Benjamin T. Langmead

September, 2012

Assistant Professor
Department of Computer Science
Johns Hopkins University
3400 North Charles St
Baltimore, MD 21218-2682

(410) 516-2033 (Office)
(443) 928-8048 (Cell)
langmea@cs.jhu.edu
www.cs.jhu.edu/~langmea

Education

Ph.D. in Computer Science — May 2012
Algorithms and High Performance Computing Approaches
for Sequencing-Based Comparative Genomics
University of Maryland (College Park, MD)
Advisor: Steven L. Salzberg

M.Sc. in Computer Science — May 2009
University of Maryland (College Park, MD)
Advisors: Steven L. Salzberg and Mihai Pop

B.A. in Computer Science — May 2003
Summa cum laude
Columbia College, Columbia University (New York, NY)

Research Interests

Computational Biology, Genomics, Sequence Alignment, Text Indexing,
Cloud Computing, Parallel & High Performance Computing

Professional Experience

Assistant Professor — July 2012 – present
Department of Computer Science, Johns Hopkins
University (Baltimore, MD)

Research Associate — June 2009 – July 2012
Department of Biostatistics, Johns Hopkins Bloomberg
School of Public Health (Baltimore, MD)

Senior Engineer — June 2003 – August 2007
Reservoir Labs, Inc. (New York, NY)

Peer Reviewed Publications

(Citation counts from Google Scholar, accessed September 16, 2012)

* Equal contribution**,** † Corresponding author

- Herb BR, Wolschin F, Hansen KD, Aryee MJ, **Langmead B**, Irizarry R, Amdam GV, Feinberg AP. Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature Neuroscience. In press.*

- Hansen KD*****, **Langmead B***, Irizarry RA. From whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology. In press.*

- **Langmead B**†, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012, 9:357-359. *(Cited 20 times)*

- Frazee A**, Langmead B**†, Leek JT†. ReCount: A multi-experiment resource of analysis-ready RNA-seq gene count datasets. *BMC Bioinformatics*. 2011, 12:449. *(Cited 3 times)*

- Hansen KD*****, Timp W*****, Corrada Bravo H*****, Sabunciyan S*****, **Langmead B***, McDonald OG, Wen B, Wu H, Liu Y, Diep D, Briem E, Zhang K, Irizarry RA, Feinberg AP. Increased methylation variation in epigenetic domains across cancer types. *Nature Genetics*. 2011 Jun 26;43(8):768-75. *(Cited 64 times)*

- Leek JT, Scharpf RB, Corrada Bravo H, Simcha D, **Langmead B**, Johnson WE, Geman D, Baggerly K, Irizarry RA. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*. 2010 Sep 14. *(Cited 95 times)*

- **Langmead B**, Hansen KD, Leek JT. Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome Biology*. 2010; 11(8):R83. *(Cited 58 times)*

- Schatz MC, **Langmead B**, Salzberg SL. Cloud computing and the DNA data race. *Nature Biotechnology*. 2010 Jul; 28(7):691-3. *(Cited 50 times)*

- **Langmead B**†, Schatz MC, Lin J, Pop M, Salzberg SL. Searching for SNPs with cloud computing. *Genome Biology*. 2009; 10(11):R134. *(Cited 101 times)*

- **Langmead B**†, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*. 2009; 10(3):R25. *(Cited 1,545 times)*

  Winner: Genome Biology award for outstanding paper in 2009.

Other Manuscripts and Publications

* Equal contribution**,** [†] Corresponding author

- Gurtowski J, Schatz MC[†], **Langmead B**[†]. Genotyping in the cloud with crossbow. *Current Protocols in Bioinformatics*. 2012 Sep; Chapter 15:Unit15.3.

- **Langmead B**. Algorithms and High Performance Computing Approaches for Sequencing-Based Comparative Genomics. 2012; Ph.D. thesis, University of Maryland.

- **Langmead B**. Cloud Computing for Data Analysis: Toward the Plateau of Productivity. *Bio IT-World*. 2010 August; Vol. 9, No. 4: 36.

- **Langmead B**. Aligning Short Sequencing Reads with Bowtie. *Current Protocols in Bioinformatics*. 2010 Dec; Chapter 11:Unit 11.7. *(Cited 24 times)*

- **Langmead B**. Highly Scalable Short Read Alignment with the Burrows-Wheeler Transform and Cloud Computing. 2009; Master's thesis, University of Maryland. *(Cited 2 times)*

Honors and Awards

- **Genome Biology Award** for outstanding paper published in 2009   2010
  BioMed Central (London, U.K.)

- **Best Poster**   2009
  Johns Hopkins 3rd Annual Young Investigator Symposium on Genomics and Bioinformatics (Baltimore, MD)

- **Best Poster** in Genetics, Genomics, Bioinformatics, and Proteomics   2008
  University of Maryland 9th Annual Bioscience Day (College Park, MD)

- **Summa Cum Laude**   2003
  Columbia College, Columbia University (New York, NY)

- **Phi Beta Kappa**, New York Delta   2003
  Columbia University (New York, NY)

Research Grants

- **Amazon Web Services in Education Research Grant**   2011
  Supporting ReCount software development ($7,500 in credits)

- **Amazon Web Services in Education Research Grant**   2009
  Supporting Myrna software development ($4,000 in credits)

- **Amazon Web Services in Education Research Grant**   2009
  Supporting Crossbow software development ($3,000 in Credits)

Invited Seminars

- Computational approaches for the DNA sequencing data deluge. Department of Computer Science. *New York, NY*. April 12, 2012.

- Computational approaches for the DNA sequencing data deluge. Courant Institute of Mathematical Sciences. *New York, NY*. March 12, 2012.

- Computational approaches for the DNA sequencing data deluge. Department of Computer Science, Princeton University. *Princeton, NJ*. March 6, 2012.

- Computational approaches for the DNA sequencing data deluge. Department of Computer Science, Johns Hopkins University. *Baltimore, MD*. March 1, 2012.

- High throughput computing approaches for comparative genomics. Cold Spring Harbor Laboratory. *Cold Spring Harbor, NY*. December 16, 2011.

- High throughput computing approaches for comparative genomics. University of Maryland School of Medicine, Institute for Genome Sciences. *Baltimore, MD*. December 1, 2011.

- Bowtie 2: Fast and sensitive read alignment using the FM Index and hardware-accelerated dynamic programming. Cold Spring Harbor Laboratory Meeting on Genome Informatics. *Cold Spring Harbor, NY*. November 3, 2011.

- Cloud-scale genomics: examples and lessons. Beyond the Genome. *Rockville, MD*. September 19, 2011.

- Alignment using the Burrows-Wheeler Transform and FM Index: techniques, open questions. Howard Hughes Medical Institute, Janelia Farm Research Campus. *Ashburn, VA*. June 17, 2011.

- Cloud computing for genome science and methods. 4th Annual Young Investigator Symposium on Genomics and Bioinformatics. *Baltimore, MD*. September 23, 2010.

- Cloud-scale genomics: examples and lessons. The 11th Annual Bioinformatics Open Source Conference (BOSC). *Boston, MA*. July 9, 2010.

- Examples of cloud-scale genomics. The Genome Informatics Alliance Meeting. *Woodinville, WA*. May 6, 2010.

- Cloud-scale genomics: using good trends in technology to fight bad ones. Johns Hopkins Genomics Seminar Series. *Baltimore, MD*. March 29, 2010.

- Bowtie: A highly scalable tool for post-genomic datasets. National Center for Biotechnology Information. *Bethesda, MD*. November 10, 2008.

Poster Presentations

- Bowtie 2: Extending Burrows-Wheeler-based read alignment to longer reads and gapped alignments. Cold Spring Harbor Laboratory Meeting on the Biology of Genomes. *Cold Spring Harbor, NY*. May 12, 2011.

- Cloud-scale RNA-sequencing differential expression analysis with Myrna. Cold Spring Harbor Laboratory Meeting on the Biology of Genomes. *Cold Spring Harbor, NY*. May 11, 2010.

- Genome-Scale Bisulfite Sequencing (co-presented with Sarven Sabunciyan). Seventh annual meeting of the CEGS grantees. *Huntsville, AL*. October 14, 2009.

- Human SNPs from short reads in hours using cloud computing. Johns Hopkins 3rd Annual Young Investigator Symposium on Genomics and Bioinformatics. *Baltimore, MD*. September 24, 2009.
  Winner: Best poster

- Human SNPs from short reads in hours using cloud computing. Workshop on Algorithms in Bioinformatics (WABI). University of Pennsylvania. *Philadelphia, PA*. September 12, 2009.

- Ultrafast and memory-efficient alignment of short reads to the human genome. 12th annual meeting on Advances in Genome Biology and Technology (AGBT). *Marco Island, FL*. February 9, 2009.

- Bowtie, an ultrafast and memory-efficient short read aligner. University of Maryland 9th Annual Bioscience Day. *College Park, MD*. November 20, 2008.
  Winner: Best poster in Genetics, Genomics, Bioinformatics, and Proteomics

Professional Activities

- Reviewer:
    - *ACM Transactions on Database Systems*
    - *BMC Bioinformatics*
    - *Bioinformatics*
    - *Computer Methods and Programs in Biomedicine*
    - *Genome Biology*
    - *Genome Research*
    - *Nature Reviews Genetics*
    - *Nature Methods*
    - *Nucleic Acids Research*
    - *PLoS ONE*

- Editorial boards:
    - *Frontiers in Bioinformatics and Computational Biology.* 2011 - present.

- Program committee:
    - *HiCOMB 2013* (12th IEEE International Workshop on High Performance Computational Biology)

Teaching and Mentoring

- Ph.D. advisees:
    - Ashok Sivakumar

- Classroom instruction:
    - Fall 2012: 600.780: Selected topics in computational genomics

- Co-mentored Biostatistics Ph.D. student Alyssa Frazee, with Jeff Leek. 2011.

- Cloud computing workshop. Beyond the Genome. *Rockville, MD*. September 19, 2011.

- Cloud computing workshop. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health. *Baltimore, MD*. August 31, 2011.

- Guest lecture. "Sequencing-based comparative genomics" for 140.688.01: Statistics for Genomics, Johns Hopkins Bloomberg School of Public Health. *Baltimore, MD*. April 27, 2011.

- Guest lecture. "Next generation sequencing technologies, alignment, base calling" for 140.688.01: Statistics for Genomics, Johns Hopkins Bloomberg School of Public Health. *Baltimore, MD*. April 28, 2010.

Software

- BSmooth (2011): Analysis pipeline for bisulfite sequencing data, including two alternate alignment pipelines. URL: rafalab.jhsph.edu/bsmooth.

- Bowtie 2 (2011): Fast and accurate tool for gapped alignment of reads to reference genomes. Based on the FM Index and Smith-Waterman accelerated with Intel SSE instructions. URL: bowtie-bio.sf.net/bowtie2. *(Downloaded 20,863 times)*

- Myrna (2010): Cloud-enabled tool for aligning mRNA reads and calculating differential gene expression. Runs on Apache Hadoop cluster or Amazon's Elastic MapReduce service. URL: bowtie-bio.sf.net/myrna. *(Downloaded 2,396 times)*

- Crossbow (2009): Cloud-enabled tool for aligning reads and detecting sequence variants. Runs on Apache Hadoop cluster or Amazon's Elastic MapReduce service. URL: bowtie-bio.sf.net/crossbow. *(Downloaded 4,853 times)*

- Bowtie (2008): Ultrafast and memory-efficient software tool for aligning short reads using the FM Index. URL: bowtie-bio.sf.net. *(Downloaded 97,692 times)*

Academic Service

- Biostatistics Information Technology Committee   June 2009 – July 2012

Society Memberships

- American Association for the Advancement of Science (AAAS)
- Association for Computing Machinery (ACM)
- Institute of Electrical and Electronics Engineers (IEEE)
- International Society of Computational Biology (ISCB)
- Phi Beta Kappa
- Society for Industrial and Applied Mathematics (SIAM)