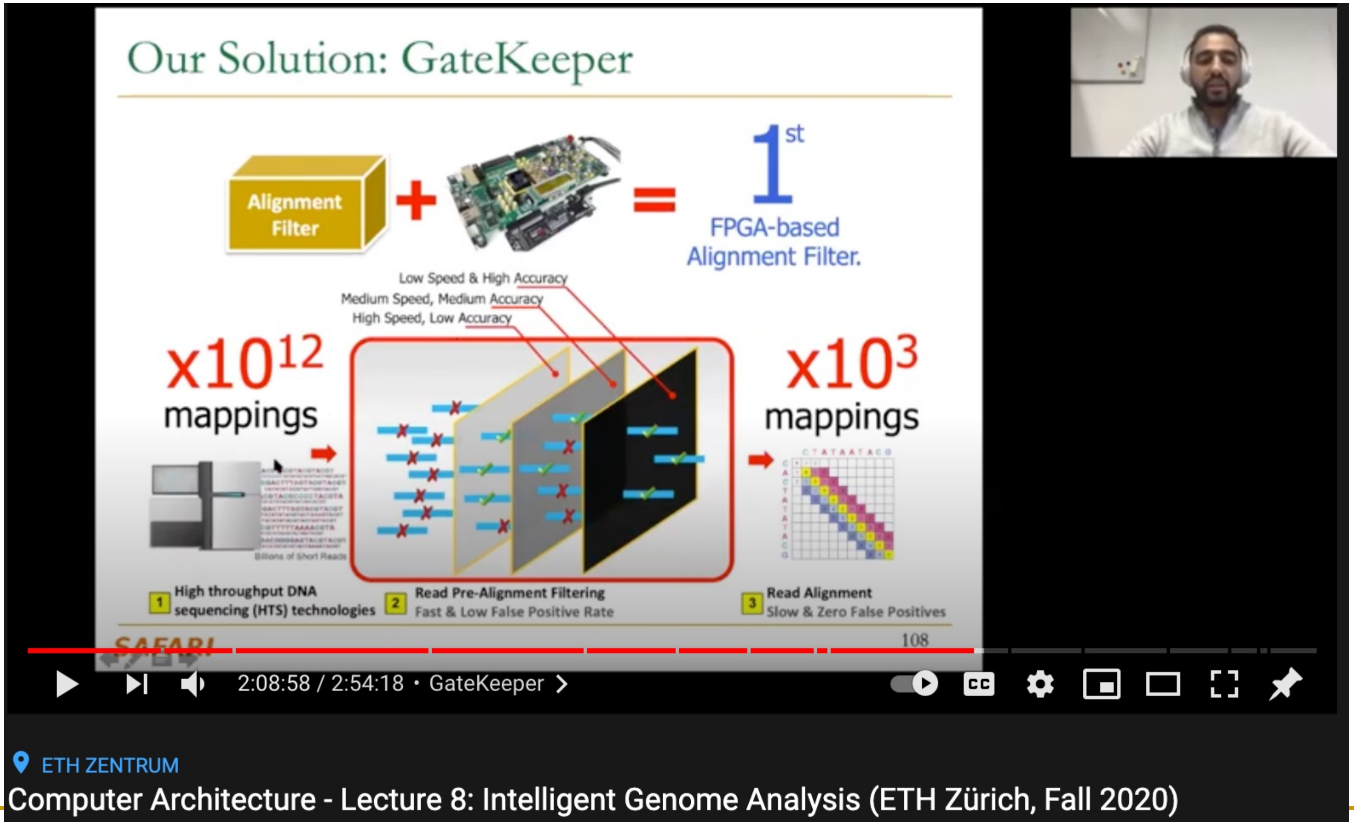
- Mohammed Alser,
  "Computer Architecture - Lecture 10: Intelligent Genome Analysis"
  ETH Zurich, Computer Architecture Course, Fall2021, Lecture 10, Virtual, 29 October 2021.
  [Slides (pptx) (pdf)]
  [Talk Video (3 hour 2 minutes, including Q&A)]
  [Related Invited Paper (at IEEE Micro, 2020)]

Computer Architecture - Lecture 10: Intelligent Genome Analysis (Fall 2021)

412 views • Streamed live on Oct 29, 2021

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)]



Our Solution: GateKeeper

Alignment Filter + = 1st FPGA-based Alignment Filter.

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

$x10^{12}$ mappings → $x10^{3}$ mappings

Billions of Short Reads

1 High throughput DNA sequencing (HTS) technologies

2 Read Pre-Alignment Filtering
Fast & Low False Positive Rate

3 Read Alignment
Slow & Zero False Positives

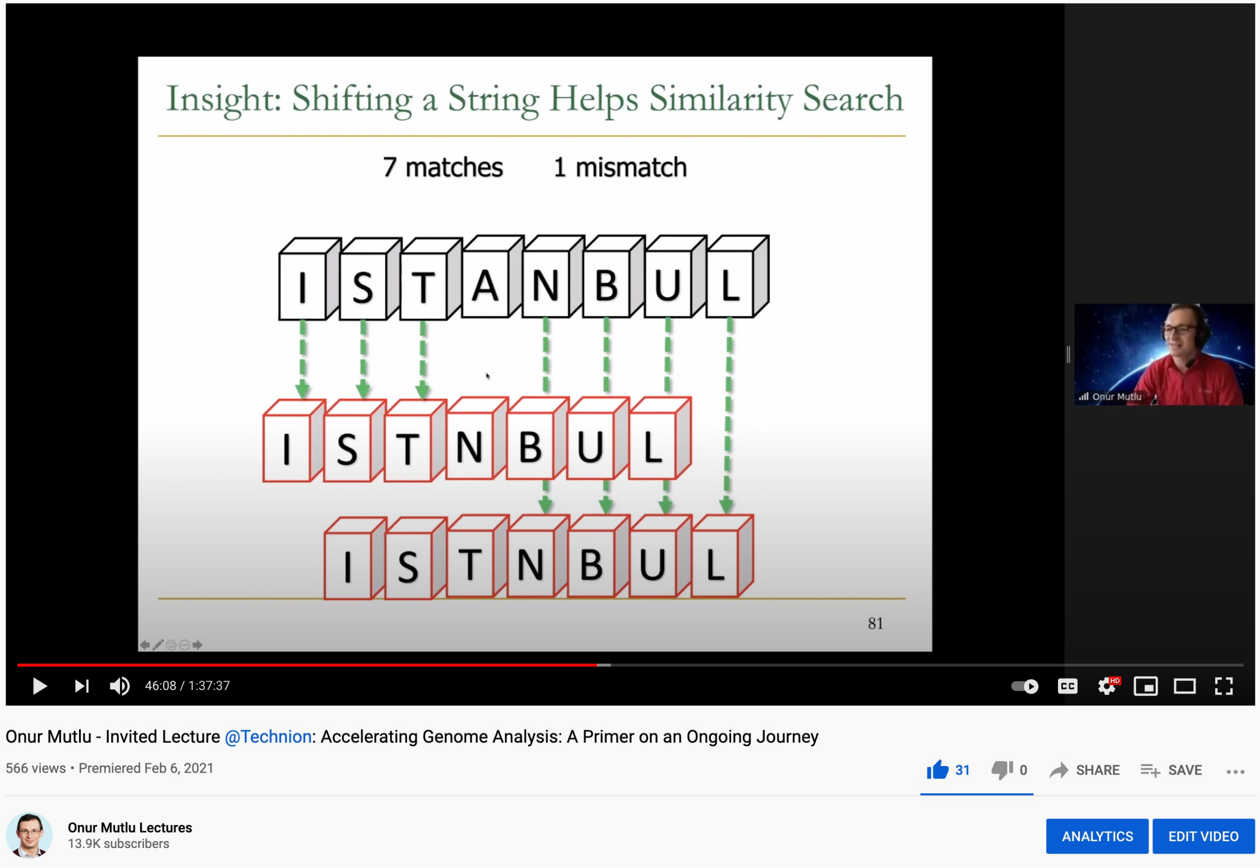
SAFARI 108

ETH ZENTRUM

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

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)]



Insight: Shifting a String Helps Similarity Search

7 matches 1 mismatch

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

566 views • Premiered Feb 6, 2021

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

P&S Mobile Genomics

Lecture 1a: Course Introduction

Can Firtina

ETH Zürich
Spring 2023
2 March 2023