



H3Africa
Human Heredity & Health in Africa



H3ABioNet
Pan African Bioinformatics Network for H3Africa



Workshop Title/name: Introduction to hands-on bioconductor

Course Overview/Description:

There will be two courses offered by the BioConductor community:

Introduction to Bioconductor Annotation Resources (Thursday, 29th April 2021 @ 8am EDT/12noon GMT/1pm WAT/ 2pm CAT/3pm EAT to 11am EDT/3pm GMT/ 4pm WAT/5pm CAT/6pm EAT)
(<https://jmacdon.github.io/Bioc2020Anno/#introduction-to-bioconductor-annotation-resources>)

There are various annotation packages provided by the Bioconductor project that can be used to incorporate additional information to results from high-throughput experiments. This can be as simple as mapping Ensembl IDs to corresponding HUGO gene symbols, to much more complex queries involving multiple data sources. In this workshop, we will cover the various classes of annotation packages, what they contain, and how to use them efficiently.

Epidemiology for Bioinformaticians (Friday, 30th April 2021 – 8am EDT/12noon GMT/1pm WAT/ 2pm CAT/3pm EAT to 11am EDT/3pm GMT/ 4pm WAT/5pm CAT/6pm EAT)
(<https://chloemirzayi.com/EpiForBioWorkshop2020/#epidemiology-for-bioinformaticians>)

Concepts of causal inference in epidemiology have important ramifications for studies across bioinformatics and other fields of health research. In this workshop, we introduce basic concepts of epidemiology, study design, and causal inference for bioinformaticians. Emphasis is placed on addressing bias and confounding as common threats to assessing a causal pathway in a variety of study design types and when using common forms of analyses such as GWAS and survival analysis. Workshop participants will have the opportunity to create their structural causal models (DAGs) and use this model to determine how to assess an estimated causal effect. Examples using DESeq2, edgeR, and limma will be used to show how multivariable models can be fitted depending on the hypothesized causal relationship.

Presented successfully at BioC2019 to 30 people, updates that material by adding a brief demonstration of ggdag, revised conceptual emphasis based on participant feedback, and applied examples using data from curatedMetagenomicData.

The introduction to hands-on Bioconductor workshop is a **virtual** two-half-day training to provide an overview of leveraging open-source Bioconductor resources for research studies. The course tools and packages relevant for this course will be hosted on Orchestra, an online platform for teaching and learning hands-on computational skills in self-contained and launchable workshop environments.

The proposed courses will lay a foundation on how to efficiently use open-source Bioconductor resources for bioinformatics analysis. The course will be in two main forms which include;

- ❖ lectures to introduce basic concepts
- ❖ demonstrations and hands-on computer practicals on analysis pipeline/workflows for mapping identifiers, querying multiple data sources, performing epidemiological causal inference, and visualizing results in high-resolution publication format.

The hands-on sessions may provide an opportunity for participants to work with their data.

Instructors for the course have experience in developing and applying methods for research analysis and also involved in developing statistical methods/algorithms and Bioconductor packages



Instructors:



Chloe Mirzayi
Data Scientist
chloe.mirzayi@sph.cuny.edu
City University of New York School of Public Health

Chloe Mirzayi, MPH is a doctoral candidate in Epidemiology at the City University of New York (CUNY) Graduate School of Public Health and Health Policy and Data Scientist at the CUNY Institute for Implementation Science in Population Health (ISPH). She has experience designing, implementing, and maintaining data collection, management, and curation tools on a variety of projects ranging from large national cohort studies to RCTs in topics related to COVID-19, microbiome, HIV, substance use, and mental health. Her research interests are focused on the role of the microbiome on human health, causal inference in microbiome research, and improving the epidemiological rigor of health and bioinformatics research.

Chloe's dissertation research will focus on the impact of antibiotics on the human gut microbiome. Her mentor and dissertation committee chair is CUNY ISPH Investigator Dr. Levi Waldron.



Levi Waldron
Associate Professor
LEVI.WALDRON@SPH.CUNY.EDU
City University of New York, Graduate School of
Public Health and Health Policy

Dr. Levi Waldron is an Associate Professor of Epidemiology and Biostatistics at the CUNY Graduate School of Public Health and Health Policy in New York. He completed a PhD at the University of Toronto and post-doctoral training at the Harvard School of Public Health and Dana-Farber Cancer Institute. His



H3Africa
Human Heredity & Health in Africa



H3ABioNet
Pan African Bioinformatics Network for H3Africa



methodological research interests focus on creating well-documented and user-friendly methods, data structures, and databases to make publicly available genomics and human microbiome data more valuable and widely used to alleviate the burden of disease.<https://waldronlab.io>



Sean Davis
Professor
seandavi@gmail.com
University of Colorado Anschutz School of Medicine

I received my B.S. in [Mechanical and Aerospace Engineering](#) from [Princeton University](#), my M.D. and Ph.D. from the University of Pittsburgh in the Schools of [Medicine](#) and [Public Health](#) (thesis advisor, [Dr. Daniel Weeks](#), [Human Genetics](#)), respectively. [Childrens Hospital and Regional Medical Center](#) was where I completed my residency in Pediatrics, followed by pediatric hematology/oncology training in the [Johns Hopkins School of Medicine](#)/National Cancer Institute [joint fellowship](#)



Benjamin Haibe-Kains
Associate Professor
benjamin.haibe.kains@utoronto.ca
Department of Medical Biophysics, Faculty of Medicine, University of Toronto

Dr. Benjamin Haibe-Kains is a Senior Scientist at the Princess Margaret Cancer Centre (PM), University Health Network, and Associate Professor in the Medical Biophysics department of the University of Toronto. Dr. Haibe-Kains earned his PhD in Bioinformatics at the Université Libre de Bruxelles (Belgium). Supported by a Fulbright Award, he did his postdoctoral fellowship at the Dana-Farber Cancer Institute and Harvard School



H3Africa
Human Heredity & Health in Africa



H3ABioNet
Pan African Bioinformatics Network for H3Africa



of Public Health (USA). Dr. Haibe-Kains' research focuses on the integration of high-throughput data from various sources to simultaneously analyze multiple facets of carcinogenesis. Dr. Haibe-Kains' team is analyzing large-scale radiological and (pharmaco)genomic datasets to develop new prognostic and predictive models to improve cancer care.

Support team

If you have specific questions about the courses, please email Francis Agamah (agmfra001@myuct.ac.za/francis.agamah@uct.ac.za), Verena Ras (verena.ras@uct.ac.za), Paballo Chauke (Paballo.Chauke@uct.ac.za)

Keywords: Bioinformatics, Epidemiology, study design, causal inference, BioConductor packages

Skill level of training: Intermediate

Language: The course and materials are offered in English

Credential awarded: Meeting participant will receive a workshop certificate

Type of training: Online

Venue of workshop: Virtual workshop hosted on Zoom

Workshop organizers: BioConductor, H3Africa, and H3ABioNet

Registration opens: 9th April 2021

Registration closes: 22nd April 2021

Notification date: 23rd April 2021

Link to the application form: <https://htraindb.h3abionet.org/node/2935>

Workshop Sponsors: H3Africa, H3ABioNet, Bioconductor

Intended Audience:

The course is designed for biomedical research graduate students (master, PhD), early-stage researchers, and bioinformaticians who would like to build their capacity in data analysis using Bioconductor resources and are familiar with R syntax and using RStudio.

Syllabus and Tools:

R statistical software
Bioconductor packages

Prerequisites:

- ❖ Computer with stable internet connection
- ❖ Basic knowledge of R syntax (<https://www.youtube.com/watch?v=V8eKsto3Ug>) and using RStudio



- ❖ Basic understanding of the various annotation sources (NCBI, EBI/EMBL) for the Introduction to Bioconductor annotation resources
- ❖ Basic background in statistics, regression, and epidemiology for the Epidemiology for Bioinformaticians

Learning objectives/outcomes:

Introduction to Bioconductor annotation resources:

- ❖ Be able to use select and mapIds methods to map between identifiers
- ❖ Be able to extract data from TxDb and EnsDb packages
- ❖ Be able to make queries using biomaRt
- ❖ Extract and utilize various data from AnnotationHub

Epidemiology for Bioinformaticians

- ❖ Assess a study design in terms of causal inference
- ❖ Learn about path blocking to prevent confounding
- ❖ Create a DAG in R using daggity and ggdag
- ❖ Identify situations when multivariate adjustment for variables is inappropriate
- ❖ Specify a model based on a DAG and then fit that model to data using DESeq2, edgeR, and limma

Workshop limitations: This workshop is not an introductory course to using R and its syntax, and will provide a foundation for continued learning in bioinformatics analysis using Bioconductor.

Workshop Programme:

Time	Topic	Trainer
29th April 2021		
2pm CAT	Introduction to Bioconductor annotation resources	Sean Davis Levi Waldron Benjamin Haibe-Kains
30th April 2021		
2pm CAT	Epidemiology for Bioinformaticians	Chloe Mirzayi Levi Waldron