

# CASmap: Splitting Short Reads Alignment with FPGA-based Streamline

## Optimization

Shaoping Ling<sup>1</sup>, Jiahao Liu<sup>2</sup>, Lingtong Hao<sup>1</sup>, Longhui Yin<sup>3</sup>, Lili Dong<sup>1</sup>, Lihua Cao<sup>1</sup>, Wei Zou<sup>1</sup>, Fen Xiao<sup>3</sup>,  
Junsuo Zhao<sup>1</sup>, Chung-I Wu<sup>1,§</sup>, Xuemei Lu<sup>1,§</sup>

<sup>1</sup>Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

<sup>2</sup>Institute of Software, Chinese Academy of Sciences, Beijing, China

<sup>3</sup>Key Laboratory of Intelligent Computing & Information Processing of Ministry of Education, Xiangtan University, Xiangtan, China

<sup>§</sup>Co-corresponding Authors: Chung-I Wu and Xuemei Lu

Email addresses:

SL: [spling@big.ac.cn](mailto:spling@big.ac.cn)

JL: [liujiahao@gmail.com](mailto:liujiahao@gmail.com)

LH: [haolt@big.ac.cn](mailto:haolt@big.ac.cn)

LY: [yinlonghui.big@gmail.com](mailto:yinlonghui.big@gmail.com)

LD: [donglili@big.ac.cn](mailto:donglili@big.ac.cn)

LC: [caolh@big.ac.cn](mailto:caolh@big.ac.cn)

WZ: [zouwei@big.ac.cn](mailto:zouwei@big.ac.cn)

FX: [xiaof@xtu.edu.cn](mailto:xiaof@xtu.edu.cn)

JZ: [junsuo@iscas.ac.cn](mailto:junsuo@iscas.ac.cn)

CW: [wuci@big.ac.cn](mailto:wuci@big.ac.cn)

XL: [luxm@big.ac.cn](mailto:luxm@big.ac.cn)

Short reads alignment, as a core computational issue in HTS (High-Throughput Sequencing) data analysis, it became a bottle-neck in HTS data real-time clinical applications. We created a new alignment system (CASmap) which implemented BWT-based alignment algorithm in a customized desktop reconfigurable computer based on FPGA reconfigurable platform. It accelerated  $\sim 30X$  and  $\sim 2X$  higher than BWA (one thread) and SOAP3 in searching read location candidates in suffix array with the power of FPGA-based streamline optimization. Multi-threading parallelization of smith-waterman algorithm was implemented in multi-core cpu of host to verify the location candidates of reads. CASmap achieved the high speed of  $\sim 310\text{Gbp/day,cpu}$  in aligning real human whole genome sequencing Hiseq pair-end reads (4 mismatches/read,  $2\times 100\text{bp}$ ) with the low power consumption of  $\sim 30\text{w/Gb}$  and high accuracy of 99%. CASmap, as an efficient reconfigurable heterogeneous computing system for short read alignment, provided a new green computing framework for HTS-driven genomic era.

**Keywords: Short Read Alignment, High-throughput Sequencing, High-performance Computing, Reconfigurable Computing, Green Computing, Heterogeneous Computing**