NHGRI Genomic Data Science Working Group (GDSWG)

Mark Craven, Ph.D.

Department of Biostatistics and Medical Informatics
Department of Computer Sciences
University of Wisconsin-Madison

September 13, 2021

NIH National Human Genome Research Institute

The Forefront of Genomics®

What is the NHGRI Genomic Data Science Working Group?

- A Subcommittee of the National Advisory Council for Human Genome Research
- Formed in Spring 2017
- Meetings are approximately every two months

Working Group Functions

- Advises NHGRI on plans related to genomic data science outside of Council meetings
- Provides input to NHGRI Director and other Institute leaders about trans-NIH issues related to data science
- Addresses broad challenges, including data management, analysis, computing and data science policy, as they relate to all areas of genomics, from basic science to genomic medicine

Current GDSWG Members

Updated September 2021

Michael Boehnke, Ph.D.
University of Michigan

Mark Craven, Ph.D.
University of Wisconsin-Madison

Trey Ideker, Ph.D.
University of California
San Diego

Gail Jarvik, M.D., Ph.D.
University of Washington

Eimear Kenny, Ph.D.
Icahn School of Medicine at Mount Sinai

Anshul Kundaje, Ph.D.
Stanford University

Christina Leslie, Ph.D.
Sloan Kettering Institute

Shannon McWeeney, Ph.D.
Oregon Health & Sciences University

Casey Overby Taylor, Ph.D.
Johns Hopkins University

Marylyn Ritchie, Ph.D.
University of Pennsylvania

Paul Sternberg, Ph.D.
California Institute of Technology

Outgoing Members:

Eric Boerwinkle

Anthony Philippakis

NHGRI Representatives: Eric Green, Carolyn Hutter, Valentina Di Francesco, Sean Garin

Workgroup Topics in 2020-2021: A Recap

- Renewal of two NHGRI CGDS FOAs: PAR-21-254 (R01), PAR-21-255(R21)
- NHGRI AnVIL project updates
- Overview of training and outreach activities for broadening and bridging genomics/data science communities

Workgroup Topics in 2020-21: A Recap

- Discussed other NIH programs that encourage the use and development of AI/ML approaches in biomed research
  - E.g. NIH Bridge2AI

- Organized and hosted the Machine Learning in Genomics Workshop, April 13-14, 2021
  - Over 1,000 participants per day
  - Over 70 countries represented

Machine Learning in Genomics Workshop: Format

- 14 invited speakers in 5 sessions
- 30 minutes of Q&A, discussion at the end of each session
- Talk videos, slide PDFs are available at https://www.genome.gov/event-calendar/Machine-Learning-in-Genomics-Tools-Resources-Clinical-Applications-and-Ethics

Machine Learning in Genomics Workshop: Sessions

- What are the opportunities and challenges for ML in genomics research?
- Algorithm development and machine learning approaches in genomics
- Ethical, Legal and Social Implications (ELSI) of machine learning in genomics
- Data and resource needs for machine learning in genomics
- Machine learning in clinical genomics

Machine Learning in Genomics Workshop: Follow Up

- The working group has drafted a report outlining recommendations along six themes:
  - Algorithm development and machine learning approaches in genomics
  - Ethical, Legal and Social Implications (ELSI) of machine learning in genomics
  - Data and resource needs for machine learning in genomics
  - Machine learning in clinical genomics
  - Training and outreach for machine learning in genomics
  - Collaboration with industry
- The working group is also planning a more detailed manuscript for publication in a journal

Machine Learning in Genomics Workshop: Data Generation and Resource Recommendations

- Support more sequencing across the evolutionary tree for ML models that are based on evolution
- Support data generation to address underrepresentation of different genetic ancestries in genomics datasets
- Enable creative ways to augment observational data sets with rationally selected, model-driven experiments, including perturbations
- Establish and support best practices for robust dataset generation with extensive, standardized metadata
- Facilitate early and easy access to both raw and processed data sets

Potential Upcoming Working Group Discussion Topics

- Development of a NHGRI strategy for leveraging advances in ML approaches in genomics, based on workshop recommendations

- Education & Training

- Organizing another genomic data science workshop in line with the NHGRI 2020 Vision

- Implementation challenges of the new NIH Data Management and Sharing Policy

Questions?

The Forefront of Genomics®