FastHASH:
A New Algorithm for Fast and Comprehensive Next-Generation Sequence Mapping

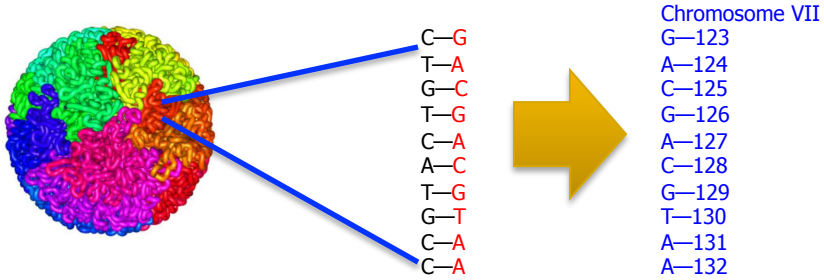
Hongyi Xin† Donghyuk Lee†
Farhad Hormozdiari‡ Can Alkan§ Onur Mutlu†

† **Carnegie Mellon University** § **University of Washington**
‡ **University of California Los Angeles**

SAFARI Carnegie Mellon

Next-Generation DNA Sequencing (NGS)

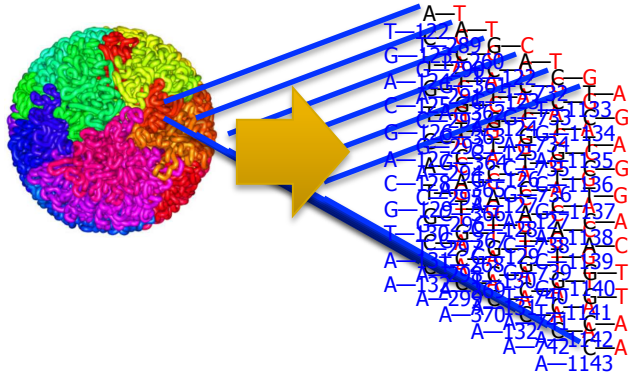
- DNA sequencing is important
- Basic Approach: Read and map



- Next-generation sequencing technologies produce a large number of short DNA fragments
  - More computationally intensive
  - Harder to map the entire genome
  - Especially when allowing polymorphism
- Goal: Design fast and comprehensive algorithms to analyze enormous amounts of NGS data

Challenge of Existing NGS Mapping Tools

- We want Fast and Comprehensive



→ need a fast tool

→ need a comprehensive tool

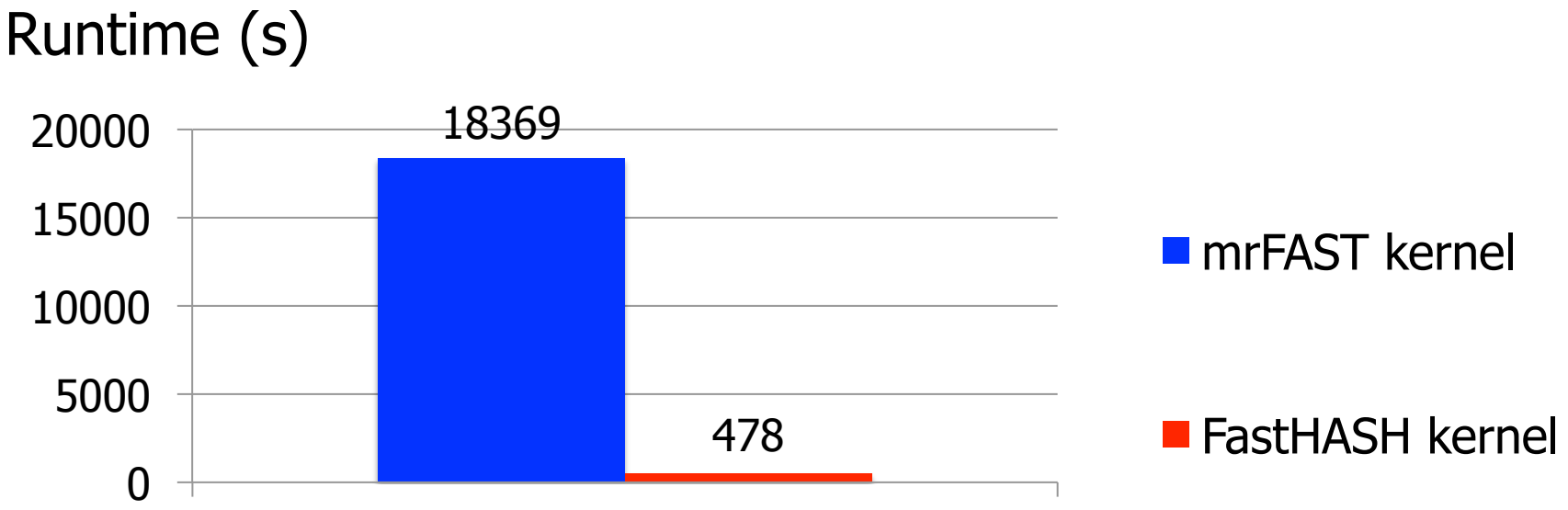
- Some tools are **only fast**
  - BWA, SOAPv2, Bowtie
  - Not comprehensive

- Some tools are **only comprehensive**
  - mrFAST, mrsFAST
  - Slow

FastHASH NGS Mapping Kernel

- **Goal**
  - Best of both worlds: Fast and Comprehensive
- **Observation**
  - Some tools do unnecessary work to guarantee comprehensiveness → main cause of slowness
- **Main Idea of FastHASH**
  - Cut down unnecessary work by taking advantage of full knowledge of reference genome
- **Two main mechanisms**
  - Adjacency Filtering and Cheap Segment Selection

Evaluation



- 38x speedup compared to state-of-the-art alignment tool
  - No sacrifice of comprehensiveness
- Also implemented on GPU for further acceleration

Thanks

- For more information, please come by our poster!

FastHASH:
A New Algorithm for Fast and Comprehensive Next-Generation Sequence Mapping

Hongyi Xin† Donghyuk Lee†
Farhad Hormozdiari‡ Can Alkan§ Onur Mutlu†

† **Carnegie Mellon University** § **University of Washington**
‡ **University of California Los Angeles**

SAFARI Carnegie Mellon