## "Representation learning of the epigenome"

## Nathan Sheffield
### Associate Professor, University of Virginia

Host: Paul Medvedev
(Professor of Computer Science and Engineering, and Biochemistry and Molecular Biology
Director of Center for Computational Biology and Bioinformatics)

**Wednesday, May 14, @ 3:00 PM**
**501 Wartik**

Zoom link available upon request

Meeting ID:
982 3347 3767

**Abstract:**

Assays of the human epigenome capture the regulatory state of cells in health and disease. With tens of thousands of experiments completed, it is now feasible to train large-scale representation models. In this talk, I will share how we approach this challenge by first abstracting all epigenome data into genomic intervals. I will describe several types of neural embedding models we use to investigate what biological questions can be addressed with epigenome region embeddings. I will also describe our efforts to curate, standardize, annotate, and share genomic interval data to make it more broadly useful for machine learning and beyond.

**Bio:**

Dr. Sheffield is an Associate Professor at the University of Virginia with appointments in Genome Sciences, Biomedical Engineering, Biochemistry and Molecular Genetics, and Data Science. He holds a PhD in Bioinformatics and Computational Biology and leads a research group specializing in computational regulatory genomics. His lab develops and applies statistical and machine learning methods to analyze bulk and single-cell epigenomic data, including chromatin accessibility, and DNA methylation data. In addition to methodological development, his group creates tools for standardization and distribution of epigenomic data and metadata. Dr. Sheffield chairs working groups within the Global Alliance for Genomics and Health (GA4GH) and the Research Data Alliance (RDA), where he leads efforts to define standards for reference genomes, epigenome metadata, and metadata schemas.

**Refreshments and snacks immediately \*afterward\* in 519 Wartik. Please BRING A MUG to help us "go green", and PLEASE WEAR A MASK during the presentation if you can.**

The semester schedule for WWWGLS is posted at www.ccbb.psu.edu/seminar

## CCBB - Pennsylvania State University

The Weekly Wednesday Wartik Genomics Lecture Series (WWWGLS) is held in 501 Wartik Lab (large commons area) on Wednesdays at 3 pm, unless noted otherwise. Coffee, tea, and snacks are usually served after the presentation. PSU faculty can nominate speakers via this form PSU affiliates can sign up to receive seminar announcements using L-WWWGLS@lists.psu.edu.

www.ccbb.psu.edu

, and is updated as information becomes available.