

Big Data Training for Cancer Research

Special Lecture Series

**Bioconductor:
software for orchestrating high-throughput genomics data analysis**

Dr. Sean Davis

May 20, 2023, 1:00 – 2:15 PM (EDT)

Abstract: Progress in biotechnology is continually leading to new types of data, resulting in data sets that are rapidly increasing in volume, resolution and diversity. The promise of unprecedented advances in our understanding of biological systems and in medicine is challenged by complexity and volume of data also challenge scientists' ability to analyze them. Meeting this challenge requires continuous improvements in analytical methods and capable, usable software tools implementing them. Bioconductor is a well-established open-source, open-development software project for the analysis and comprehension of high-throughput data in genomics and molecular biology. The project aims to enable interdisciplinary research, collaboration and rapid development of scientific software. Based on the statistical programming language R, Bioconductor comprises more than 2000 interoperable packages from a large, diverse community of scientists. These packages undergo a formal initial review and continuous automated testing. Each package includes documentation and working example use cases. Bioconductor supports many types of high-throughput sequencing data and associated annotation resources; contains mature facilities for microarray analysis; and covers proteomic, metabolomic, flow cytometry, quantitative imaging, cheminformatic and other high-throughput data. Package interoperability enables the rapid creation of workflows combining and integrating multiple data types and tools for statistical inference, regression, network analysis, machine learning and visualization at all project stages from data generation to publication. A growing community of researchers and users contributes to ongoing development, online support, and education. The influence of the project is evidenced by more than 250,000 downloads per year and tens of thousands of citations in the literature. I will present an overview of the project for prospective users and contributors.



Speaker Bio: Dr. Davis is a pediatric oncologist and biomedical data scientist. He joined the University of Colorado School of Medicine and Comprehensive Cancer Center in January, 2021. Previously, he spent 19 years at the National Institutes of Health in the Intramural Research Program at NHGRI and NCI. His research interest is in the development and application of software, tools, and databases for comprehending genomic and other high-throughput biological data.