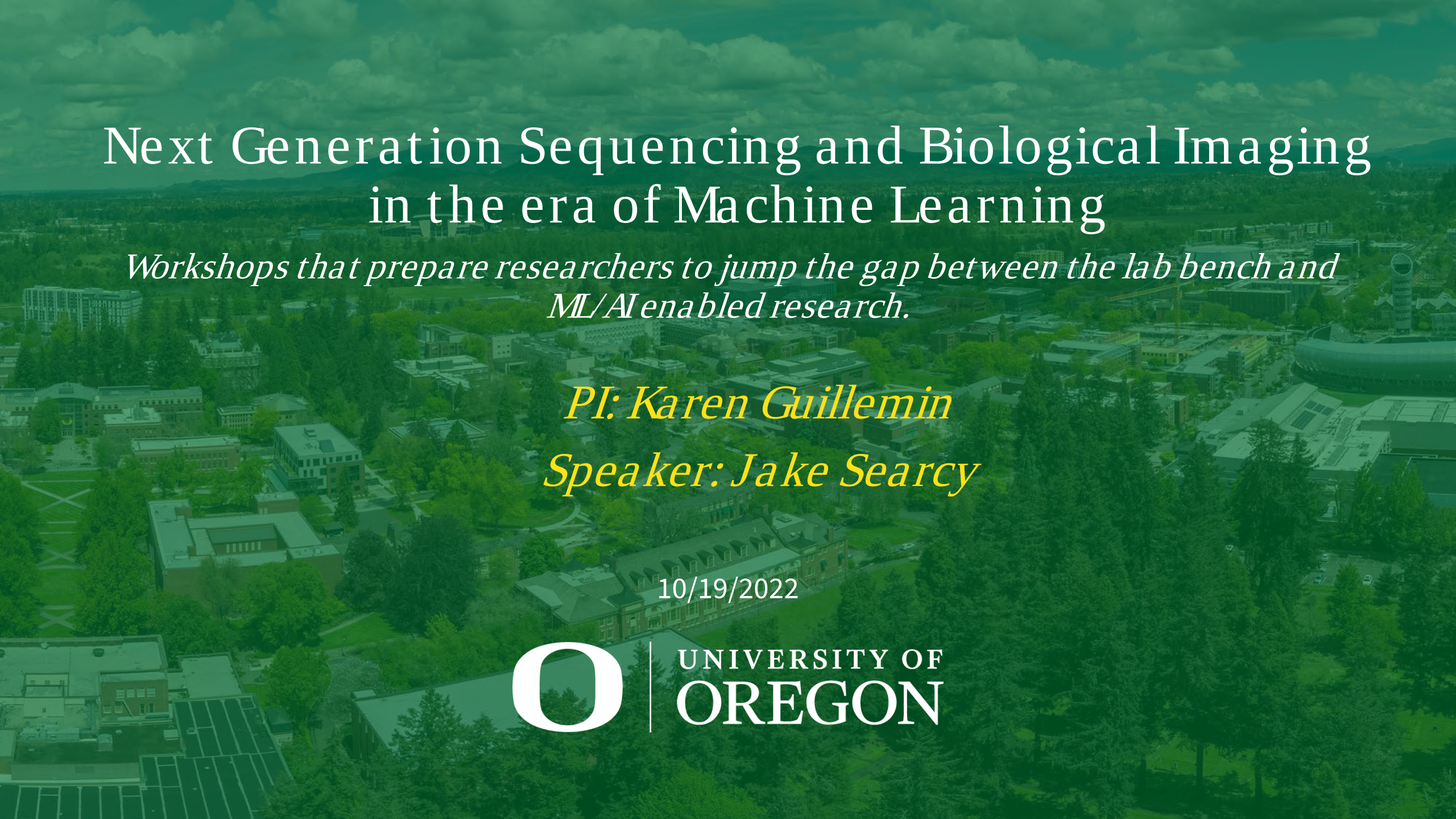
Next Generation Sequencing and Biological Imaging in the era of Machine Learning

Workshops that prepare researchers to jump the gap between the lab bench and ML/AI enabled research.

PI: Karen Guillemin

Speaker: Jake Searcy

10/19/2022



UNIVERSITY OF OREGON

Summary

- AI and ML is driven by high quality data
- It's not always clear to bench scientists what 'high quality' data means
- Our goal was to help span this barrier by teaching the basics of machine learning (ML), what makes datasets AI ready and foundational data manipulation skills needed to create an AI ready dataset
- Developed 3, 1-week workshops
  - Introduction to Data for ML (Gateway week)
  - Sequencing Data for ML
  - Imaging Data for ML

Instructors

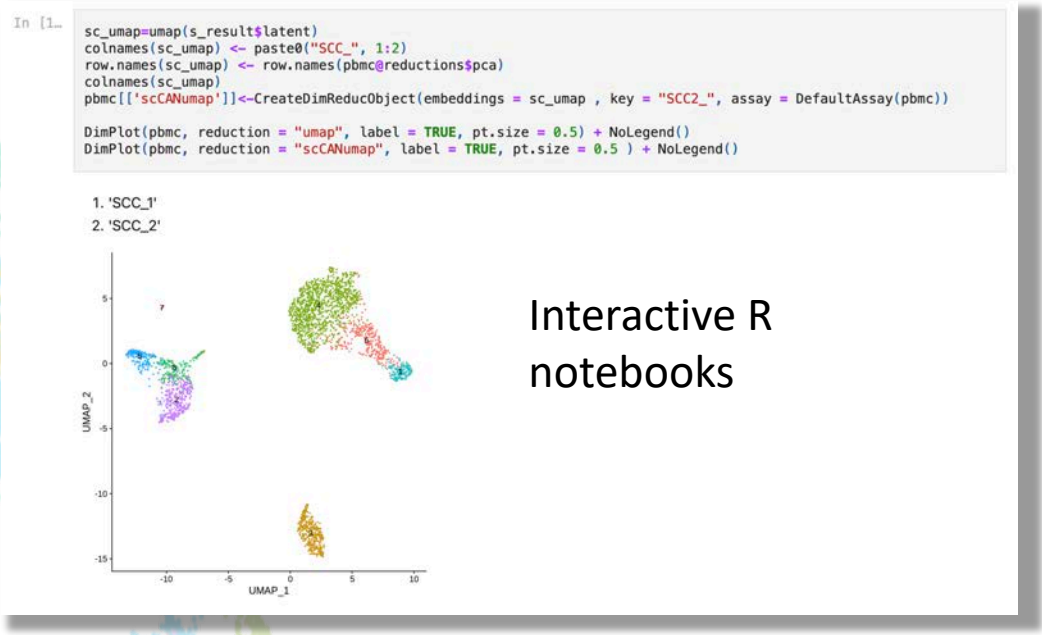
Jake Searcy

Emily Beck

UNIVERSITY OF OREGON

Highlights

- Our workshops were a mix of lectures, interactive programing exercises, and guest speakers
  - Workshop Page
- All material is available for re-use on github
  - Github Link
- Post Survey Results:
  - **All respondents report feeling more comfortable with using the techniques taught** with a variety of topics being reported as ‘most useful’
    - Bash Scripting, Basic ML techniques in R, Data management and harmonization



Interactive R notebooks

Guest Speakers

Bill Cresko | Lillian Aoki | Kaye Shek | Doug TurnBull

UNIVERSITY OF OREGON

Challenges

- Wide range of incoming skill sets and goals
  - Participants ranged from advanced undergraduates to faculty
  - Obtaining Introductory level bash and R skills was a goal of many participants
  - Most participants interested in ML were already familiar with these tools
  - ***Important to be flexible, optional challenge exercises helped keep more advanced students engaged***
- Maintaining attendance across all three workshops was a challenge
  - 3 weeks was a large commitment for a co-curricular activities
  - Due to low registration, we moved our third week (Image week) to an asynchronous format with material and videos being available to students, with office hours for questions and discussions.

Future Work

- Experimenting with different delivery methods
  - Shorter Bite-sized Chucks for co-curricular delivery
  - Adding material directly to new and existing courses for more comprehensive training
- Prepare students for success
  - More frequent introductory training in bash, R, and python
- Building material collections
  - Working with data librarians to collate these materials with others into a searchable online database, to make them easier to reuse by students and other educators.