

## Cloud-scale genomics: examples and lessons

Ben Langmead<sup>1,3</sup>, Michael C. Schatz<sup>1</sup>, Jimmy Lin<sup>2</sup>, Mihai Pop<sup>1</sup>, Steven L. Salzberg<sup>1</sup>, Kasper Hansen<sup>3</sup>, Jeff Leek<sup>3</sup>

1. Center for Bioinformatics and Computational Biology, University of Maryland, College Park, MD 20742, USA
2. The iSchool, College of Information Studies, University of Maryland, College Park, MD 20742, USA
3. Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, Maryland 21205, USA

Email: [blangmea@jhsph.edu](mailto:blangmea@jhsph.edu)

Project URL: <http://bowtie-bio.sf.net/myrna>

Download URL: <https://sourceforge.net/projects/bowtie-bio/files/myrna>

License: Artistic License

In the race between DNA sequencing throughput and computer speed, sequencing is winning by a mile. While sequencing throughput has increased at a rate of about 5-fold per year, computer speed generally follows “Moore’s Law,” doubling every 18 or 24 months. As this gap widens, the question of how to design higher-throughput analysis pipelines becomes critical. Cloud computing, which allows any researcher equipped with a credit card to tap into the vast economies of scale afforded by some of the world’s largest data centers, is increasingly seen as one way to combat this trend.

This talk examines and draws lessons from two tools that were designed from the ground up to run on a commercial cloud service. These tools are: Crossbow, a cloud-based tool for SNP genotyping from short reads, and Myrna, a cloud-based tool for calculating differential gene expression from large RNA-seq datasets. The design and performance characteristics of these tools provide some insight into how readily a range of comparative genomics applications can be adapted to the commercial cloud. The discussion will include assessments and advice regarding ease-of-use ease-of-development of cloud applications. The discussion will also highlight ways in which the development process for tools like Crossbow and Myrna could be improved in the future.