BioCompute: A Standardized Method to Communicate Bioinformatic Workflow Information and Ease Organizational Burden

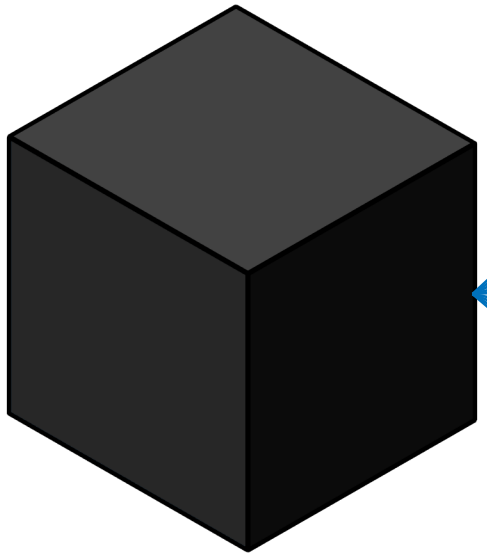
Jonathon Keeney, Ph.D.
keeneyjg@gwu.edu
Assistant Research Professor, George Washington University
Managing Director, BioCompute Executive Steering Committee

Computational Toxicology Communities of Practice
July 25th, 2019

NGS Data Flows

BioCompute Objects

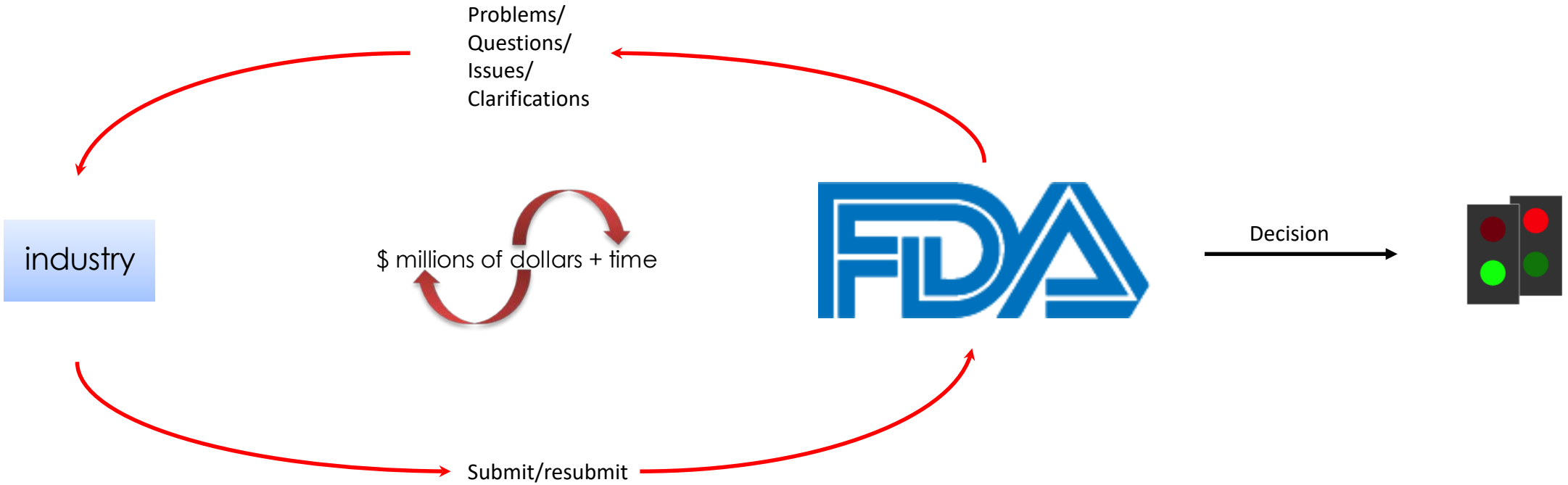

```
$ fastq-dump -X 2 SRR001666 --split-3
$ fastq-dump -X 2 SRR001666 --split-3
$ fastq-dump -X 2 SRR001666 --split-3
$ fastq-dump -X 2 SRR001666 --split-3
Read 2 spots for SRR001666
Written 2 spots for SRR001666
$ head SRR001666_1.fastq SRR001666_2.fastq
==> SRR001666_1.fastq <==
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
IIIIIIIIIIIIIIIIIIIIIIIIIIIIII9IG9IC
@SRR001666.2 071112_SLXA-EAS1_s_7:5:1:801:338 length=36
GTTCAGGGATACGACGTTTGTATTTTAAGAATCTGA
+SRR001666.2 071112_SLXA-EAS1_s_7:5:1:801:338 length=36
IIIIIIIIIIIIIIIIIIIIIIIIIIIIII6IBI

==> SRR001666_2.fastq <==
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
AAGTTACCCTTAACAACTTAAGGGTTTTCAAATAGA
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
IIIIIIIIIIIIIIIIIIIIDIIIIIII>IIIIII/
@SRR001666.2 071112_SLXA-EAS1_s_7:5:1:801:338 length=36
AGCAGAAGTCGATGATAATACGCGTCGTTTTATCAT
+SRR001666.2 071112_SLXA-EAS1_s_7:5:1:801:338 length=36
IIIIIIIIIIIIIIIIIIIIIIGII>IIIII-I)8I
```



- Ancestry
- Cancer
- Microbiome
- Disease correlation
- Agriculture
- Synthetic biology
- Livestock
- Metagenomics
- Personalized medicine

Wasted Time and Money

BioCompute Objects



This is not a Guidance Document

DRAFT: Please provide comments and suggestions

Submitting Next Generation Sequencing Data to the Division of Antiviral Products
Experimental Design and Data Submission

Acceptable Next Generation Sequencing Platforms

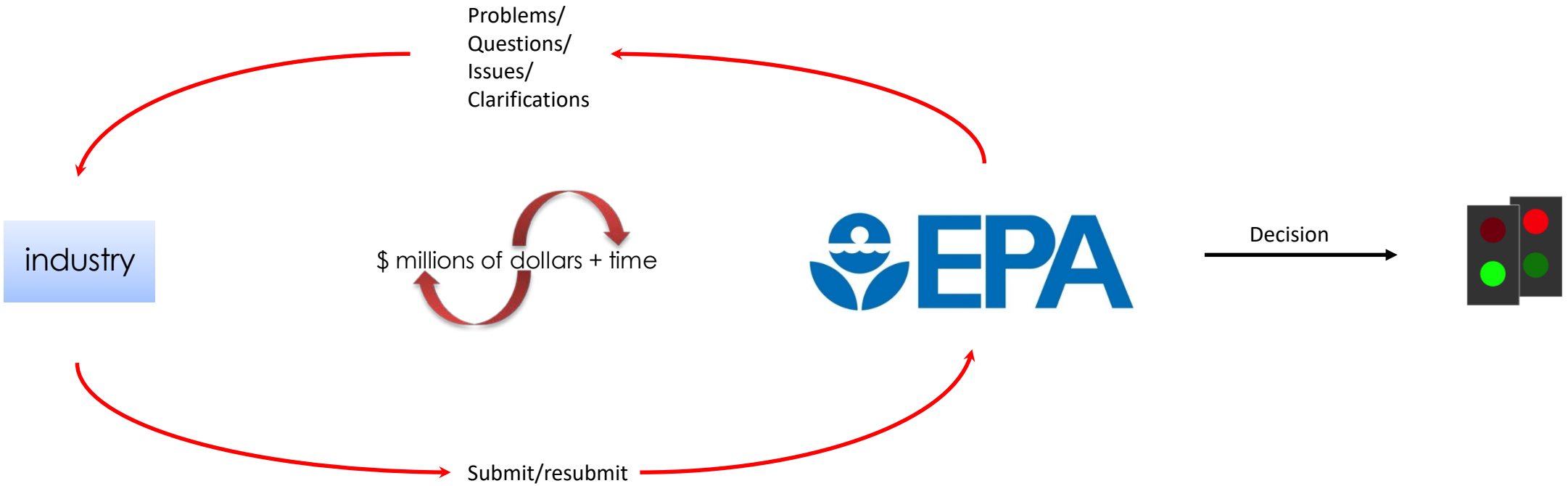
The division will accept Next Generation sequencing data generated from most standard Next Generation Sequencing (NGS) platforms provided the sponsor supplies the appropriate details for the sequencing platform, the protocols to be used for sample preparation, the raw NGS data, and the methods used to analyze the data. We recommend communicating with the division early in the process and providing these details prior to submitting the sequencing data. Please consider the following information when preparing your NGS submissions.

Data Transfer

1. **Portable hard drive**
    a. The raw NGS data in the fastq format should be sent to the division on a secured, portable hard drive following the guidelines outlined in this Guidance: http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM163567.pdf
    b. Please note that only the raw NGS data, the frequency table, and a table of contents should be contained on the hard drive. Additional files, such as those with a .exe extension may result in rejection of the submission. In addition, if the hard drive is password protected (not required or recommended at this time), please consult with the division ahead of time to ensure that the password is provided to the appropriate personnel in the document room.
    c. All additional data should be submitted via the electronic document gateway

Wasted Time and Money

BioCompute Objects



NATIONAL ARCHIVES

FEDERAL REGISTER

The Daily Journal of the United States Government

Notice

Use of High Throughput Assays and Computational Tools; Endocrine Disruptor Screening Program; Notice of Availability and Opportunity for Comment

A Notice by the Environmental Protection Agency on 06/19/2015

PUBLISHED DOCUMENT

AGENCY:

Environmental Protection Agency (EPA).

ACTION:

Notice.

SUMMARY:

This document describes how EPA is planning to incorporate an alternative scientific approach to screen chemicals for their ability to interact with the endocrine system. This will improve the Agency's ability to fulfill its statutory

DOCUMENT DETAILS

Printed version:
PDF

Publication Date:
06/19/2015

Agency:
Environmental Protection Agency

Dates:
Comments must be received on or before August 18, 2015.

Comments Close:
08/18/2015

Document Type:

Alternative Test Methods and Strategies to Reduce Vertebrate Animal Testing

The Toxic Substances Control Act (TSCA), as amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act, directs EPA to:

- reduce and replace, to the extent practicable and scientifically justified, the use of vertebrate animals in the testing of chemical substances or mixtures; and
- promote the development and timely incorporation of alternative test methods or strategies that do not require new vertebrate animal testing.

TSCA also requires EPA to develop a strategic plan on this topic and provide a progress report on the implementation of the plan to Congress every five years since the date of the enactment of the Lautenberg Chemical Safety Act, i.e. beginning in 2021.

In 2018, EPA published its Strategic Plan to Promote the Development and Implementation of Alternative Test Methods within the TSCA Program. The Strategic Plan incorporated input from two public meetings and written comments submitted on the draft strategic plan.

A solution should…

BioCompute Objects

- Be human readable: like a GenBank sequence record
- Be computer readable: structured information with predefined fields and associated meanings of values
- Contain enough information to understand the computational pipelines, interpret information, maintain records, and reproduce experiments
- Have a way to be sure the information has not been altered: immutable

Solution: BioCompute

BioCompute Objects

- Standard for communicating computational analysis workflows
- Acts like an envelope for entire pipeline
  - Can incorporate other standards (e.g. CWL, FHIR Genomics)
- Built with FDA
- Human and machine readable
  - Written in JSON
- Categorized by domains
- Adheres to F.A.I.R. principles
- Adaptable
- Preserves data provenance
- Unique IDs for versioning

802.11 Analogy

BioCompute Objects

BioCompute Objects

```
{
    "id": "obj.1277",
    "name": "Human[tax:9606] CAG repeat [SO:0000726] detection for Huntington's disease [DOID:12858]",
    "version": "1.0",
    "createdby": "hadley_king@gwmail.gwu.edu",
    "digital_signature": "QdB7vu4rLeLtMFmPHQ9S",
    "verification_status": "unreviewed",
    "publication_status": "draft",
    "usability_domain": ["Determine CAG repeats [SO:0000726] in a sequence for the diagnosis of
    Huntington's disease [DOID:12858] for HTT gene [HGNC:4851]: 10 - 35 CAG trinucleotide repeats: No
    Risk, 36 - 40 CAG trinucleotide repeats: Moderate Risk, CAG trinucleotide repeats: High Risk...],
    40 "authors": [{"orcid": "0000-0002-9920-565X}],
    "description_domain": {
        "keywords": ["disease", "human disease"...],
        "xref": ["SO:0000726", "tax:9606", "DOID:12858", "ICD:G10"...],
        "pipeline_steps": {"HIVE_hexagon":{...}, "HIVE_heptagon":{...}}},
    "execution_domain": {
        "pipeline_version": "1.0",
        "platform": "HIVE",
        "driver": "hive://hive-driver",
        "script": "hive://workflows/huntingtons_disease_detection.hwf",
        "prerequisites": [
            {"name":"HIVE_hexagon","version":"1.3"},
            {"name":"HIVE_heptagon","version":"1.3}],
        "env_parameters": ["HIVEv1.3]},
    "parametric_domain": {
        "divergence_threshold_percent": "10",
        "alignment_seed": "14 letters",
        "alignment_min_match_len": "50",
        "repeat_calling_min_coverage": "20",
        "repeat_calling_frequency_cutoff": "0.2},
    "io_domain": {
        "reference_uri_list": [
            "http://www.ncbi.nlm.nih.gov/nuccore/SRR4243395",
            "http://www.ncbi.nlm.nih.gov/nuccore/NG_009378.1],
        "input_uri_list": ["hive://nuc-read/514683", "hive://nuc-read/514682],
        "output_uri_list": ["hive://data/509234/repeat-profile.csv]},
    "error_domain": ["false negative discovery < 0.000001]
}
```

Metadata
(not needed for computation)

parametrized pipeline with all internal arguments well defined

workflow – defined in HIVE

parametric domain

Verification Kit

input domain

output domain

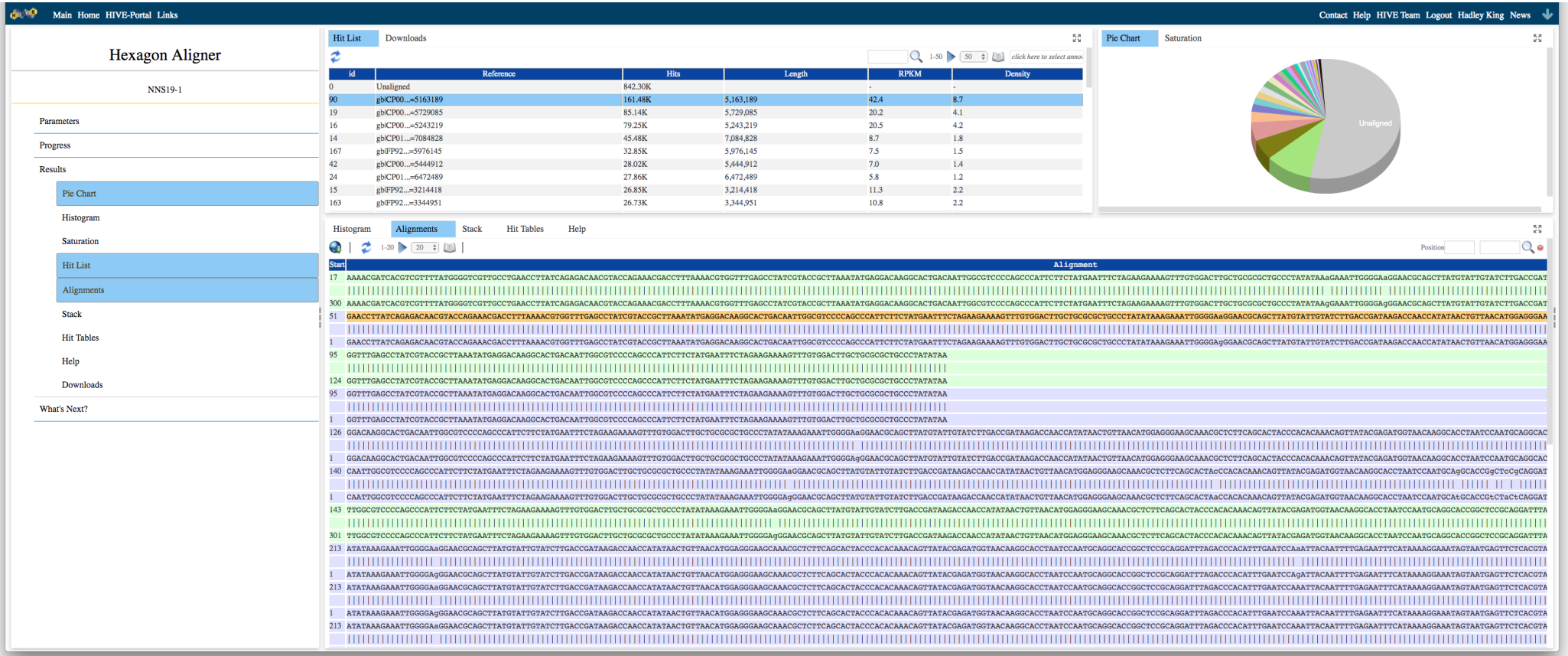
BCO Example: HCV-1 drug resistance

BioCompute Objects

- Goal: Identify SNPs, insertions, and deletions that correlate with reduced ledipasvir antiviral drug efficacy in Hepatitis C virus subtype 1
  - Genome sequencing data from a drug resistant cohort

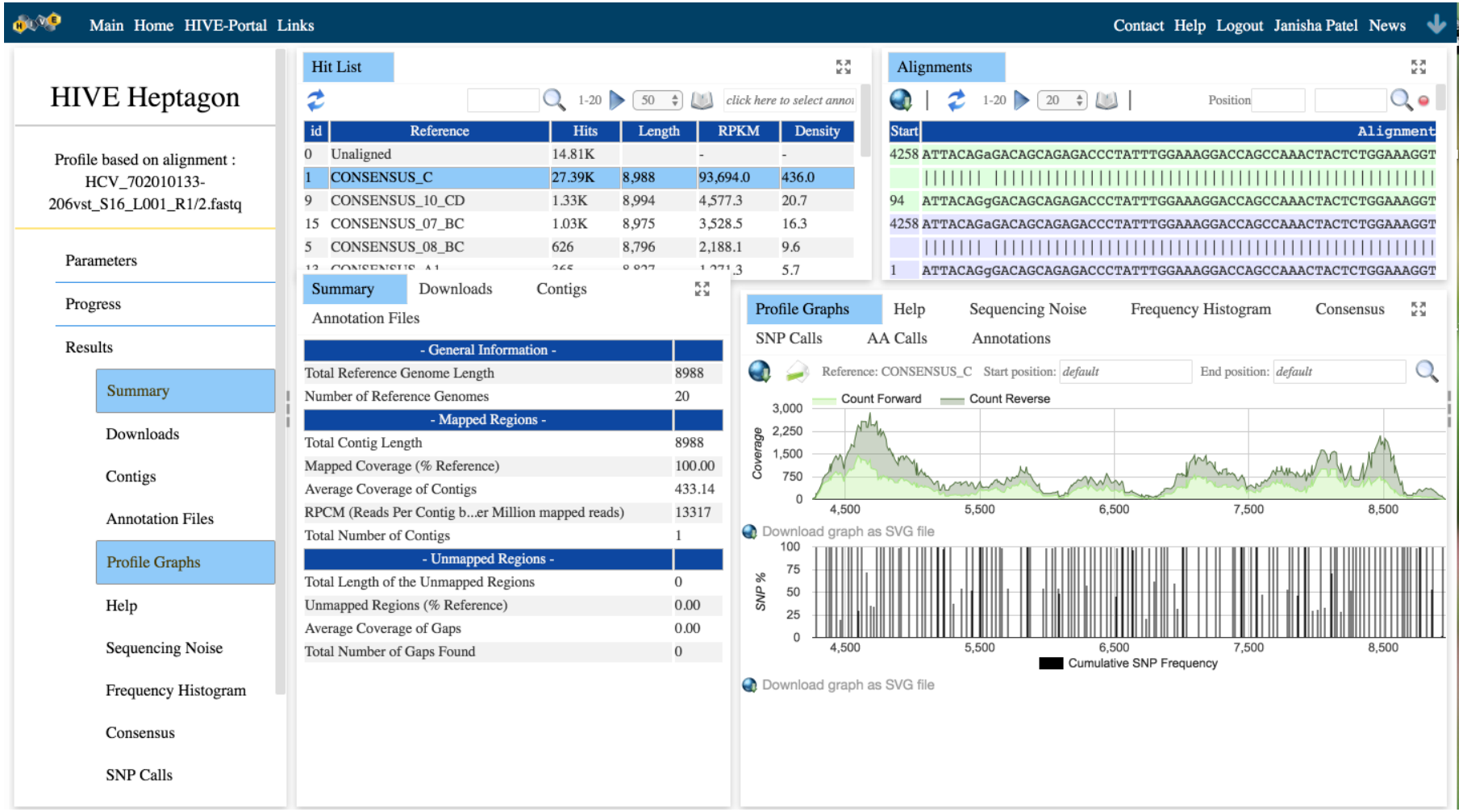
BCO Example: HCV-1 drug resistance

BioCompute Objects



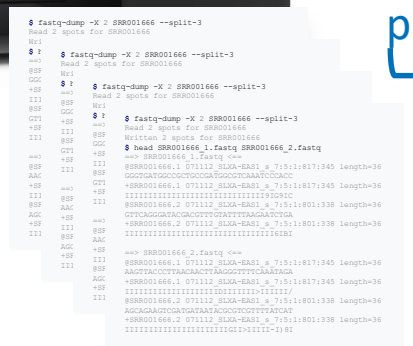
BCO Example: HCV-1 drug resistance

BioCompute Objects

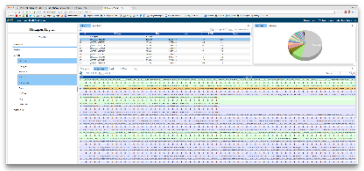


BCO Example: HCV-1 drug resistance

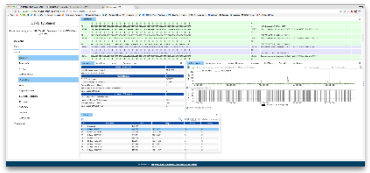
BioCompute Objects



Parameters
Scripts/drivers/
prerequesites



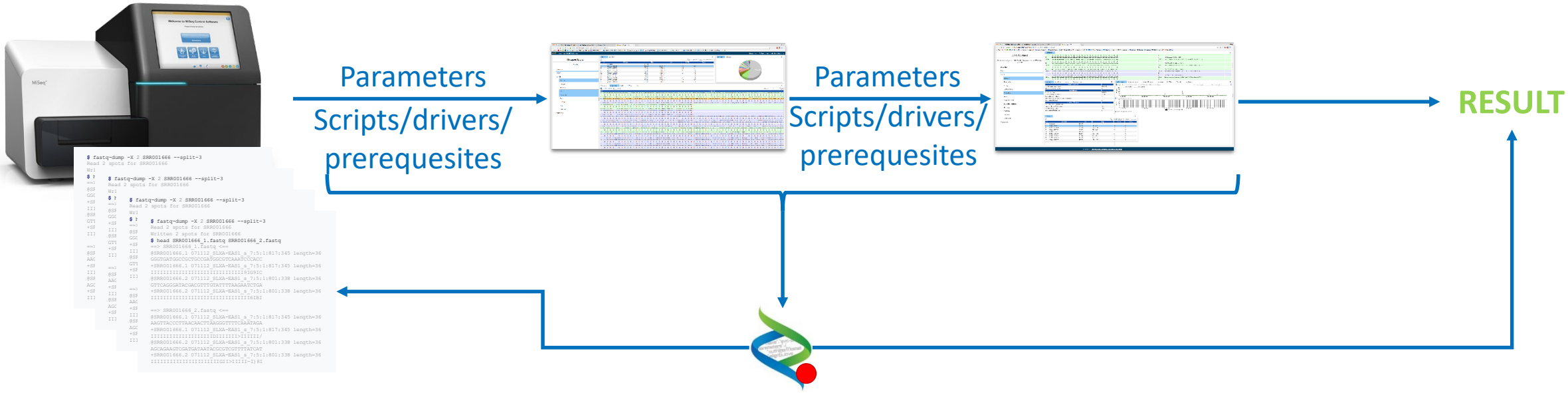
Parameters
Scripts/drivers/
prerequesites



RESULT

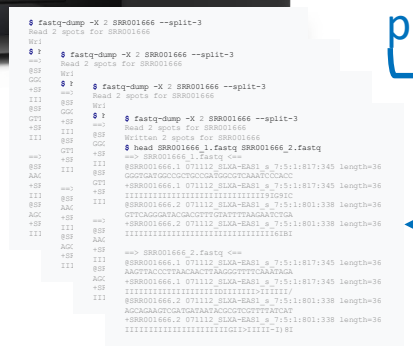
BCO Example: HCV-1 drug resistance

BioCompute Objects

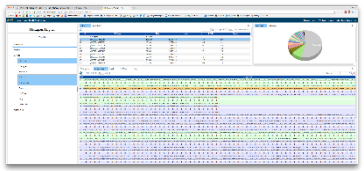


BCO Example: HCV-1 drug resistance

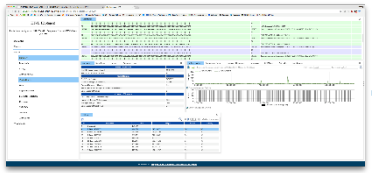
BioCompute Objects



Parameters
Scripts/drivers/
prerequesites



Parameters
Scripts/drivers/
prerequesites

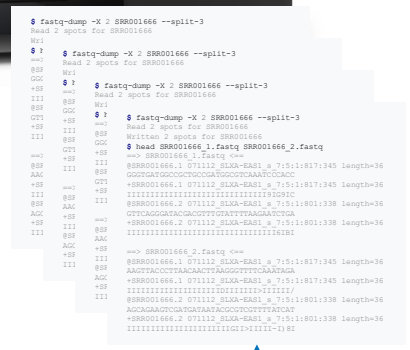


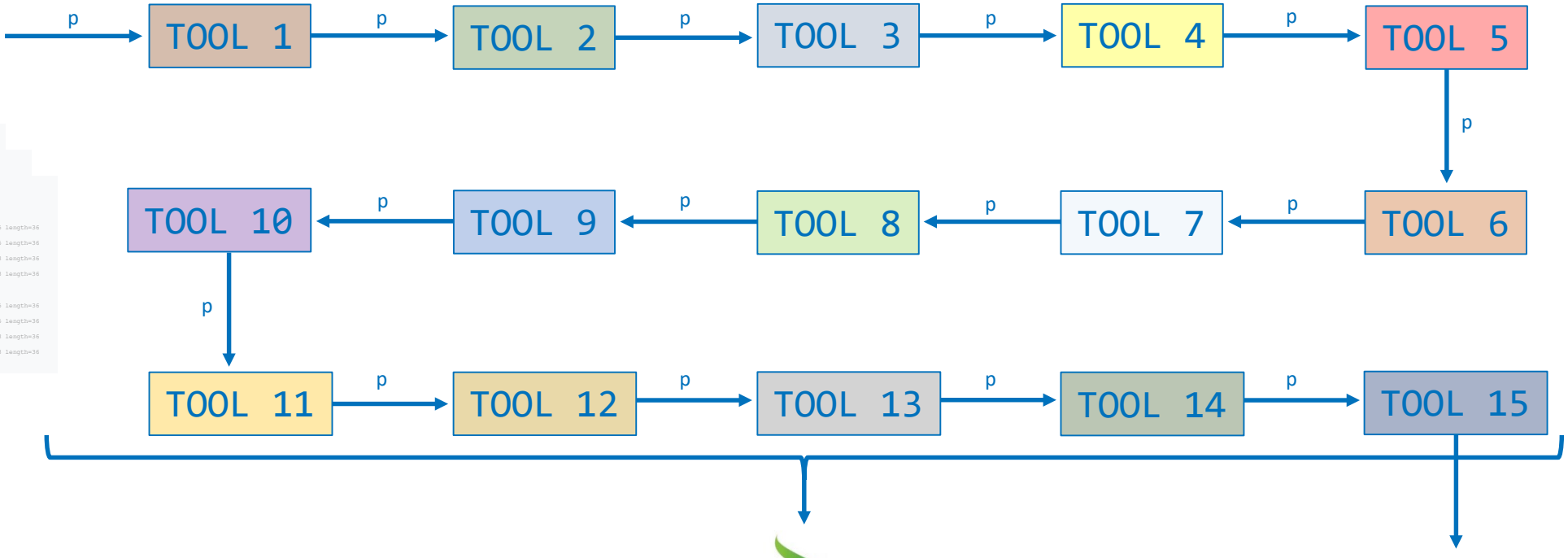
RESULT


```
   "usability_domain": [
       "Identify baseline single nucleotide polymorphisms SNPs [SO:0000694], insertions [SO:0000667], and deletions [SO:0000045] that correlate with reduced
ledipasvir [PubChem:67505836] antiviral drug efficacy in Hepatitis C virus subtype 1 [taxID:31646]",
       "Identify treatment emergent amino acid substitutions [SO:0000048] that correlate with antiviral drug treatment failure",
       "Determine whether the treatment emergent amino acid substitutions [SO:0000048] identified correlate with treatment failure involving other drugs
against the same virus",
       "GitHub CWL example: https://github.com/mr-c/hive-cwl-examples/blob/master/workflow/hive-viral-mutation-detection.cwl#L20"
   ],
   "extension_domain":{
       "github_extension": {
           "github_URL": "https://github.com/common-workflow-language/hive-cwl-examples",
           "gitHub_URI": {
```

BCO Example: HCV-1 drug resistance

BioCompute Objects






```
  "usability_domain": [
      "Identify baseline single nucleotide polymorphisms SNPs [SO:0000694], insertions [SO:0000667], and deletions [SO:0000045] that correlate with reduced
ledipasvir [PubChem:67505836] antiviral drug efficacy in Hepatitis C virus subtype 1 [taxID:31646]",
      "Identify treatment emergent amino acid substitutions [SO:0000048] that correlate with antiviral drug treatment failure",
      "Determine whether the treatment emergent amino acid substitutions [SO:0000048] identified correlate with treatment failure involving other drugs
against the same virus",
      "GitHub CWL example: https://github.com/mr-c/hive-cwl-examples/blob/master/workflow/hive-viral-mutation-detection.cwl#L20"
  ],
  "extension_domain":{
      "github_extension": {
          "github_URL": "https://github.com/common-workflow-language/hive-cwl-examples",
          "gitHub_URI": {
```

```
{
   "object_id": 1270,
   "type": "antiviral_resistance_detection",
   "provenance_domain": {
       "name": "HCV1a ledipasvir resistance SNP detection",
       "structured_name": "HCV1a [taxID:31646] ledipasvir
[PubChem:67505836] resistance SNP [SO:0000694] detection",
       "version": "1.1",
       "digital_signature": "905d7fce3f3ac64c8ea86f058ca71658",
       "review": [
           {
               "status": "approved",
               "comment": "Approved by GW staff. Waiting for approval from
FDA Reviewer",
               "reviewer": {"type": "orcid", "orcid":
"http://orcid.org/0000-0003-1409-4549" }
           },
           {
               "status": "approved",
               "comment": "The revised BCO looks fine",
               "reviewer": {
                   "type": "entity",
                   "entity": {
                       "name": "Eric Donaldson",
                       "affiliation": "FDA",
                       "email": "Eric.Donaldson@fda.hhs.gov"
                   }
               }
           }
       ],
       "derived_from" : null,
       "obsolete" : null,
       "embargo" : {
           "start_time": null,
           "end_time": null
       },
       "created": "2017-01-24T09:40:17-0500",
       "modified": "2018-03-21T18:31:48-0400",
       "contributors": [
           {
               "type": "entity",
               "entity": {
                   "name": "Eric Donaldson",
                   "affiliation": "FDA",
                   "email": "Eric.Donaldson@fda.hhs.gov",
                   "contribution": "authoredBy"
               }
           },
           {
               "type": "orcid",
               "orcid": "http://orcid.org/0000-0003-1409-4549",
               "contribution": "createdBy"
           }
       ],
      "license":{
          "address": "https://creativecommons.org/licenses/by/4.0/",
          "access_time": "2017-01-24T09:40:17-0500"
      }
   },
   "usability_domain": [
       "Identify baseline single nucleotide polymorphisms SNPs [SO:0000694], insertions
[SO:0000667], and deletions [SO:0000045] that correlate with reduced ledipasvir
[PubChem:67505836] antiviral drug efficacy in Hepatitis C virus subtype 1 [taxID:31646]",
       "Identify treatment emergent amino acid substitutions [SO:0000048] that correlate
with antiviral drug treatment failure",
       "Determine whether the treatment emergent amino acid substitutions [SO:0000048]
identified correlate with treatment failure involving other drugs against the same
virus",
       "GitHub CWL example: https://github.com/mr-c/hive-cwl-
examples/blob/master/workflow/hive-viral-mutation-detection.cwl#L20"
   ],
   "extension_domain":{
       "github_extension": {
          "github_URL": "https://github.com/common-workflow-language/hive-cwl-examples",
          "gitHub_URI": {
            "pipeline_steps": {
               "tool": [
                   {
                       "step_number": "1",
                       "name": "HIVE-hexagon",
                       "description": "Alignment of reads to a set of references",
                       "version": "1.3",
                       "prerequisite": [
                           {
                               "name": "Hepatitis C virus genotype 1",
                               "URI": {
                                   "address":
"http://www.ncbi.nlm.nih.gov/nuccore/22129792",
                                   "access_time": "2017-01-24T09:40:17-0500"
                               }
                           },
                           {
                               "name": "Hepatitis C virus type 1b complete genome",
                               "URI": {
                                   "address":
"http://www.ncbi.nlm.nih.gov/nuccore/5420376",
                                   "access_time": "2017-01-24T09:40:17-0500"
                               }
                           },
                           {
                               "name": "Hepatitis C virus (isolate JFH-1) genomic RNA",
                               "URI": {
                                   "address":
"http://www.ncbi.nlm.nih.gov/nuccore/13122261",
                                   "access_time": "2017-01-24T09:40:17-0500"
                               }
                           },
                           {
                               "name": "Hepatitis C virus clone J8CF, complete genome",
                               "URI": {
                                   "address":
"http://www.ncbi.nlm.nih.gov/nuccore/386646758",
                                   "access_time": "2017-01-24T09:40:17-0500"
                               }
                           },
                           {
                               "name": "Hepatitis C virus S52 polyprotein gene",
                               "URI": {
                                   "address":
"http://www.ncbi.nlm.nih.gov/nuccore/295311559",
                                   "access_time": "2017-01-24T09:40:17-0500"
                               }
                           }
                       ],
                       "input_uri_list": [
                           {
                               "address":
"https://hive.biochemistry.gwu.edu/dna.cgi?cmd=objFile&ids=514683",
                               "access_time": "2017-01-24T09:40:17-0500"
                           },
                           {
                               "address":
"https://hive.biochemistry.gwu.edu/dna.cgi?cmd=objFile&ids=514682",
                               "access_time": "2017-01-24T09:40:17-0500"
                           }
                       ],
                       "output_uri_list": [
                           {
                               "address":
"https://hive.biochemistry.gwu.edudata/514769/allCount-aligned.csv",
                               "access_time": "2017-01-24T09:40:17-0500"
                           }
                       ]
                   },
                   {
                       "step_number": "2",
                       "name": "HIVE-heptagon",
                       "description": "variant calling",
                       "version": "1.3",
                       "prerequisites": null,
                       "input_uri_list": [
                           {
                               "address":
"https://hive.biochemistry.gwu.edudata/514769/dnaAccessionBased.csv",
                               "access_time": "2017-01-24T09:40:17-0500"
                           }
                       ],
                       "output_uri_list": [
                           {
                               "address":
"https://hive.biochemistry.gwu.edudata/514801/SNPProfile.csv",
                               "access_time": "2017-01-24T09:40:17-0500"
                           },
                           {
                               "address":
"https://hive.biochemistry.gwu.edudata/14769/allCount-aligned.csv",
                               "access_time": "2017-01-24T09:40:17-0500"
                           }
                       ]
                   }
               ]
           }
       },
       "execution_domain": {
           "script_type": "text",
           "script": "https://hive.biochemistry.gwu.edu/dna.cgi?cmd=home",
           "script_driver": "manual",
           "pipeline_version": "2.0",
           "platform": "hive",
           "software_prerequisites": [
               {
                   "name": "HIVE-hexagon",
                   "version": "babajanian.1",
                   "uri": {
                       "address":
"https://hive.biochemistry.gwu.edu/dna.cgi?cmd=dna-hexagon&cmdMode=-",
                       "access_time": "2017-01-24T09:40:17-0500"
                   }
               },
               {
                   "name": "HIVE-heptagon",
                   "version": "albinoni.2"
               }
           ],
           "domain_prerequisites": [
               {
                   "name": "access HIVE",
                   "url": "https://hive.biochemistry.gwu.edu/dna.cgi?cmd=login"
               },
               {
                   "name": "access to ftp",
                   "url": "ftp://:22/"
               },
               {
                   "name": "access to e-utils",
                   "url": "http://eutils.ncbi.nlm.nih.gov/entrez/eutils/"
               }
           ],
           "env_parameters": {
               "HOSTTYPE" :"x86_64-linux"
           }
       },
       "parametric_domain": {
           "HIVE-hexagon" : {
               "seed": "14",
               "minimum_match_len": "66",
               "divergence_threshold_percent": "30"
           },
           "HIVE-heptagon": {
               "minimum_coverage": "15",
               "freq_cutoff": "0.10"
           }
       },
       "io_domain": {
           "input_subdomain": {
               "subject": [
                   {
```

BCO applications for computational toxicology

BioCompute Objects

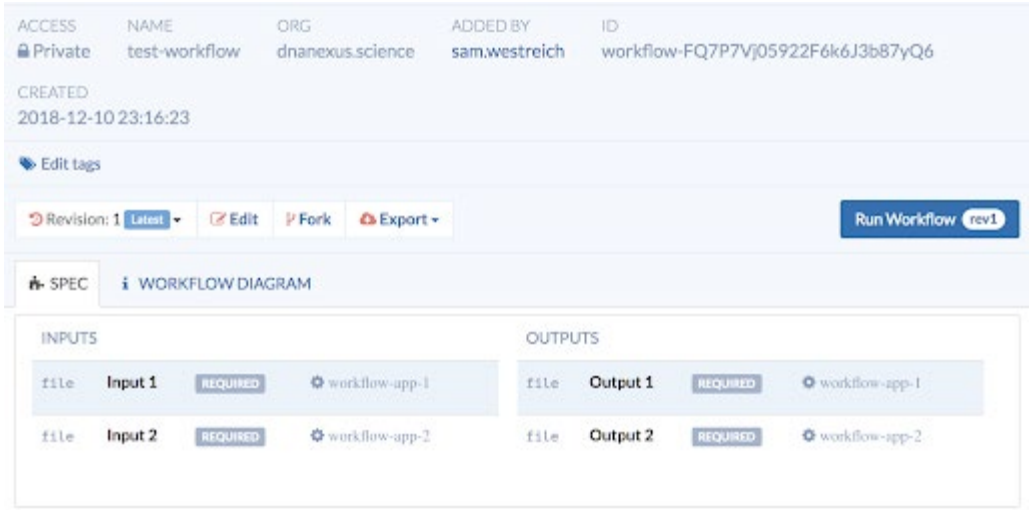
- Create dose-response modeling extensions to capture both BMDS and tcpl pipelines
- For tcpl
  - Each BCO would be the platform-specific steps used to analyze the results

All Chemicals in Assay Endpoint: ACEA_ER_80hr

Excel

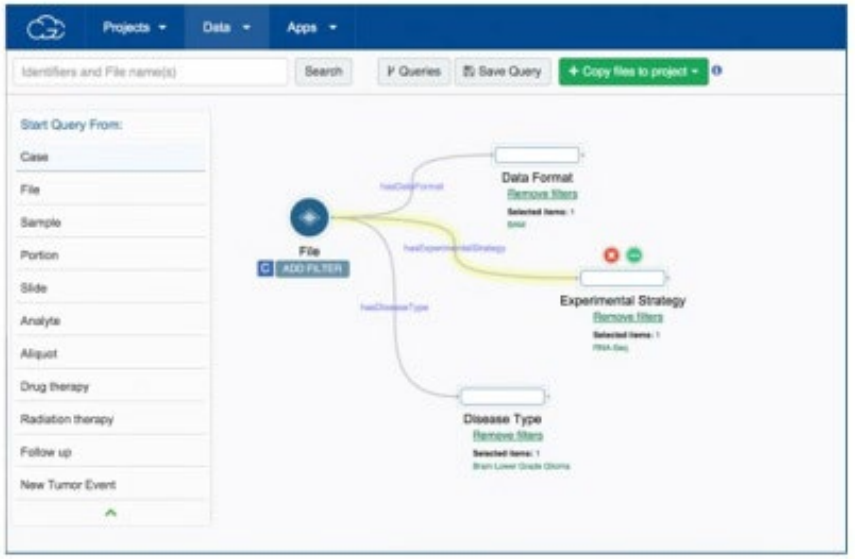
Annotations | Citations | tcpl Processing | Reagents | AOPs

	Assay Run Type	Level Applied	Method Name	Description
1	MULTI	2	none	apply no level 2 method
2	MULTI	3	pval.apid.medpcbyconc.max	plate-wise median response of positive control (max)
3	MULTI	3	resp.pc	response percent activity
4	MULTI	3	bval.apid.nwllslowconc.med	Take the median cval of the n wells and the first two concentrations, by apid
5	MULTI	3	resp.shiftneg.3bmad	Make values below baseline zero.
6	MULTI	4	bmad.aeid.lowconc.twells	bmad based on two lowest concentration of treatment wells
7	MULTI	5	bmad3	Add a cutoff value of 3*bmad.
8	MULTI	5	pc20	Add a cutoff value of 20.
9	MULTI	6	singlept.hit.high	Look for single point hits with activity only at the highest conc tested
10	MULTI	6	singlept.hit.mid	Look for single point hits with activity not at highest conc tested
11	MULTI	6	multipoint.neg	Look for inactives with multiple medians above baseline
12	MULTI	6	noise	Look for noisy curves, relative to the assay
13	MULTI	6	border.hit	Look for actives with borderline activity
14	MULTI	6	border.miss	Look for inactives with borderline activity
15	MULTI	6	modlga.lowconc	AC50 less than lowest concentration tested
16	MULTI	6	gnls.lowconc	Look for low concentration gnls winners
17	MULTI	6	overfit.hit	Flag hit-calls that would get changed after doing the small N correction to the aic values.
18	MULTI	6	efficacy.50	Flag hit-calls with efficacy values less than 50% -- intended for biochemical assays.



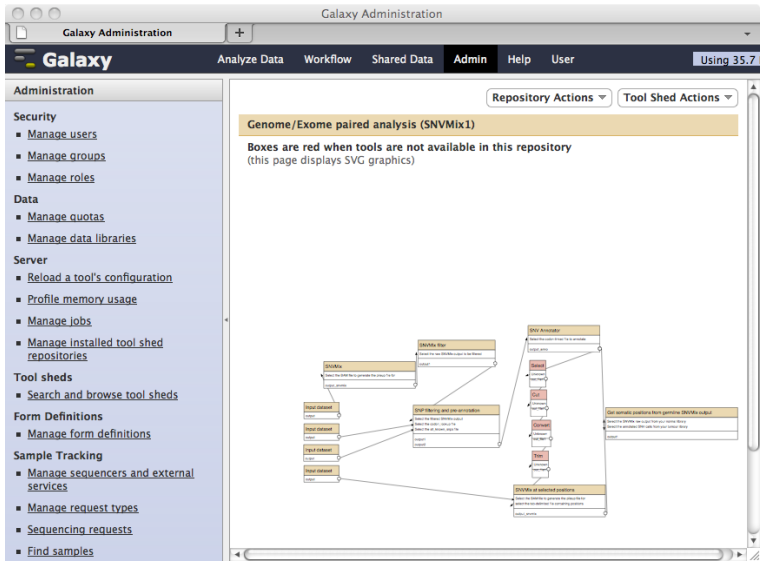
BioCompute Objects

Seven Bridges

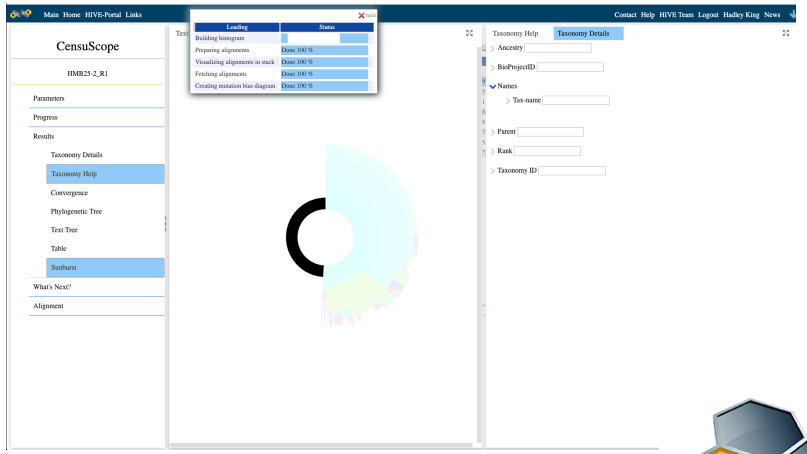


precisionFDA

DNAnexus
powering the genomics revolution™

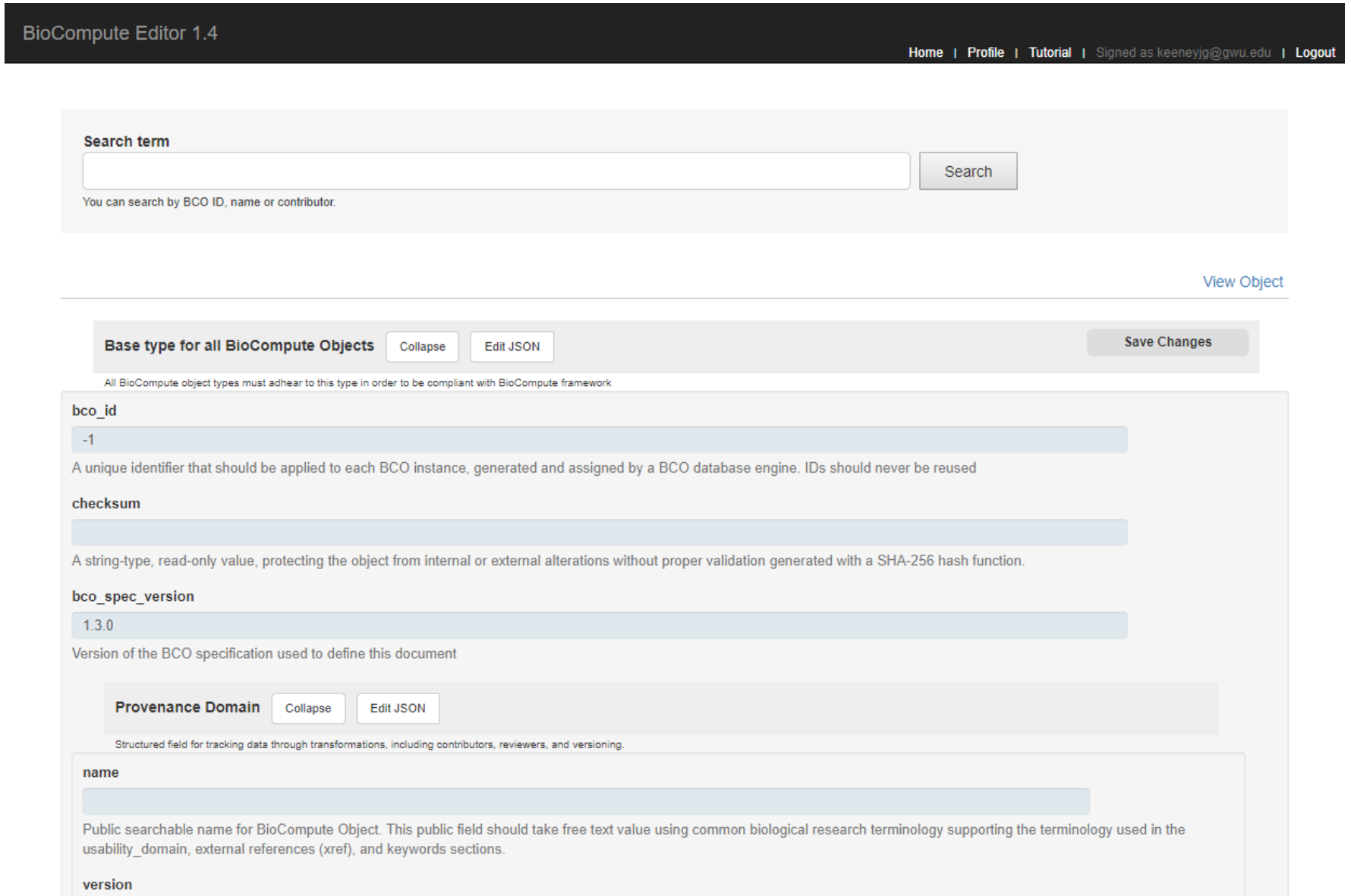


Galaxy PROJECT



HIVE

BioCompute Objects



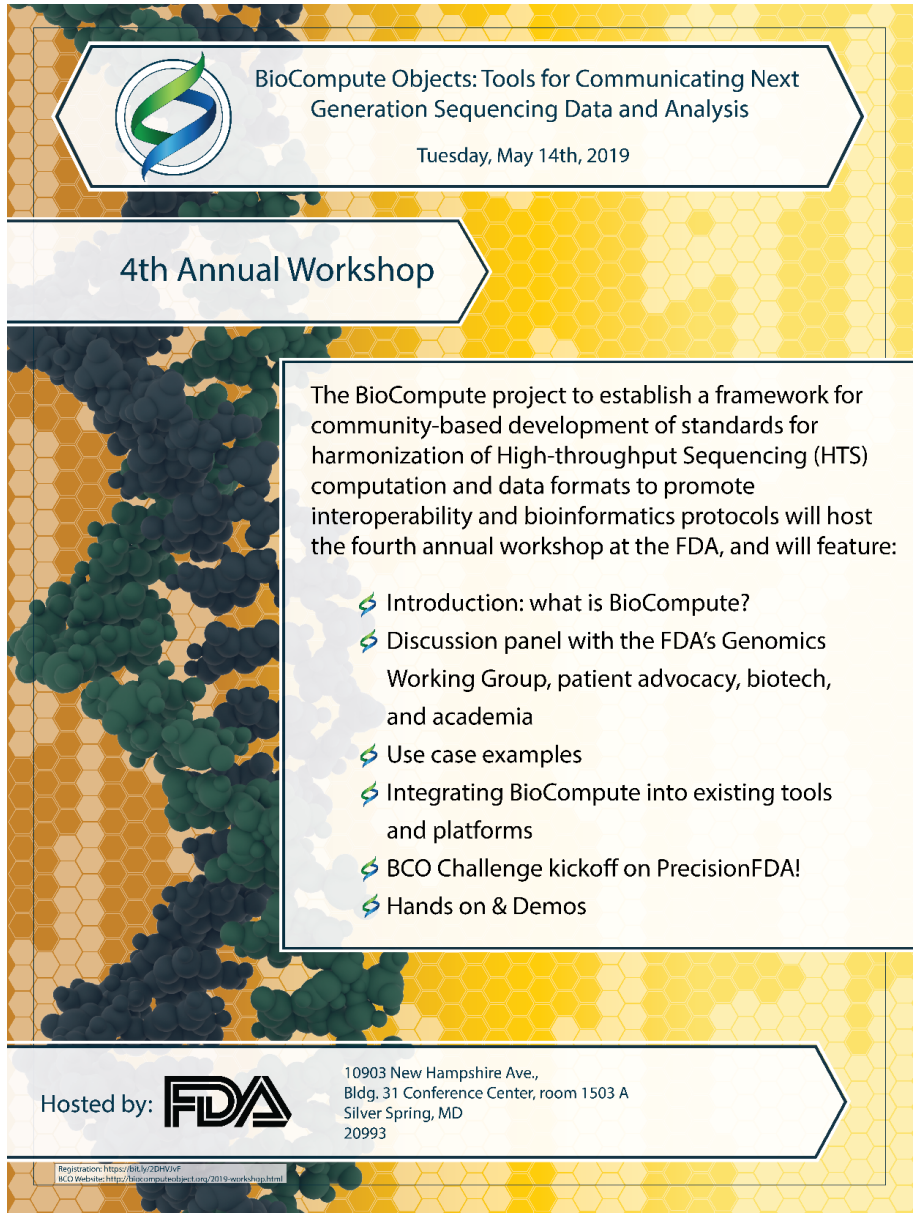
https://www.biocomputeobject.org/bco_editor/


Children's National
precisionFDA
ATTAIN
Global Alliance for Genomics & Health
DISEASE ONTOLOGY
EPINOMICS
NIH NATIONAL CANCER INSTITUTE
CUROVERSE
Forum for Collaborative HIV Research
Google
CDC CENTERS FOR DISEASE CONTROL AND PREVENTION
EMBL-EBI
DNAnexus
HARVARD UNIVERSITY
NATIONAL LIBRARY OF MEDICINE
NCBI
FHIR
FHIR Genomics
paradigm4
Oxford e-Research Centre
UNIVERSITY OF OXFORD
Seven Bridges
NIST
Galaxy PROJECT
NEW YORK GENOME CENTER
GENOME NEXT
CRITICAL PATH INSTITUTE
NIH
research object.org
CLC bio
HL7
CWL
FLATIRON
MANCHESTER 1824 The University of Manchester
elixir UNITED KINGDOM
CUNY

2019 Workshop

BioCompute Objects

- Discussion panel
  - How is workflow communication currently handled?
  - What are the challenges?
  - FDA, academia, private sector, patient advocacy
- Use case examples
- Demos and hands on
- BioCompute App-a-thon on PrecisionFDA



Use Case Examples

BioCompute Objects

- Tuberculosis Detection
  - Tuberculosis (TB) is top infectious killer in the world
  - WHO is adopting ReSeqTB pipeline to address the many challenges of detecting TB
  - Requires lineage identification, prediction of antibiotic resistance, recurrence of TB in previously treated patients
- Test Submission
  - HCV-1a use case using synthesized data
  - What data are necessary to make a regulatory decision?
  - Are summary data from one analysis pipeline sufficient?
  - How will the analysis pipeline be validated?
- Embleema
  - Embleema is a platform that allows users to take control of their own data
  - Marketplace for directly selling personal genome data
  - Aggregator for Real World Evidence

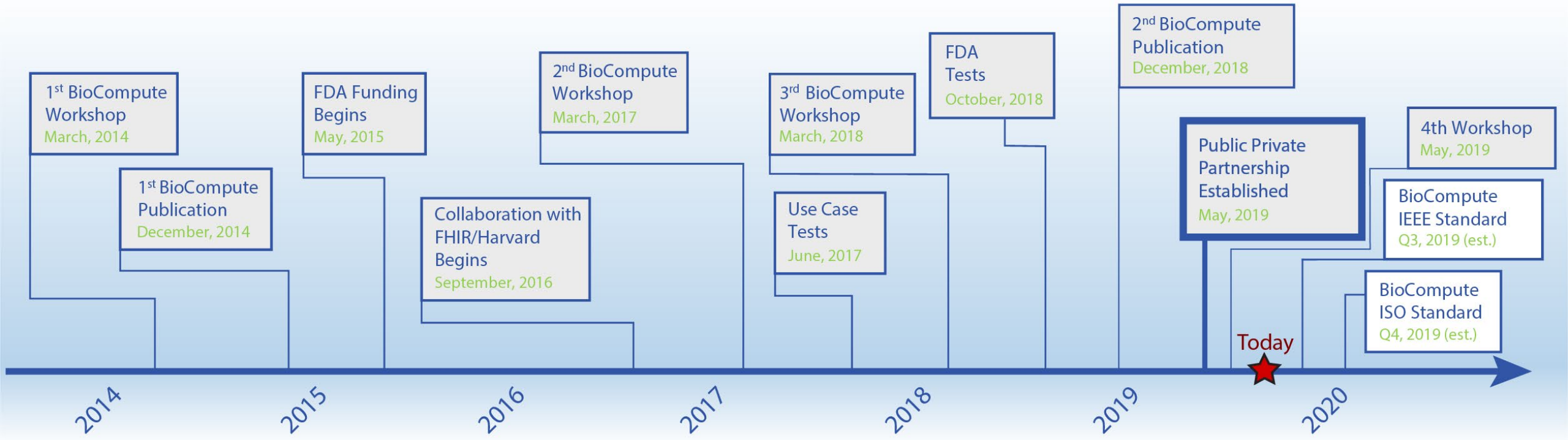
Demos

BioCompute Objects

- BioCompute standalone Editor
- BioCompute integration into HIVE
- BioCompute integration into Galaxy
- BioCompute integration into Seven Bridges

Timeline of Major Events

BioCompute Objects



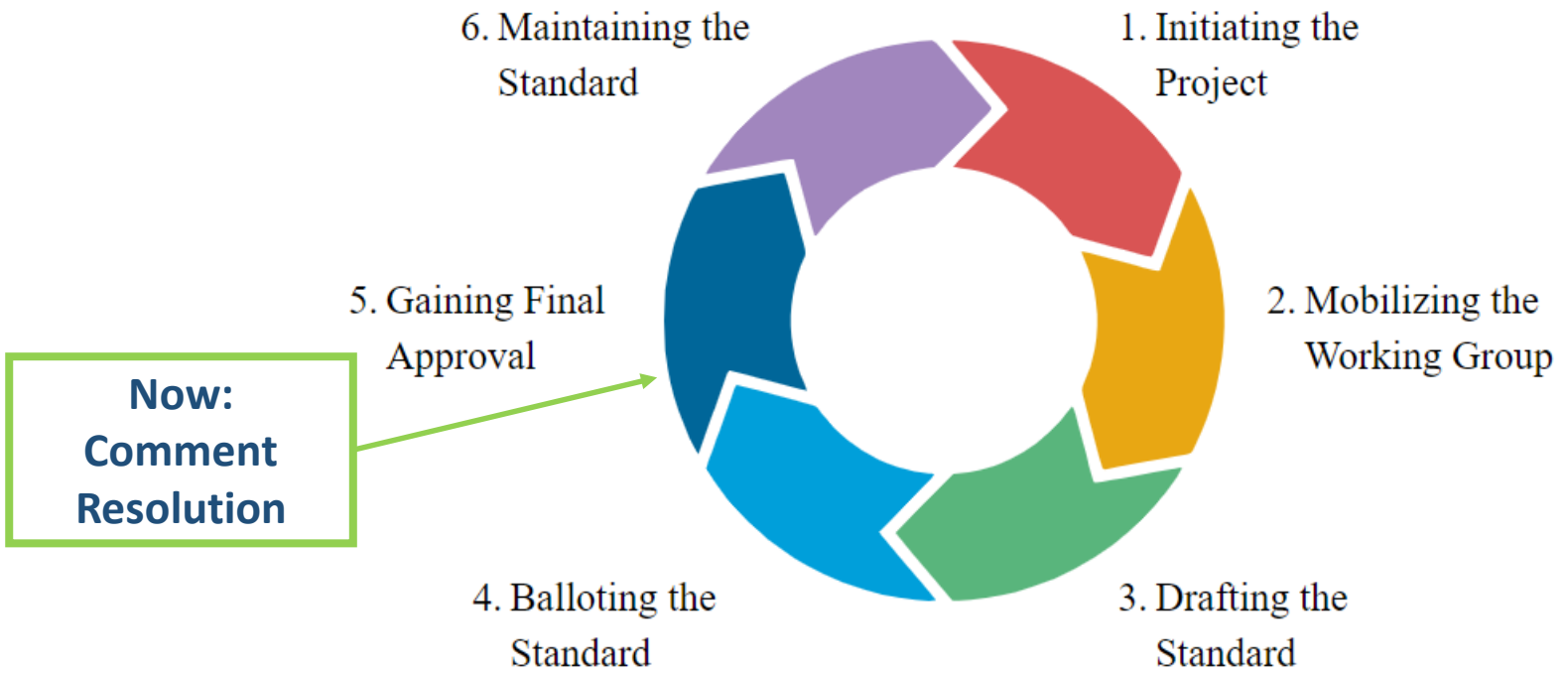
Future Plans

BioCompute Objects

- BCO writer integrated into more platforms
- BCO reader functionality
- BCO execute
- BCO cross-platform functionality
  - Requires containerization
- BCO databases
  - Cancer specific database

BioCompute Objects

IEEE Standards Development Process



BioCompute Objects

International Standards Organization

- Increase reach
- IEEE and ISO joint agreement
  - 7 month estimated timeline, if accepted
  - SC32 WG2 currently reviewing
    - Convened by Denise Warzel

BioCompute Public-Private Partnership

- Develop a community of stakeholders with interests in creating a versatile data harmonization framework that allows the standardized definition of platform-independent bioinformatics pipelines for execution, and is easily read by humans and machines.
- Facilitate the development of tools and facilities implementing data typing, instantiation, deposition, storage, and distribution of validated BioCompute Objects through a BioCompute database, in order to enable reproducible scientific research and regulatory submissions of data and computations.
- Facilitate portability of pipelines for execution.

Raja Mazumder (Chair)

Dennis Dean

Carole Goble

Jonathon Keeney (Managing Director)

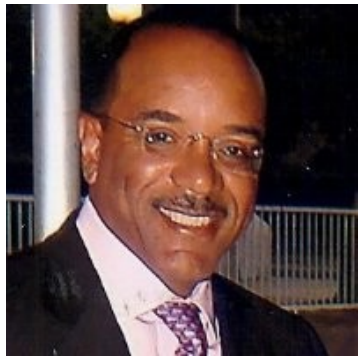
Vahan Simonyan

Gil Alterovitz

Jeremy Goecks

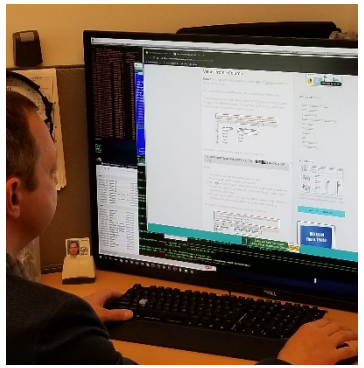
Executive Steering Committee

Hadley King (Chair)



Dan Taylor

Jonas Almeida



Anton Golikov

Nicola Soranzo

Elaine Thompson

Stian Soiland-Reyes

Jason Travis

Ntino Krampis

Michael R. Crusoe

Konstantinos Karagiannis

Technical Steering Committee

BioCompute Object (BCO) App-a-thon

PrecisionFDA partnered with George Washington University and the FDA HIVE to launch the BioCompute Object (BCO) App-a-thon

- Participants are be given the opportunity to innovate and standardize BCO creation and conformance
- Beginning and advanced tracks will allow individuals from varying levels to participate

App-a-thon launching May 14th 2019!

BioCompute Object (BCO) App-a-thon

BioCompute Objects

App-a-thon Goals:

- Attract novices and advanced programmers alike (beginner and advanced tracks)
- Provide exposure to those unfamiliar with BioCompute, and highlight its utility in communicating bioinformatics pipelines
- Build examples of ways that the BCO specification can be integrated into other tools (e.g., a platform)
- Create a database of BCOs that others can freely use

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

precisionFDA

* Complete conformance check details are available on challenge site

BioCompute Object (BCO) App-a-thon

May 14 through October 18

	Beginner Track: Create BCOs	Advanced Track: Create BCO software tools	
BCOs and Tools	 Complete BCO	 Conformance Tool  Writer Tool	**May 14 – October 18, 2019** Participants create BCOs and BCO tools
Conformance Check*	 Ability of an FDA reviewer to understand the BCO	 Ability to generate accurate BCO; Ability to conduct conformance check on BCOs	**Deadline: October 18, 2019** Evaluation begins
Results			

App-a-thon Launch: May 14, 2019
Challenge details are made available

HIVE precisionFDA THE GEORGE WASHINGTON UNIVERSITY WASHINGTON, DC

BioCompute Object (BCO) App-a-thon

BioCompute Objects

App-a-thon Goals:

- Attract novices and advanced programmers alike (beginner and advanced tracks)
- Provide exposure to those unfamiliar with BioCompute, and highlight its utility in communicating bioinformatics pipelines
- Build examples of ways that the BCO specification can be integrated into other tools (e.g., a platform)
- Create a database of BCOs that others can freely use

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

precisionFDA

BioCompute Object (BCO) App-a-thon

BioCompute Objects

Submissions to the Beginner Track must:

- Documents a multi-step bioinformatics pipeline in JSON file format
- Conforms to current BioCompute Specifications
  - https://github.com/biocompute-objects/BCO_Specification/tree/1.3.0
- Can be shown publicly

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

precisionFDA

BioCompute Object (BCO) App-a-thon

BioCompute Objects

Submissions to the Advanced Track must:

- A tool with the ability to do the following:
  - Create a BCO
  - Check a BCO for conformance
  - Display a BCO
- A detailed User manual (README)
- The application may be uploaded to precisionFDA or submitted as code

THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

precisionFDA

BioCompute Objects

App-a-thon site

https://precision.fda.gov/challenges/7

precisionFDA contact information:

Elaine Johanson, Acting Director

Office of the Chief Scientist/Office of Health Informatics

precisionfda@fda.hhs.gov

precisionFDA