Advancing the Regulatory Science of Omics via Crowdsourcing on the precisionFDA Platform

Elaine Johanson, BS[1] , Ruth Bandler, MS[1], Zivana Tezak, PhD[2], Heike Sichtig, PhD[2], Yi Yan, PhD[2], Errol Strain, PhD[3], John Didion, PhD[4], Ezekiel Maier, PhD[5]
[1] FDA Office of Health Informatics (OHI), [2] FDA Center for Devices and Radiological Health (CDRH), [3] FDA Center for Veterinary Medicine (CVM), [4] DNAnexus, [5] Booz Allen

Background

The rapid advancement of next-generation sequencing (NGS) technology has fueled innovation of omics-based diagnostic tests for guiding more individualized care – known as precision medicine. Ensuring the safety and efficacy of precision medicine treatments includes the understanding of the strengths and limitations of omics-based diagnostic tests.

The precisionFDA platform, a secure cloud-based collaborative environment, is using crowdsourcing as one way to advance the regulatory science of omics tests. Six community challenges completed on precisionFDA have generated a total of 425 responses from 176 participants. These challenges can help enable the evaluation of novel algorithms, and the development of best practices.

PrecisionFDA

Vision: Platform for regulatory science exploration of the tools and techniques for analyses of large biological datasets, while encouraging collaboration, interaction, and data sharing.

Community Members Can:

Benchmark Bioinformatic Pipelines	Build Workflows and Run Pipelines	Partner with FDA to Conduct Challenges	Contribute to a Library of Resources	Sponsor App-a-Thons	Engage Globally with Experts

FDA Scientists Can:

Collaborate with External Stakeholders

4,000 users
50 Terabytes of data
3,800 publicly available apps

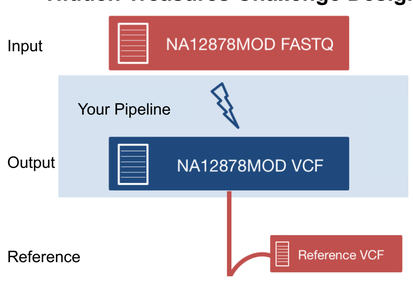
Future:

Conduct the Interactive Portion of a Regulatory Review

Consistency, Truth, and Hidden Treasures Challenges

Motivation: Encourage innovation, benchmark, and understand the strengths and limitations of bioinformatics pipelines for identifying genetic variants.

Hidden Treasures Challenge Design



Engagement:

- Consistency: 21 submissions, 17 participants
- Truth: 35 submissions, 25 participants
- Hidden Treasures: 86 submissions, 30 participants

Results:

1. Development, piloting and publication of best practices for benchmarking human germline small-variants
2. Benchmarking of a novel deep neural network-based variant caller
3. Evaluation of an optimized version of a best practice variant calling pipeline

Conclusions: (1) accuracy of variant calling pipelines differs depending on the type and size of genetic variants being assessed, (2) in silico injected variants can be useful to assess accuracy of unknown variant calling.

precisionFDA
Precision.FDA.gov

Six community challenges have been completed, which have generated 425 submissions from 176 participants.

Public challenges encourage innovation and enable independent benchmarking of algorithms. Cross-agency collaboration amplifies the benefits of crowdsourcing.

The current BioCompute Object App-a-thon is focused on improving the reproducibility of computational pipelines.

Future challenge topics include somatic variant calling, tumor mutational burden calculation, and brain tumor biomarkers.

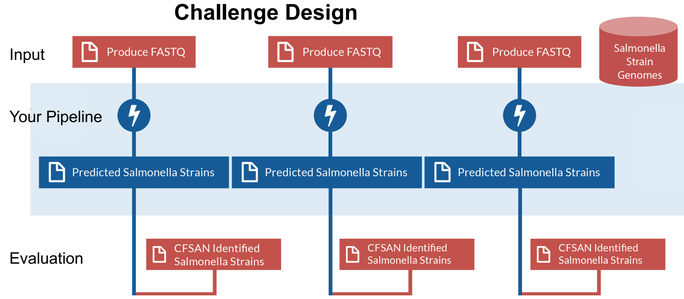
The precisionFDA vision has expanded to include multi-omics and artificial intelligence related to precision medicine.

precisionFDA
Precision.FDA.gov

CFSAN Pathogen Detection Challenge

Motivation: Reduce illness and economic loss from foodborne pathogens by innovating and benchmarking tools for pathogen detection in quasi-metagenomics samples.

Challenge Design



Engagement: Received 26 submissions from 11 participants.

Results:

1. 5 submitters identified serotypes in 9 of 13 *Salmonella* samples
2. 2 submitters identified MLST types in 9 of 13 *Salmonella* samples

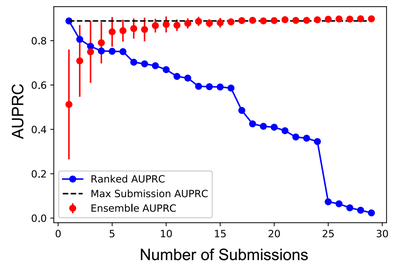
Conclusions: (1) Accuracy was higher on synthetic samples, and (2) taxonomic accuracy differences were larger in culture positive samples.

CDRH ID-NGS Diagnostics Biothreat Challenge

Motivation: Advance early microorganism detection during pathogen outbreaks by evaluating algorithms for identifying and quantifying emerging pathogens from their genomic fingerprints.

Challenge Design

Result



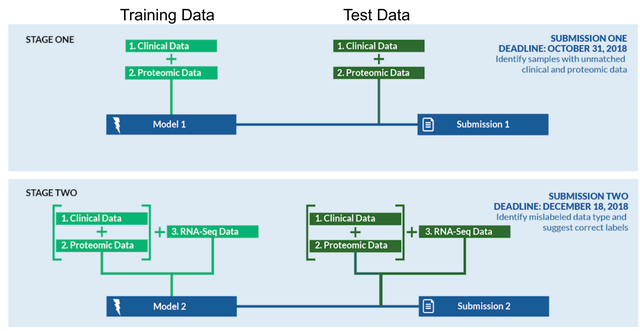
Engagement: Received 29 submissions from 11 participants.

Conclusion: In silico evaluations may be a poor proxy for mock clinical evaluations.

NCI-CPTAC Sample Mislabeling Correction Challenge

Motivation: Develop and benchmark algorithms for detecting and correcting sample mislabeling (accidental swapping of patient samples), a problem which contributes to irreproducible results.

Challenge Design



Engagement: Received 230 submissions from 80 participants.

Next Steps:

1. Draft and submit a challenge results scientific manuscript
2. Create a software package to distribute the top performing methods

The views expressed in this poster are those of the authors and do not necessarily reflect the official policy or position of the FDA, HHS, or U.S. Government