Building a Cloud Based Infrastructure for Transparent and Reproducible Analysis of Omic Data

Abstract: Big data is a termed often used to describe data sets that are too large to be analyzed with commodity computer hardware running end-user software. Until recently, high complexity big data sets from genomics and proteomics were obtained by domain experts and analyzed by dedicated bioinformatics staff. Rapid advances in next-generation sequencing and mass spectrometry have driving down the cost of genomic and proteomic experiments making the acquisition of these complex data sets available to a much broader research community. More and more investigators are seeking answers from next gen sequencing and proteomics in lieu of traditional bench-top molecular biology. The adoption of omics in the lab has been further aided by a fervent community of scientists creating user friendly data analysis software. Investigators now have a dizzying array of options for analyzing short read sequencing or tandem mass spectrometry data. Given these trends in omic science, the call for transparent and reproducible data analysis pipelines is gaining increased momentum. This talk will explore examples of integrative genomic and proteomic workflows from the perspective a bench scientist attempting to address the ongoing challenges for developing transparent and reproducible scientific workflows. A strategy for how data analysis pipelines can be efficiently disseminated and reproduced across research labs will be presented.

About the Speaker: Michael A. Freitas, PhD, is the Director of the OSU Comprehensive Cancer Center Proteomics Shared Resource and Director of the OSU Wexner Medical Center's Biomedical Sciences Graduate Program. He is an Associate Professor in the Department of Cancer Biology and Genetics with an adjunct appointment in the Department of Biomedical Informatics. Dr. Freitas has been working in the field of mass spectrometry for >30 years and has spent the last 20+ years developing and improving mass spectrometry and bioinformatic methods in the fields of cancer proteomics and chromatin biology. He is a recipient of the Camille and Henry Dreyfus New Faculty Award and the American Society for Mass Spectrometry Research Award. His research is focused on development and application of multiomic methods to understand chromatin regulatory networks in cancer. The lab has extensive expertise in characterizing protein modifications and identifying protein:protein interactions by mass spectrometry. Dr. Freitas is also a cofounder of MassMatrix Inc, a software company that develops data analysis solutions for BioPharma. Dr. Freitas is a leader in bioinformatics education at OSU. He is the founder and primary faculty leader of the OSU Data Science for Scientists club, a club that meets weekly to discuss topics in programming, quantitative data analytics and machine learning.

Michael A. Freitas, PhD
Friday, September 25th, 11:00am-12:00pm
Carmen Zoom