Crowdsourcing Advances Reproducibility Standards and Biomarker Development

Holly Stephens[1], Elaine Johanson[2], Emily Boja[2], Elaine Thompson[3*], Yuriy Gusev[4], Krithika Bhuvaneshwar[4], Subha Madhavan[4], Jonathon Keeney[5], Hadley King[5], Janisha Patel[5], Raja Mazumder[5], Sean Watford[1], Ezekiel Maier[1]

[1] Booz Allen, [2] FDA Office of Health Informatics (OHI), [3] FDA Center for Biologics Evaluation and Research (CBER), * Current: Henry M. Jackson Foundation, [4] Georgetown University, [5] George Washington University

BioCompute Object App-a-thon - *Advancing Standards for Computational Pipeline Reproducibility*

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Why Should I Care?

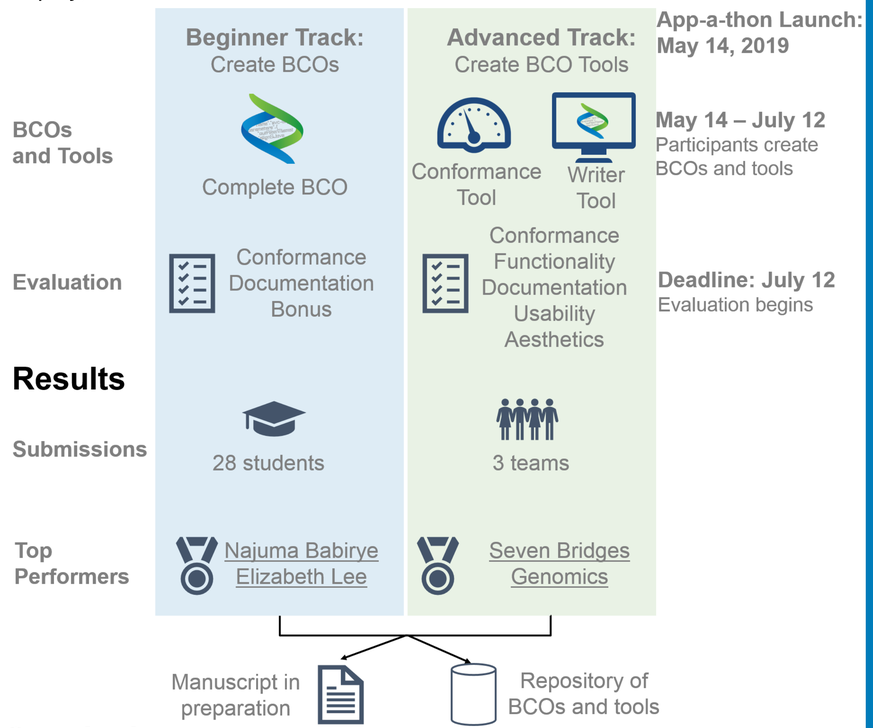
Nonstandard reporting of bioinformatic analysis limits reproducibility, hindering scientific and regulatory progress. The BioCompute specification is an IEEE approved standard for documenting computational pipelines. For more information see BioComputeObject.org

BioCompute Objects

- ✓ JSON-based text documentation
- ✓ Stores parameters, versions, steps…
- ✓ Domains group similar information
- ⚠ Not independently runnable

Methods

App-a-thon participants created BCOs or developed software that can create, display, and validate BCO conformance.



Conclusions

- **Novice bioinformatics students were an ideal cohort** because of their lack of familiarity with BioCompute Objects, broad diversity of research interests, and motivation to submit high quality work.
- Most beginner track scores were high, indicating that most users did not struggle with the specification, **demonstrating the approachability of the BCO specification for bioinformatics novices**

precisionFDA

Precision.FDA.gov

PrecisionFDA offers a high-performance computing platform, access to a community of experts, a library of tools and applications, a competition framework, and virtual lab workspaces to allow FDA scientists and reviewers to collaborate with external partners.

COMING SOON - precisionFDA is adding capabilities to enable data scientists to easily use Shiny R apps, Python, and Jupyter Notebooks.

More than a dozen community challenges and app-a-thons have been run on precisionFDA.

Insights gleaned from the BioCompute Object app-a-thon, such as the explicit demarcation of required fields, were incorporated into BioCompute specification version 1.4.

Multi-omic biomarkers of brain tumor prognosis were identified by the participants of the Brain Cancer Predictive Modeling and Biomarker Discovery Challenge.

precisionFDA
Precision.FDA.gov

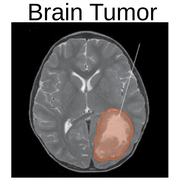
Brain Cancer Predictive Modeling and Biomarker Discovery Challenge - *Discovering Biomarkers Predictive of Brain Caner Patient Survival*

Georgetown | Lombardi COMPREHENSIVE CANCER CENTER

GEORGETOWN icbi THE INNOVATION CENTER FOR BIOMEDICAL INFORMATICS

Why Should I Care?

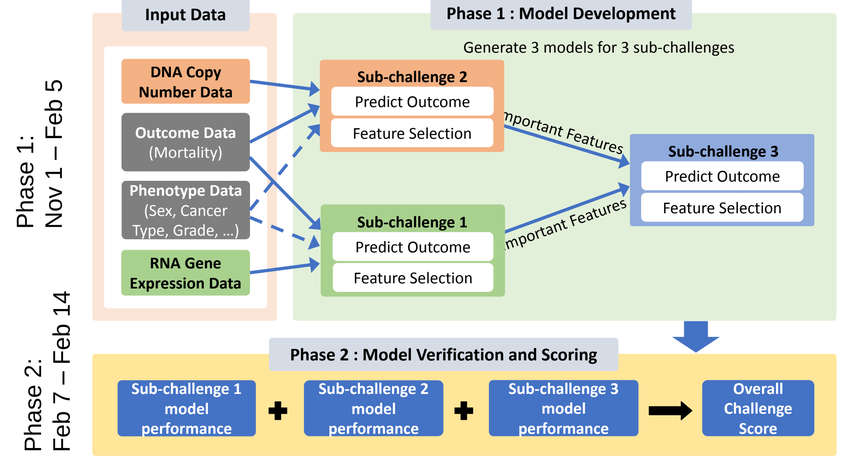
- ~80K primary brain tumor diagnoses occur each year
- There is a high cost of care and limited treatment options
- Large multi-omics datasets, such as REMBRANDT and TCGA, enable the identification of novel biomarkers



Credit : NCI Connect

Methods

Participants develop AI/ML models to identify biomarkers and predict outcomes



Results

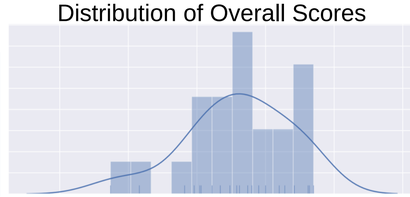
Submissions

Phase 1: 30 → Phase 2: 22

Common Models

Gradient Boosting
Ensemble
Support Vector Machine
Logistic Regression

Distribution of Overall Scores



The top performing model, submitted by the Sentieon team, used 46 features including 40 genes, 4 cytobands and 2 clinical attributes.

Conclusions

- Sub-challenge 2 may be more challenging than sub-challenge 1 because DNA copy number was more complex compared to gene expression data.
- ***TGF-beta regulation of extracellular matrix*** **pathway was enriched** among the 40 genes selected by the top performing model. TGF-β signaling is dysregulated and promotes tumor progression in cancer.