NCI Cancer Data Science Community
Training Resources

Listed below you will find cancer data science training resources and events available to the NCI community. This includes seminars, in-person and online courses, workshops, email lists, SIGs, and more. *Note:* this list reflects current, known resources. If you are aware of other training opportunities, please let us know so we can add to this list. (NCICBIITDataScienceTraining@mail.nih.gov)

TOPICS

Genomics and Bioinformatics

- **Bioinformatics User Forum (**https://abcsfrederick.info/BUF**):** The Bioinformatics User Forum is a place to build the bioinformatics community within the NCI and Frederick National Laboratory as well as among our collaborators at NIH, Ft Detrick and beyond. Anyone interested in the bioinformatics challenges in these communities is welcome to join us (search for "Bioinformatics-User-Forum" at list.nih.gov to sign up for our list serve).

BACK TO TOP

NCI Cancer Data Science Community
Training Resources

- Bioinformatics Support Service at the NIH Library. The NIH Library offers bioinformatics support services for NIH staff and select HHS agencies. This service is free and features bioinformatics consulting and training, complemented by analysis tools and databases hosted on high-performance computing workstations. https://go.usa.gov/xdPcX.

- **Statistics for Lunch** (https://abcsfrederick.info/Stats4Lunch)**:** Statistics for Lunch focuses on exploring topics in statistics at an accessible, intuitive level. We usually provide a high level, easy to understand seminar followed by a practical, hands-on session to practice what is covered in the seminar.

- **Programmer's Corner** (https://abcsfrederick.info/ProgrammersCorner): Programmer's Corner is a quarterly meeting of programmers to discuss topics of interest to the group. We usually have 2 or 3 short talks centered around a common topic, followed by discussion.

- Bioinformatics Training and Education Program (BTEP) https://btep.ccr.cancer.gov/ Contact: NCIBTEP@mail.nih.gov

- CBIIT Computational Genomics and Bioinformatics Group. Contact Dr. Daoud Meerzaman

- **Cancer Prevention and Control (CPFP) Colloquia Series:** upcoming presentations

- **CBIIT Cancer Data Science Workshops** Past Presentations

- Ingenuity Pathway Analysis - Software to integrate and streamline solutions for users in the discovery, translational, and clinical spaces

Image Processing

- **Data Science Bowl -** a collection of tutorials on image segmentation https://www.kaggle.com/c/data-science-bowl-2017/kernels

- **NCI's Cancer Imaging Archive (TCIA)** http://www.cancerimagingarchive.net/

- **Assay Guidance Manual**
  - **Assay Development Guidelines for Image-Based High Content Screening, High Content Analysis and High Content Imaging** https://www.ncbi.nlm.nih.gov/books/NBK100913/

  - **Advanced Assay Development Guidelines for Image-Based High Content Screening and Analysis** https://www.ncbi.nlm.nih.gov/books/NBK126174/

Data Science

- **Big Data to Knowledge (BD2K) Training Coordination Center** https://bigdatau.ini.usc.edu

BACK TO TOP

- ○ Provides an educational resource discovery index to explore, aggregate and organize biomedical and data science educational content. The database contains more than 10,000 videos, documents, assignments, books and courses for personal and professional use.

- **NLM's University-based Biomedical Informatics and Data Science Research Training Programs** https://www.nlm.nih.gov/ep/GrantTrainInstitute.html

  - ○ The U.S. National Library of Medicine supports research training in biomedical and data science at sixteen educational institutions in the United States. This website provides information related to Informatics Training and a list of academic institutions participating in the NLM training program.

- **HHS Data Science CoLab.** Brings together learners, practitioners, and collaborators to participate in *Reimagine Data*, building a collaborative community around data science, and learning cohorts. The curriculum includes: GitHub, R programming, data visualization, foundational statistics, linear regression, and Machine Learning. https://www.hhs.gov/cto/initiatives/data-science-colab/index.html

- **NIH Data Science Distinguished Seminar Series**. A dynamic lecture series that brings high-profile researchers and experts to the NIH main campus to exemplify the roles that integration of the computational and the quantitative sciences play in the today's most innovative biomedical research, health research policy development, and sustainable research practices.https://datascience.nih.gov/community/datascience-at-nih/seminar

- **Cold Spring Harbor Laboratories Courses**. A list of courses, and workshops sponsored by the Cold Spring Harbor Laboratories in cancer research, cell biology, omics, neuroscience, genetics, immune/infect, neural data science, imaging structure, programming for biology, and computational genomics among other topics. https://meetings.cshl.edu/courseshome.aspx

- **Data Science SIG Slack Workspace.** In an effort to enhance data scientist collaboration, communication, and information exchange, the NIH Data Science Special Interest Group organizers have created a new Slack workspace to discuss all things data science. The community engagement and interest has been substantial, as we have had 70 data scientists join the new workspace in 12 hours. Feel free to join us. http://bit.ly/nih-datascience-slack

Machine Learning

- **The Exascale CANcer Distributed Learning Environment (CANDLE)** https://cbiit.cancer.gov/ncip/hpc/candle

  - ○ Using the challenges within each JDACS4C pilot to shape priorities for the CANDLE project, the DOE laboratories are drawing upon their strengths in high-performance computing (HPC), machine learning and data analytics, coupled with domain strengths at NCI and Frederick National Laboratory for Cancer Research to establish the foundations for CANDLE. Exploiting exascale

technologies and capabilities anticipated for deep and machine learning, the project is scheduled to deliver critical technologies to the community that will be used to advance precision oncology.

- CANDLE Workshops at NIH - https://wiki.nci.nih.gov/display/HPC/CANDLE+Workshops

Links to NIH, NCI at Frederick, and HHS Training Resources

- **NIH Library training** - https://www.nihlibrary.nih.gov/training/calendar

- **NIH HPC Training** - https://hpc.nih.gov/training

- **Foundation for Advanced Education in the Sciences** (FAES). Graduate school for the NIH (https://faes.org)

  - Serves NIH sites in Bethesda, Frederick, Rockledge, Baltimore, Arizona, North Carolina, and Montana. Offers 150 academic for-credit courses per year, including over 60 biotechnology training workshops (Advanced Studies in Bioinformatics and Data Science)

  - **Course Catalog of all FAES Courses** https://faes.org/content/explore-courses

- **HHS Data Science Co-Lab** - Offers occasional NIH-wide training https://www.hhs.gov/idealab/dscolab/

- NCI@Frederick Data Science Initiative - a collaboration between NCI and the University of Maryland. Offers a series of four workshops to provide a meeting ground for informal exchange of knowledge, ideas and networking to facilitate collaborative projects that will be carried out by UMD data science graduate and undergraduate students under the joint supervision of UMD and NCI researchers.

Groups and Listservs

- **NIH.AI** This website is a forum for trans-NIH efforts related to Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL). https://nciphub.org/groups/nihai

- **NIH Data Science Listserv** https://list.nih.gov/cgi-bin/wa.exe?SUBED1=nih-datascience-I&A=1

- **Bioinformatics Listserv** https://list.nih.gov/cgi-bin/wa.exe?SUBED1=BIOINFO-GENERAL-NCI&A=1

- **NIH Statistical Code Review Group** https://datascience.nih.gov/community/workforce/resources#training-code

- **NIH Bioinformatics Listserv** https://list.nih.gov/cgi-bin/wa.exe?SUBED1=BIOINFO-GENERAL-NCI&A=1

BACK TO TOP

- **Bioinformatics User Forum Listserv**
  https://list.nih.gov/cgi-bin/wa.exe?SUBED1=BIOINFORMATICS-USER-FORUM&A=1

Other Programs and Initiatives

- NIH Biomedical Information Science and Technology Initiative (BISTI). Coordinates research grants, training opportunities, and scientific symposia associated with biomedical computing. Regular monthly meetings are conducted to discuss program status, future needs and directions, and topics of interest to the bioinformatics community.

- NCI Information Technology for Cancer Research (ITCR). Training and Outreach Coordination Center (still in concept stage). For information contact: nciitcr@mail.nih.gov

Fellowship Training Opportunities

- **Accelerating Therapeutics for Opportunities in Medicine (ATOM) Consortium.**
  https://atomscience.org ATOM Fellowship Program
  https://pharm.ucsf.edu/smdc/programs-projects/atom-faqs

- **Big Data-Scientist Training Enhancement Program (BD-STEP).** A collaboration between the Department of Veterans Affairs (VA) and the National Cancer Institute
  https://cssi.cancer.gov/bd-step

- **Center for Strategic Scientific Initiatives (CSSI)**.
  https://www.va.gov/oaa/specialfellows/programs/sf_bdstep.asp

NCI Training Offerings Across DOCs (partial list)

- **Center for Cancer Research** Bioinformatics Training and Education Program (BTEP)

- **Division of Cancer Epidemiology and Genetics (DCEG) Training Resources for Fellows and Staff**
  https://dceg.cancer.gov/fellowship-training/training-resources-for-fellows-and-staff

- **caDSR Training Wiki: caDRS, EVS, and the Semantic Infrastructure**
  https://wiki.nci.nih.gov/display/COREtraining/caDSR+Training+Wiki

- **NCI Cancer Imaging Archiving Tutorial** http://www.cancerimagingarchive.net/

- **NCI at Frederick Computer Software & Statistical Services Training**
  https://ncifrederick.cancer.gov/Services/Computer/Training/default.aspx

BACK TO TOP

NIH High-Performance Computing (HPC) Group / Biowulf

- The [NIH High-Performance Computing (HPC) Group](#) plans, manages, and supports HPC systems for the NIH intramural community. If users need assistance, walk-in consults are held once a month on the NIH Main Campus. For information, contact: staff@hpc.nih.gov.

    - [Introduction to Biowulf](#): an online, self-paced course that provides an introduction to the [Biowulf system](#). Recommended for new and experienced Biowulf users.

        - https://irp.nih.gov/supercomputing

        - (https://hpc.nih.gov/training/intro_biowulf/)

NIH Graduate Data Science Summer Program (GDSSP)

- The NIH will introduce the [Graduate Data Science Summer Program](#) in summer 2019. The program is a partnership between the NIH Office of Intramural Training & Education and the newly formed Office of Data Science Strategy. The GDSSP is designed for master's students in computer and data science with an interest in the biomedical research enterprise. GDSS students will spend the summer working at the NIH side-by-side with some of the world's leading biomedical researchers. In addition to working in a research group, GDSSP students will participate in a customized curriculum that will help them to explore the many uses of data science in biomedical research and to improve their leadership skills, including self-awareness, resiliency, conflict management, effective mentoring, and emotional intelligence. The GDSSP is a cohort based program where students can build peer networks and lifelong friendships.

KEY CONTACTS

Carl McCabe, Lead, NCI CBIIT Scientific Computing Program: Carl.Mccabe@nih.gov

Janelle Cortner, Lead, Scientific Data Management and Storage: Janelle.Cortner@nih.gov

Daoud Meerzaman, Lead, Computational Genomics and Biomedical Informatics Group; Scientific Software License Portfolio: Daoud.Meerzaman@nih.gov

George Zaki, High-Performance Computing (HPC): george.zaki@nih.gov

[BACK TO TOP](#)

Lynn Borkon, Strategic Collaborations & Workshops: lynn.borkon@nih.gov