National Cancer Institute – Department of Energy Collaborations: Extending Frontiers of Predictive Oncology and Exascale Computing

Emily Greenspan[1], Amy Gryshuk[2], Eric Stahlberg[3], Christine Chalk[4], Thuc Hoang[5]

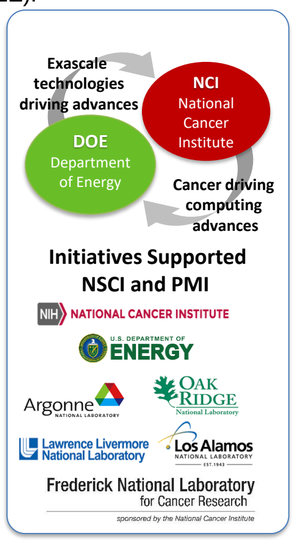
[1]Center for Biomedical Informatics & Information Technology, National Cancer Institute; [2]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory;; [3]Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc.; [4]Office of Science, Department of Energy;[5]National Nuclear Security Administration, Department of Energy

NATIONAL CANCER INSTITUTE
Center for Biomedical Informatics & Information Technology

Abstract

In 2016, the National Cancer Institute and the US Department of Energy established an inter-agency collaboration to accelerate precision oncology and shape the future for emerging exascale computing.

The initial partnership program, *Joint Design of Advanced Computing Solutions for Cancer (JDACS4C)*, is driven and enabled by several government initiatives, and explores three co-designed pilot efforts where exascale computing capabilities and computational approaches join precision oncology priorities. The partnership is also developing new cross-cutting technologies including uncertainty quantification (UQ) methods to evaluate the level of confidence or certainty in AI model predictions and a scalable, open source deep learning environment (CANDLE).

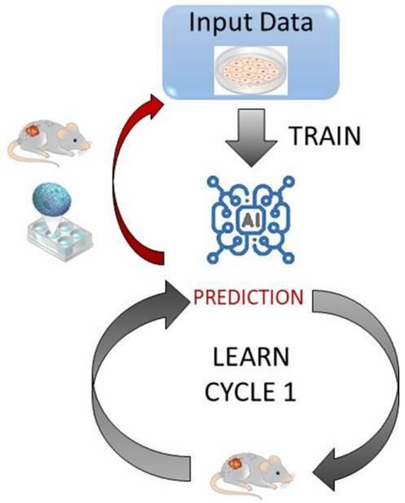


To build upon the nascent predictive oncology community driven by JDACS4C, both agencies have recognized the opportunity to establish and grow an *Envisioning Computational Innovations for Cancer Challenges (ECICC)* community focused on developing new cancer challenge areas that push the limits of current cancer research computational practices and compel development of innovative computational technologies.

Joint Design of Advanced Computing Solutions for Cancer (JDACS4C)

Pilot 1: Predictive Modeling for Pre-clinical Screening

Opportunity: Development of ML based predictive models of drug response to improve pre-clinical drug screening and development of new targeted therapies



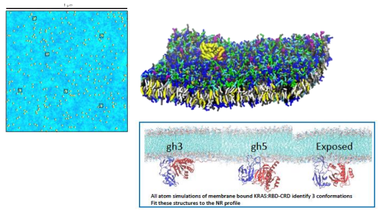
Impact:

- Demonstration of transfer learning: Phase 1 PDX drug study designed and underway based on initial predictions from ML models trained on cell line data
- Standards for data collection & annotation across PDXnet

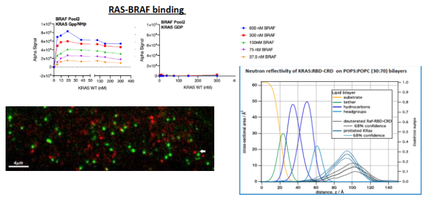
Pilot 2: RAS Biology on Membranes

Opportunity: Development of large-scale ML-enabled computer simulations to extend understanding of RAS-membrane interactions supporting new therapeutic opportunities targeting RAS

SIMULATION



BIOPHYSICAL CHARACTERIZATION

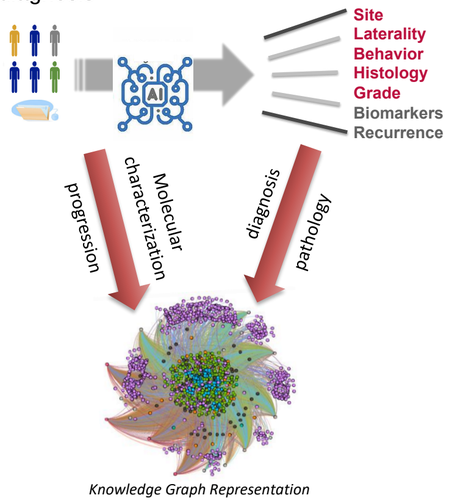


Impact:

Close coupling of experimentation and computation is delivering more complete picture of RAS biology – capturing behavior that cannot be detected experimentally

Pilot 3: Population Information Integration, Analysis and Modeling for Precision Cancer Surveillance

Opportunity: NLP deep learning models to enable real time incidence reporting and clinical trial eligibility assessment for population cancer surveillance supporting informed decision making at time of diagnosis

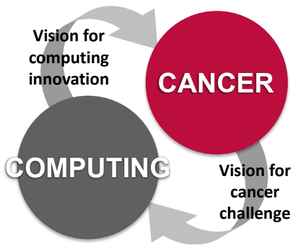


Knowledge Graph Representation

Impact: Natural language processing DL model able to automatically abstract 5 data elements correctly from 43% of path reports across 12 registries (~80% of reports have 4 of 5 elements correctly abstracted)

Computational Oncology Community Building

Opportunity: Recognition that a team science driven, multi-disciplinary approach can stimulate moving from descriptive analytics to data-driven predictive analytics. There is an exciting opportunity to support the establishment and growth of an emerging computational oncology community identify those research areas that, when focused on by large, motivated, collaborative teams, will move this developing field forward.



Impact: Envisioning Computational Innovations for Cancer Challenges (ECICC) community – a broad group of cancer and computational scientists at various career stages from multiple organizations. Several interactive, collaborative in-person and virtual events have been held to develop new challenge areas at the intersection of cancer and computing.

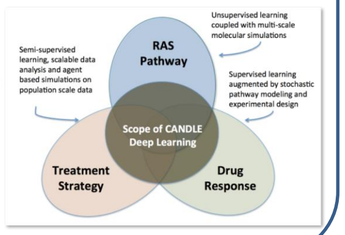
Cross-cutting Capabilities

CANDLE

Open source, scalable, collaboratively developed software platform for providing deep learning methodologies

Impact:

- Deployed on NIH Biowulf
- 150+ NIH staff trained in deep learning at workshops
- Code/benchmarks for deep learning publicly available
- In use by NCI intramural investigators



Uncertainty Quantification

General methods to improve confidence/level of certainty of results from predictive computational models developed by JDACS4C pilots

Impact:

- Ability to make ML predictions actionable by giving associated measure of certainty
- Techniques for reducing noise of large, heterogeneous, real-world data
- Prioritization of experimental resources and increased generalizability of computational predictive models (transfer learning)

JDASC4C Scientific Leadership

Pilot 1: James Doroshow (NCI), Yvonne Evrard (FNLCR), Rick Stevens (ANL)
Pilot 2: Dwight Nissley (FNLCR), Fred Streitz (LLNL)
Pilot 3: Lynne Penberthy (NCI), Gina Tourassi (ORNL)
UQ: Tanmoy Bhattacharya (LANL)
CANDLE: Rick Stevens (ANL), et al

Acknowledgments

This work has been supported in part by the Joint Design of Advanced Computing Solutions for Cancer (JDACS4C) program established by the U.S. Department of Energy (DOE) and the National Cancer Institute (NCI) of the National Institutes of Health. This work was performed under the auspices of the U.S. Department of Energy by Argonne National Laboratory under Contract DE-AC02-06-CH11357, Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, Los Alamos National Laboratory under Contract DE-AC5206NA25396, and Oak Ridge National Laboratory under Contract DE-AC05-00OR22725. LLNL-POST-793938.