

**Early
Detection
Research
Network**



**Biomarkers AI and Bioinformatics Workshop
Planning
Meeting #2**

January 23, 2024



Hold a three-day workshop followed by a hackathon at Caltech in Spring/Summer 2024 to gather researchers, practitioners, and experts in cancer science, informatics and related fields, particularly at the intersection of biomarkers and AI.



Through a series of keynotes, contributed talks, and extensive discussions around use cases, needs, and capabilities, develop recommendations for the next few years to support advancement and application of data-driven approaches to cancer biomarker research leveraging datasets from the EDRN and related consortiums.

Proposed Tasks

1. Agree on goals and purpose – [Meeting #1](#) (November 3, 2023)
2. Agree on high level sessions – [Meeting #1](#) (November 3, 2023)
3. Workshop write up – Google document distributed
4. Identify additional program committee members – Recv'd suggestions
5. Identify dates for the workshop which includes ensuring Caltech facilities are open – Meeting #2 (January 23, 2024)
6. Setup website on EDRN portal – Draft developed
7. Distribute invitation and call for posters – Early February
8. Assign program committee members to – Meeting #3 (TBD)
 - Chair sessions
 - Chair posters
 - Organize the hack-a-thon
9. Build schedule and identify key speakers – Meeting #4 (TBD)

Potential Sessions

- *In Silico and Real World Biomarker Discovery and AI* - methodology, application and verification
 - *Cancer biomarker discovery -- use cases and opportunities for AI*
 - *Biomarker data (knowledge) Integration*: Integration of imaging, radiomics, high dimensional data (Proteomics's, genomics, epigenetics, etc.)
 - *Showcasing Datasets for AI/ML* - real world data, synthetic data, simulated data
- *Biomarker Data Preparation, Computation, and Analysis*
 - *Uncertainty, Statistical Rigor, Bias Considerations in AI/ML*
 - *Statistical methods vs AI considerations*
 - *Explainability, Reproducibility, Interpretability, etc of results*
 - *Creating AI ready datasets*: Data Preparation, Harmonization and standardization
 - *Data capture, sharing, federation, and scalable computation* for cancer biomarkers
- *Novel / emerging applications in image analysis using AI?*
- *Ethical considerations* with data and AI/ML
- *Emerging Capabilities and Methodologies in AI* - LLMs/Foundation Models, Generative AI, Federation/Federated Learning
 - *Application to biomarker research and biomarker science*
- *Academic-Industry Partnerships* in AI and Bioinformatics

Potential Dates – Summer 2024

Month	Monday – Wednesday	Tuesday - Thursday	
June	Mon, Jun 3 – Wed, Jun 5	Tue, Jun 4 – Thu, Jun 6	
	Mon, Jun 10 – Wed, Jun 12	Tue, Jun 11 – Thu, Jun 13	
	Mon, Jun 17 – Wed, Jun 19	Tue, Jun 18 – Thu, Jun 20	
July	Mon, Jul 8 – Wed, Jul 10	Tue, Jul 9 – Thu, Jul 11	
	Mon, Jul 29 – Wed, Jul 31	Tue, Jul 20 – Thu, Aug 1	
August	Mon, Aug 5 – Wed, Aug 7	Tue, Aug 6 – Thu, Aug 8	4/5
	Mon, Aug 12 – Wed, Aug 14	Tue, Aug 13 – Thu, Aug 15	5/5
	Mon, Aug 26 – Wed, Aug 28	Tue, Aug 27 – Thu, Aug 29	

Hack-a-thon

Leverage data captured from EDRN along with public datasets.

Include ideation/brainstorming about combining diverse datasets for specific big questions that may have been considered out of reach just a few years ago.

Include bioinformatics and AI/ML experts PIs, junior investigators and all those interested in learning more about the application and use of various bioinformatics and AI/ML tools and methods to cancer biomarker research.

Explore use of synthetic data and ML foundation models

Website Under Development



[EDNRN](#) / [News and Events](#) / [Upcoming Meetings](#) / [AI & Bioinformatics Workshop](#)

AI & Bioinformatics Workshop

Welcome! This page contains information about the upcoming artificial intelligence and bioinformatics workshop, projected to be held in the Spring/Summer of 2024.

- [Registration Form](#)
- [Call for Posters](#)

Purpose of the Workshop

Some of the major advances in the field of cancer over the last few years include a shift towards more data intensive and computational workflows through a variety of tools. This is no different for cancer biomarkers. The increase of data and computation coupled with more and more emerging capabilities in bioinformatics, machine learning (ML), and artificial intelligence (AI), is unlocking new opportunities to apply data-driven methods to scientific discovery and validation. Rapid advances in areas such as foundation models and generative AI are providing more and more capability in areas such as image segmentation and identification. As research consortia—such as the Early Detection Research Network (EDRN)—continue to capture more data, there are numerous opportunities to leverage and apply shared tools, algorithms, and capabilities within the community. In order to leverage these capabilities, there is a need to ensure that diverse datasets are captured, structured, and readily accessible for analysis. Similarly, there is a need to provide more and more packaged tools and services to support the analysis.

A three-day workshop followed by a hackathon is being planned at Caltech for Spring/Summer 2024 to gather researchers, practitioners, and experts in cancer science, informatics and related fields, particularly at the intersection of biomarkers and AI. Through a series of keynotes, contributed talks, and extensive discussions around use cases, needs, and capabilities, we plan to develop recommendations for the next few years to support advancement and application of data-driven approaches to cancer biomarker research leveraging datasets from the EDRN and related consortia.

Workshop Goals

- Define use cases and needs for cancer biomarker discovery and validation
- Discuss the state-of-the-art and current gaps in applying AI/ML and bioinformatics methods to cancer biomarker research
- Discuss challenges around reproducibility, interpretability, and explainability
- Discuss current informatics capabilities and infrastructure needs to support use cases
- Develop a set of goals and recommendations to address these needs in cancer biomarker research for the next few years

Hackathon

A hackathon is being planned which will leverage data captured in EDRN's LabCAS biomarker data commons, and leveraging public datasets. Rather than being just about programming/hacking, the hackathon will include ideation/brainstorming about combining diverse datasets for specific big questions that may have been considered out of reach just a few years ago. The high level workflows required for this means that besides bioinformatics and AI/ML experts, others who will find a role to play include PIs, junior investigators and all those interested in learning more about the application and use of various bioinformatics and AI/ML tools and methods to cancer biomarker research.

Two breast cancer datasets are available, although more datasets from more organs (such as lung, pancreas, prostate, etc.) are welcome. We are working on making available additional reference sets.

Participants will work with the data through LabCAS, accessing data from S3, running workflows on AWS instances that are capable of interfacing with standard ML aids like Jupyter notebooks. It will be expected that virtual environments are used. While there is no restriction on which language to use, expected primary languages are Python, R

Program Committee

Dan Crichton, NASA/JPL

Heather Kincaid, NASA/JPL

Ashish Mahabal, Caltech

Ziding Feng, Fred Hutchinson Cancer Research Center

Chad He, Fred Hutchinson Cancer Research Center

Matt Schabath, Moffitt Cancer Research Center

Anirban Maitra, MD Anderson

Jennifer Beane, Boston University

Eugene Koay, MD Anderson

Paul Boutros, UCLA

Zhen Zhang, Johns Hopkins University

Sudhir Srivastava, National Cancer Institute

Christos Patriotis, National Cancer Institute

Amanda Skarlupka, National Cancer Institute

We will continue to add more program committee members over time

Goals

- Define use cases and needs for cancer biomarker discovery and validation
- Discuss the state-of-the-art and current gaps in applying AI/ML and bioinformatics methods to cancer biomarker research
- Discuss challenges around reproducibility, interpretability, and explainability
- Discuss current informatics capabilities and infrastructure needs to support use cases
- Develop a set of goals and recommendations to address these needs in cancer biomarker research for the next few years

Proposed Audience

- EDRN Site Investigators and their bioinformatics/data science/computational science researchers and teams
- Other NCI Programs, including CBIIT
- Academic partners and other researchers at the intersection of cancer and AI/bioinformatics