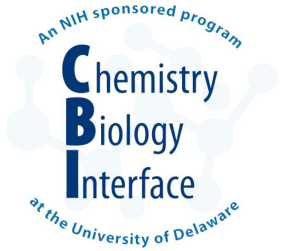
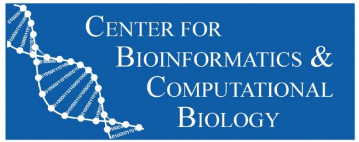
FAIR Data Practices for Omics Analysis Workshop

Welcome and Overview

University of Delaware

April 18, 2022 (Day 1)





So…who are you?

CBCB Bioinformatics Core

Mission

" . . . to provide the expertise & computational infrastructure support necessary to empower life science researchers at UD and our partner institutions to pursue bioinformatics and data science-enabled lines of research."

help@bioinformatics.udel.edu

CBCB Bioinformatics Core

Consultation

- Project Planning
- Grant Proposals

Project Support

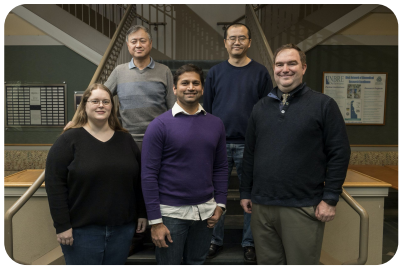
- 'Omics Analysis
- Literature/Data Mining
- Web/Database Support
- Scientific Programming
- Application Development/Hosting

Training Workshops

Infrastructure Support

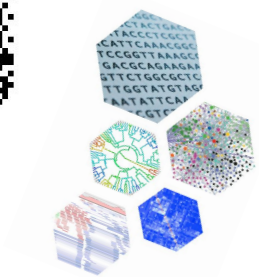
Fee-based & Collaborative Project Models

How Can We Help?



help@bioinformatics.udel.edu

core.bioinformatics.udel.edu



CBCB Bioinformatics Core

help@bioinformatics.udel.edu

Shawn Polson
Director, CBCB Bioinformatics Core Facility
Director, Bioinformatics Network of Delaware (Delaware INBRE)
Associate Professor (CISC, PLSC, BISC)
Expertise: *Genomics, Transcriptomics, Metagenomics, Phylogenetics, Genomic Technologies, High Performance Computing, Bioinformatics Programming*

Jaysheel Bhavsar
Bioinformatics Programmer
Adjunct Instructor
Expertise: *Bioinformatics Programming, Web Design, High Performance Computing, Transcriptomics, Metagenomics*

Amelia Harrison
Bioinformatics Training Coordinator
Expertise: *Microbial/Viral Community Analysis, Ecology, Phylogenetics, Metagenomics*

Madolyn MacDonald
Associate Bioinformatics Scientist
Expertise: *Genome Annotation, Bioinformatics Programming, Biopharmaceutical, Adventitious Virus Detection*

Hongzhan Huang
Associate Professor (CISC)
Expertise: *Proteomics, Pathway/Enrichment Analysis, Biostatistics*

Data Science

Chuming Chen
Associate Professor (CISC)
Expertise: *Data Management and Data Integration, Cloud Computing, Big Data Analytics, Deep Learning, Bioinformatics, Semantic Web and Ontology Engineering*

Julie Cowart
Bioinformatics Software Engineer
Expertise: *Software Engineering, Data Management, Web Development*

DBI BioIT

Karol Miaskiewicz
Manager, DBI Computing Operations
Expertise: *High Performance Computing, Systems Administration, Computational Chemistry*

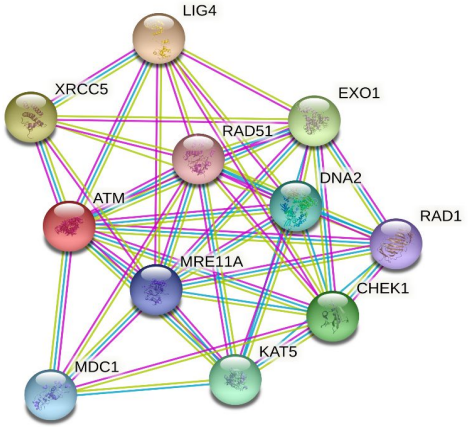
‘Omics Analyses

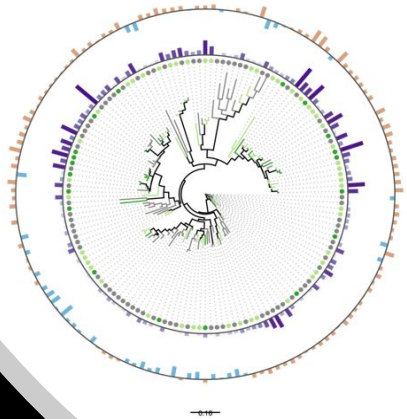
Approaches (Illumina, PacBio, other sources)

- RNA-Seq (mRNA, mi/smRNA, lincRNA)
- ChIP-Seq, ATAC-Seq
- Variant Analysis (SNP, InDel, Structural)
- *De novo* Genome Assembly
- Genome Annotation
- Metagenomics/Metatranscriptomics
- Microbial Community Analysis (16S, others)
- Proteomics Data Analysis
- Others

Interpretation

- Multi-omics Integration
- Pathway, Gene Ontology Annotation
- Statistical Analyses
- Enrichment Analysis
- Data Visualization





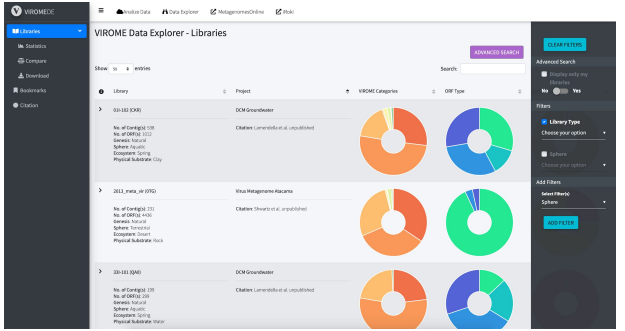
help@bioinformatics.udel.edu

Application Development

Developing an effective way to display and explore project data can enhance the ability to biologically interpret and leverage the results.

We can work with you to develop custom web-based applications to explore your data.

These applications can be restricted to your own internal use . . . or used as a way to share dynamic views of your data with collaborators or the scientific community in general.



help@bioinformatics.udel.edu

CBCB/DBI Cyberinfrastructure

High Performance Computing: BioMix

- Computational Cluster for Bioinformatics Applications
  - 2772 cpu cores, >22TB of RAM (256GB – 2TB per node)
  - User storage allocations
  - Slurm Management: interactive, batch, docker
  - Common bioinformatics software maintained and supported
- Open access for UD/INBRE Education and Research

Data Center

- >3.5PB of storage including 2.4PB Dell Isilon Storage Cluster (NIH S10)
- Database design and hosting; web portal hosting
- Development and Production Web and Database Servers

Software Support

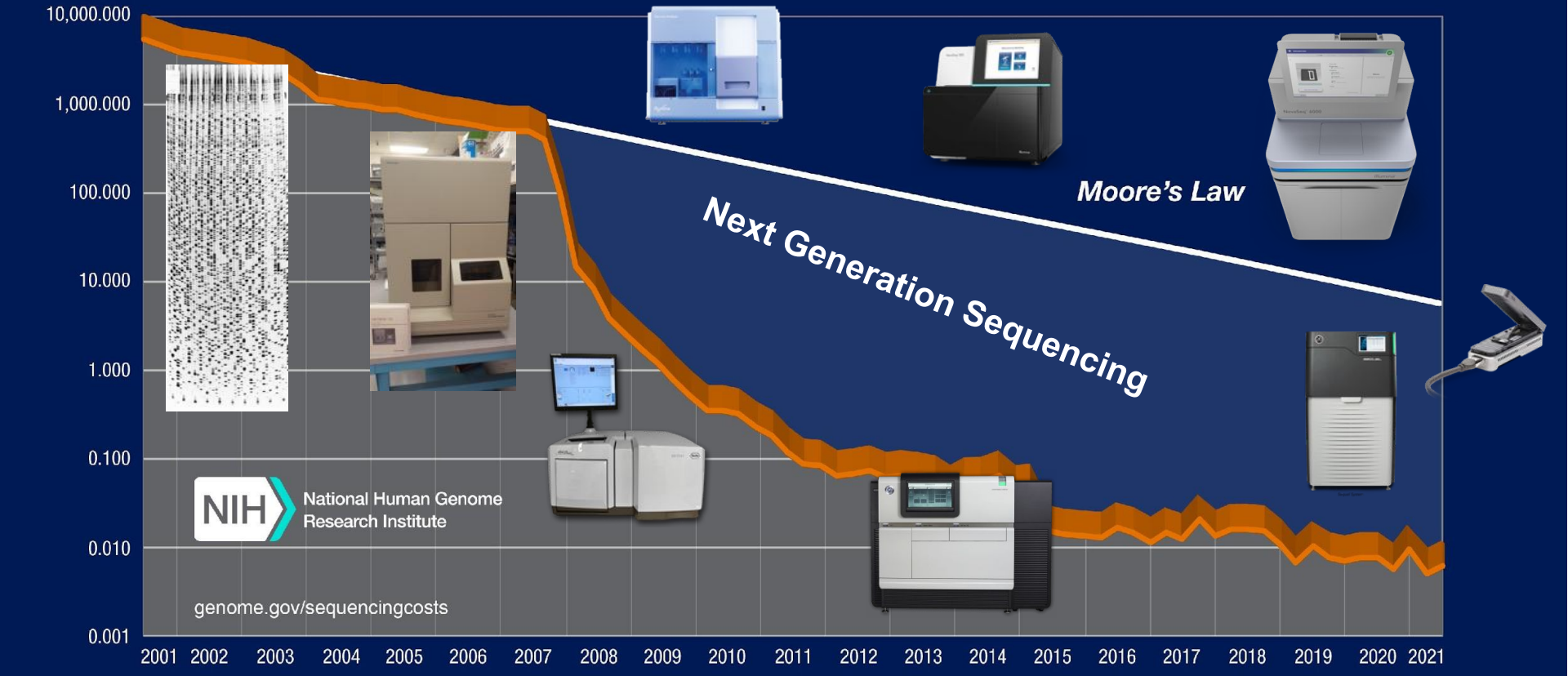
- Commercial (CLC) and open-source software tools

OK…and why are we here?

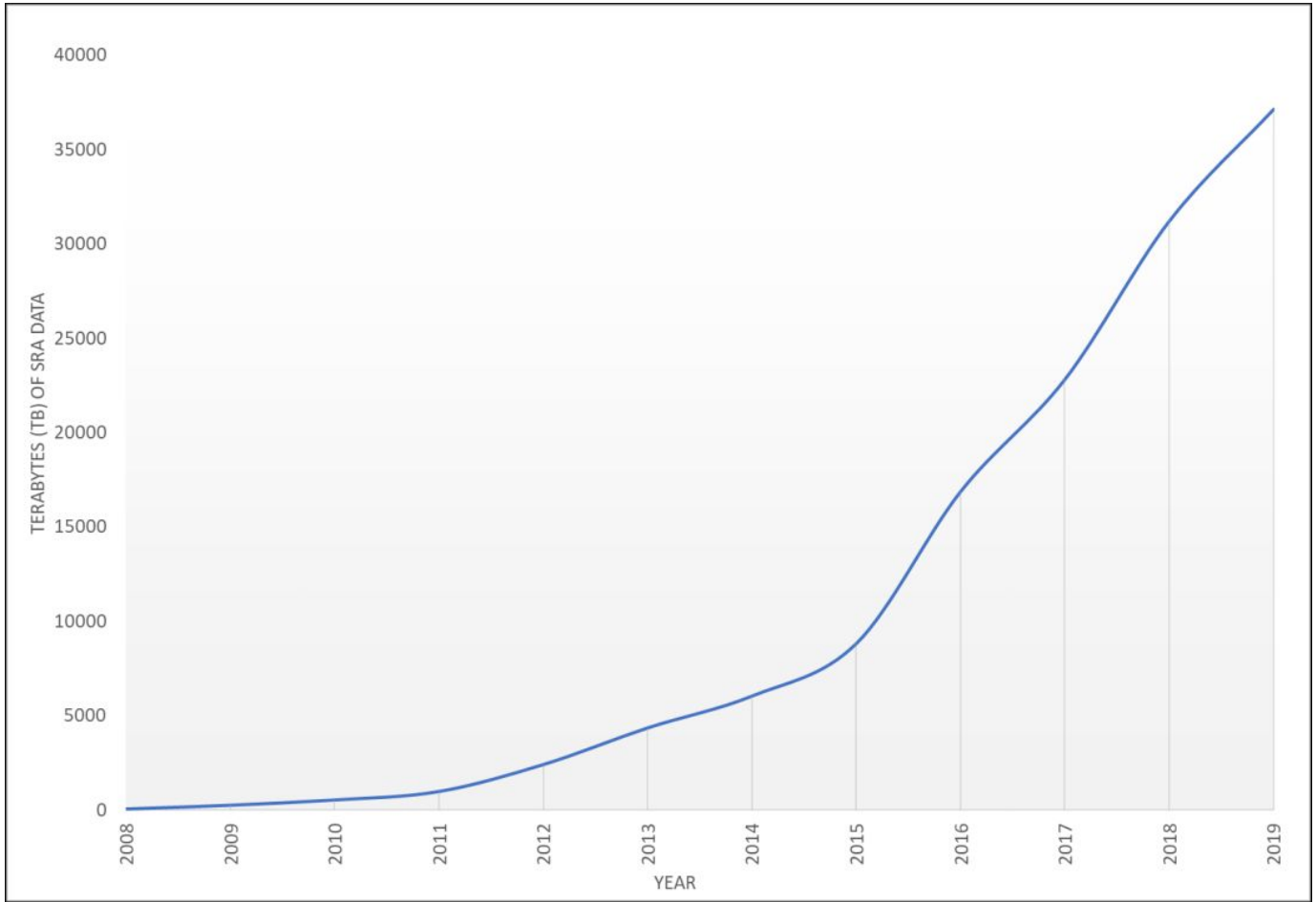
Its an increasingly data-centric society: from social media, to wearable devices and the internet of things, to data-driven science research

…And whether you are a data scientist or not – we all need to understand how this will impact our lives and work

Cost per Raw Megabase of DNA Sequence



Growth of NCBI Sequence Read Archive



What can we do with all this data?

- With so much sequence and other biological data available…it is now possible to do “meta-analyses” that combine publicly available data and ask new questions of it

- In the near future, it is likely that many of us will be leveraging such datasets to answer questions relevant to our own research

- **BUT!** This data is only useful if the original investigators adequately described its context

FAIR Data Practices for Omics Analysis

- Learn about the importance of applying FAIR data practices to data generated in the laboratory

- Gain hands-on experience with executing and recording information about an omics analysis workflow on a high performance computational cluster

- Practical application of FAIR data principles for submitting results to publicly-accessible data repositories

Practical AI/ML for Computational Biology and Chemistry Workshop

June 13 - 17, 2022

1:00pm – 5:00pm

Seats are limited, register ASAP.

Registration is limited to UD students, faculty, and staff.

Questions should be directed to:
workshop@bioinformatics.udel.edu

We are excited to announce "Practical AI/ML for Computational Biology and Chemistry" Workshop. This workshop will provide participants with a conceptual understanding of various AI/ML approaches, practical applications of AI/ML in computational biology and chemistry, and hands-on exercises that emphasize the importance of data preparation and readiness for AI/ML.

We will give participants opportunity with first-hand experiences on the issues in dealing with data that is not well-prepared. We will cover various data formats, processing and wrangling techniques to get the data into a form where it can be utilized by AI/ML algorithms. We will also teach different visualization techniques to better understand the data at hand. Basic Python programming is recommended but not required. **We will offer a Python training June 1-3 to help meet this need.**

The workshop will be held in-person on June 13 to 17 from 1:00pm to 5:00pm (see Syllabus) in the Ammon Pinizzotto Biopharmaceutical Innovation Center conference room 140 on STAR Campus.

Let’s get to know each other!

Tell us:

- Your name
- Who you are
- Something about your research (elevator pitch)
- What you’d like to take from the workshop
- 2 Truths and a Lie about yourself

Draft Schedule

Monday, April 18, 1-5pm

- Welcome, Introductions, and Workshop Overview
- **Activity**: Technology/Software Orientation and Test
- **Lecture**: Overview of FAIR Data Principles (Harrison)
- **Activity**: Good/bad data practice discussion (Harrison)
- **Tutorial:** Web-based resources (Polson)
- **Assignment:** Web-based bioinformatics exercise

Tuesday, April 19, 1-5pm

- **Activity:** Assignment report out
- **Lecture:** Computational Notebooks (Harrison)
- **Activity**: Documenting computational work (Harrison)
- **Lecture**: Overview of Omics Analysis (Polson)
- **Tutorial:** HPC and Biomix (Polson)
- **Assignment**: Biomix exercise

Draft Schedule

Wednesday, April 20, 1:30-5:30pm

- **Lecture:** Intro to Biomedical Ontologies (Arighi)
- **Lecture**: Intro to SNP Project (Harrison)
- **Tutorial:** Exploring NCBI SRA (Harrison)
- **Activity**: SRA dataset exploration
- **Lecture**: Environment Ontology (Ferrell)
- **Assignment**: SRA/ontology assignment

Thursday, April 21, 1-5pm

- **Lecture**: Genomic Variant (SNP) Analysis (Polson)
- **Tutorial:** SNP 1: QC and Trimming (Polson)
- **Tutorial:** SNP 2: Reference Mapping & Variant Detection (Polson)
- **Lecture**: Mis-annotations (Harrison)
- **Lecture:** Interoperability (Harrison)
- **Assignment:** Complete Variant Detection

Friday, April 22, 1:30-5:30pm

- **Tutorial**: SNP 3: Comparing Variants (Polson)
- **Lecture**: Iroki Overview (Harrison)
- **Tutorial:** SNP 4: Viewing and Decorating your Tree (Polson)
- **Demonstration**: Depositing sequence data in NCBI repositories (Polson)
- **Lecture:** Other data repositories (Harrison)
- **Discussion/Q&A** (feel free to bring questions about your research)