Daehwan Kim

June 20, 2017

Post-doctoral Research Fellow
Center for Computational Biology
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University, School of Medicine
1900 E. Monument St. Rm 101
Baltimore, MD 21205

(240) 997-3411
infphilo@gmail.com
www.ccb.jhu.edu/people/infphilo

Education

Ph.D. in Computer Science May 2013
University of Maryland, College Park, MD, USA
Thesis: RNA-sequencing Analysis: Read Alignment and Discovery and Reconstruction of Fusion Transcripts
Advisor: Steven L. Salzberg
GPA: 4.0/4.0

B.Eng. in Computer Science and Engineering February 2004
Chung-Ang University, Seoul, South Korea
Summa Cum Laude

Research Interests

Computational Biology, DNA and RNA Sequence Alignment, Graph Alignment to Population of Genomes, Genotyping, Personalized Medicine

Professional Experience

Post-doctoral Research Fellow June 2013 – Present
Center for Computational Biology
Johns Hopkins University, School of Medicine, Baltimore, MD

My research projects are related to Computational Biology and Genomics. I develop algorithms and software to rapidly and accurately analyze data generated by high-throughput next generation sequencers. I collaborate with biostatisticians, medical researchers, and other faculty at Johns Hopkins University School of Medicine.

Chief Technology Officer October 2003 – July 2008
Nexon DD, Seoul, South Korea (Game Development Company)

I led the development of several personal computer-based games as a team manager and main developer. I established and continually improved software development and quality assurance processes. My responsibilities also included hiring and managing developers.

Peer Reviewed Publications

(Citation count from Google Scholar, accessed June 20, 2017)
* Equal contribution, † Corresponding author

- **Daehwan Kim***†, Li Song*, Florian Breitwieser*, and Steven L. Salzberg. Centrifuge: rapid and accurate classification of metagenomic sequences. *Genome Research* 2016. 26: 1721-1729. *(Cited 13 times)*

- Zhenguo Sun†, Xiquan Ke, Steven L. Salzberg, **Daehwan Kim**, Valentin Antonescu, Yulan Cheng, Binbin Huang, Jee Hoon Song, John M. Abraham, Sariat Ibrahim, Hui Tian, and Stephen J. Meltzer†. The novel fusion transcript NR5A2-KLHL29 is generated by an insertion at the KLHL29 locus. *Cancer* 10.1002/cncr.30510.

- Mihaela Pertea, **Daehwan Kim**, Geo Pertea, Jeffrey T. Leek, and Steven L. Salzberg†. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie, and Ballgown. *Nature Protocols* 11, 1650-1667, 2016. *(Cited 26 times)*

- **Daehwan Kim**†, Ben Langmead†, and Steven L. Salzberg†. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12, 357-360, 2015. *(Cited 244 times)*

- **Daehwan Kim**†, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* 2013, 14:R36. *(Cited 3,126 times)*

- Cole Trapnell†, Adam Roberts, Loyal Goff, Geo Pertea, **Daehwan Kim**, David R. Kelley, Harold Pimentel, Steven L. Salzberg, John L. Rinn, and Lior Pachter. Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks. *Nature Protocols* 7, 562-578, 2012. *(Cited 3,512 times)*

- **Daehwan Kim**† and Steven L. Salzberg. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biology* 2011, 12:R72. *(Cited 324 times)*

Other Manuscripts

- **Daehwan Kim** and Steven L. Salzberg. Next-Generation Genotyping: graph alignment, and ultrafast and accurate genotyping. *In preparation*.

Honors and Awards

- **Best Poster** 2011
  Johns Hopkins 5th Annual Young Investigator Symposium on Genomics and Bioinformatics
  (Baltimore, MD)

- **Block Grant Fellowship** 2008, 2009
  University of Maryland, College Park, MD

- **Employee of the Year** 2005, 2006
  Nexon DD, Seoul, South Korea
  Awarded $10,000 prize each year

Invited Seminars

- Next-Generation Analysis of Our Genomes.
  Carnegie Mellon University, PA. March 2017.

- Next-Generation Analysis of Our Genomes.
  University of Texas Southwestern Medical Center, TX. March 2017.

- Next-Generation Analysis of Our Genomes.
  University of Washington, Seattle, WA. March 2017.

- Next-Generation Analysis of Our Genomes.
  University of Virginia, Charlottesville, VA. February 2017.

- HISAT-genotype: practical approach for analyzing human variation on a personal computer.
  Pennsylvania State University, PA. September 2016.

- HISAT-genotype: practical approach for analyzing human variation on a personal computer.
  Seoul National University (V. Narry Kim's lab), Seoul, South Korea. June 2016.

- HISAT-genotype: practical approach for analyzing human variation on a personal computer.
  Korea Institute of Science and Technology (KIST), Seoul, South Korea. May 2016.

- HISAT-genotype: practical approach for analyzing human variation on a personal computer.
  Mini-Symposium of Informatics, University of Colorado, Denver, CO. April 2016.

- Graph-based alignment of next-generation sequencing reads to a population of human genomes.
  Broad Institute, Boston, MA. December 2015.

- Graph-based alignment of next-generation sequencing reads to a population of human genomes.
  Seven Bridges Genomics, Boston, MA. December 2015.

- Graph-based alignment of next-generation sequencing reads to a population of human genomes.
  Genome Informatics conference, Cold Spring Harbor Laboratory, NY. October 2015.

Poster Presentations

- HISAT-genotype: fast software for analyzing human genomes on a personal computer. Cold Spring Harbor Laboratory Meeting on the Biology of Genomes. Cold Spring Harbor, NY. May 2016.

- Centrifuge: rapid and accurate classification of metagenomic sequences. Cold Spring Harbor Laboratory Meeting on the Biology of Genomes. Cold Spring Harbor, NY. May 2015.

- SAWHI (now HISAT): Spliced Alignment With Hierarchical Indexing. Cold Spring Harbor Laboratory Meeting on the Biology of Genomes. Cold Spring Harbor, NY. May 2014.

- Reconstruction and Estimation of Fusion Transcripts from RNA-Sequencing. Cold Spring Harbor Laboratory Meeting on the Biology of Genomes. Cold Spring Harbor, NY. May 2013.

- TopHat2: Parallel mapping of transcriptomes to detect indels, gene fusions, and more. Cold Spring Harbor Laboratory Meeting on the Biology of Genomes. Cold Spring Harbor, NY. May 2012.

- TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Cold Spring Harbor Laboratory Meeting on the Biology of Genomes. Cold Spring Harbor, NY. May 2011.

Mentoring

- Mentored Howard Huang, senior undergraduate student at JHU (Biomedical Engineering) for a ten-week internship. 2016.

- Mentored Raymon Cao, sophomore undergraduate student at JHU (Biomedical Engineering) for a ten-week internship. 2016.

- Mentored Joe Paggi, senior undergraduate student at MIT (Computational Biology) for a ten-week internship. 2015.

- Mentored over twenty programmers when I was chief technology officer at Nexon DD. 2004 – 2008.

Software

- HISAT-genotype (2017): our next generation platform that enables rapid and accurate genomic analysis of our genomes using next-generation sequencing data on a desktop within a few hours. The platform currently supports HLA typing, discovery of novel HLA alleles, DNA fingerprinting analysis, and other functionalities. URL: www.ccb.jhu.edu/hisat-genotype

- Centrifuge (2016): a very rapid and memory-efficient system for the classification of next-generation sequences from microbial samples, with better sensitivity than and comparable accuracy to other leading systems. URL: www.ccb.jhu.edu/software/centrifuge

- HISAT2 (2015): a fast and sensitive alignment program for mapping next-generation sequencing reads (both DNA and RNA) to a population of human genomes. Based on an extension of the Burrows-Wheeler transform for graphs, I designed and implemented a graph FM index (GFM), an original approach and its first implementation. URL: www.ccb.jhu.edu/software/hisat2

- HISAT (2014): a fast and sensitive spliced alignment program for mapping RNA-seq reads. In addition to a global FM index that represents the whole genome, HISAT uses a large set of small FM indexes that collectively cover the genome in smaller segments. These small indexes (called local indexes) combined with several alignment strategies enable effective alignment of reads. URL: www.ccb.jhu.edu/software/hisat

- TopHat2 (2013): a splice junction mapper for RNA-Seq reads. TopHat2 aligns RNA-seq reads to mammalian-sized genomes using the ultra high-throughput short read aligner Bowtie, and then analyzes the mapping results to identify splice junctions between exons. URL: www.ccb.jhu.edu/software/tophat

- TopHat-Fusion (2011): an enhanced version of TopHat with the ability to align reads across fusion points, which result from the breakage and re-joining of two different chromosomes, a common occurrence in some types of cancer. URL: www.ccb.jhu.edu/software/tophat/fusion_index.shtml

Additional Background and Skills

- Mathematics: performed independent study of undergraduate-level mathematics including Number Theory, Differential Equation, Abstract Algebra, Vector Calculus, Mathematical Analysis, Topology, and Statistics

- Biology and Chemistry: independently studied General Biology, General Chemistry, Molecular Biology, Organic Chemistry, and Biochemistry

- Programming: >17 years of experience in C/C++, >9 years in Python

- Operating Systems and Clusters: excellent working knowledge of Linux, Mac, Windows, and SLURM

- Completed New England BioLabs' annual two-week *Molecular Biology Summer Workshop* in 2016, principally covering cloning, gene expression, Quantitative RT-PCR, RNA interference, CRISPR/Cas9, and DNA fingerprinting.

- Wet-lab experience in the lab of Prof. Taekjip Ha at Johns Hopkins University: cloning vectors, protein interaction, nucleosome repositioning, and smFRET (Single Molecule Florescence Resonance Energy Transfer). August 2016 – Present.

Other Interests

- Marathon runner. I have participated in thirteen full marathons including the Boston Marathon, the New York City Marathon, and the Chicago Marathon. 2008 – Present.