

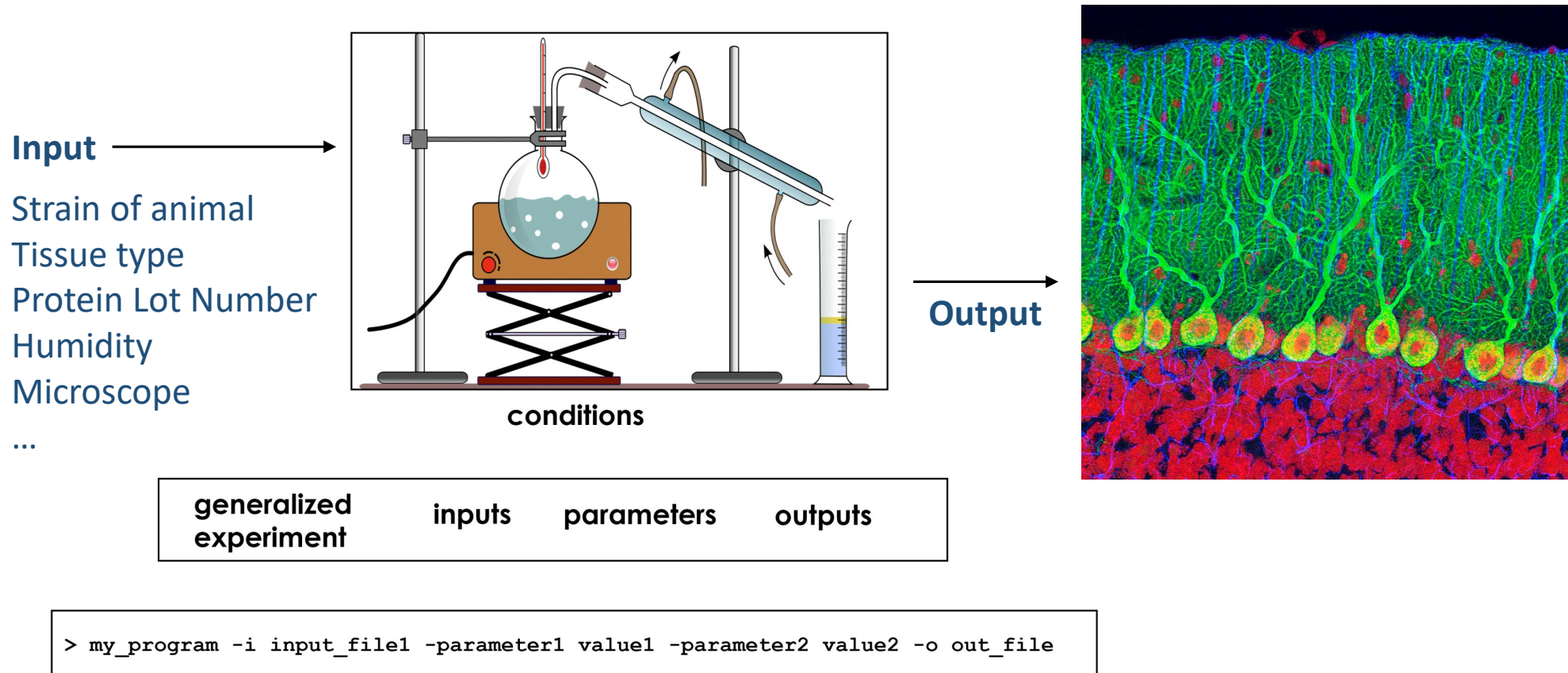
BioCompute: A Standardized Method to Communicate Bioinformatic Workflow Information and Ease Organizational Burden

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

Webinar: MODSIM Interest Group
April 21st, 2020

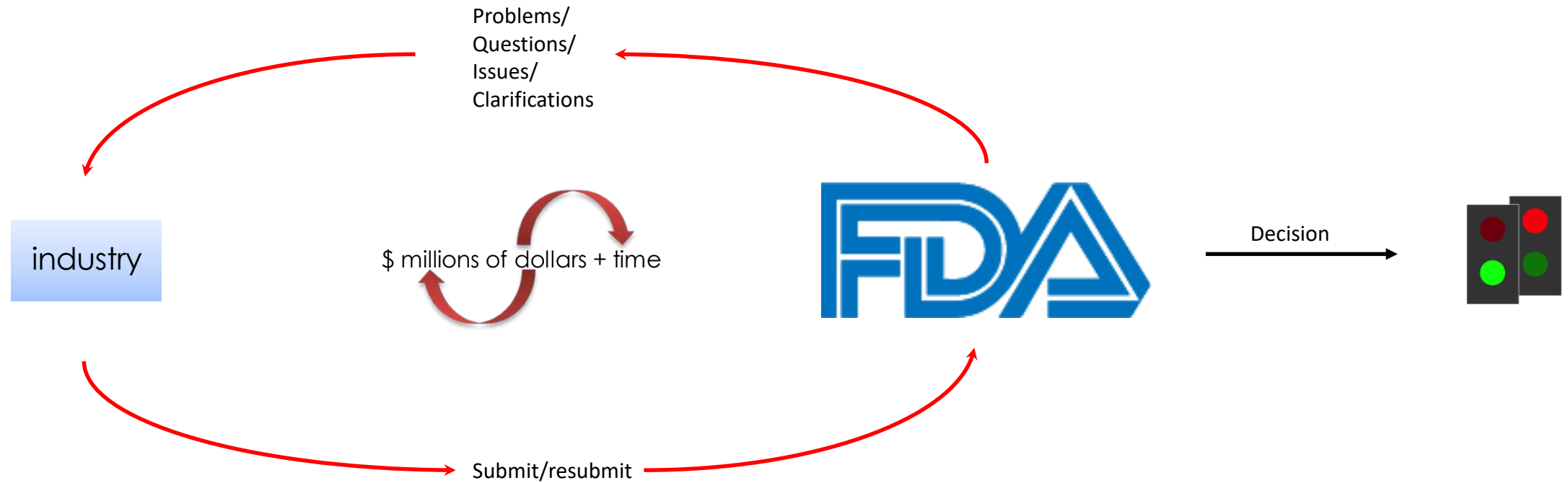
keeneyjg@gwu.edu

Challenge: Workflow Communication



Analogy: wet lab experiments

Wasted Time and Money



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.

A Solution Should...

- 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

- Standard for communicating genomic 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





<p>Top Level BCO ID: https://w3id.org/biocompute/1.3.0/examples/FDA-NA-TestsBreastCancer Checksum: 06DACE70679F35BA87A3DD6FFFD4ED24A4F5B8C2571264C37E5F1B3ADE04A31 Specification: https://w3id.org/biocompute/1.3.0/</p>	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Metadata</div>
<p>Provenance Domain Name: FDA-NA-TestsBreastCancer Version: 1.0 Review: approved: Natalie Abrams, NIH ; createdBy Created: 2018-05-24T09:40:17-0500 Modified: 2018-06-21T14:06:14-0400 Embargo: Start: 2000-09-26T14:43:43-0400 End: 2000-09-26T14:43:45-0400 Contributors: Janisha Patel (http://orcid.org/0000-0002-8824-4637), George Washington University; createdBy, modifiedBy Dara Baker, George Washington University; authoredBy License: https://spdx.org/licenses/CC-BY-4.0.html --> licensing is inferred by OncoMX licensing. Pub=</p>	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Parametric domain</div>
<p>Usability Domain FDA-approved or cleared nucleic acid-based human biomarker tests for breast cancer The .xlsx file FDA-NA-TestsBreastCancer.xlsx contains FDA-approved human biomarker tests for breast cancer. Each row represents one gene linked to its respective test. Genes are identified by UniProtKB, HgncName, EDNR number Tests are distinguished by manufacturer, FDA submission ID(s), clinical trial ID(s) and PubMed ID(s).</p> <p>Extension Domain Dataset Extension: Comment: Unique column headers for the dataset Test_disease_use: FDA-listed disease corresponding to approved test test_trade_name: FDA-listed product name test_manufacturerfee: FDA-listed patent company for the approved test sest_submission: FDA submission ID(s), web links; FDA-listed patent ID associated with test test_is_panel: A single biomarker or biomarker panel? Y for yes, N for no gene_symbol: HGNC ID from https://www.genenames.org uniprotKB_ac: UniProtKB from https://www.uniprot.org biomarker_id: Matched to EDNR IDs based on HGNC Name biomarker_origin: Characteristic that makes this a biomarker; molecular abnormalities that can lead to cancer ncit_biomarker: Searchable terms for gene/Biomarker from NCI Thesaurus (NCIt)</p>	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Usability domain</div> <div style="border: 1px solid black; padding: 5px; display: inline-block;">Extension domain</div>
<p>Description Domain Keywords: cancer, breast cancer, biomarker, biomarker test, FDA, UniProtKB, EDNR External References: (Name, Namespace, Ids) PubMed; pubmed; UniProt; accession; EDNR; EDNR number; HGNC; HgncName; GTR; GTR terms; Platform: Manual Pipeline Steps: Step 1: Download FDA-approved tests Description: FDA-approved tests were downloaded a list of FDA-approved or cleared nucleic acid based tests Input List: https://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm330711.htm Output List: ~/FDA-approved-or-cleared-NA-based-tests</p>	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Description domain</div>
<p>Execution Domain Scripts: none Script Driver: manual Software Prerequisites: None External Data Endpoints: Name In Vitro Diagnostics > Nucleic Acid Based Tests URL https://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm330711.htm Name NCBI Genetic Testing Registry URL https://www.ncbi.nlm.nih.gov/gtr/ Environment Variables: None</p>	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Execution domain</div>
<p>Parametric Domain N/A</p>	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Parametric domain</div>
<p>Input/Output Domain Input Subdomain: Filename: Multiple test files from "Nucleic Acid Based Tests: List of Human Tests" Access Time: 2018-10-10T11:34:02-5:00 URI: https://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm330711.htm Output Subdomain: Filename: FDA-NA-TestsBreastCancer.xlsx Media Type: xlsx/csv Access Time: 2018-10-10T11:37:02-5:00 URI: https://docs.google.com/spreadsheets/d/1xUY7WJNEZHyCgH5sYpxEuqAbtgVUuWgR2oc0IwhH28Y/edit#gid=1492026303</p>	<div style="border: 1px solid black; padding: 5px; display: inline-block;">IO domain</div>
<p>Error Domain</p>	<div style="border: 1px solid black; padding: 5px; display: inline-block;">Error domain</div>

BCO Example: HCV-1 drug resistance

- **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

Hexagon Aligner

NNS19-1

Parameters

Progress

Results

- Pie Chart
- Histogram
- Saturation
- Hit List
- Alignments
- Stack
- Hit Tables
- Help
- Downloads
- What's Next?

Hit List

id	Reference	Hits	Length	RPKM	Density
0	Unaligned	842.30K		-	-
90	gb CP00...=5163189	161.48K	5,163,189	42.4	8.7
19	gb CP00...=5729085	85.14K	5,729,085	20.2	4.1
16	gb CP00...=5243219	79.25K	5,243,219	20.5	4.2
14	gb CP01...=7084828	45.48K	7,084,828	8.7	1.8
167	gb FP92...=5976145	32.85K	5,976,145	7.5	1.5
42	gb CP00...=5444912	28.02K	5,444,912	7.0	1.4
24	gb CP01...=6472489	27.86K	6,472,489	5.8	1.2
15	gb FP92...=3214418	26.85K	3,214,418	11.3	2.2
163	gb FP92...=3344951	26.73K	3,344,951	10.8	2.2

Alignments

Stack

Hit Tables

Help

Downloads

What's Next?

Pie Chart

Saturation

Alignment

```
17 AAACGATCACGTCGTTTATFGGGTCTGCTGAACTTATCAGAGAACCTACAGAACGACCTTAAACGTGGTTGAGCTATCTGACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
300 AAACGATCACGTCGTTTATFGGGTCTGCTGAACTTATCAGAGAACCTACAGAACGACCTTAAACGTGGTTGAGCTATCTGACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
51 GAACCTATCAGAGAACCTACAGAACGACCTTAAACGTGGTTGAGCTATCTGACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
1 GAACCTATCAGAGAACCTACAGAACGACCTTAAACGTGGTTGAGCTATCTGACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
95 GGTTCGAGCTATCTGACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
124 GGTTCGAGCTATCTGACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
95 GGTTCGAGCTATCTGACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
1 GGTTCGAGCTATCTGACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
126 GGACAAGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
1 GGACAAGCACTGACAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
140 CAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
1 CAATGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
143 TTGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
301 TTGGCGTCCCGCCATCTCTCTATGAATTCFAGAAGAAAAGTTGTGGACTTGTGCGCGCTGCCATATAAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
213 ATATAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
1 ATATAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
213 ATATAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
1 ATATAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
213 ATATAAGAAATGGGGAAGAACGAGCTTATGTATGTATCTTGACCGAT
```

BCO Example: HCV-1 drug resistance

Main Home HIVE-Portal Links Contact Help Logout Janisha Patel News

HIVE Heptagon

Profile based on alignment :
HCV_702010133-206vst_S16_L001_R1/2.fastq

Parameters

Progress

Results

- Summary
- Downloads
- Contigs
- Annotation Files
- Profile Graphs
- Help
- Sequencing Noise
- Frequency Histogram
- Consensus
- SNP Calls

Hit List

id	Reference	Hits	Length	RPKM	Density
0	Unaligned	14.81K		-	-
1	CONSENSUS_C	27.39K	8,988	93,694.0	436.0
9	CONSENSUS_10_CD	1.33K	8,994	4,577.3	20.7
15	CONSENSUS_07_BC	1.03K	8,975	3,528.5	16.3
5	CONSENSUS_08_BC	626	8,796	2,188.1	9.6
12	CONSENSUS_11	266	8,977	1,271.3	5.7

Alignments

Start	Alignment
4258	ATTACAGaGACAGCAGAGACCCCTATTTGGAAAGGACCAGCCAAACTACTCTGGAAAGGT
94	ATTACAGgGACAGCAGAGACCCCTATTTGGAAAGGACCAGCCAAACTACTCTGGAAAGGT
4258	ATTACAGaGACAGCAGAGACCCCTATTTGGAAAGGACCAGCCAAACTACTCTGGAAAGGT
1	ATTACAGgGACAGCAGAGACCCCTATTTGGAAAGGACCAGCCAAACTACTCTGGAAAGGT

Summary

Annotation Files

- General Information -	
Total Reference Genome Length	8988
Number of Reference Genomes	20
- Mapped Regions -	
Total Contig Length	8988
Mapped Coverage (% Reference)	100.00
Average Coverage of Contigs	433.14
RPCM (Reads Per Contig b...er Million mapped reads)	13317
Total Number of Contigs	1
- Unmapped Regions -	
Total Length of the Unmapped Regions	0
Unmapped Regions (% Reference)	0.00
Average Coverage of Gaps	0.00
Total Number of Gaps Found	0

Profile Graphs

SNP Calls AA Calls Annotations

Reference: CONSENSUS_C Start position: default End position: default

Download graph as SVG file

Download graph as SVG file

```
{
  "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"
                },
                {
                  "address":
"https://hive.biochemistry.gwu.edudata/514801/SNPPProfile.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": [
                    {
                      "address":
"https://hive.biochemistry.gwu.edudata/514769/dnaAccessionBased.csv",
                      "access_time": "2017-01-24T09:40:17-0500"
                    },
                    {
                      "address":
"https://hive.biochemistry.gwu.edudata/514801/SNPPProfile.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"
                    }
                  ]
                }
              }
            }
          ]
        }
      }
    ]
  }
}
```

```

{
  "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 [S0:0000694] detection",
    "version": "1.1",
    "digital_signature": "905d7fce3f3ac64c8ea86f058ca71658",
    "review": [
      {
        "status": "approved",
        "comment": "Approved by GW staff. Waiting for approval from
FDA Reviewer",
        "url": "http://www.ncbi.nlm.nih.gov/nucleotide/22129792",
        "access_time": "2017-01-24T09:40:17-0500"
      }
    ]
  },
  "usability_domain": [
    {
      "Identify baseline single nucleotide polymorphisms SNPs [S0:0000694], insertions
[S0:0000667], and deletions [S0:0000045] that correlate with reduced ledipasvir
[PubChem:67505836] antiviral drug efficacy in Hepatitis C virus subtype 1 [taxID:31646]",
      "Identify treatment emergent amino acid substitutions [S0:0000048] that correlate
with antiviral drug treatment failure",
      "Determine whether the treatment emergent amino acid substitutions [S0:0000048]
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": {
            "address": "https://hive.biochemistry.gwu.edudata/514769/dnaAccessionBased.csv",
            "access_time": "2017-01-24T09:40:17-0500"
          }
        },
        "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": [
              {
                "step_number": "2",
                "name": "HIVE-heptagon",
                "description": "variant calling",
                "version": "1.3",
                "prerequisites": null,
                "input_uri_list": [
                  {
                    "address": "https://hive.biochemistry.gwu.edudata/514769/allCount-aligned.csv",
                    "access_time": "2017-01-24T09:40:17-0500"
                  }
                ]
              }
            ]
          }
        }
      }
    }
  ]
}

```

```
{
  "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 [S0:0000694] 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 [S0:0000694], insertions
[S0:0000667], and deletions [S0:0000045] that correlate with reduced ledipasvir
[PubChem:67505836] antiviral drug efficacy in Hepatitis C virus subtype 1 [taxID:31646]",
    "Identify treatment emergent amino acid substitutions [S0:0000048] that correlate
with antiviral drug treatment failure",
    "Determine whether the treatment emergent amino acid substitutions [S0:0000048]
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"
          }
        }
      ]
    }
  ]
}
```

```
http
http
```

```
http
http
```

```
http
http
http
http
```

```
{
  "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/SNPPProfile.csv",
    "access_time": "2017-01-24T09:40:17-0500"
  },
  {
    "address":
"https://hive.biochemistry.gwu.edudata/14769/allCount-aligned.csv",
```

```
"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
```

```
{
  "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 [S0:0000694] 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": [
    {
      "obsolete": false,
      "embargo": false,
      "start": null,
      "end": null,
      "created": null,
      "modified": null,
      "contributor": null
    }
  ],
  "license": {
    "address": null,
    "access_time": null
  },
  "usability_domain": {
    "Identify base pairs in a sequence
[S0:0000667], and determine if a sequence
[PubChem:67505836] is a match for a
"Identify tree structure of a sequence
with antiviral drug resistance
"Determine whether a sequence is
identified correlate with a
virus",
    "GitHub CWL examples/blob/master
examples/blob/master",
    "extension_domain": {
      "github_extensions": [
        "github_extensions",
        "github_extensions"
      ]
    }
  },
  "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": null,
                "access_time": null
              }
            }
          ]
        }
      ]
    }
  ],
  "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": [
          {
            "address":
"https://hive.biochemistry.gwu.edudata/514769/dnaAccessionBased.csv",
            "access_time": "2017-01-24T09:40:17-0500"
          },
          {
            "address":
"https://hive.biochemistry.gwu.edudata/514801/SNPPProfile.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": [
          {
            "address":
"https://hive.biochemistry.gwu.edudata/514769/dnaAccessionBased.csv",
            "access_time": "2017-01-24T09:40:17-0500"
          },
          {
            "address":
"https://hive.biochemistry.gwu.edudata/514801/SNPPProfile.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"
          }
        ]
      }
    }
  }
}
```



```
{
  "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 [S0:0000694] 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": null,
    "modified": null,
    "contributor": null
  },
  "license": {
    "address": null,
    "access": null
  },
  "usability_domain": {
    "identify_treatment_emergent_amino_acid_substitutions": [
      {
        "description": "Identify treatment emergent amino acid substitutions [S0:0000045] that correlate with reduced ledipasvir
[PubChem:67505836] antiviral drug efficacy in Hepatitis C virus subtype 1 [taxID:31646]",
        "detection": "Identify treatment emergent amino acid substitutions [S0:0000048] that correlate
with antiviral drug treatment failure",
        "determination": "Determine whether the treatment emergent amino acid substitutions [S0:0000048]
identified correlate with treatment failure involving other drugs against the same
virus",
        "example": "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": null
      }
    }
  },
  "parametric_domain": {
    "HIVE-hexagon": {
      "seed": "14",
      "minimum_match_len": "66",
      "divergence_threshold_percent": "30"
    },
    "HIVE-heptagon": {
      "minimum_coverage": "15",
      "freq_cutoff": "0.10"
    }
  },
  "pipeline_steps": [
    {
      "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"
        }
      ]
    },
    {
      "step_number": "2",
      "name": "HIVE-heptagon",
      "description": "variant calling",
      "version": "1.3",
      "prerequisites": null,
      "input_uri_list": [
        {
          "address":
"https://hive.biochemistry.gwu.edudata/514769/allCount-aligned.csv",
          "access_time": "2017-01-24T09:40:17-0500"
        }
      ]
    }
  ],
  "output_uri_list": [
    {
      "address":
"https://hive.biochemistry.gwu.edudata/514769/dnaAccessionBased.csv",
      "access_time": "2017-01-24T09:40:17-0500"
    },
    {
      "address":
"https://hive.biochemistry.gwu.edudata/514801/SNPPProfile.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"
    }
  },
  "io_domain": {
    "input_subdomain": {
      "subject": [
        {
          "parametric_domain": {
            "HIVE-hexagon": {
              "seed": "14",
              "minimum_match_len": "66",
              "divergence_threshold_percent": "30"
            },
            "HIVE-heptagon": {
              "minimum_coverage": "15",
              "freq_cutoff": "0.10"
            }
          }
        }
      ]
    }
  }
}
```

Key Feature of a BCO

- **Abstract away workflow based on commonalities**
 - Platform/tool/protocol independent
- **Usability Domain**
 - Free text description
- **Data provenance**
 - Data manifest, track files from beginning to end
 - Track user attribution (authored by, contributed by, reviewed by, etc.)
- **Validation Kit**
 - Error Domain + IO Domain
 - Sanity check: given the input files and the inherent error, is the output this analysis claims to have gotten valid?
- **Extensible**
 - Extension Domain
 - Open source repository
- **Embargo Domain**
 - Prevent others from viewing a BCO for any amount of time

BioCompute Schema Files

<https://gitlab.com/IEEE-SA/2791/ieee-2791-schema>

- IEEE 2791 Schema
- Project overview
- Details
- Activity
- Releases
- Repository
- Issues 1
- Merge Requests 1
- CI / CD
- Security & Compliance
- Packages
- Analytics
- Wiki
- Snippets

README BSD 3-clause "New" or "Revised" License

Name	Last commit	Last update
.gitignore	Creates initial release of BioCompute Object Schema in prep for ball...	1 year ago
2791object.json	Addressing Stian's comment: 12) (editorial) Comment in this line us...	4 months ago
AUTHORS	Rename all BCO → IEEE-2791 Object (and similar) Transformations ...	6 months ago
CONTRIBUTORS	Rename all BCO → IEEE-2791 Object (and similar) Transformations ...	6 months ago
LICENSE	Rename all BCO → IEEE-2791 Object (and similar) Transformations ...	6 months ago
README.md	Rename all BCO → IEEE-2791 Object (and similar) Transformations ...	6 months ago
description_domain.json	Update id lines to new link - removed `/schemas` from line 3 in all s...	6 months ago
error_domain.json	Update id lines to new link - removed `/schemas` from line 3 in all s...	6 months ago
execution_domain.json	SSR Comments Fixes to address Comments from SSR on 2791 sche...	6 months ago
io_domain.json	SSR Comments Fixes to address Comments from SSR on 2791 sche...	6 months ago
parametric_domain.json	SSR Comments Fixes to address Comments from SSR on 2791 sche...	6 months ago
provenance_domain.json	Minor Fixes Removed name and updates 'object_id' definition	4 months ago
usability_domain.json	SSR Comments Fixes to address Comments from SSR on 2791 sche...	6 months ago

BioCompute Schema Files

<https://gitlab.com/IEEE-SA/2791/ieee-2791-schema>

The screenshot shows the GitLab interface for the IEEE 2791 Schema repository. The top navigation bar includes the GitLab logo, navigation links (Projects, Groups, Snippets, Help), a search bar, and a 'Sign in / Register' button. The left sidebar contains a navigation menu with options like Project overview, Details, Activity, Releases, Repository, Issues (1), Merge Requests (1), CI / CD, Security & Compliance, Packages, Analytics, Wiki, and Snippets. The main content area displays the repository's file structure, including a README and a BSD 3-clause license. A table lists the files in the repository, with columns for Name, Last commit, and Last update. The file '2791object.json' is highlighted with a red box.

Name	Last commit	Last update
.gitignore	Creates initial release of BioCompute Object Schema in prep for ball...	1 year ago
2791object.json	Addressing Stian's comment: 12) (editorial) Comment in this line us...	4 months ago
AUTHORS	Rename all BCO → IEEE-2791 Object (and similar) Transformations ...	6 months ago
CONTRIBUTORS	Rename all BCO → IEEE-2791 Object (and similar) Transformations ...	6 months ago
LICENSE	Rename all BCO → IEEE-2791 Object (and similar) Transformations ...	6 months ago
README.md	Rename all BCO → IEEE-2791 Object (and similar) Transformations ...	6 months ago
description_domain.json	Update id lines to new link - removed `/schemas` from line 3 in all s...	6 months ago
error_domain.json	Update id lines to new link - removed `/schemas` from line 3 in all s...	6 months ago
execution_domain.json	SSR Comments Fixes to address Comments from SSR on 2791 sche...	6 months ago
io_domain.json	SSR Comments Fixes to address Comments from SSR on 2791 sche...	6 months ago
parametric_domain.json	SSR Comments Fixes to address Comments from SSR on 2791 sche...	6 months ago
provenance_domain.json	Minor Fixes Removed name and updates 'object_id' definition	4 months ago
usability_domain.json	SSR Comments Fixes to address Comments from SSR on 2791 sche...	6 months ago

ACCESS: Private | NAME: test-workflow | ORG: dnanexus.science | ADDED BY: sam.westreich | ID: workflow-FQ7P7Vj05922F6k6J3b87yQ6

CREATED: 2018-12-10 23:16:23

Edit tags

Revision: 1 | Latest | Edit | Fork | Export | Run Workflow rev1

SPEC | WORKFLOW DIAGRAM

INPUTS

file	Input 1	REQUIRED	workflow-app-1
file	Input 2	REQUIRED	workflow-app-2

OUTPUTS

file	Output 1	REQUIRED	workflow-app-1
file	Output 2	REQUIRED	workflow-app-2



Projects | Data | Apps

Identifiers and File name(s) | Search | Queries | Save Query | Copy files to project

Start Query From:

- Case
- File
- Sample
- Portion
- Slide
- Analyte
- Aliquot
- Drug therapy
- Radiation therapy
- Follow up
- New Tumor Event

Workflow diagram showing steps: File (ADD FILTER) -> Data Format (Remove filters) -> Experimental Strategy (Remove filters) -> Disease Type (Remove filters)



Galaxy Administration

Galaxy Administration | Administration | Security | Data | Server | Tool sheds | Form Definitions | Sample Tracking

Repository Actions | Tool Shed Actions

Genome/Exome paired analysis (SNVMix1)

Boxes are red when tools are not available in this repository (this page displays SVG graphics)

Workflow diagram showing tool dependencies and execution flow.

Main Home | HIVE Portal | Links

CensusScope

HMB25-2_R1

Parameters | Progress | Results | Taxonomy Details | Taxonomy Help | Convergence | Phylogenetic Tree | Tree View | Table | Subtree | What's Next? | Alignment

Loading Status

Building histogram	Done 100%
Preparing alignments	Done 100%
Visualizing alignments in track	Done 100%
Fetching alignments	Done 100%
Creating mutation heat diagram	Done 100%

Taxonomy Help | Taxonomy Details

Alcemy | BioProjectID | Name | Taxname | Parent | Rank | Taxonomy ID

Donut chart visualization showing taxonomic distribution.



BCO Portal

BioCompute Editor

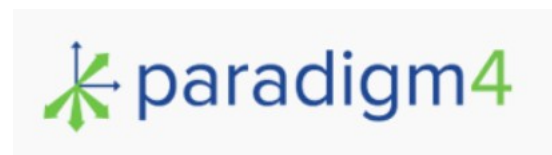
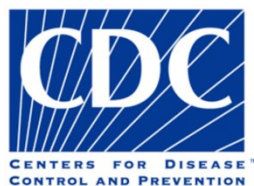
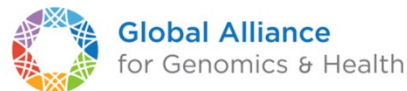


Sign in

SIGN IN NOW

Don't have an account? [Sign up](#)

<https://portal.aws.biochemistry.gwu.edu/sign-in>



Use-Case Examples

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

BioCompute Object (BCO) App-a-thon

May 14 through October 18



Formal Standardization

- Institute of Electrical and Electronics Engineers Standard (approved January 2020)
 - Anticipated publication date: April 10th, 2020

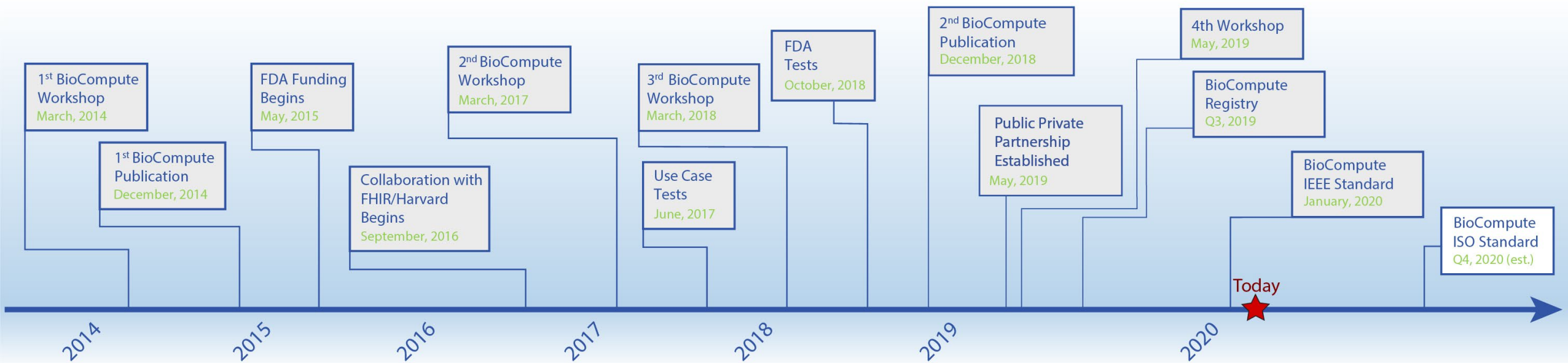


Integrating with Other Standards

- International Standards Organization certification expected by Q4 2020 through joint agreement
 - TC215 Healthcare Informatics WG3 (Semantic Content)



BCO Timeline



Expansion Beyond Life Sciences

- ISO
 - SC32 (Data Management and Interchange) WG2 (Metadata)
- Collaborative explorations
 - Atin Basuchoudhary, Virginia Military Institute
 - Economics
 - Dan Chrichton, Jet Propulsion Laboratories
 - Data modeling

Modeling Simulation Examples



- **BioGaming**

- Similar to “war gaming”
- Data driven simulations of a pandemic
- Provides insights into how interventions or pathogen changes may affect pandemic course

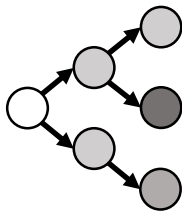
13

Al

Aluminum
26.981538

- **Aluminum Efflux**

- CBER algorithm for creating aluminum pharmacokinetic profiles
 - Mitkus et al., 2011
- Safety prediction of aluminum containing infant vaccines
- Predicts Aluminum safety threshold
 - Based on level set by the Agency for Toxic Substances and Disease Registry



- **nSARS-CoV-2 Evolution Tracking**

- CDER-based pipeline for identifying variants
- Also used to inform glycospike protein

Jonathon Keeney, Ph.D.
Assistant Research Professor
The George Washington University
keeneyjg@gwu.edu