P&S Accelerating Genomics
Introduction & Course Logistics

Dr. Mohammed Alser
@mealser

ETH Zurich
Fall 2022
27 October 2022
Intelligent Genomic Analyses (Fall 2022)

https://www.youtube.com/watch?v=nA41964-9r8
Review

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

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

ETH Zurich, Gloriastrasse 35, 8092 Zürich, Switzerland
Two P&S Genomics Courses

1. Mobile Genomics Course - Meeting 1: Course Introduction & Project Proposals (Spring 2022)
   - Onur Mutlu Lectures

2. Mobile Genomics Course - Meeting 2: Introduction to Sequencing (Spring 2022)
   - Onur Mutlu Lectures

3. Mobile Genomics Course - Meeting 3: Read Mapping (Spring 2022)
   - Onur Mutlu Lectures

4. Mobile Genomics Course - Meeting 4: GateKeeper (Spring 2022)
   - Onur Mutlu Lectures

5. Mobile Genomics Course - Meeting 5: MAGNET & Shouji (Spring 2022)
   - Onur Mutlu Lectures

Livestream - P&S Genome Sequencing on Mobile Devices (Spring 2022)

- 11 videos • 213 views • Last updated on 5 Jun 2022

https://www.youtube.com/playlist?list=PL5Q2soXY2Zi8NrPDgOR1yRU_Cxxjw-u18
Genomics (Spring 2022)

- **Spring 2022 Edition:**
  - [https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile](https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile)

- **Youtube Livestream:**
  - [https://www.youtube.com/playlist?list=PL5Q2soXY2Zi8NrPDgOR1yRU_Cxxjw-u18](https://www.youtube.com/playlist?list=PL5Q2soXY2Zi8NrPDgOR1yRU_Cxxjw-u18)

- Project course
  - Taken by Bachelor’s/Master’s students
  - Genomics lectures
  - Hands-on research exploration
  - Many research readings
## Course Materials

### Spring 2022 Meetings/Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Livestream</th>
<th>Meeting</th>
<th>Learning Materials</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>8.3</td>
<td>Live</td>
<td>M1: P&amp;S Mobile Genomics Course</td>
<td>Required Materials</td>
<td>Recommended Materials</td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>Introduction &amp; Project Proposals</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W2</td>
<td>15.3</td>
<td>Live</td>
<td>M2: Introduction to Sequencing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W3</td>
<td>22.3</td>
<td>Live</td>
<td>M3: Read Mapping</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W4</td>
<td>29.3</td>
<td>Live</td>
<td>M4: GateKeeper</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W5</td>
<td>5.4</td>
<td>Live</td>
<td>M5: MAGNET &amp; Shouji</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>YouTube Video</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W6</td>
<td>12.4</td>
<td>Live</td>
<td>M6: SneakySnake</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W7</td>
<td>26.4</td>
<td>Live</td>
<td>M7: GenStore</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W8</td>
<td>03.05</td>
<td>Premiere</td>
<td>M8: GRIM-Filter</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W9</td>
<td>10.05</td>
<td>Live</td>
<td>M9: Genome Assembly</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W10</td>
<td>17.05</td>
<td>Live</td>
<td>M10: Genomic Data Sharing Under Differential Privacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W11</td>
<td>07.06</td>
<td>Premiere</td>
<td>M11: Accelerating Genome Sequence Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tue.</td>
<td></td>
<td>(PDF) [link] (PPT)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile](https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile)
Past Semesters (P&S Accelerating Genomics)

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

- **Spring 2022:**
  https://safari.ethz.ch/projects_and_seminars/spring2022/doku.php?id=bioinformatics

- **Fall 2021:**
  https://safari.ethz.ch/projects_and_seminars/fall2021/doku.php?id=bioinformatics

- **Spring 2021:**
  https://safari.ethz.ch/projects_and_seminars/spring2021/doku.php?id=bioinformatics

- **Fall 2020:**
  https://safari.ethz.ch/projects_and_seminars/fall2020/doku.php?id=bioinformatics
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: Accelerating Genomics

- We will cover the **basics** of **genome analysis** to understand the **computational steps** of the entire pipeline and find the computational bottlenecks.

- Students will learn about the **existing efforts** for accelerating one or more of these steps.

- Student will carry out a **hands-on project** to implement and improve these acceleration efforts.
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.

- Digital Design and Computer Architecture (or equivalent course)

- A good knowledge in C programming language is required.

- Experience in at least one of the following is highly desirable: FPGA implementation and GPU programming.

- Interest in making things efficient and solving problems
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](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? (I)

Juan Gómez Luna
Senior Researcher and Lecturer

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

Can Firtina
PhD Student

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

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/](https://safari.ethz.ch/safari-group/)
Course Info: Who Are We? (II)

Nika Mansourighiasi
PhD Student
Processing-In-Memory | Emerging Memory & Processing Technologies

Max Rumpf
Masters Student
Bioinformatics | Computational Genomics | Sequence Analysis & Alignment | Machine Learning

Julien Eudine
Masters Student
Computational Genomics | High-Performance Computing | Heterogeneous Computing Systems | Hardware acceleration

Get to know them and their research: [https://safari.ethz.ch/safari-group/](https://safari.ethz.ch/safari-group/)
Course Info: Who Are We? (III)

Arvid Gollwitzer
Masters Student
Bioinformatics | Computational Genomics | Sequence Analysis & Alignment | Medical Applications | Clinical Metagenomics

Younjoo Lee
Masters Student
Computer architecture | FPGA-acceleration | Bioinformatics | Processing in Memory.

Luca Blum
Masters Student
Machine Learning | Computational Genomics | Bioinformatics | High-Performance Computing.

Get to know them and their research: https://safari.ethz.ch/safari-group/
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/fall2022/doku.php?id=bioinformatics](https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=bioinformatics)

- Useful information for the course

- We will also use Glip, Moodle, and E-mail for Q&A, announcements, ..

- Check your Glip, Moodle, and E-mail frequently for announcements
Project Assignment

- We ask you to **study the learning materials** before next meetings!

- We gave you **5 days** to enter your preferences,

- Then, we **match** your interests, skills, and background with a suitable project.

- We assign you a project and put you in touch with a **mentor** in the second week.
Review

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

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

ETH Zurich, Glorstrasse 35, 8092 Zürich, Switzerland
Basic Information

P&S Genomics - Fall 2022 @ ETH Zurich
We would like to warmly welcome all of you to the P&S Genomics course :)

Because we care about you, we would like to know more about you and your preferences to serve you better in this course and plan for a fruitful course moving forward ;)

The questionnaire is due on 4 October 2022, 23:59 Zurich time.

* Required

Full name *

Your answer

This is a required question

Email address *

Your answer
Assignment 2: Project Preferences

Which of the following technical skills do you have? *
Choose what you feel comfortable to use throughout your project. Choose all applicable.

<table>
<thead>
<tr>
<th></th>
<th>Stay away</th>
<th>Basic knowledge (only took a course)</th>
<th>Good knowledge (did a few projects)</th>
<th>Did several projects</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>C/Verilog/VHDL programming for FPGAs</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>CUDA programming for GPUs</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>C programming</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>C++ programming</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Python programming</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Surveying a topic &amp; benchmarking tools</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I am fine with using any programming language and learning new skills</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Your Responsibilities

- 1 Lecture every week
  - Monday 4-5 PM
    - [https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile](https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile)
  - Thursday 10-11 AM
    - [https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=bioinformatics](https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=bioinformatics)

- Attendance is mandatory

- Working on your project for ~6 hours per week

- Meeting your mentors weekly is required

- Replying on time to Glip, Moodle, and E-mail messages.
What is Intelligent Genome Analysis?

- Fast genome analysis
  - *Real-time analysis*

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

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

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

- Avoiding erroneous analysis
  - *E.g., your father is not your father*

Bandwidth

Energy-efficiency & Latency

Privacy

Scalability

Accuracy
Next Meetings

1. Obtaining Genomic Sequencing Data

   A. DNA Extraction
   B. DNA Fragmentation
   C. Library Preparation

2. Basecalling

   - Illumina
     - Generating Sequencing Data: Multiple images
     - Sequencing: .BCL/.CBCL
     - Downloading Real Sequencing Data: .FASTA/.FNA

   - ONT
     - Generating Sequencing Data: Squiggle
     - Sequencing: .FAST5
     - Downloading Real Sequencing Data: .SRA

   - PacBio
     - Generating Sequencing Data: 30-hour movie
     - Sequencing: .BAM
     - Downloading Real Sequencing Data: .EN_A

3. Quality Control

   - Illumina
   - ONT
   - PacBio
     - Quality Control: .FASTQ
   - Read
   - Trimmed read
   - Quality Control: .FASTQ
Next Meetings

4 Read Mapping

**A Indexing**
- Reference Genome
- Seeds

<table>
<thead>
<tr>
<th>Seed content</th>
<th>Seed locations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1, 4, 6</td>
</tr>
<tr>
<td></td>
<td>3, 5, 12</td>
</tr>
<tr>
<td></td>
<td>50, 52</td>
</tr>
<tr>
<td></td>
<td>2, 100</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**B Seeding**
- Read
- Seeds

**C Pre-alignment Filtering**
- Filter 1
- Filter 2
- Filter N

**D Sequence Alignment**
- Read
- Dynamic Programming Matrix

<table>
<thead>
<tr>
<th>Reference Subsequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>.SAM/.BAM</td>
</tr>
</tbody>
</table>

5 Variant Calling

**A Processing Mapping Data**
- Removing read duplicates and cleaning read mapping results

**B Variation Classification**

**C Generating Variant Calls**

<table>
<thead>
<tr>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>T</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aligned Reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>T</td>
</tr>
<tr>
<td>A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>List of variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHR</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>Chr1</td>
</tr>
</tbody>
</table>

27
Next Meetings

- **1 Lecture every week**
  - **Monday 4-5 PM**
    - [https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile](https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=genome_seq_mobile)
  - **Thursday 10-11 AM**
    - [https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=bioinformatics](https://safari.ethz.ch/projects_and_seminars/fall2022/doku.php?id=bioinformatics)
<table>
<thead>
<tr>
<th>#</th>
<th>Title</th>
<th>Duration</th>
<th>Lecturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Mobile Genomics Course - Meeting 2: Introduction to Sequencing (Fall 2021)</td>
<td></td>
<td>Onur Mutlu Lectures</td>
</tr>
<tr>
<td>4</td>
<td>Mobile Genomics Course - Meeting 3: Read Mapping (Fall 2021)</td>
<td></td>
<td>Onur Mutlu Lectures</td>
</tr>
<tr>
<td>5</td>
<td>Mobile Genomics Course - Meeting 4: GateKeeper (Fall 2021)</td>
<td></td>
<td>Onur Mutlu Lectures</td>
</tr>
<tr>
<td>6</td>
<td>Mobile Genomics Course - Lecture 5: MAGNET &amp; Shouji (Fall 2021)</td>
<td></td>
<td>Onur Mutlu Lectures</td>
</tr>
<tr>
<td>7</td>
<td>Computer Architecture - Lecture 8: Intelligent Genome Analysis (ETH Zürich, Fall 2020)</td>
<td></td>
<td>Onur Mutlu Lectures</td>
</tr>
<tr>
<td>Page</td>
<td>Topic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td><strong>Topics To Be Covered (II)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td><strong>GRIM-Filter: Fast Seed Location Filtering in DNA Read Mapping w/ Processing-in-Memory</strong> - Jeremie Kim</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Onur Mutlu Lectures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td><strong>Comp. Architecture - Lecture 9a: GenASM: Approx. String Matching Accelerator</strong> (ETH Zürich, Fall 2020)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Onur Mutlu Lectures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td><strong>Mobile Genomics Course - Lecture 8: Genome Assembly</strong> (Fall 2021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Onur Mutlu Lectures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td><strong>Mobile Genomics Course - Lecture 9: GRIM-Filter</strong> (Fall 2021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Onur Mutlu Lectures</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Mobile Genomics Course - Lecture 10: Genomic Data Sharing Under Differential Privacy</strong> (Fall 2021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Onur Mutlu Lectures</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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]

Review

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

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

ETH Zurich, Gloriastrasse 35, 8092 Zürich, Switzerland
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”
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]
Feedback From Our Community!

James Ferguson
@Psy_Fer_

This is awesome! I've got my evening reading sorted.

Stéphane Le Crom
@slecom

Very complete article on the evolution of read alignment algorithms. #NGS #genomics

Svetlana Gorokhova
@SGorokhova

An impressive overview of read alignment methods over the last three decades

BContrerasMoreira @BrunoContrerasM · Sep 10
Replying to @mealser @GenomeBiology and 3 others
Buen hilo de repaso sobre la evolución de los algoritmos de alineamiento de secuencias a medida que ha mejorado la tecnología de secuenciación

https://twitter.com/mealser/status/1435223377644503040
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"

[Lightning Talk Video (1.5 minutes)]
[Lightning Talk Slides (pptx) (pdf)]
[Talk Video (18 minutes)]
[Slides (pptx) (pdf)]
GenStore: A High-Performance In-Storage Processing System for Genome Sequence Analysis

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
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”

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 Cali¹ Konstantinos Kanellopoulos² Joël Lindegger² Zülal Bingöl³ Gurpreet S. Kalsi⁴ Ziyi Zuo⁵ Can Firtina² Meryem Banu Cavlak² Jeremie Kim² Nika Mansouri Ghiasi² Gagandeep Singh² Juan Gómez-Luna² Nour Almadhoun Alserr² Mohammed Alser² Sreenivas Subramoney⁴ Can Alkan³ Saugata Ghose⁶ Onur Mutlu²

¹Bionano Genomics ²ETH Zürich ³Bilkent University ⁴Intel Labs ⁵Carnegie Mellon University ⁶University of Illinois Urbana-Champaign
GenASM Framework [MICRO 2020]

- Damla Senol Cali, Gurpreet S. Kalsi, Zulal Bingol, Can Firtina, Lavanya Subramanian, Jeremie S. Kim, Rachata Ausavarungrunrun, 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"


[Lightning Talk Video (1.5 minutes)]
[Lightning Talk Slides (pptx) (pdf)]
[Talk Video (18 minutes)]
[Slides (pptx) (pdf)]
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“
[Source Code]
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
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\textsuperscript{1}, Amir Nassereldine\textsuperscript{1}, Mohammed Alser\textsuperscript{2}, Juan Gómez Luna\textsuperscript{2}, Onur Mutlu\textsuperscript{2}, Izzat El Hajj\textsuperscript{1}

\textsuperscript{1}American University of Beirut, Lebanon \hspace{1cm} \textsuperscript{2}ETH 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"
[Source Code]
[Online link at Bioinformatics Journal]
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]

Sequence alignment

Shouji: a fast and efficient pre-alignment filter for sequence alignment

Mohammed Alser¹,²,³,*, Hasan Hassan¹, Akash Kumar², Onur Mutlu¹,³,*, and Can Alkan³,*

¹Computer Science Department, ETH Zürich, Zürich 8092, Switzerland, ²Chair for Processor Design, Center For Advancing Electronics Dresden, Institute of Computer Engineering, Technische Universität Dresden, 01062 Dresden, Germany and ³Computer Engineering Department, Bilkent University, 06800 Ankara, Turkey
In-Memory Sequence Analysis GRIM-Filter


**BMC Genomics**

Research | Open Access | Published: 09 May 2018

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

Jeremy 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
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=qPliiwUVFug
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)]
More on Intelligent Genome Analysis …

- Mohammed Alser,
  "Computer Architecture - Lecture 10: Intelligent Genome Analysis"
  ETH Zurich, Computer Architecture Course, Fall 2021, 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)]
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)]
More on Fast Genome Analysis …

  - Slides (pptx) (pdf)
  - Talk Video (1 hour 37 minutes, including Q&A)
  - Related Invited Paper (at IEEE Micro, 2020)
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](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](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](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](https://www.youtube.com/watch?v=rgjl8ZyLsAg&list=PL5Q2soXY2Zi9E2bBVAgCqLgwiDRQDTyId)

[https://www.youtube.com/onurmutlulectures](https://www.youtube.com/onurmutlulectures)
Prior Research on Genome Analysis (1/2)


- Alser + "SneakySnake: A Fast and Accurate Universal Genome Pre-Alignment Filter for CPUs, GPUs, and FPGAs.", *Bioinformatics*, 2020.


P&S Accelerating Genomics
Introduction & Course Logistics

Dr. Mohammed Alser
@mealser

ETH Zurich
Fall 2022
27 October 2022