Biomarkers AI and Bioinformatics Workshop Planning
Meeting #4

April 12, 2024

Agenda

1. Review task list

2. Advertising Slide for the EDRN Steering Committee Meeting

3. Session / Theme Chairs

4. Speakers

5. Hack-a-thon

   1. Planning Team

   2. Datasets and Exercises

6. Logistics

Proposed Tasks

1. ~~Goals and purpose~~
2. ~~High level sessions~~
3. ~~Workshop description drafted~~
4. ~~Identify additional program committee members~~
5. ~~Identify dates for the workshop which includes ensuring Caltech facilities are open~~
6. ~~Setup website on EDRN portal~~
7. ~~Distribute invitation and call for posters~~
8. Program committee member assignments
   - Session / Theme Chairs
   - Hack-a-thon Chairs
   - Key speakers
9. Venue – meals and support
10. Post skeleton structure (early May)
11. Review abstracts (May)
12. Open registration
13. Build schedule and continue to discuss speakers

Sessions/Themes

In Silico and Real World Biomarker Discovery and AI

      Cancer biomarker discovery — use cases and applications of AI

      Biomarker data (knowledge) Integration: imaging, radiomics, high dimensional data (proteomics, genomics, epigenetics, etc.)

      Showcasing datasets for AI/ML — real world data, synthetic data, and simulated data

Biomarker Computation and Methodology Considerations for applying AI

      Uncertainty, statistical rigor, bias considerations in AI/ML

      Statistical methods vs. AI: optimal approaches to enable data analysis

      Trust in ML methodologies for cancer biomarkers: explainability, reproducibility, and interpretability of results

Considerations in Data Preparation, Sharing, and Analysis

      Creating AI ready datasets: data preparation, harmonization, and standardization

      Making data usable: data capture, sharing, federation, and scalable computation for cancer biomarkers

      Using Open Science to link AI algorithms and datasets to publications

      Ethical considerations with application of AI/ML

Emerging Capabilities and Methodologies in AI

      Novel and emerging methods for image analysis

      LLMs/Foundation Models, Generative AI, Federation/Federated Learning

Academic-Industry Partnerships in AI and Bioinformatics

      Technology transfer

      Structuring mutually beneficial relationships

Sessions/Themes Assignments

Session/Themes	Organizer(s)	Moderator(s)	Speakers
In Silico and Real World Biomarker Discovery and AI	Zhen, Matt		
Biomarker Computation and Methodology Considerations for applying AI	Steve, Zhen		
Considerations in Data Preparation, Sharing, and Analysis	Eugene, Jen		
Emerging Capabilities and Methodologies in AI	Ashish, Matt		
Academic-Industry Partnerships in AI and Bioinformatics	Eugene, Chad		

Speaker Suggestions (to date)

Potential Speaker	Institution	
Anima Anandkumar	Caltech	Bren Professor of CS/AI; Nvidia
Fei-Fei Li	Stanford	Professor of CS/AI; Director of Stanford AI Lab
Bo Wang	University of Toronto	Lead Scientist for AI
Ying Yuan	MD Anderson	Professor of Biostatistics
Thomas Fuchs	Mount Siani and Paige.AI	Director of Computational Pathology; Founder
Faisal Mahmood	Harvard University	Professor of Pathology
Ned Sharpless	UNC	Director of Cancer Policy & Innovation; Former NCI Director
Regina Barzilay	MIT	Professor of AI & Health
Louis Staudt	NCI	Director, Center for Cancer Genomics
Nikolaus Schultz	Sloan Kettering	Director, Cancer Data Science Initiative;
Niven R. Narain	BERG Health	Chairman, President & CEO
Andrea Califano	Columbia and Chan Zuckerburg	Professor and President, CZ Biohub New Yor
Susan Gregurick	NIH	Director, NIH Data Science
Changhuei Yang	Caltech	Professsor of Biophysics
Ryan Scott	NASA Ames	Science Lead for the NASA Ames Life Sciences Data
Yolanda Gil	USC/ISI	Director of Data Science; National Academy Chair for AI and Science
Alex Szalay	JHU	Director of Insittue for Data Intensive Science for Astronomy and Biology

https://docs.google.com/spreadsheets/d/1bEcyaWSDFOIJpBI-fHHNlwLx8_pXAwqpDPoLOK0faRA/edit#gid=0

Hackathon related timeline

Finalize hackathon planning team and setup call (April 19) - need a couple of volunteers

Finalize datasets (Tue, April 30) - need one more dataset

Problem definitions for hackathon (Tue, 21 May)

Techniques/libraries identified (Tue, 11 Jun)

Initial notebooks created (Tue, 2 Jul)

Datasets and problems and libraries and notebooks loaded (Tue, 23 Jul)

Meeting starts (Tue, 13 Aug)

Work continues (Aug 15++)

During the workshop

Day 1: Overview and problem description (poster?) - discussions around open problems / opportunities

Day 2: Informal discussions encouraged; rooms available for meetings

Hackathon Day 3: Short presentations: problem; data; focused exercises with AI ready datasets and configured notebooks

EDRN Datasets

Can use:

- Automated quantitative breast density measures (Moffitt; Dr. J Heine)
- Breast Cancer Biomarker Analysis (Moffitt; Dr. J Heine)

We are exploring a few other datasets but open to suggestions

Venue/Meals

Plan for morning and afternoon refreshments

Potential late afternoon/evening poster session and social

Lunch options

- Allow participants to pre-order for lunch brought in

- Lunch "on your own"