



Platform Based BioCompute Object Advances and BioCompute Controlled Database Access

Dennis A. Dean, II, PhD

September 13, 2022

Agenda

Motivation

Platform Enabled BioCompute Objects

- Platform Perspective
 - BCO App with Database Publishing

BCO Creation

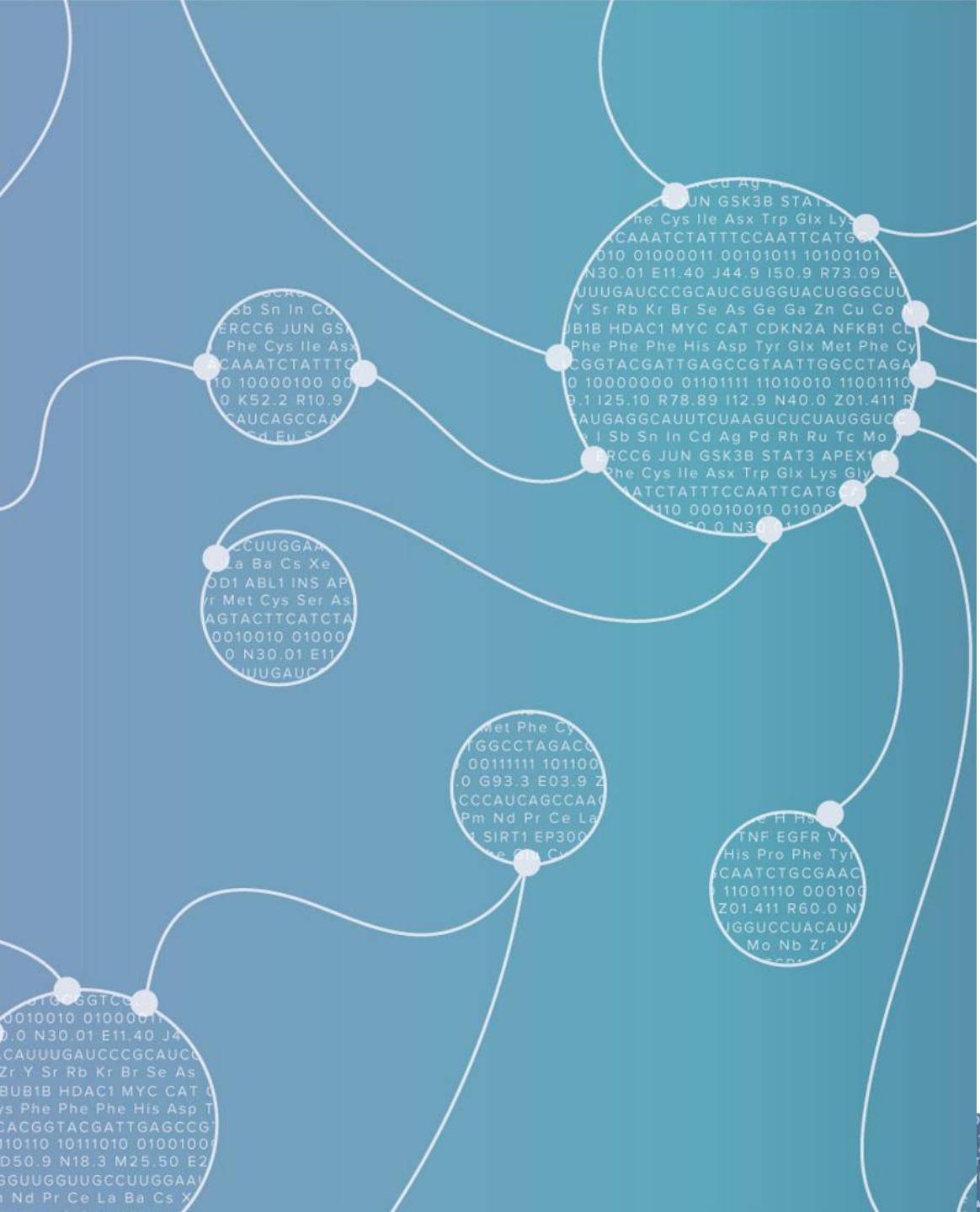
- BCO APP and Example Availability
 - BCO APP Demonstration - Workflow
 - BCO APP Demonstration - Task
 - Advance Topics – Executable BCOs

Towards Best Practices and Regulatory Submission Best Practices

- ## Best Practices

Motivation

Seven Bridges



Bringing Experimental Rigor to Bioinformatics

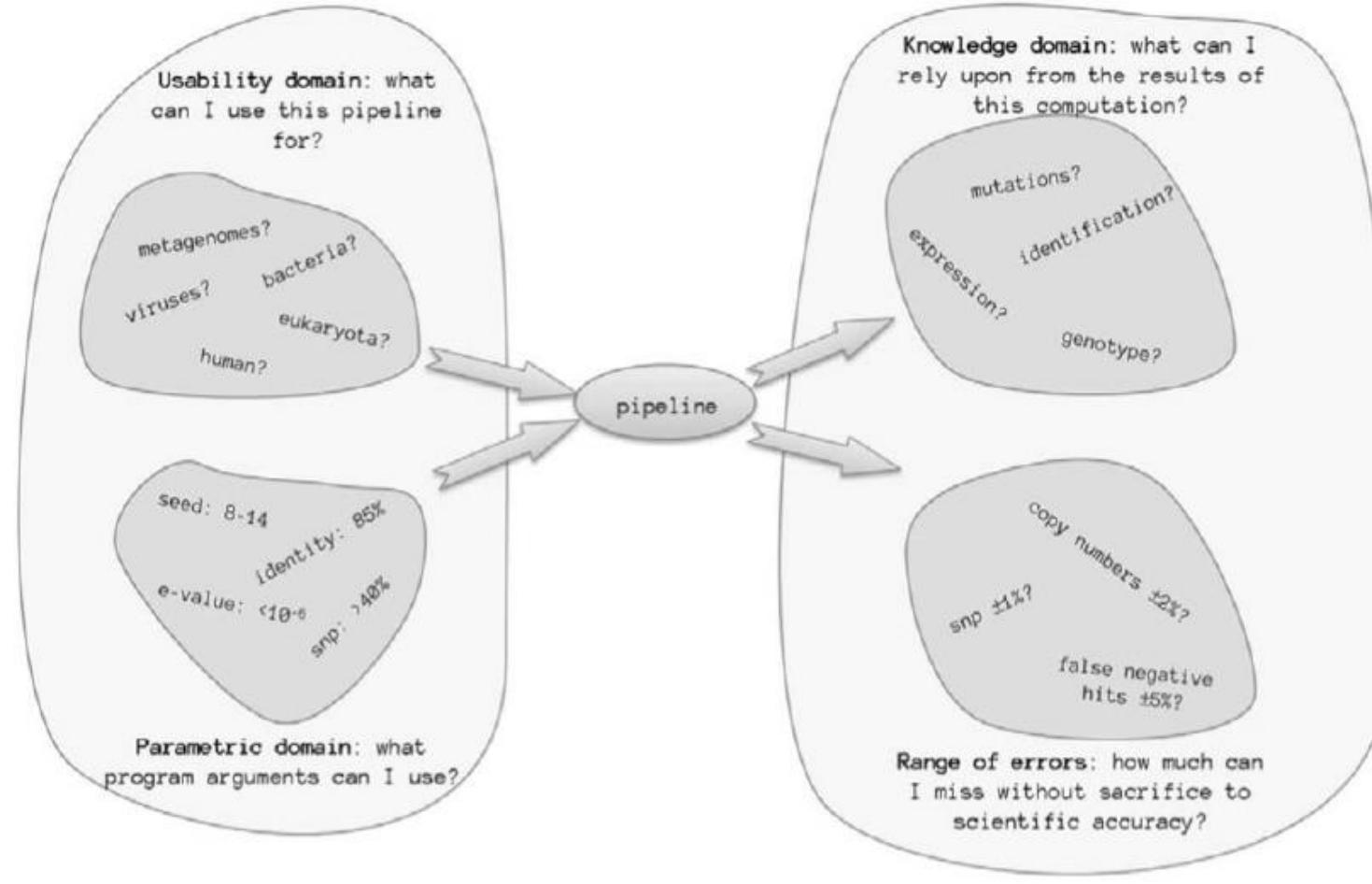


Figure 3.

Illustration of an experimental protocol that will have a specific set of input and output domains.

Simonyan et al 2017

Specify Appropriate Parameter Set

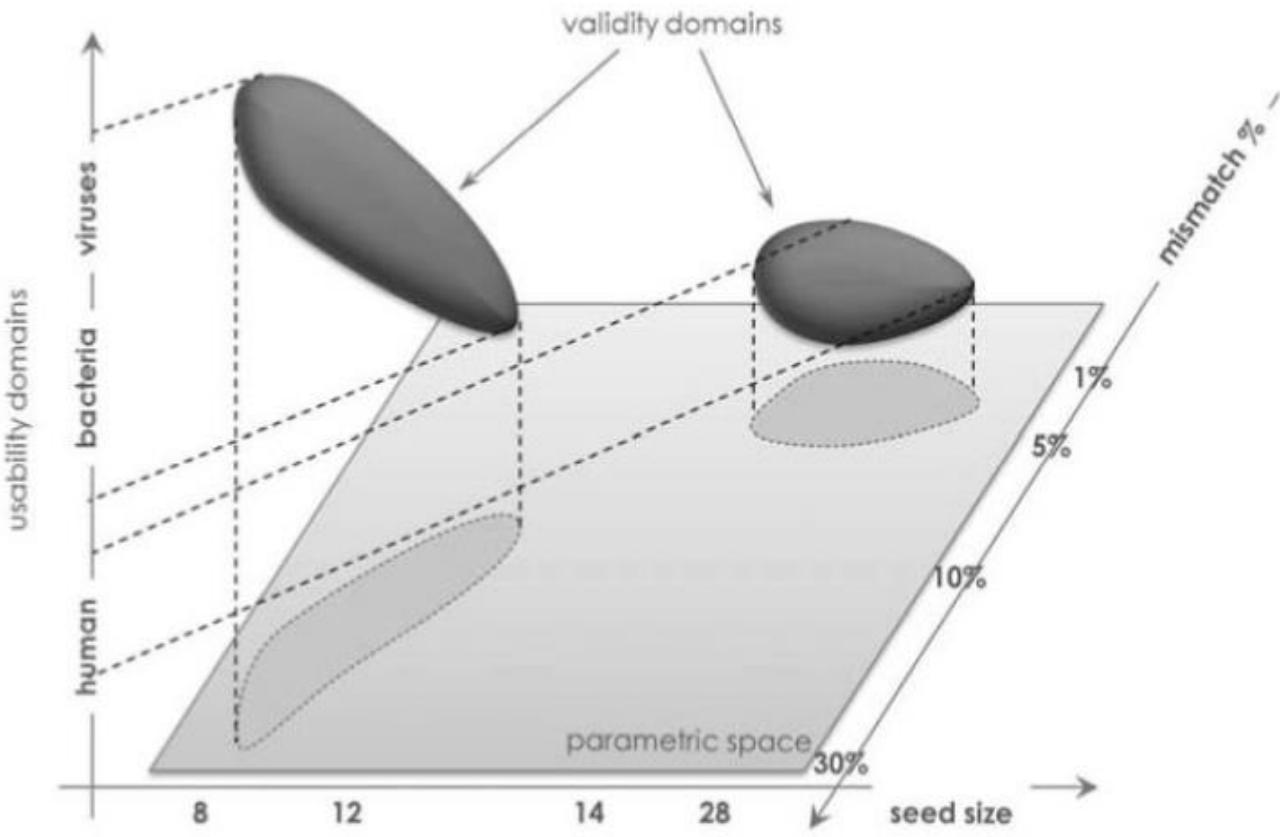


Figure 4.

Illustration of validity domains of a pipeline in three dimensions. The parametric space (3 axis) in this case is expansive, but only a small subset represented by the dark blue shapes are valid in the pipeline. Parameters that fall outside of these areas are not within the scope of the pipeline and not guaranteed to result in usable outputs.

Simonyan et al 2017

BioCompute Object Standard Aims to be Inclusive, Extendable with a Low Barrier to Entry

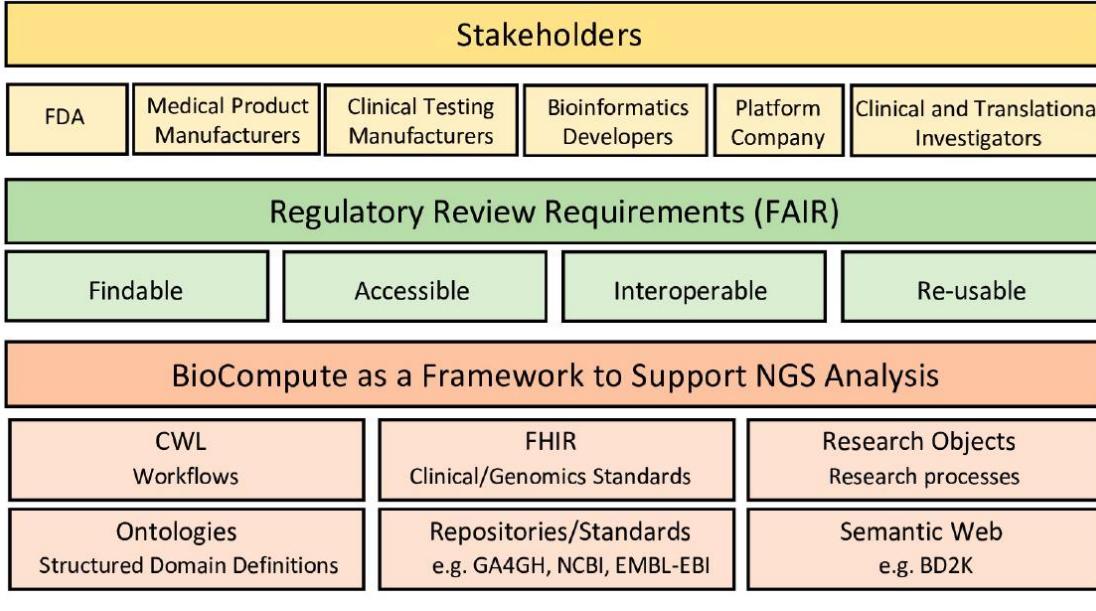


Fig 3. General FHIR platform schematic with proposed BCO integrations and extensions. BCO, BioCompute Object; BD2K, Big Data to Knowledge; Desc., description; EMBL-EBI, European Molecular Biology Laboratory-European Bioinformatics Institute; Env., environmental; FDA, Food and Drug Administration; FHIR, Fast Healthcare Interoperability Research; GA4GH, Global Alliance for Genomics and Health; ID, identification; IO, input/output; NCBI, National Center for Biotechnology Information; NGS, Next-Generation Sequencing; Prereq., prerequisite; PROV, provenance specification; RO, research object; URI, uniform resource identifier; W3C, World Wide Web Consortium; Xref, external reference.

Alterovitz et al 2019

BioCompute Object Standard Aims to be Inclusive, Extendable with a Low Barrier to Entry

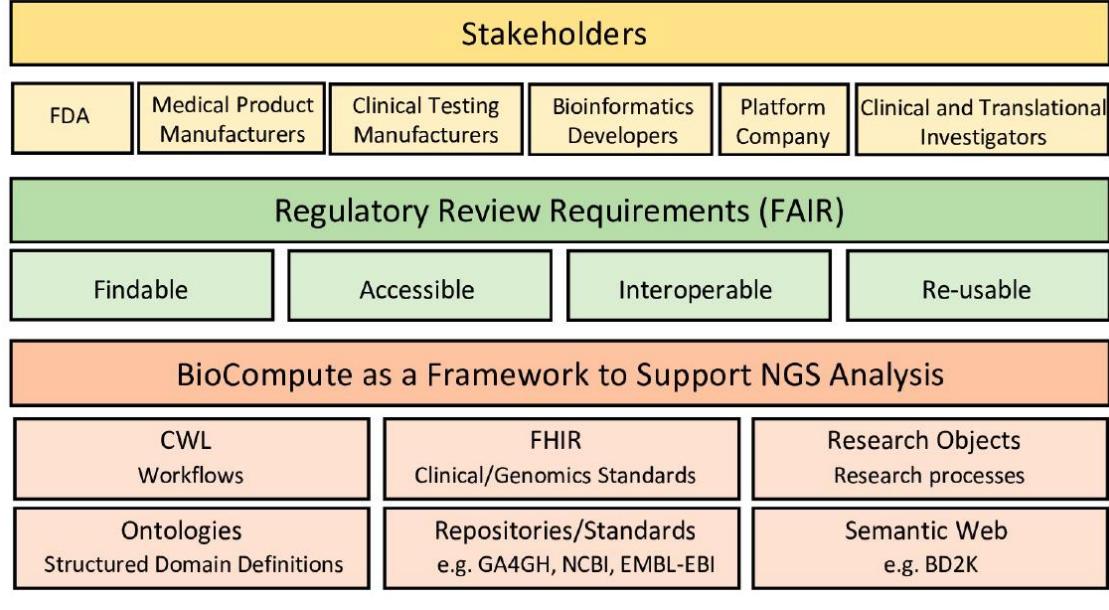
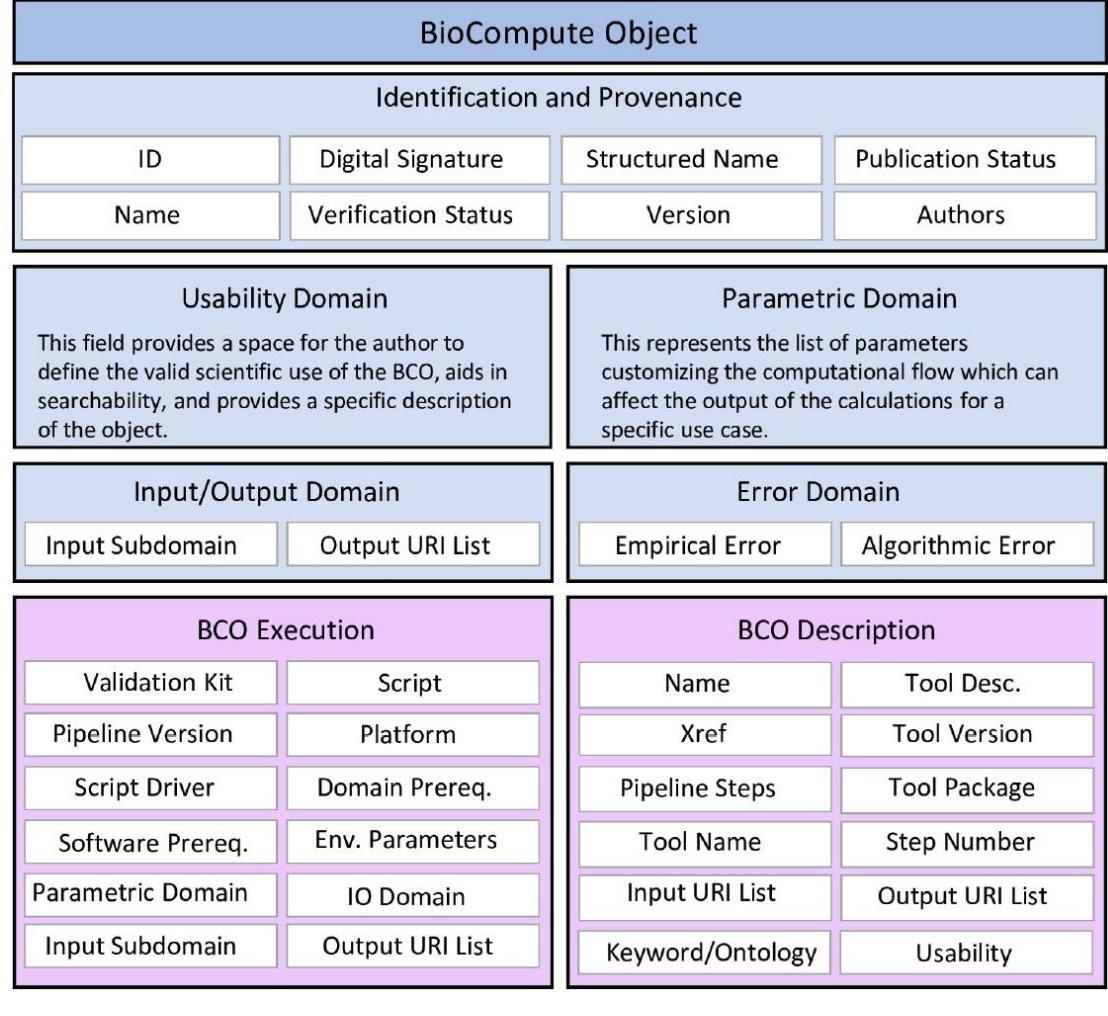


Fig 3. Generic HTS platform schematic with proposed BCO integrations and extensions. BCO, BioCompute Object; BD2K, Big Data to Knowledge; Desc., description; EMBL-EBI, European Molecular Biology Laboratory-European Bioinformatics Institute; Env., environmental; FDA, Food and Drug Administration; FHIR, Fast Healthcare Interoperability Research; GA4GH, Global Alliance for Genomics and Health; ID, identification; IO, input/output; NCBI, National Center for Biotechnology Information; NGS, Next-Generation Sequencing; Prereq., prerequisite; PROV, provenance specification; RO, research object; URI, uniform resource identifier; W3C, World Wide Web Consortium; Xref, external reference.



Alterovitz et al 2019

BioCompute Object Success Rests on Integrated Solutions

BioCompute Object JSON in Sublime Text Window

```
111-00100111-01110011-11010000-011100001-01000000-10100110-110011010-10000100-00110110-101110010-01000  
0-G93-3 E03-9 L79-899 E78-2 D64-9 E11-65 E55-9 148-91 R82-5 R30-0 K52-2 R10-9 D50-9 N'8-3 M25-80
```

BioCompute Object Success Rests on Integrated Solutions

```
rna-seq-alignment-star.bco.json UNREGISTERED
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
1 characters selected
Spaces: 2 JSON
```

BioCompute Object JSON in Sublime Text Window

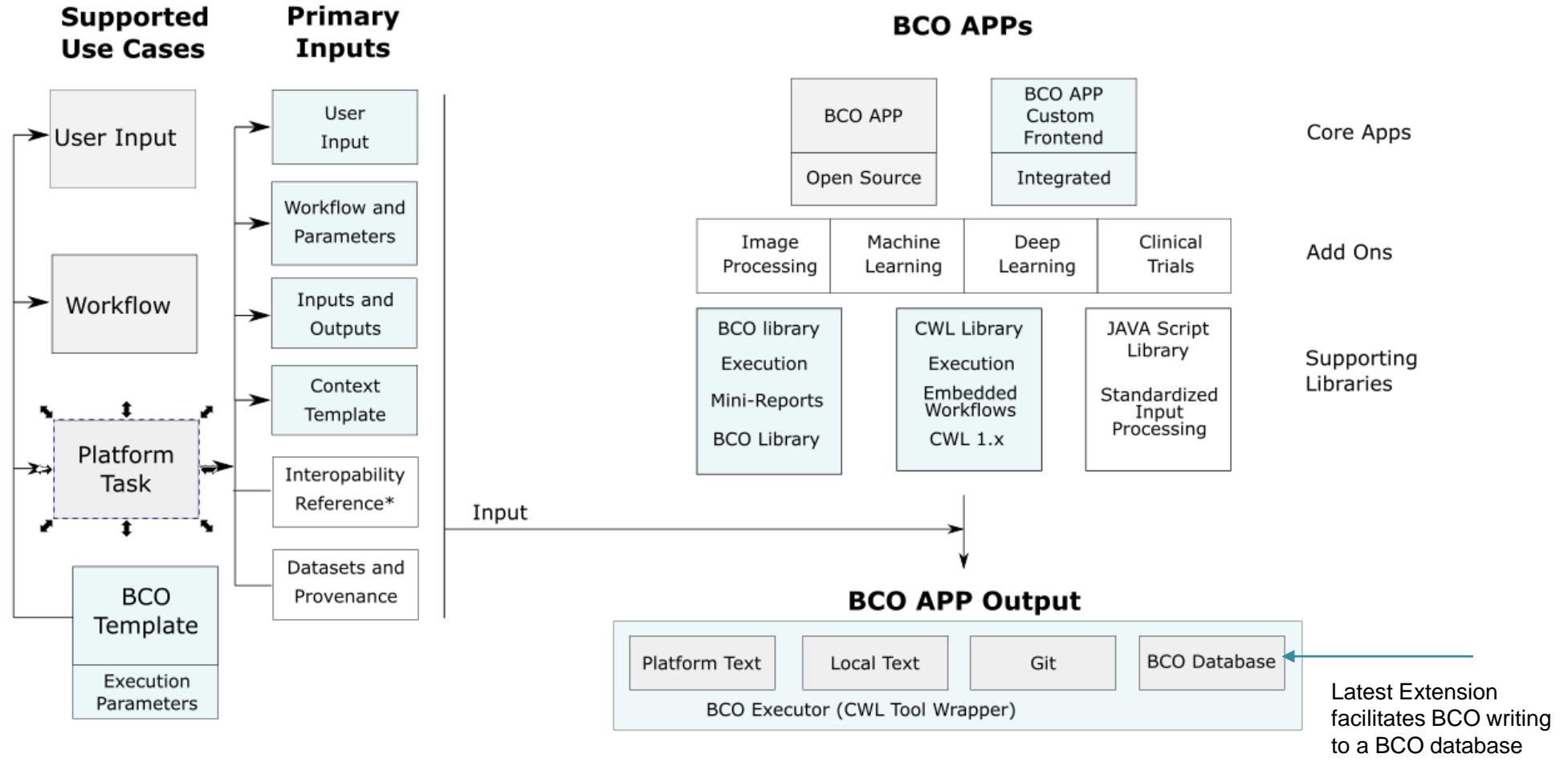
The Seven Bridges BCO App provides tools for generating BioCompute Objects and tools for working with BioCompute Objects. The BCO App can generate BioCompute Objects from user text input, workflows written in Common Workflow Language (CWL), and from CWL/task information generated on the Cancer Genomics Cloud. The BCO App also includes tools for reviewing and validating BioCompute Objects. This version of the BCO App extends our initial prototype developed for the precisionFDA BioCompute Object Challenge.

The BioCompute Object standard aims to facilitate the communication of Next Generation Sequencing Analyses by providing a structured way for communicating complex analyses as commonly seen in bioinformatics and precision medicine applications. See the BioCompute Object webpage for additional information (<https://biocomputeobject.org>).

Text Composer CWL Composer Platform Composer Need Help?

BioCompute Object Generator for Seven Bridges Platforms

BioCompute Object Vision for Platform Solutions



*Example interoperability references include the Tool Registry Service (TRS), the Workflow Execution Service (WES), and the Data Repository Service (DRS)

Platform-Based BCO Generation and Publishing

10000000 01101111 11C10010 11001110 00010010 01000011 00101011 10100101 11000101 01001000 01110011 01110111 00010110 00111111 10110010 00001001 00101010 10000001
1.125.10 R78.89 I12.9 N40.0 Z01.411 R60.0 N30.0 E1.40 J44.9 I50.9 R73.9 E87.6 E11.29 Z15.2 I10 Z12.4 E11.9 E78.5 N39.0 G93.3 E03.9 Z79.899 E78.2 D64.9 E11.65 E55.1
AUGAGGCAUUTCUAGACUCUUAUGGUCCUACAUUUAGUCCGCAUCGGGUACUGGGCUUAAGCGAUACACCUAAUAGGUACUGGGCACGGACCCA
Te 1 Sb Sn In Cd Ag Pd Rh Tr Tc Mo Nb Zr Y Sr Br Kr Se As Ge Ga Zn Cu Co Ni Fe Mn Cr V Ti Sc Ca Ar K Cl S P Si Al Sm Pm h
P ERCC6 JUN GSK3B STAT3 APEX1 ESR1 BUB18 HDAC1 MYC CAT CDKN2A NFKB1 CLU IGF1R PML CREB1 E2F1 RBL UBE2IMD NR3C1 SH
Phe Cys Ho Asx Trp Glx Lys Gly Arg Lys Pro Phe His Tyr Gix Met Phe Cys Leu Ser Asn Ho Glu Val Asx Trp Gly Pro Arg Phe Glu Cys Gin Val Phe Thr His Pro Phe Tyr Met C
NACAACTATCTTCTTCAATCATCGATGAGGCCACGGTACGATTGAGCGTAATTGGCTCATGACGCTTTCTCCGGCTTCGGCAGGCCAGAACGACTGGGTGGCAATCTCGAACCG



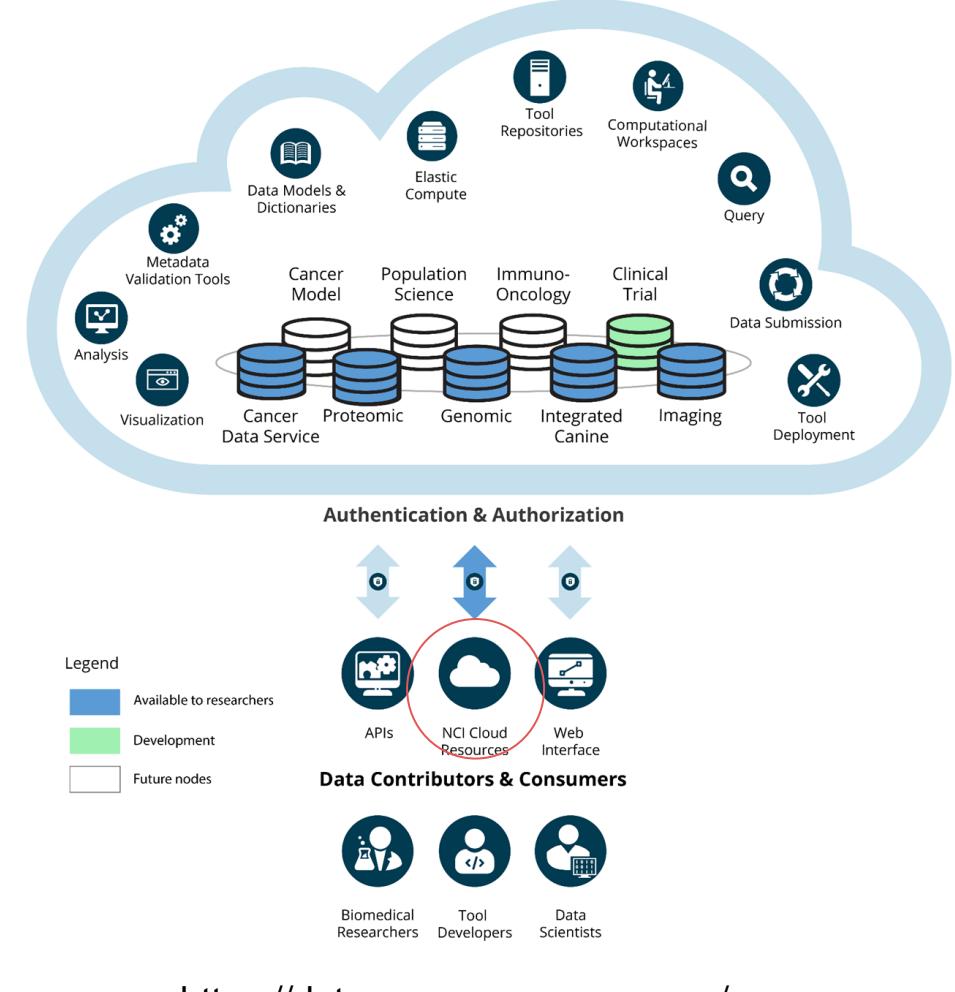
CANCER GENOMICS CLOUD SEVEN BRIDGES

Provides powerful, yet easy to use interfaces to empower cancer researchers to draw new insights from petabyte scale data.

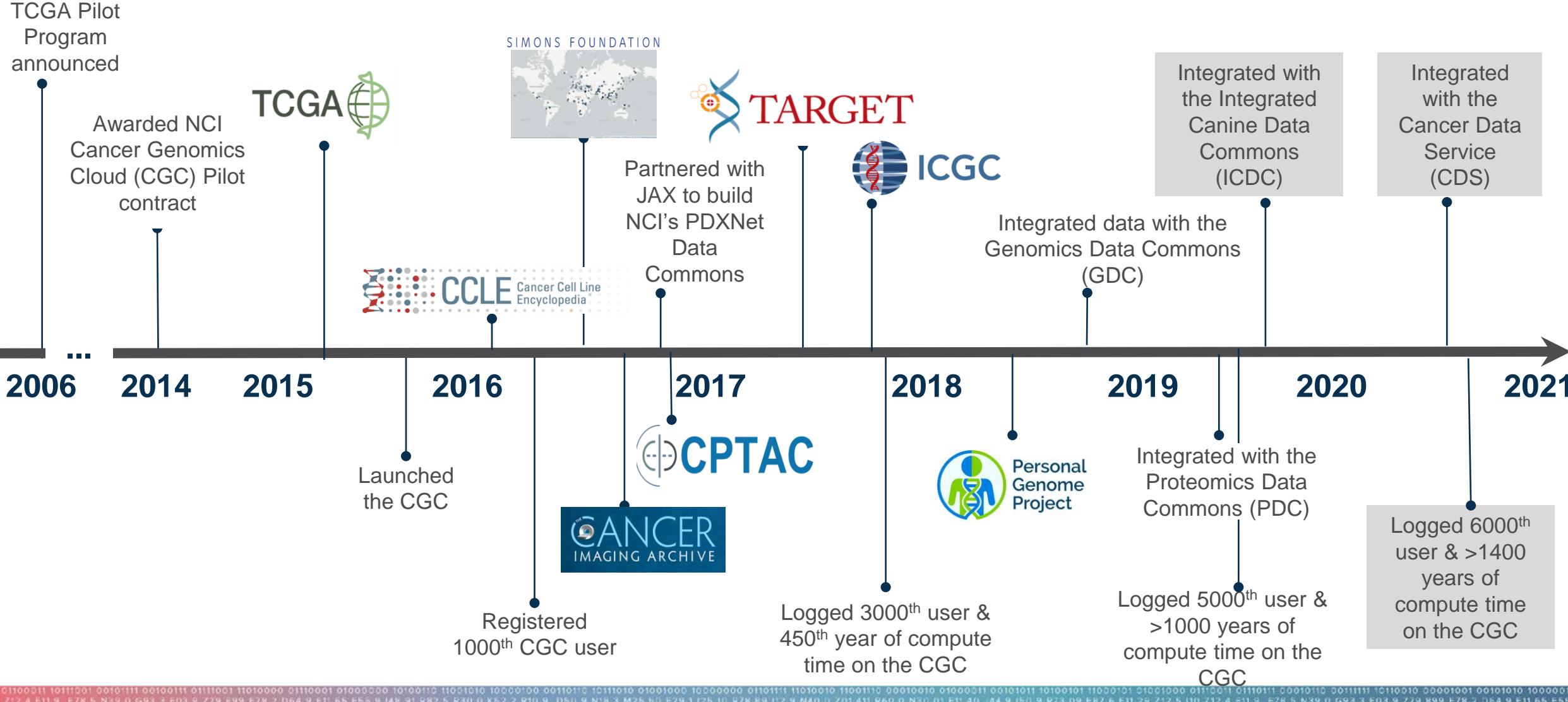
Stable, secure, and highly customizable cloud storage and computing platform

3+	760	1500	7800	80000
Petabytes Public Data	Public Tools & workflows	Years of Compute	Users	User-created workflows

NCI Cancer Research Data Commons (CRDC)



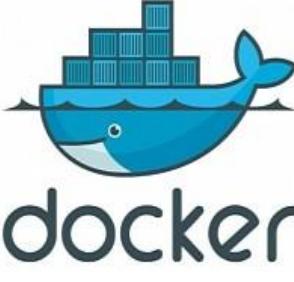
Petabytes of public data, at your fingertips



Owning our CWL Centric View of the World

Tools for Reproducibility

- Use of Docker-containerized bioinformatics pipelines to provide a consistent computational environment
- Automatically generated and accessible logs for every task



COMMON
WORKFLOW
LANGUAGE

◀ Back to tasks

COMPLETED Fusion Transcript Detection - ChimeraScan run - 01-02-16 19:51:42 ↗

Executed on Jan. 2, 2016 14:54 by arsenijae
Price: \$8.78 | Duration: 18 hours, 44 minutes
App: fusion-transcript-detection-chimerascan

Inputs	App Settings	Outputs
#... TCGA_JUNCID_2642364.b23ad2ad-d6d2-4e5a-96d7-424cd50...	Min.support 10	html_file _1_bedpe4oncofuse.txt_oncofuse.html
reference ucsc.hg19.fasta	FilterList type annotated.genes	R_workspace _2_chimeras.RData
genes human_hg19_genes_2014.gtf		oncofuse_out _2_bedpe4oncofuse.txt_oncofuse
false_positives hg19_bodymap_false_positive_chimeras.txt		chimeras_html _3_chimeras.html
		ind... _4_ucsc.hg19_human_hg19_genes_2014.genePred_index.tar...
		circos_pdf _2_circos.pdf
		filtered_fusions _2_Chimera.filtered.fusions.txt
		detected_fusions _2_Chimera.fusions.txt

- App versions & parameters
- Input & output files
- Information required to ensure your tools work reproducibly across platforms

Platform Viewpoint

- . Primary Viewpoint and Focus
 - BioCompute Object Operational Usability
 - Efficient Generation of Complex Workflows
- . Secondary Viewpoint and Focus
 - Integrate with Distributed infrastructure
 - Reproducible and Portable Computation
 - Data, Workflow, and Task Provenance
 - Integrate with BioCompute Databases

Implementation: BioCompute Templates

Microsoft Word – Google Doc - Confluence

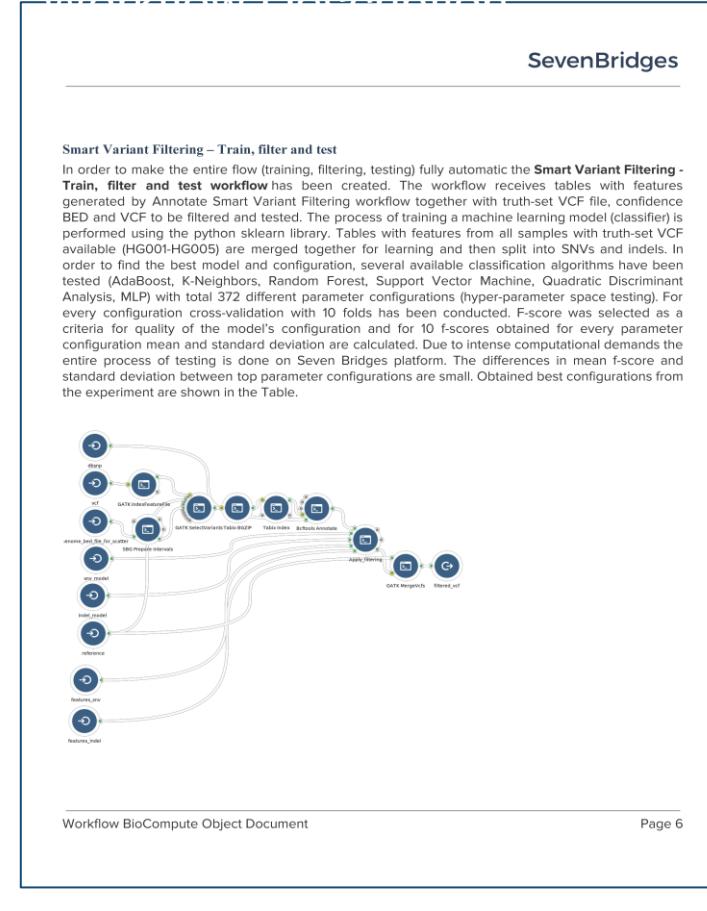
SevenBridges

Table of Contents

Project Description	4
Overview	4
Workflow Schematic	4
Apply Smart Variant Filtering Parallel	5
Smart Variant Filtering – Train, filter and test	6
Smart Variant Filtering - Prepare features	7
BioCompute Object Field Entries	8
Top Level Field	8
BioCompute Object Identifier	8
Type	8
Digital Signature	8
BCO Version	8
Provenance Domain	8
Name	8
Version	8
Structured Name	8
Review	8
Inheritance/derivation	8
Obsolescence	8
Embargo	8
Created	9
Modification	9
Contributors	9
License	9
Usability Domain	9
Usability Domain	9
Extension Domain	10
Extension Domain	10
Extension to External References:SMART on FHIR Genomics	10
Extension to External Reference:GitHub	10

Workflow BioCompute Object Document

Page 1



SevenBridges

Appendix 1: BioCompute Object Specification (ver. 1.1)

BioCompute Objects		
	ID	Description
Top Level Fields		
BioCompute Object Identifier	BCO_id	Unique identifier that should be applied to each BCO instance. Assigned by a BCO database engine, like URL. It never be reused.
Type Digital Signature	type digital_signature	As any object of the type, it has its own fields. A string-type, read-only generated and stored by a BCO database, protecting the object from internal or external alterations without proper validation. It can be used for validation, downloading, and transferring BCOs.
BCO Version	Bco_spec_version	The version of the BCO specification used to define this document.
Provenance Domain		
Name	name	Name of the BCO.
Structured name	structured_name	Computable text field designed to represent a BCO instance name in visible interfaces
Version	version	Records the versioning of this BCO instance object. A change in the BCO affecting the outcome of the computation should be deposited as a new BCO, not as a new version.
Review	review	Describes the status of an object in the review process. Status flags: unreviewed, in-review, approved, suspended, rejected.
Inheritance/derivation	derived_from	If the object is derived from another, this field will specify the parent object, in the form of the objectid. If it is null, if inherits only from the base BioCompute Object or a type definition.
Obsolescence	obsolete	If the object has an expiration date this field will specify that using the datetime type.
Embargo	embargo	If the object has a period of time that it is not public, that range can be specified using these fields. Using the datetime type a start and end time are specified for the embargo.
Created	created	Using the datetime type the time of initial creation of the BCO is recorded.
Modification	modified	Using the datetime type the time of most recent modification of the BCO is recorded.
Contributors	contributors	List to hold contributor identifiers and a description of their type of contribution, including a field for ORCIDs to record author information, as they allow for the author to curate their information after submission.

Workflow BioCompute Object Document

Page 18

Working documents/templates to support BioCompute content review and collaborations

Implementation: R Toolkits for CWL Parsing and BCO Generation

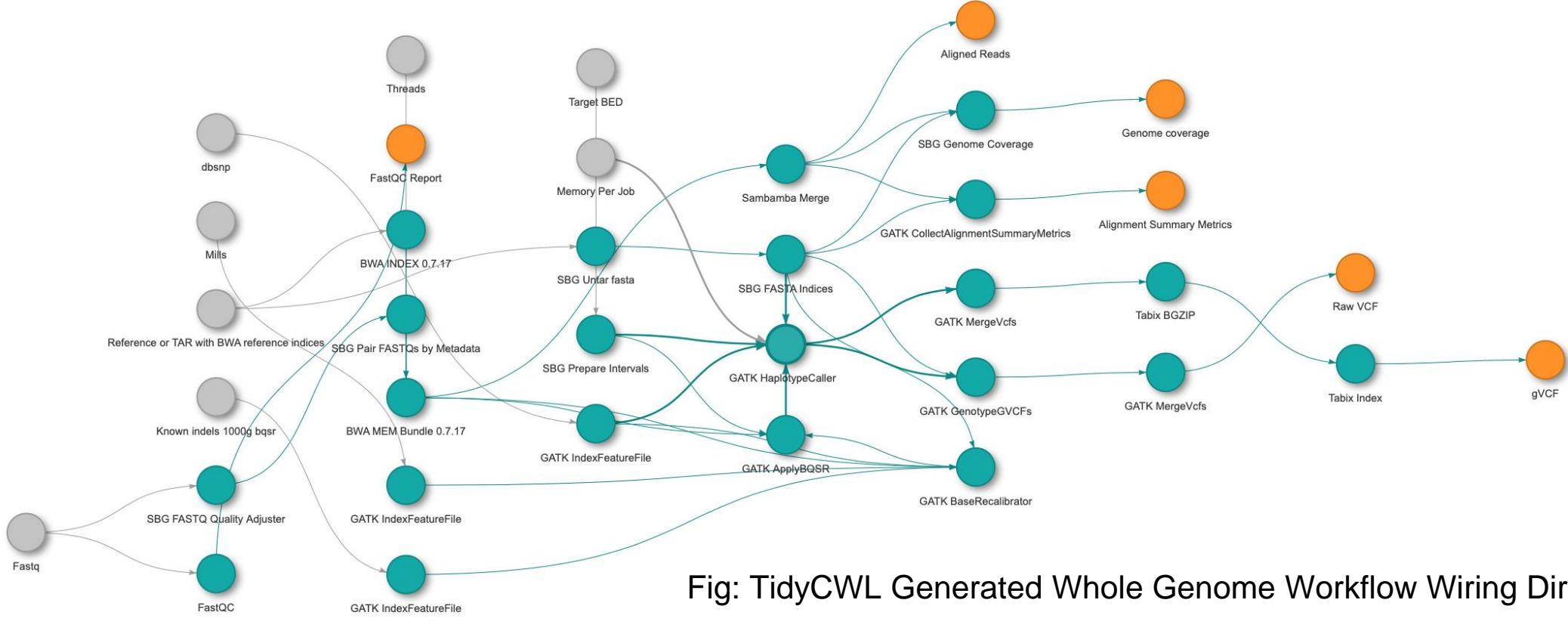


Fig: FlyCWL Generated Whole Genome Workflow Wiring Direction

Goals

- Prepopulate BioCompute Object Fields: Inputs, Outputs Parameters
 - Support Machine and Human Readable Output (JSON and PDF)

BCO APP Availability

10000000 01101111 1101001C 11001110 00010010 01000011 00101011 10100101 11000101 01001000 01110011 01110111 00010110 00111111 10110010 00001001 00101010 1000000111 1.125.10. R78.89 12.7. N4.0 Z01.411. R60.0 N30.0. E1.10. J4.49. I5.09 R73.09 E87.6 E11.29 Z12.5 I10 Z12.4 E11.9 E78.5 N39.0 G93.3 E03.9 Z79.899 E78.2 D64.9 E11.65 E55.1 SAGAGGCAUUCUAAUGUCUCUAUGGUCCUACAUUUGAACCGCAUCGUGGUCAUAGCGAUACCUAAGUACGGUGCACGGACCCA
Te I Sb Sn In Cd Ag Pd Rh Ru Te Mo Nb Zr Y Sr Br Kr Se As Ge Ga Zn Cu Co Ni Fe Mn Cr V Ti Sc Ca Ar K Cl S P Si Al Sm Pm h p ERCC6 JUN GSK3B STAT3 APEX1 ESR1 BUB1B HDAC1 MYC CAT CDKN2A NFKB1 CLU IGF1R PML CREB1 E2F1 RBL UBE2E1 IMDR3C1 SII FA Phe Cys lle Asx Trp Glx Lys Gly Arg Lys Phe Pro His Pro Tyr Glx Met Pro Cys Lys Ser Asn Ile Glu Val Arx Trp Gly Pro Arg Phe Glu Cys Gin Val Phe Thr His Pro Phe Tyr Met Cys
NAAAATCTATTCTTCAATCATGCATGGGCCACGGTAGCTTGTGGCTTAATGACGCTTTCTCTCGCCGCTTCGCGACCAGAACGACTGGGTCGGCAATCTCGAACCG

BCO APP Manuscript

- BCO App Manuscript published at F1000Research is a good starting point
 - Manuscript Include
 - Overview
 - Links to source code
 - Links to examples
 - Links to container
 - BCO Database Writing Functionality in next release
 - Pending a GIT Security Update

SOFTWARE TOOL ARTICLE

 Check for updates

BCO App: tools for generating BioCompute Objects from next-generation sequencing workflows and computations [version 1; peer review: 2 approved]

Nan Xiao  , Soner Koc  , David Roberson, Phillip Brooks  , Manisha Ray,  Dennis Dean 

 Author details

Abstract

The BioCompute Object (BCO) standard is an IEEE standard (IEEE 2791-2020) designed to facilitate the communication of next-generation sequencing data analysis with applications across academia, government agencies, and industry. For example, the Food and Drug Administration (FDA) supports the standard for regulatory submissions and includes the standard in their Data Standards Catalog for the submission of HTS data. We created the BCO App to facilitate BCO generation in a range of computational environments and, in part, to participate in the Advanced Track of the precisionFDA BioCompute Object App-a-thon. The application facilitates the generation of BCOs from both workflow metadata provided as plaintext and from workflow contents written in the Common Workflow Language. The application can also access and ingest task execution results from the Cancer Genomics Cloud (CGC), an NCI funded computational platform. Creating a BCO from a CGC task significantly reduces the time required to generate a BCO on the CGC by auto-populating workflow information fields from CGC workflow and task execution results. The BCO App supports exporting BCOs as JSON or PDF files and publishing BCOs to both the CGC platform and to GitHub repositories.

 ALL METRICS
1089
 VIEWS
106
 DOWNLOADS

 Get PDF
 Get XML
 Cite
 Export
 Track
 Email
 Share

BCO App Available Across Our Platforms

Cloud Projects Data Public Apps Public Projects Developer

Dashboard Files Apps Tasks pdx-workflows-rna ⓘ Interactive Analysis Notes

Public interactive apps maintained by Cancer Genomics Cloud

Custom interactive apps

Genome Browser
Visualize alignments, SNV/Indels, annotation tracks, check coverage and mismatch, assess alignments and variants
3 files Open

Data Cruncher
Analyze and explore data using JupyterLab or RStudio
Open

Variant Browser
Filter and interpret your annotated data
0 files Open

bco-app
BCO App provides tools for generating, reviewing, and validating BioCompute Objects (BCOs). BCO App can generate BCOs from user text input, workflows written in CWL, and from
Open

BCO App Available Across Our Platforms

 Projects ▾ Data ▾ Public Apps Public Projects ▾ Developer ▾

Dashboard Files Apps Tasks

Public interactive apps

maintained by Cancer Genomics Cloud



Genome Browser

Visualize alignments, SNV/Indels, annotation tracks, check coverage and mismatch, assess alignments and variants

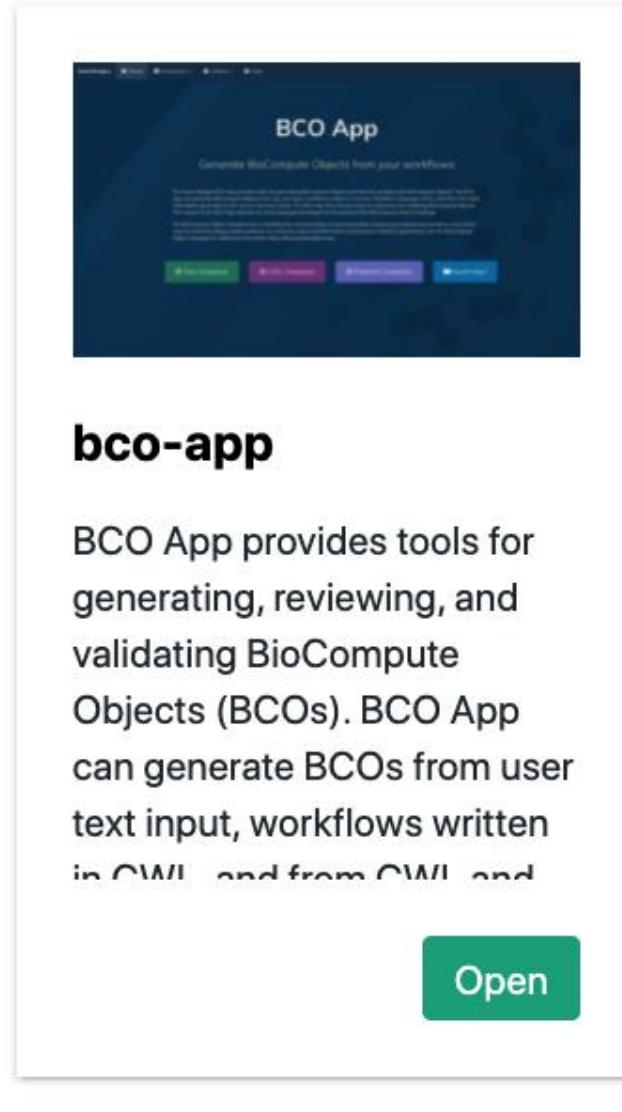
3 files

Open



Data Cruncher

Analyze and using Jupyter RStudio



Dennis

[Interactive Analysis](#) [Notes](#)

Custom interactive apps

A screenshot of the BCO App interface. The title 'BCO App' is at the top. Below it is a sub-header 'Generate BioCompute Objects from your workflows'. There is a large text area with several lines of code or text. At the bottom are four colored buttons: green, purple, blue, and orange.

bco-app

BCO App provides tools for generating, reviewing, and validating BioCompute Objects (BCOs). BCO App can generate BCOs from user text input, workflows written in CWL, and from

[Open](#)

SevenBridges®

BCO Tool Demonstration

CGC BCO Examples Project

CGC BCO Examples Project

Projects ▾ Data ▾ Public Apps ▾ Public Projects ▾ Developer ▾ Dennis ▾

Dashboard Files Apps Tasks Data Studio BCO-CWL Examples ⓘ Interactive Browsers Settings Notes

Description

Overview
Workflows used to demonstrate BCO APP functionality on the Cancer Genomics Clouds. Project includes several workflows in support of multiple BCO use cases. We assembled these workflows in support of the "BioCompute Workshop: Building BioCompute Objects on three popular Bioinformatics platforms."

BCO APP Software

Please see our BCO App Manuscript published at ResearchF1000. An upcoming BCO App release will include a BCO Data Base update feature and an SSH support for submitting BCOs to GIT pages.

- Xiao N, Koc S, Roberson D, Brooks P, Ray M, Dean D. BCO App: tools for generating BioCompute Objects from next-generation sequencing workflows and computations. F1000Res. 20

Workflow Examples

- PDX RNS Expression Estimation Workflow
- RNA Expression Workflow for Patient Tumor
- GATK Best Practice Data Pre-Processing
- Differential Expression - Salmon DESeq2
- Neopeptope Analysis
- RNA Seq

References

- Patel, J, Dean II, DA, King, CH, Keeney, J, Xiao, N, Koc, S, Minina, E, Golikov, A, Brooks, P, Kahsay, R, Ray, M, Roberson, D, Mazumder, R, Development and validation of novel tools for BioCompute Object Creation, Database, 2021, <https://doi.org/10.1093/database/baab008>
- Stephens, SH, King, H, Watford, S, Patel, J, Dean II, DA, Koc, S, Xiao, N, Donaldson, E, Crusoe, M, Thompson, E, Purkayastha, A, Mazumder, R, Johanson, E, Keeney, J, Strengthening the BioCompute Standard by Crowdsourcing on PrecisionFDA, 2020, bioRxiv
- Xiao, N, Koc, S, David Roberson, D, Brooks, P, Manisha Ray, M, Dean II, DA, BCO App: Tools for Generating BioCompute Objects from Next-Generation Sequencing Workflows and Computations, 2020, F1000 Research, (<https://doi.org/10.12688/f1000research.25902.1>)
- Alterovitz, G, Dean II, DA, Goble, C, Crusoe, M, Soiland-Reyes, S, Bell, S, Hayes, A, King, CH, Johanson, E, Thompson, EE, Donaldson, E, Tsang, HS, Goecks, J, Yao, J, Almeida, JS, Guo, L, Walderhaug, M, Walsh, P, Kahsay, R, Bloom, T, Lai, Y, Simonyan, V, Mazumder, R, Enabling Precision Medicine via Standard Communication of NGS Provenance, Analysis, and Results, PLOS Biology, 2018. PMID: PMC6338479. 20 Sep 16;9:1144. doi: 10.12688/f1000research.25902.1. PMID: 33299553; PMCID: PMC7702177.

Members

phil_webster OWNER Copy, Write, Execute, Admin
Dennis ADMIN Copy, Write, Execute, Admin
soner ADMIN Copy, Write, Execute, Admin
saul_acevedo ADMIN Copy, Write, Execute, Admin

Manage members Leave project

Analysis

Search

Your executions will appear here. Before you start, learn more about them.

* Planning to make a CGC Public Project

CGC BCO Examples

Projects ▾ Data ▾ Public Apps ▾ Public Projects ▾ Developer ▾ Dennis ▾

Dashboard Files Apps Tasks Data Studio BCO-CWL Examples 0 Interactive Browsers Settings Notes

Description Tags Members Email notifications

Overview

Workflows used to demonstrate BCO APP functionality on the Cancer Genomics Cloud cases. We assembled these workflows in support of the "BioCompute Workshop: Building BCO APP Software"

Please see our BCO App Manuscript published at ResearchF1000. An upcoming BCO support for submitting BCOs to GIT pages.

- Xiao N, Koc S, Roberson D, Brooks P, Ray M, Dean D. BCO App: tools for generation and computations. F1000Res. 2020; 9:100.

BCO APP Software

Workflow Examples

- PDX RNS Expression Estimation Workflow
- RNA Expression Workflow for Patient Tumor
- GATK Best Practice Data Pre-Processing
- Differential Expression - Salmon DESeq2
- Neoepitope Analysis
- RNA Seq

References

- Patel, J, Dean II, DA, King, CH, Keeney, J, Xiao, N, Koc, S, Minina, E, Golikov, A, et al. Development and validation of novel tools for BioCompute Object Creation, Data Integration, and Standardization. bioRxiv preprint doi: https://doi.org/10.1101/2020.06.10.212446; 2020.
- Stephens, SH, King, H, Watford, S, Patel, J, Dean II, DA, Koc, S, Xiao, N, Donaldson, E, Crusoe, M, Thompson, L, Purkayastha, A, Mazumder, R, Johansson, E, Keeney, J. Strengthening the BioCompute Standard by Crowdsourcing on PrecisionFDA, 2020, bioRxiv
- Xiao, N, Koc, S, David Roberson, D, Brooks, P, Manisha Ray, M, Dean II, DA, BCO App: Tools for Generating BioCompute Objects from Next-Generation Sequencing Workflows and Computations, 2020, F1000 Research, (<https://doi.org/10.12688/f1000research.25902.1>)
- Alterovitz, G, Dean II, DA, Goble, C, Crusoe, M, Soiland-Reyes, S, Bell, S, Hayes, A, King, CH, Johansson, E, Thompson, EE, Donaldson, E, Tsang, HS, Goecks, J, Yao, J, Almeida, JS, Guo, L, Walderhaug, M, Walsh, P, Kahsay, R, Bloom, T, Lai, Y, Simonyan, V, Mazumder, R. Enabling Precision Medicine via Standard Communication of NGS Provenance, Analysis, and Results, PLOS Biology, 2018. PMID: PMC6338479. 20 Sep 16:9:1144. doi: 10.12688/f1000research.25902.1. PMID: 33299553; PMCID: PMC7702177.

Workflow Examples

Before you start, learn more about them.

* Planning to make a CGC Public Project

BCO Tool Demonstration

Workflow Example



BCO App

Generate BioCompute Objects from your workflows

The Seven Bridges BCO App provides tools for generating BioCompute Objects and tools for working with BioCompute Objects. The BCO App can generate BioCompute Objects from user text input, workflows written in Common Workflow Language (CWL), and from CWL/task information generated on the Cancer Genomics Cloud. The BCO App also includes tools for reviewing and validating BioCompute Objects. This version of the BCO App extends our initial prototype developed for the precisionFDA BioCompute Object Challenge.

The BioCompute Object standard aims to facilitate the communication of Next Generation Sequencing Analyses by providing a structured way for communicating complex analyses as commonly seen in bioinformatics and precision medicine applications. See the BioCompute Object webpage for additional information (<https://biocomputeobject.org>).

Text Composer

CWL Composer

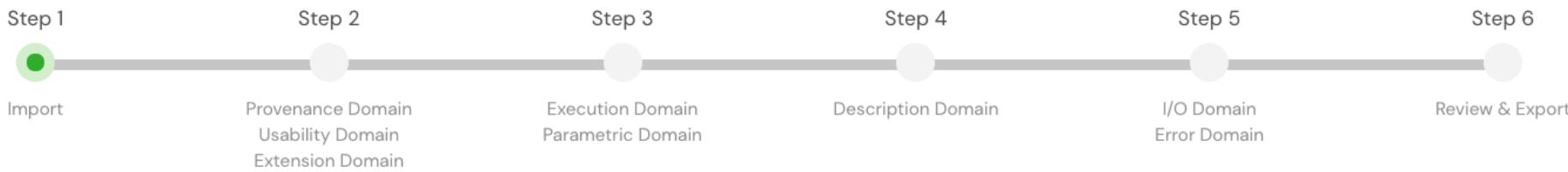
Platform Composer

Need Help?

precisionFDA

BioCompute Object App-a-thon
Advance Track Winner

Step 1/6 - Import



Import from Seven Bridges Platforms

Platform

Cancer Genomics Cloud (CGC)

Paste the Auth Token from the [CGC Developer Dashboard](#)

.....

Choose Project

soner/pdx-workflows-rna

Choose Task in Project (Optional – If Exist)

Choose App to Import CWL Workflow

rna-seq-alignment-star-2-5-4b-cwl-1-0

Previous

Next

Step 1/6 - Import

Step 1



Import

Step 2



Step 3



Step 4



Step 5



Step 6



Review & Export

Import from

Platform

Cancer Genomics

Paste the Auth Token

.....

Choose Project

soner/pdx-workflows-rna

Choose App to Import CWL Workflow

rna-seq-alignment-star-2-5-4b-cwl-1-0

Platform

Cancer Genomics Cloud (CGC)

Paste the Auth Token from the CGC Developer Dashboard

.....

Choose Project

soner/pdx-workflows-rna

Choose App to Import CWL Workflow

rna-seq-alignment-star-2-5-4b-cwl-1-0

⊕ Previous

⊕ Next

SevenBridges Home Generators Utilities Help

Step 2/6 - Provenance / Usability / Extension

Step 1 Step 2 Step 3 Step 4 Step 5 Step 6

Import Provenance Domain Usability Domain Extension Domain Execution Domain Parametric Domain Description Domain I/O Domain Error Domain Review & Export

Workflow Visualization

Legend

- input
- output
- start
- end

1. Provenance Domain

Name for the BCO: RNA Expression Estimation Workflow for Patient Tumor

Version of the BCO instance object: 1.0.0

Inheritance or derivation from: <https://cgc-api.sbggenomics.com/v2/apps/soner/pdx-workflows-ma/ma-expression-estimation-workflow-for-patient-tumor/0/raw/>

License: <https://spdx.org/licenses/CC-BY-4.0.html>

BCO initial creation date: 2021-07-27

BCO modification date:

SevenBridges

- [Home](#)
- [Generators](#) ▾
- [Utilities](#) ▾
- [Help](#)

Step 2/6 - Provenance / Usability / Extension

Step 1 Step 2 Step 3 Step 4 Step 5 Step 6

Import Provenance Domain
Usability Domain
Extension Domain Execution Domain
Parametric Domain Description Domain
I/O Domain
Error Domain Review & Export

Workflow Visualization

Workflow Visualization

```

graph LR
    A[RNASeqMetrics] --> B[Forward probability]
    A --> C[Picard CollectRNaseqMetrics report]
    B --> D[BAM in transcript coordinates]
    C --> E[Integrated QC report]
    E --> F[FastQC]
    F --> G[FASTQC HTML report]
    F --> H[FASTQC reports ZIP archive]
    F --> I[RSEM Plot Model]
    F --> J[RSEM model plot]
    F --> K[metadata_annot]
    F --> L[perc_correct_strand]
    F --> M[perc_useable_bases]
    G --> N[Somaller Sites File]
    H --> O[FASTA Reference]
    I --> P[RSEM model plot]
    J --> Q[Somaller Extract]
    K --> R[Isotom level expression estimates]
    L --> S[Somaller Extract File]
    M --> T[QC Integrate RNA - no xonome]
    N --> U[SBG Decompressor]
    O --> V[SBG Decompressor]
    P --> W[SBG Decompressor]
    Q --> X[SBG Decompressor]
    R --> Y[SBG Decompressor]
    S --> Z[SBG Decompressor]
    T --> A
    U --> C
    V --> C
    W --> C
    X --> C
    Y --> C
    Z --> C
  
```

Legend

- input (orange circle)
- output (blue circle)
- step (green circle)

2021-07-27

BCO modification date

Step 2/6 - Provenance / Usability / Extension



Workflow Visualization



1. Provenance Domain

Name for the BCO

RNA Expression Estimation Workflow for Patient Tumor

Version of the BCO instance object

10

Inheritance or derivation from

<https://cgc-api.sbggenomics.com/v2/apps/soner/pdx-workflows-rna/rna-expression-estimation-workflow-for-patient-tumor/Orai>

✓

License

<https://spdx.org/licenses/CC-BY-4.0.html>

	https://cgc-api.sanger.ac.uk/v2/apps/somer/pdx-workflows-ma/ma-expression-estimation-workflow-for-patient-tumor/Oj/raw	
License	https://spdx.org/licenses/CC-BY-4.0.html	
BCO initial creation date	2021-07-27	
BCO modification date		

Step 3/6 - Execution / Parametric



4. Execution Domain

Script

<https://cgc-api.sbggenomics.com/v2/apps/soner/pdx-workflows-rna/rna-expression-estimation-workflow-for-patient-tumor/O/raw/>



Script driver

Seven Bridges Common Workflow Language Executor



Algorithmic tools and software prerequisites

New Edit Delete Copy

Show 10 entries

Search:

name	version	uri	access_time	sha1_chksum
Seven Bridges Platform	2021-07-27	https://igor.sbggenomics.com/	2021-07-27	

Showing 1 to 1 of 1 entries Previous 1 Next

External data endpoints

New Edit Delete Copy

Show 10 entries

Search:

name	url
No data available in table	

Showing 0 to 0 of 0 entries Previous Next

Step 3/6 - Execution / Parametric



4. Execution Domain

4. Execution Domain

Script

<https://cgc-api.sbggenomics.com/v2/apps/soner/pdx-workflows-rna/rna-expression-estimation-workflow-for-patient-tumor/O/raw>



Script driver

Seven Bridges Common Workflow Language Executor



Algorithmic tools and software prerequisites

New | Edit

Edit

Delete

Copy

Show 10 entries

Search:

1

Page 1

No data available in table

Step 4/6 - Description



6. Description Domain

step_number	name	description	version
1 1	#SBG_Genome_Coverage	SBG Genome Coverage extends BEDTools Genome Coverage. The Genome Coverage calculates histograms, per-base reports and BedGraph summaries of feature coverage (aligned sequences for example) for a given genome. This extended version additionally extracts and creates a text file containing summary coverage stats. Note: This tool should be used for genome data only.	
2 2	#SBG_Untar_fasta	SBG Untar fasta is primarily used for extracting FASTA file from TAR with more different files. For example, extracting FASTA from BWA-MEM's index TAR. **Inputs** TAR (containing FASTA). **Outputs** FA/FASTA/FA.GZ/FASTA.GZ from TAR. Important: TAR.GZ format is not supported. If only decompress operation is required use SBG Decompressor tool.	v1.0
3 3	#Sambamba_Merge	Sambamba Merge is used for merging several sorted BAM files into one. The sorting order of all the files must be the same, and it is maintained in the output file.	0.5.9
4 4	#SBG_FASTQ_Quality_Adjuster	This app detects quality score format used in input FASTQ file. FASTQ quality score is then converted to standard Sanger quality score if conversion is required. It is basically a compact merged version of "SBG Fastq Quality Detector" and "SBG Fastq Quality Converter", created to speed up the execution of pipelines. Supported source formats are: Solexa, Illumina 1.3, Illumina 1.5 and Illumina 1.8.	
Tabix indexes a TAB-delimited genome position file in.tab.bgz and creates an index file in.tab.bgz.tbi when region is absent from the command-line. The input data file must be position sorted and compressed by bgzip which has a gzip(1) like interface. After			

2 2	#SBG_Untar_fasta	SBG Untar fasta is primarily used for extracting FASTA file from TAR with more different files. For example, extracting FASTA from BWA-MEM's index TAR. **Inputs** TAR (containing FASTA). **Outputs** FA/FASTA/FA.GZ/FASTA.GZ from TAR. Important: TAR.GZ format is not supported. If only decompress operation is required use SBG Decompressor tool.	v1.0
3 3	#Sambamba_Merge	Sambamba Merge is used for merging several sorted BAM files into one. The sorting order of all the files must be the same, and it is maintained in the output file.	0.5.9
4 4	#SBG_FASTQ_Quality_Adjuster	This app detects quality score format used in input FASTQ file. FASTQ quality score is then converted to standard Sanger quality score if conversion is required. It is basically a compact merged version of "SBG Fastq Quality Detector" and "SBG Fastq Quality Converter", created to speed up the execution of pipelines. Supported source formats are: Solexa, Illumina 1.3, Illumina 1.5 and Illumina 1.8.	
Tabix indexes a TAB-delimited genome position file in.tab.bgz and creates an index file in.tab.bgz.tbi when region is absent from the command-line. The input data file must be position sorted and compressed by bgzip which has a gzip(1) like interface. After			

Step 5/6 - IO / Error



7. Input/Output Domain

Input subdomain

[New](#) [Edit](#) [Delete](#) [Copy](#)

Show 10 entries

Search:

filename

uri

access_time

No data available in table

Showing 0 to 0 of 0 entries

[Previous](#) [Next](#)

Output subdomain

[New](#) [Edit](#) [Delete](#) [Copy](#)

Show 10 entries

Search:

mediatype

uri

access_time

No data available in table

Showing 0 to 0 of 0 entries

[Previous](#) [Next](#)

8. Error Domain

Empirical error subdomain

[New](#) [Edit](#) [Delete](#) [Copy](#)

Show 10 entries

Search:

key

value

No data available in table

Showing 0 to 0 of 0 entries

[Previous](#) [Next](#)

Step 6/6 - Review and Export



Top Level Fields

BCO ID

<https://biocompute.sbggenomics.com/bco/58ffb98f-2e7d-43fd-9e2c-2fabebb9f686>



Review & Export

► Generate & Preview BCO

```
3 *   "provenance_domain": {  
4 *     "name": "POX RNA Expression Estimation Workflow",  
5 *     "version": "1.0.7",  
6 *     "review": [],  
7 *     "derived_from": "https://cgc-api.sbggenomics.com/v2/apps/pdxnet/pdx-wf-commit2/pdx-rna-expression-estimation-workflow/7/raw/",  
8 *     "obsolete_after": "2022-09-13T00:00:00+0000",  
9 *     "embargo": ["2022-09-13T00:00:00+0000", "2022-09-13T00:00:00+0000"],  
10 *    "created": "2022-09-13T00:00:00+0000",  
11 *    "modified": "2022-09-13T00:00:00+0000",  
12 *    "contributors": [],  
13 *    "license": "https://spdx.org/licenses/CC-BY-4.0.html"  
14 *  },  
15 *  "usability_domain": "This RSEM workflow (RSEM 1.2.31) for quantifying gene expression uses the STAR aligner and is optimized to work with FASTQ input files.\n\nTo process multiple samples, please consider running batch tasks with this workflow.",  
16 *  "extension_domain": {  
17 *    "fhir_extension": {  
18 *      "fhir_endpoint": "",  
19 *      "fhir_version": "",  
20 *      "fhir_resources": {}  
21 *    },  
22 *    "scm_extension": {  
23 *      "scm_repository": "",  
24 *      "scm_type": "git",  
25 *    }  
26 *  },  
27 *}
```

Export as JSON/PDF

Export as JSON

Export as PDF

Save to CGC Platform, BCO Database, or GitHub Project

Push to GitHub

Upload to biocomputeobject.org

Upload to Platform

Step 6/6 - Review and Export



Top Level Fields

BCO ID

<https://biocompute.sqgenomics.com/bco/afcf3c33-b097-44d0-a516-79303a8d2124>

 Export as JSON

 Export as PDF

[Save to CGC Platform, BCO Database, or GitHub Project](#)

 Push to GitHub

 Upload to biocomputeobject.org

 Upload to Platform

Export as JSON/PDF

 Export as JSON

 Export as PDF

Save to Platform or GitHub Project

 Push to GitHub

Top Level Fields

BCO ID

<https://biocompute.sbggenomics.com/bco/58ffb98f-2e7d-43fd-9e2c-2fabeb9>

Review & Export

```
1 + [
2   "spec_version": "https://w3id.org/biocompute/1.4.2",
3   "object_id": "https://biocompute.sbggenomics.com/bco/58ffb98f-2e7d-43fd-9e2c-2
4   "etag": "362e8707c456f3e8fcfa87f1b99826e616366eb4ec99603368707a002c194607c",
5   "provenance_domain": {
6     "name": "PDX RNA Expression Estimation Workflow",
7     "version": "1.0.7",
8     "review": [],
9     "derived_from": "https://cgc-api.sbggenomics.com/v2/apps/pdxnet/pdx-wf-commi
10    "obsolete_after": "2022-09-13T00:00:00000",
11    "embargo": ["2022-09-13T00:00:00000", "2022-09-13T00:00:00+00000"],
12    "created": "2022-09-13T00:00:00000",
13    "modified": "2022-09-13T00:00:00000",
14    "contributors": [],
15    "license": "https://spdx.org/licenses/CC-BY-4.0.html"
16  },
17  "usability_domain": "This RSEM workflow (RSEM 1.2.31) for quantifying gene ex
18  "extension_domain": {
19    "fhir_extension": {
20      "fhir_endpoint": "",
21      "fhir_version": "",
22      "fhir_resource": ""
23    }
24  }
```

Export as JSON/PDF

Export as JSON

Save to CGC Platform, BCO Database, or GitHub Project

Push to GitHub

Upload to biocomputeobject.org

Push your generated BCO file to biocomputeobject.org

Note: we don't record any of your account information.

Token:

.....

BCO Prefix:

BCO

BCO Schema:

IEEE

BCO Owner Group:

bco_drafter

BCO return message:

Successful!

Pushed to the biocomputeobject.org DB with Status
Code: 200 !

Push

Open BioCompute DB

Close

Export as PDF

Upload to Platform

BioCompute Object (BCO) Domains

All Domains Mandatory Domains

Top Level Fields

Provenance Domain

Usability Domain

Extension Domains

Description Domai

Execution Domai

Parametric Domain

I/O Domain

Error Domain

BCO Execution Domain

ID	Description
Execution Domain	
Execution Domain	execution_domain The fields required for execution of the BCO have been encapsulated together in order to clearly separate information needed for deployment, software configuration, and running applications in a dependent environment.
Script	script Points to an internal or external reference to a script object that was used to perform computations for this BCO instance. This may be reference to Galaxy Project or Seven Bridges Genomics pipeline, a Common Workflow Language (CWL) object in GitHub, HIVE computational service or any other type of script.
Script Driver	script_driver The reference to an executable that can be launched in order to perform a sequence of commands described in the script. For example if the pipeline is driven by a HIVE script, the script driver is the hive execution engine. For CWL based scripts specify cwl-runner. Another very general commonly used in Linux based operating systems is shell.
Algorithmic tools and Software Prerequisites	software_prerequisites Field listing the minimal necessary prerequisites, library, tool versions needed to successfully run the script to produce BCO.
External Data Endpoints	external_data_endpoints An optional multi-value field listing the minimal necessary domain specific external data source access in order to successfully run the script to produce BCO.
Environmental Variables	environment_variables This is an array of key-value pairs useful to configure the execution environment on the target platform.

BCO Tool Demonstration

Task Example

Generating a BioCompute Object for Task Information

SevenBridges Home BCO Composer BCO Validator Help Sign out

Step 1/6 - Import

Step 1 Step 2 Step 3 Step 4 Step 5 Step 6

Import Provenance Domain Usability Domain Extension Domain Execution Domain Parametric Domain Description Domain I/O Domain Error Domain Review & Export

Import from Seven Bridges Platforms

Create new BCO from a CWL workflow Edit existing BCO

Platform SBPLA (US)

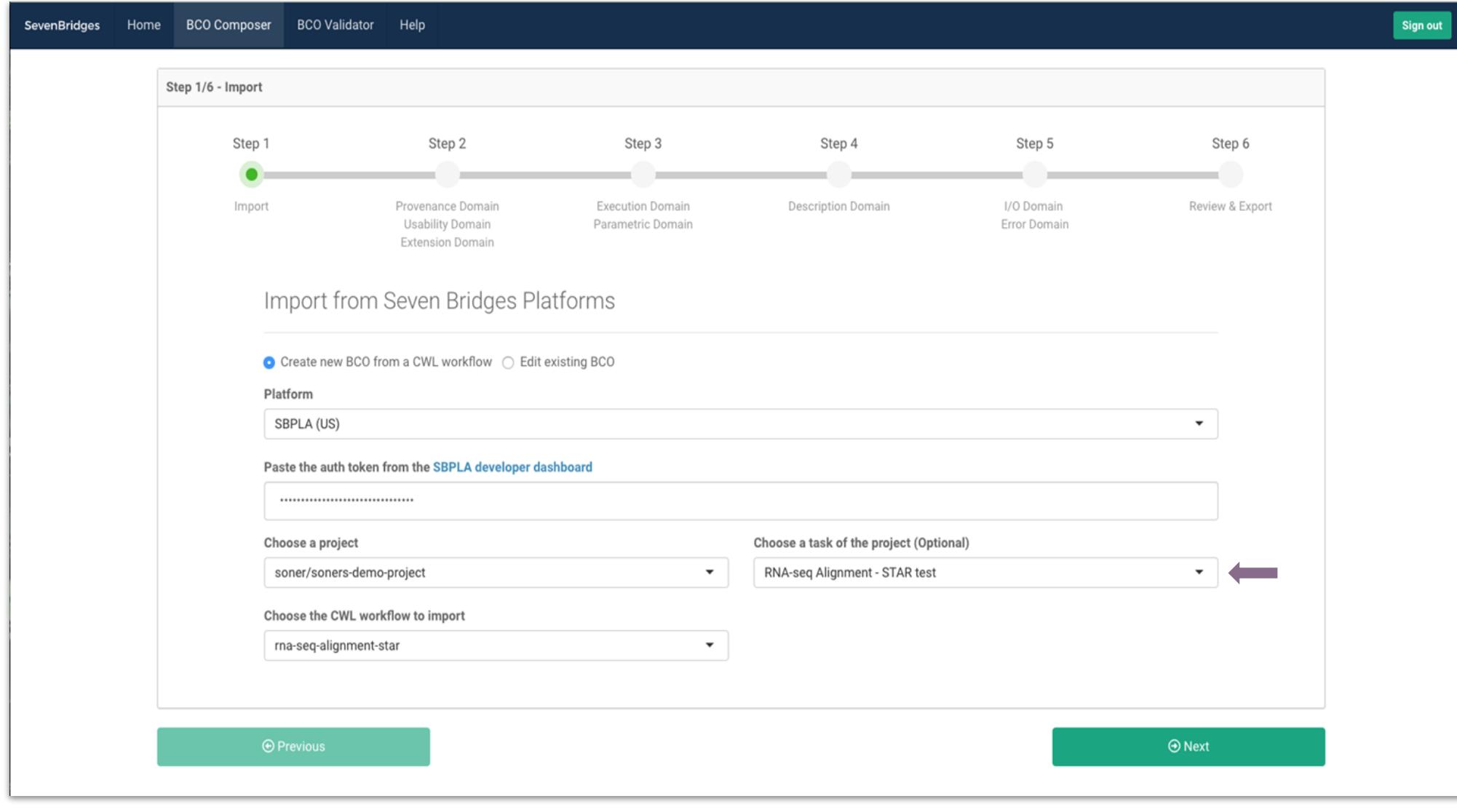
Paste the auth token from the [SBPLA developer dashboard](#)

Choose a project soner/soners-demo-project

Choose a task of the project (Optional) RNA-seq Alignment - STAR test

Choose the CWL workflow to import rna-seq-alignment-star

Previous Next



Generating a BioCompute Object for Task Information

SevenBridges Home BCO Composer BCO Validator Help Sign out

Step 1/6 - Import

Step 1 Step 2 Step 3 Step 4 Step 5 Step 6

Import Provenance Domain Usability Domain Extension Domain Execution Domain Parametric Domain Description Domain I/O Domain Error Domain Review & Export

Import from Seven Bridges Platforms

Choose a task of the project (Optional)

RNA-seq Alignment - STAR test

Choose a project

soner/soners-demo-project

Choose a task of the project (Optional)

RNA-seq Alignment - STAR test

Choose the CWL workflow to import

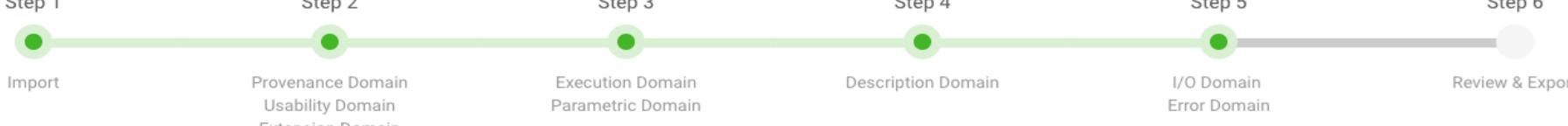
rna-seq-alignment-star

Previous Next



Archive a Task (Run) with a BioCompute Object

Step 5/6 - IO / Error



Step 1 Import Step 2 Provenance Domain
Usability Domain
Extension Domain Step 3 Execution Domain
Parametric Domain Step 4 Description Domain Step 5 I/O Domain
Error Domain Step 6 Review & Export

7. Input/Output Domain

Input subdomain

Show 10 entries Search:

filename	uri	access_time
1 chr20.fa	https://igor.sbggenomics.com/u/soner/soners-demo-project/files/58063598e4b0c58b0f206692	2016-10-18T14:45:49Z
2 Sample2_RNASeq_chr20.pe_2.fastq	https://igor.sbggenomics.com/u/soner/soners-demo-project/files/58063598e4b0c58b0f20658e	2016-10-18T14:45:49Z
3 Sample2_RNASeq_chr20.pe_1.fastq	https://igor.sbggenomics.com/u/soner/soners-demo-project/files/58063598e4b0c58b0f2065c6	2016-10-18T14:45:49Z
4 chr20_annotations.gtf	https://igor.sbggenomics.com/u/soner/soners-demo-project/files/58063598e4b0c58b0f2065e3	2016-10-18T14:45:49Z

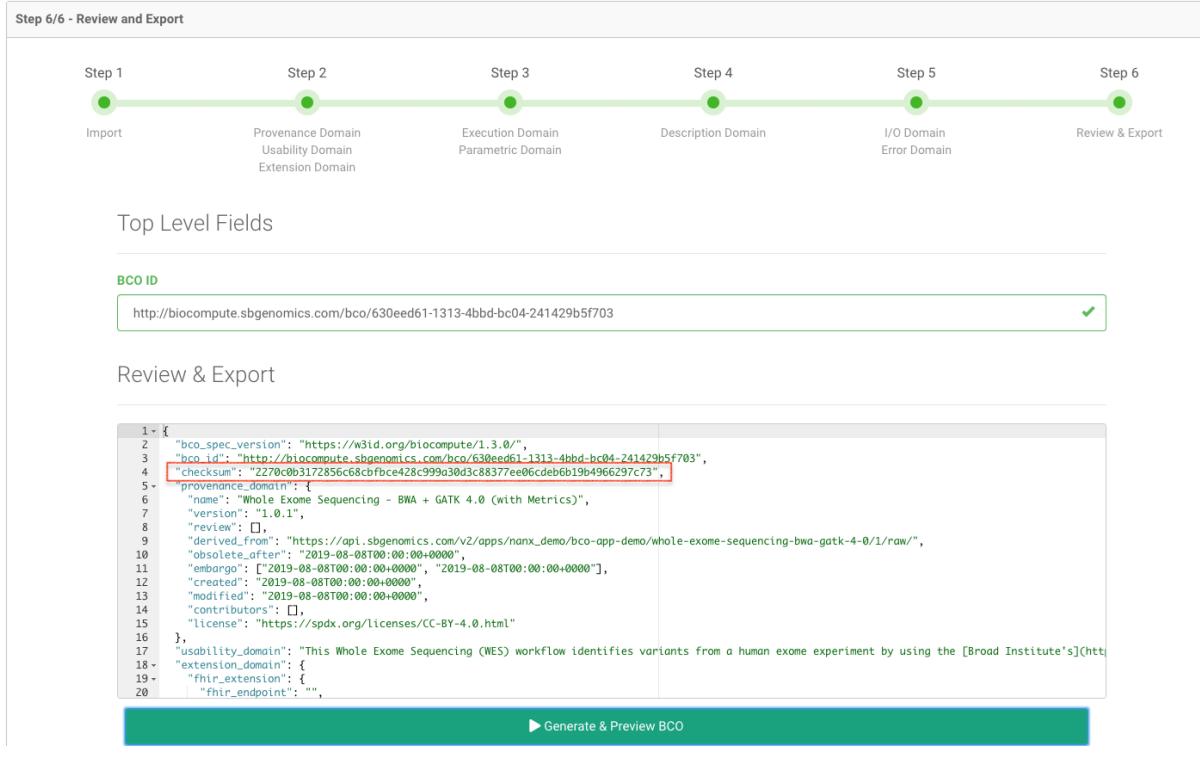
Showing 1 to 4 of 4 entries Previous 1 Next

Output subdomain

Show 10 entries Search:

mediatype	uri	access_time
1 out	https://igor.sbggenomics.com/u/soner/soners-demo-project/files/58063598e4b0c58b0f2065f6	2016-10-18T14:45:49Z
2 out	https://igor.sbggenomics.com/u/soner/soners-demo-project/files/58063598e4b0c58b0f206610	2016-10-18T14:45:49Z
3 out	https://igor.sbggenomics.com/u/soner/soners-demo-project/files/58063598e4b0c58b0f20667c	2016-10-18T14:45:49Z
4 fastq	https://igor.sbggenomics.com/u/soner/soners-demo-project/files/58063598e4b0c58b0f2065d9	2016-10-18T14:45:49Z

Check Sum Computed



BioCompute Object Generation Window

CONFIDENTIAL

SevenBridges

5 Appendix 2: The Complete BioCompute Object

```

{
  "bco_spec_version": "https://w3id.org/biocompute/1.3.0/",
  "bco_id": "http://biocompute.sbggenomics.com/bco/f8489c29-d7ed-477e-85c8-2dc0b7832041",
  "checksum": "c85a8cb045244513ec53323e20f2f002e0fc4d6448244f4124ef4d7f913456cd",
  "provenance_domain": {
    "name": "RNA-seq Alignment - STAR",
    "version": "1.0.0",
    "review": [],
    "derived_from": "https://api.sbggenomics.com/v2/apps/soner/soners-demo-project/rna-seq-align",
    "obsolete_after": "2019-08-08T00:00:00+0000",
    "embargo": ["2019-08-08T00:00:00+0000", "2019-08-08T00:00:00+0000"],
    "created": "2019-08-08T00:00:00+0000",
    "modified": "2019-08-08T00:00:00+0000",
    "contributors": [],
    "license": "https://spdx.org/licenses/CC-BY-4.0.html"
  },
  "usability_domain": "Alignment to a reference genome and transcriptome presents the first step of
  "extension_domain": {
    "fhir_extension": {
      "fhir_endpoint": "",
      "fhir_version": "",
      "fhir_resources": {}
    },
    "scm_extension": {
      "scm_repository": "",
      "scm_type": "git",
      "scm_commit": "",
      "scm_path": "",
      "scm_preview": ""
    }
  }
}

```

BioCompute Object – Generated by Biocore

GCCU
Ip HS
TP53

SevenBridges®

Check Sum Computed

Step 6/6 - Review and Export

Step 1 Step 2 Step 3 Step 4 Step 5 Step 6

Import Provenance Domain Usability Domain Extension Domain Execution Domain Parametric Domain Description Domain I/O Domain Error Domain Review & Export

Top Level Fields

```
1 {  
2   "bco_spec_version": "https://w3id.org/biocompute/1.3.0/",  
3   "bco_id": "http://biocompute.sbggenomics.com/bco/630eed61-1313-4bbd-bc04-241429b5f703",  
4   "checksum": "2270c0b3172856c68cbfbce428c999a30d3c88377ee06cdeb6b19b4966297c73",  
5   "provenance_domain": {  
6     "name": "Whole Exome Sequencing - BWA + GATK 4.0 (with Metrics)",  
7     "version": "1.0.1",  
8     "review": □.  
9     "derived_from": "https://api.sbggenomics.com/v2/apps/nanx_demo/bco-app-demo/whole-exome-sequencing-bwa-gatk-4-0/1/raw/",  
10    "obsolete_after": "2019-08-08T00:00:00+0000",  
11    "embargo": ["2019-08-08T00:00:00+0000", "2019-08-08T00:00:00+0000"],  
12    "created": "2019-08-08T00:00:00+0000",  
13    "modified": "2019-08-08T00:00:00+0000",  
14    "contributors": □,  
15    "license": "https://spdx.org/licenses/CC-BY-4.0.html"  
16  },  
17  "usability_domain": "This Whole Exome Sequencing (WES) workflow identifies variants from a human exome experiment by using the [Broad Institute's](  
18  "extension_domain": {  
19  "fhir_extension": {  
20    "fhir_endpoint": "",  
21  }  
22  }  
23  }  
24  }  
25  }  
26  }  
27  }  
28  }  
29  }  
30  }  
31  }  
32  }  
33  }  
34  }  
35  }  
36  }  
37  }  
38  }  
39  }  
40  }  
41  }  
42  }  
43  }  
44  }  
45  }  
46  }  
47  }  
48  }  
49  }  
50  }  
51  }  
52  }  
53  }  
54  }  
55  }  
56  }  
57  }  
58  }  
59  }  
60  }  
61  }  
62  }  
63  }  
64  }  
65  }  
66  }  
67  }  
68  }  
69  }  
70  }  
71  }  
72  }  
73  }  
74  }  
75  }  
76  }  
77  }  
78  }  
79  }  
80  }  
81  }  
82  }  
83  }  
84  }  
85  }  
86  }  
87  }  
88  }  
89  }  
90  }  
91  }  
92  }  
93  }  
94  }  
95  }  
96  }  
97  }  
98  }  
99  }  
100  }  
101  }  
102  }  
103  }  
104  }  
105  }  
106  }  
107  }  
108  }  
109  }  
110  }  
111  }  
112  }  
113  }  
114  }  
115  }  
116  }  
117  }  
118  }  
119  }  
120  }  
121  }  
122  }  
123  }  
124  }  
125  }  
126  }  
127  }  
128  }  
129  }  
130  }  
131  }  
132  }  
133  }  
134  }  
135  }  
136  }  
137  }  
138  }  
139  }  
140  }  
141  }  
142  }  
143  }  
144  }  
145  }  
146  }  
147  }  
148  }  
149  }  
150  }  
151  }  
152  }  
153  }  
154  }  
155  }  
156  }  
157  }  
158  }  
159  }  
160  }  
161  }  
162  }  
163  }  
164  }  
165  }  
166  }  
167  }  
168  }  
169  }  
170  }  
171  }  
172  }  
173  }  
174  }  
175  }  
176  }  
177  }  
178  }  
179  }  
180  }  
181  }  
182  }  
183  }  
184  }  
185  }  
186  }  
187  }  
188  }  
189  }  
190  }  
191  }  
192  }  
193  }  
194  }  
195  }  
196  }  
197  }  
198  }  
199  }  
200  }  
201  }  
202  }  
203  }  
204  }  
205  }  
206  }  
207  }  
208  }  
209  }  
210  }  
211  }  
212  }  
213  }  
214  }  
215  }  
216  }  
217  }  
218  }  
219  }  
220  }  
221  }  
222  }  
223  }  
224  }  
225  }  
226  }  
227  }  
228  }  
229  }  
230  }  
231  }  
232  }  
233  }  
234  }  
235  }  
236  }  
237  }  
238  }  
239  }  
240  }  
241  }  
242  }  
243  }  
244  }  
245  }  
246  }  
247  }  
248  }  
249  }  
250  }  
251  }  
252  }  
253  }  
254  }  
255  }  
256  }  
257  }  
258  }  
259  }  
260  }  
261  }  
262  }  
263  }  
264  }  
265  }  
266  }  
267  }  
268  }  
269  }  
270  }  
271  }  
272  }  
273  }  
274  }  
275  }  
276  }  
277  }  
278  }  
279  }  
280  }  
281  }  
282  }  
283  }  
284  }  
285  }  
286  }  
287  }  
288  }  
289  }  
290  }  
291  }  
292  }  
293  }  
294  }  
295  }  
296  }  
297  }  
298  }  
299  }  
300  }  
301  }  
302  }  
303  }  
304  }  
305  }  
306  }  
307  }  
308  }  
309  }  
310  }  
311  }  
312  }  
313  }  
314  }  
315  }  
316  }  
317  }  
318  }  
319  }  
320  }  
321  }  
322  }  
323  }  
324  }  
325  }  
326  }  
327  }  
328  }  
329  }  
330  }  
331  }  
332  }  
333  }  
334  }  
335  }  
336  }  
337  }  
338  }  
339  }  
340  }  
341  }  
342  }  
343  }  
344  }  
345  }  
346  }  
347  }  
348  }  
349  }  
350  }  
351  }  
352  }  
353  }  
354  }  
355  }  
356  }  
357  }  
358  }  
359  }  
360  }  
361  }  
362  }  
363  }  
364  }  
365  }  
366  }  
367  }  
368  }  
369  }  
370  }  
371  }  
372  }  
373  }  
374  }  
375  }  
376  }  
377  }  
378  }  
379  }  
380  }  
381  }  
382  }  
383  }  
384  }  
385  }  
386  }  
387  }  
388  }  
389  }  
390  }  
391  }  
392  }  
393  }  
394  }  
395  }  
396  }  
397  }  
398  }  
399  }  
400  }  
401  }  
402  }  
403  }  
404  }  
405  }  
406  }  
407  }  
408  }  
409  }  
410  }  
411  }  
412  }  
413  }  
414  }  
415  }  
416  }  
417  }  
418  }  
419  }  
420  }  
421  }  
422  }  
423  }  
424  }  
425  }  
426  }  
427  }  
428  }  
429  }  
430  }  
431  }  
432  }  
433  }  
434  }  
435  }  
436  }  
437  }  
438  }  
439  }  
440  }  
441  }  
442  }  
443  }  
444  }  
445  }  
446  }  
447  }  
448  }  
449  }  
450  }  
451  }  
452  }  
453  }  
454  }  
455  }  
456  }  
457  }  
458  }  
459  }  
460  }  
461  }  
462  }  
463  }  
464  }  
465  }  
466  }  
467  }  
468  }  
469  }  
470  }  
471  }  
472  }  
473  }  
474  }  
475  }  
476  }  
477  }  
478  }  
479  }  
480  }  
481  }  
482  }  
483  }  
484  }  
485  }  
486  }  
487  }  
488  }  
489  }  
490  }  
491  }  
492  }  
493  }  
494  }  
495  }  
496  }  
497  }  
498  }  
499  }  
500  }  
501  }  
502  }  
503  }  
504  }  
505  }  
506  }  
507  }  
508  }  
509  }  
510  }  
511  }  
512  }  
513  }  
514  }  
515  }  
516  }  
517  }  
518  }  
519  }  
520  }  
521  }  
522  }  
523  }  
524  }  
525  }  
526  }  
527  }  
528  }  
529  }  
530  }  
531  }  
532  }  
533  }  
534  }  
535  }  
536  }  
537  }  
538  }  
539  }  
540  }  
541  }  
542  }  
543  }  
544  }  
545  }  
546  }  
547  }  
548  }  
549  }  
550  }  
551  }  
552  }  
553  }  
554  }  
555  }  
556  }  
557  }  
558  }  
559  }  
560  }  
561  }  
562  }  
563  }  
564  }  
565  }  
566  }  
567  }  
568  }  
569  }  
570  }  
571  }  
572  }  
573  }  
574  }  
575  }  
576  }  
577  }  
578  }  
579  }  
580  }  
581  }  
582  }  
583  }  
584  }  
585  }  
586  }  
587  }  
588  }  
589  }  
590  }  
591  }  
592  }  
593  }  
594  }  
595  }  
596  }  
597  }  
598  }  
599  }  
600  }  
601  }  
602  }  
603  }  
604  }  
605  }  
606  }  
607  }  
608  }  
609  }  
610  }  
611  }  
612  }  
613  }  
614  }  
615  }  
616  }  
617  }  
618  }  
619  }  
620  }  
621  }  
622  }  
623  }  
624  }  
625  }  
626  }  
627  }  
628  }  
629  }  
630  }  
631  }  
632  }  
633  }  
634  }  
635  }  
636  }  
637  }  
638  }  
639  }  
640  }  
641  }  
642  }  
643  }  
644  }  
645  }  
646  }  
647  }  
648  }  
649  }  
650  }  
651  }  
652  }  
653  }  
654  }  
655  }  
656  }  
657  }  
658  }  
659  }  
660  }  
661  }  
662  }  
663  }  
664  }  
665  }  
666  }  
667  }  
668  }  
669  }  
670  }  
671  }  
672  }  
673  }  
674  }  
675  }  
676  }  
677  }  
678  }  
679  }  
680  }  
681  }  
682  }  
683  }  
684  }  
685  }  
686  }  
687  }  
688  }  
689  }  
690  }  
691  }  
692  }  
693  }  
694  }  
695  }  
696  }  
697  }  
698  }  
699  }  
700  }  
701  }  
702  }  
703  }  
704  }  
705  }  
706  }  
707  }  
708  }  
709  }  
710  }  
711  }  
712  }  
713  }  
714  }  
715  }  
716  }  
717  }  
718  }  
719  }  
720  }  
721  }  
722  }  
723  }  
724  }  
725  }  
726  }  
727  }  
728  }  
729  }  
730  }  
731  }  
732  }  
733  }  
734  }  
735  }  
736  }  
737  }  
738  }  
739  }  
740  }  
741  }  
742  }  
743  }  
744  }  
745  }  
746  }  
747  }  
748  }  
749  }  
750  }  
751  }  
752  }  
753  }  
754  }  
755  }  
756  }  
757  }  
758  }  
759  }  
760  }  
761  }  
762  }  
763  }  
764  }  
765  }  
766  }  
767  }  
768  }  
769  }  
770  }  
771  }  
772  }  
773  }  
774  }  
775  }  
776  }  
777  }  
778  }  
779  }  
780  }  
781  }  
782  }  
783  }  
784  }  
785  }  
786  }  
787  }  
788  }  
789  }  
790  }  
791  }  
792  }  
793  }  
794  }  
795  }  
796  }  
797  }  
798  }  
799  }  
800  }  
801  }  
802  }  
803  }  
804  }  
805  }  
806  }  
807  }  
808  }  
809  }  
810  }  
811  }  
812  }  
813  }  
814  }  
815  }  
816  }  
817  }  
818  }  
819  }  
820  }  
821  }  
822  }  
823  }  
824  }  
825  }  
826  }  
827  }  
828  }  
829  }  
830  }  
831  }  
832  }  
833  }  
834  }  
835  }  
836  }  
837  }  
838  }  
839  }  
840  }  
841  }  
842  }  
843  }  
844  }  
845  }  
846  }  
847  }  
848  }  
849  }  
850  }  
851  }  
852  }  
853  }  
854  }  
855  }  
856  }  
857  }  
858  }  
859  }  
860  }  
861  }  
862  }  
863  }  
864  }  
865  }  
866  }  
867  }  
868  }  
869  }  
870  }  
871  }  
872  }  
873  }  
874  }  
875  }  
876  }  
877  }  
878  }  
879  }  
880  }  
881  }  
882  }  
883  }  
884  }  
885  }  
886  }  
887  }  
888  }  
889  }  
890  }  
891  }  
892  }  
893  }  
894  }  
895  }  
896  }  
897  }  
898  }  
899  }  
900  }  
901  }  
902  }  
903  }  
904  }  
905  }  
906  }  
907  }  
908  }  
909  }  
910  }  
911  }  
912  }  
913  }  
914  }  
915  }  
916  }  
917  }  
918  }  
919  }  
920  }  
921  }  
922  }  
923  }  
924  }  
925  }  
926  }  
927  }  
928  }  
929  }  
930  }  
931  }  
932  }  
933  }  
934  }  
935  }  
936  }  
937  }  
938  }  
939  }  
940  }  
941  }  
942  }  
943  }  
944  }  
945  }  
946  }  
947  }  
948  }  
949  }  
950  }  
951  }  
952  }  
953  }  
954  }  
955  }  
956  }  
957  }  
958  }  
959  }  
960  }  
961  }  
962  }  
963  }  
964  }  
965  }  
966  }  
967  }  
968  }  
969  }  
970  }  
971  }  
972  }  
973  }  
974  }  
975  }  
976  }  
977  }  
978  }  
979  }  
980  }  
981  }  
982  }  
983  }  
984  }  
985  }  
986  }  
987  }  
988  }  
989  }  
990  }  
991  }  
992  }  
993  }  
994  }  
995  }  
996  }  
997  }  
998  }  
999  }  
9999  }
```

► Generate & Preview BCO

BioCompute Object Generation Window

CONFIDENTIAL

SevenBridges

5 Appendix 2: The Complete BioCompute Object

```
{  
  "bco_spec_version": "https://w3id.org/biocompute/1.3.0/",  
  "bco_id": "http://biocompute.sbggenomics.com/bco/f8489c29-d7ed-477e-85c8-2dc0b7832041",  
  "checksum": "c85a8cb045244513ec53323e20f2f002e0fc4d6448244f4124ef4d7f913456cd",  
  "provenance_domain": {  
    "modified": "2019-08-08T00:00:00+0000",  
    "contributors": [],  
    "license": "https://spdx.org/licenses/CC-BY-4.0.html"  
  },  
  "usability_domain": "Alignment to a reference genome and transcriptome presents the first step of  
  "extension_domain": {  
    "fhir_extension": {  
      "fhir_endpoint": "",  
      "fhir_version": "",  
      "fhir_resources": {}  
    },  
    "scm_extension": {  
      "scm_repository": "",  
      "scm_type": "git",  
      "scm_commit": "",  
      "scm_path": "",  
      "scm_preview": ""  
    }  
  },  
  "RNA-seq Alignment - STAR":  
  "Page 14 of 16"
```

BioCompute Object – Generated PDF

Publishing a BioCompute Object to a GIT Page

BioCompute Object Examples Page on Git

README.md

BCO Examples

A repository for BCO example flat files.

Table of Contents:

- [HCV1a](#) - This BCO was developed with the [Reproducibility and Interpretation use case](#) in mind. This is the archetypal BCO example and is in the [BCO Specification](#) repository.
- [glycosylation-sites-UniCarbKB](#) - This BCO was developed with the [Data integration use case](#) in mind. The full repository is available [here](#)
- [UVP](#) - This BCO was developed with the [Accountability use case](#) in mind. The full repository is available [here](#)
- [HIVE_metagenomics](#) - This BCO was developed with the [Reusability Use Case](#) in mind. The full repository is available [here](#)

Kallisto-demo CWL on Git

Branch: master → workflows / workflows / kallisto-demo.cwl Find file Copy path

mr-c final update of old draft-3 to v1.0 0367a4d on Aug 14, 2017 1 contributor

Executable File | 37 lines (29 sloc) | 594 Bytes Raw Blame History

```
#!/usr/bin/env cwl-runner
cwlVersion: v1.0
class: Workflow

# Inspired by
# https://github.com/InSilicoDB/pipeline-kallisto/blob/develop/main.nf

hints:
DockerRequirement:
dockerPull: insilicodb/kallisto

inputs:
transcripts: File[]
reads: File[]

steps:
indexing:
run: ../../tools/kallisto-index.cwl
in:
fasta-files: transcripts
out:
- index
quantifying:
run: ../../tools/kallisto-quant.cwl
in:
index: indexing/index
fastqs: reads
out:
- quantification

outputs:
quants:
type: File
outputSource: quantifying/quantification
```

HCV1a BioCompute Object Examples on Git

examples/HCV1a.json at master

GitHub, Inc. [US] | [github.com/biocompute-objects...](#) Search Sign in Sign up

biocompute-objects / examples

Branch: master → examples / HCV1a.json Find file Copy path

HadleyKing Update HCV1a.json 3647b05 on Apr 22 1 contributor

349 lines (347 sloc) | 13.4 KB Raw Blame History

```
{ "bcfo_id": "https://w3id.org/biocompute/1.3.0/examples/HCV1a.json", "checksum": "66DACE70679F35B8A7A3D06FFFED4ED244AF5B8C251264C37E5F1B3ADE04A31", "bcfo_spec_version": "https://w3id.org/biocompute/1.3.0/", "provenance_domain": { "name": "HCV1a ledipasvir resistance SNP detection", "url": "https://w3id.org/biocompute/1.3.0/examples/HCV1a.json" } }
```

Generated PDF Output

BioCompute Object for Regulatory Review

BCO Title: RNA-seq Alignment - STAR

ECU Generation Date: August 08, 2019

BCC Specification Version: V1.3.0

DCU Generator. Seven bridges

Contents

1 BioCompute Object Domain Entries	1
1.1 Top Level Fields	1
1.2 Provenance Domain	1
1.3 Usability Domain	1
1.4 Extension Domain	2
1.5 Description Domain	3
1.6 Execution Domain	5
1.7 Parametric Domain	7
1.8 Input/Output Domain	9
1.9 Error Domain	9
2 Funding	10
3 References	10
4 Appendix 1: BioCompute Object Specification v1.3.0	11
5 Appendix 2: The Complete BioCompute Object	14

CONFIDENTIAL

1 BioCompute Object Domain Entries

5.1 Top Level Fields

```
[ "https://www.ncbi.nlm.nih.gov/biocompute/1.3.0/",  
  "http://biocompute.abigenomics.com/bco/fb4899c29-d7ed-477e-85c8-2dc0b7832041"  
  "c56a8cb046244513ec53323e0f20f2002e0fc4d6448244f4124ef4d7f913456cd" ]
```

9.4. PERFORMANCE METRICS

```
        "name": "RNA-seq Alignment - STAR",
        "version": "1.0.0",
        "review": [],
        "derived_from":
        "https://api.abgenomics.com/v2/apps/somer/somers-demo-project/rna-seq-alignment-star/0/raw/*",
        "obsolete_after": "2019-08-08T00:00:00+0000",
        "embargo": ["2019-08-08T00:00:00+0000",
                    "2019-08-08T00:00:00+0000"],
        "created": "2019-08-08T00:00:00+0000",
        "modified": "2019-08-08T00:00:00+0000",
        "contributors": [],
        "license": "https://spdx.org/licenses/CC-BY-4.0.html"
    }
}
```

www.ijerpi.org

the first step of RNA-Seq analysis. This pipeline uses STAR, an ultrafast RNA-seq aligner capable of mapping full length RNA sequences and detecting de novo canonical junctions, non-canonical splices, and chimeric (fusion)



BCO Advance Topics

Executable BCOS

SILICO FCM

Computational Platform for In Silico Clinical Trials of Familial Cardiomyopathy

Overview

- Computational Platform for In Silico Clinical Trials or Familial Cardiomyopathy
 - genetic, biological, pharmacologic, clinical, imaging and patient specific cellular aspects
 - Seven Bridges GRAF
 - European Union's Horizon 2020 Funding



SILICO FCM

Computational Platform for In Silico Clinical Trials of Familial Cardiomyopathy

Overview

- Computational Platform for In Silico Clinical Trials or Familial Cardiomyopathy
 - genetic, biological, pharmacologic, clinical, imaging and patient specific cellular aspects
 - Seven Bridges GRAF
 - European Union's Horizon 2020 Funding

Approach

- SILICO FCM Use Case
 - Individualized treatment computed genomics information and computational simulations
 - Computing across multiple platforms
 - Submitting computational procedure for regulatory review
 - Augment existing infrastructure
 - Adding execution information to existing BCO fields
 - Use open source CWL Executors
 - Develop light weight BCO execution wrapper

SILICO FCM – GATK Best Practice Data Processing Workflow

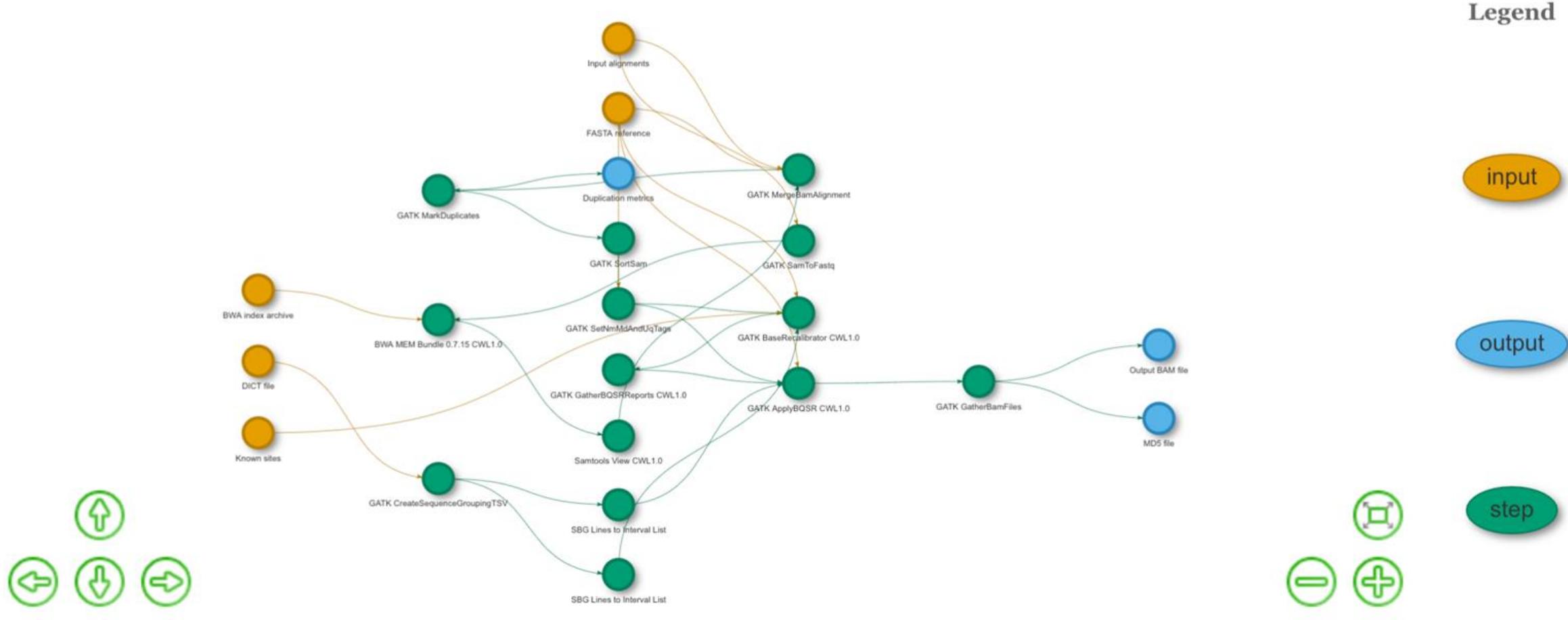


Figure Generated From BCO App

BCO Execution Domain

Store CWL Execution Information

Script

<https://cgc-api.sbggenomics.com/v2/apps/jeffrey.grover/cgc-testing/broad-best-practice-data-pre-processing-workflow-4-1-O-O/raw/>

Script driver

Seven Bridges Common Workflow Language Executor

Algorithmic tools and software prerequisites

New

Edit

Delete

Copy

Show 10 entries

Search:

name	version	uri	access_time	sha1_chksum
Seven Bridges Platform	2021-07-21	https://igor.sbggenomics.com/	2021-07-21	

Showing 1 to 1 of 1 entries

Previous Next

BCO App Editor

Executing BCOs Locally

Workflow Selection

Select a Local Workflow

Connect to SevenBridges Project Select a Workflow

Create BCO from Scratch without Workflow

Generate BioCompute Object via BCO App



Execution of BCO

Clone the Repository and Install the Environment

Run the Script with your Input BCO

Evaluate the CWL Run Results

Save Generated BCO directly via BCO App to:

- Local
- GitHub Project
- SevenBridges Project

Executing BCOs Locally

Workflow Selection

Select a Local Workflow

Connect to SevenBridges Project Select a Workflow

Create BCO from Scratch without Workflow

Generate BioCompute Object via BCO App



Save Generated BCO directly via BCO App to:

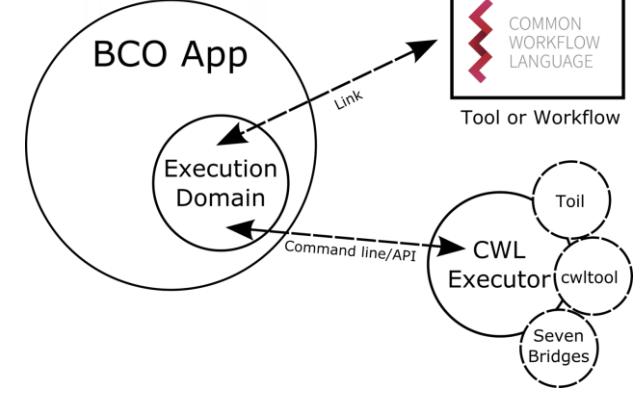
- Local
- GitHub Project
- SevenBridges Project

Execution of BCO

Clone the Repository and Install the Environment

Run the Script with your Input BCO

Evaluate the CWL Run Results



Executing BCOs Locally

Workflow Selection

Select a Local Workflow

Connect to SevenBridges Project Select a Workflow

Create BCO from Scratch without Workflow

Generate BioCompute Object via BCO App



Save Generated BCO directly via BCO App to:

- Local
- GitHub Project
- SevenBridges Project

Status and Outcomes

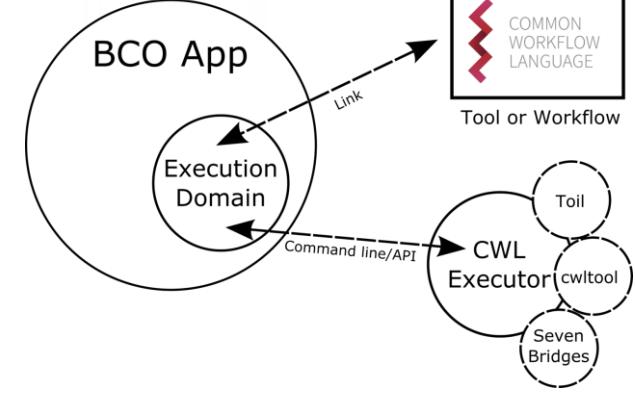
- Existing BCO fields support portable execution
- CWL 1.x greatly facilitates portable execution
- Demonstrate ability to execute workflow described within BCO
- Will facilitate a more thorough review of workflows

Execution of BCO

Clone the Repository and Install the Environment

Run the Script with your Input BCO

Evaluate the CWL Run Results



Grover et al. 2021

<https://github.com/skoc/BCO-Execution>

BCO Tool Demonstration

Summary and Findings

BCO Generation Prototype Project Outcomes

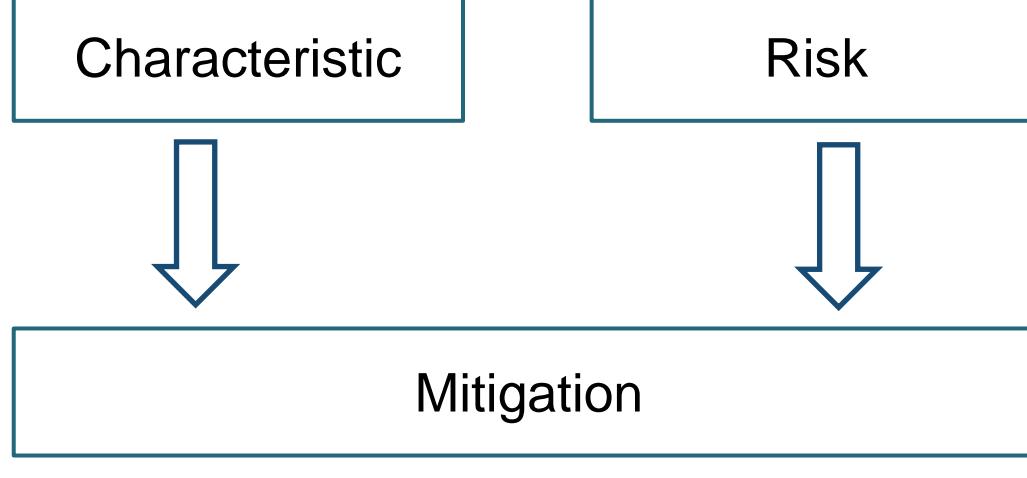
- Primary Outcomes
 - Multiple Implementations Required to Support a Range of Team Members
 - Paper, Word Processing, Confluence and Shiny App
 - Building on CWL Infrastructure Can Reduce Time to Generate
 - Integrate project best practices (Project Descriptions vs. CWL workflow description)
 - Documenting a task run may have direct applications for regulatory submissions
- Towards an Interoperable Platform
 - Integrating BCO Generation across platforms with Apps
 - Creating a customizable BCO generation experience



Towards BCO Best Practices

Advising on BCO Content - Framework

Proposed Content Framework



Advising on BCO Content – BCO Exchange Risk

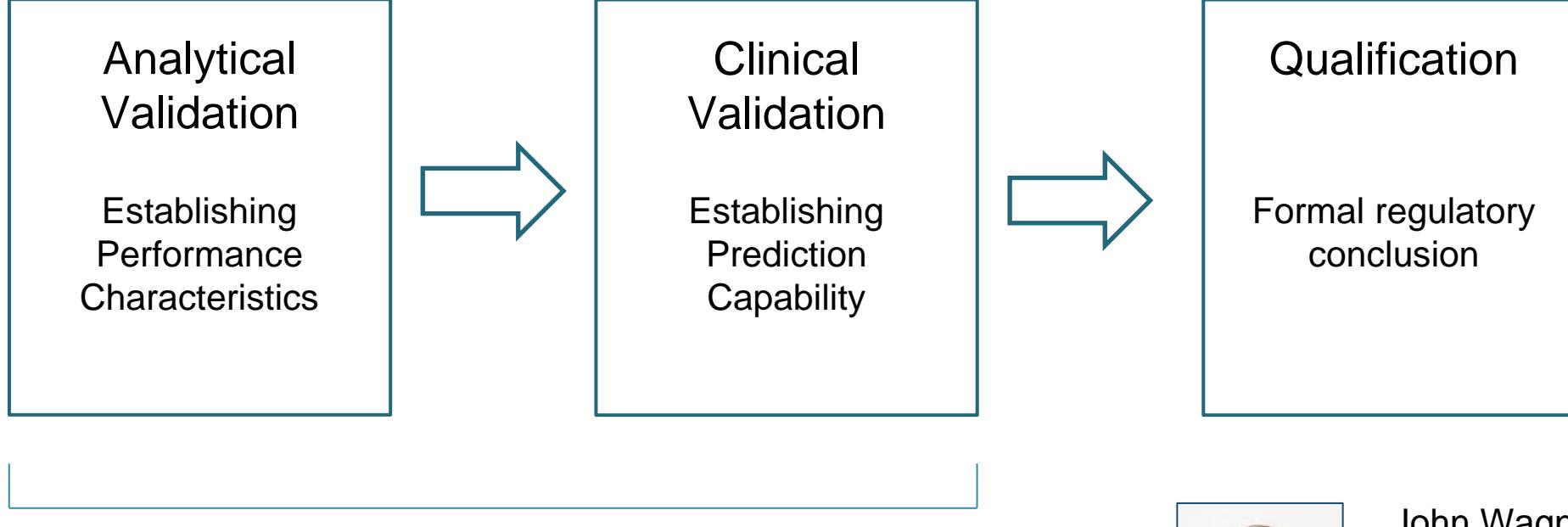
Use Case	Characteristics	Risk	Mitigation
Between Two Users	Unfettered communication	A short delay in clarifying details	Include use context in the description
	Similar Execution Environment	Minimal execution risk	Include execution domain details and test case information with results
	Common Application Context	Slight risk of workflow misuse	Context with required proprietary information included
Between Organizations	Communication is limited or contracted.	Delay in establishing a contract with a potential loss of IP	Include use limitations within BCO (license information)
	The execution environment is different but modifiable	Substantial execution risk is possible	Include execution requirements within BCO
	Application Context may differ.	Inappropriate use of workflow	Describe contextual factors within the description domains
FDA Regulatory Submission	Formal communication protocol	Long delays in clarifying details	Include links to supplementary sources and BCOs links
	Execution environment likely to differ	Not possible to replicate the results	Include permanent resource identifiers
	Novel aspects may require additional context.	Risk of Patient adverse event	Include global identifiers to proprietary information

Advising on BCO Content – Application Risks

Workflow Use	Characteristic	Risk	Mitigation
Applications Issues			
Cloud NEO: A cloud Pipeline for identifying patient-specific tumor neoantigens	Strategies for personalizing care can differ widely	Essential details of the personalization strategy are not communicated	Emphasize relationship to best practices and novel components. Include external links to best practices
	Both methods and implementation can be quite complex	Results may differ by methods (algorithm) or parameter set	Include algorithm and parameter justification. Include evidence for novel methods or nonstandard parameters
	Transcript Identification accuracy may differ by ethnicity	Transcript may not be appropriate for individuals with a specific characteristic	Include validation study data linked including demographics
ML Classifier Tumor Classifier Model (TCI 1)	Cancer-related machine learning models can be complex to understand	Misuse is due to the domain and implementation complexity	Provide equal space for the domain and the modeling information
	Resources in a National Laboratory environment are not likely to match commercial or academic settings.	Resources in National Laboratory are not available for commercial or academic use.	Document unique implementation resources
	The training set is complex (30K columns - repurposed from studies)	The training set is complex, which may obfuscate the ability to identify performance bias	Document training set along with the data and tools used to generate.

Towards Biomarker Qualification

FDA Public Meeting: Identification of Concepts and Terminology for Multi-Component Biomarkers

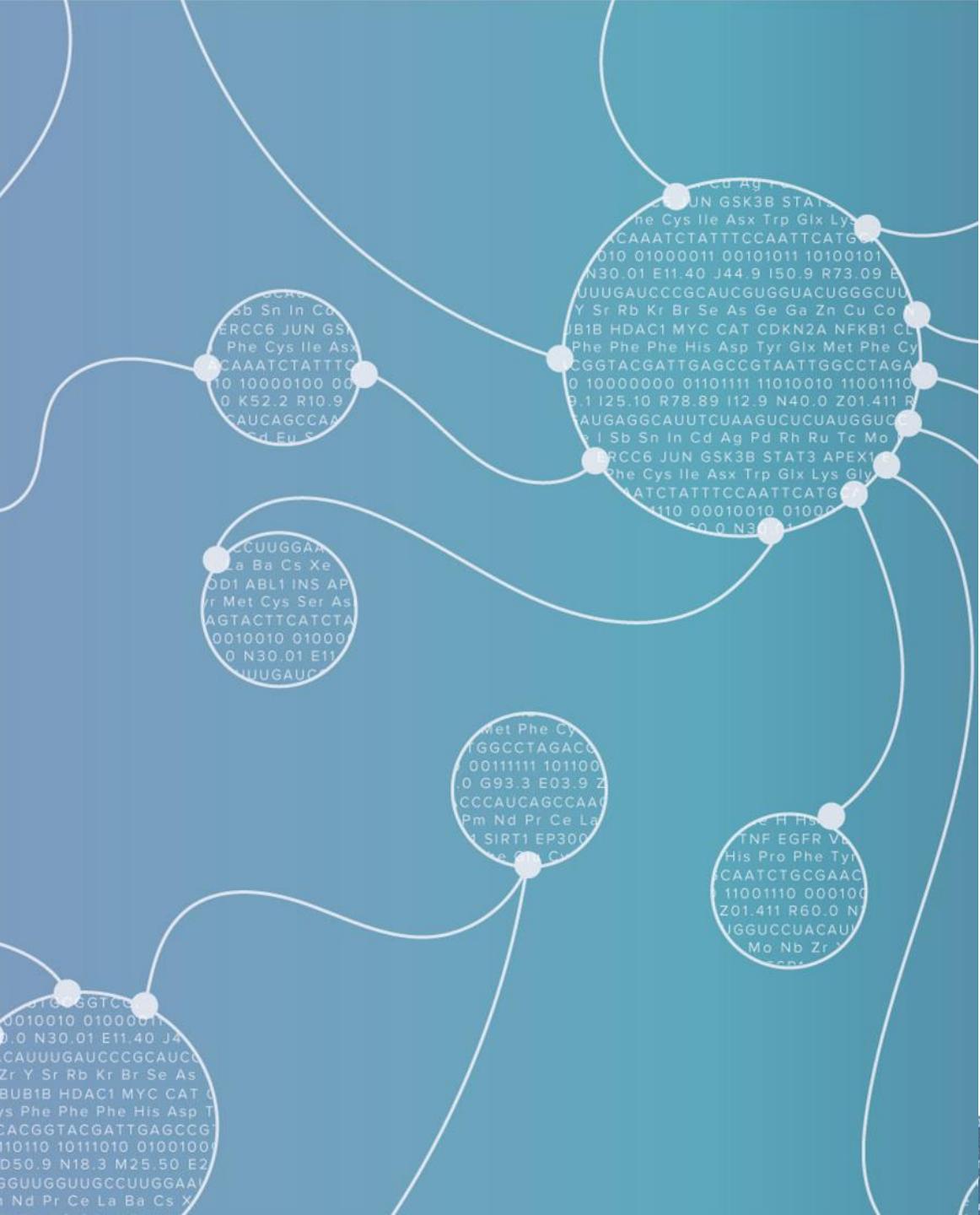


John Wagner, MD, PhD
Koneska Health

Keynote Address
March 23, 2022

Summary

SevenBridges



Summary

Platform Enabled BioCompute Objects

- Supporting an open source framework within a platform setting
- Streamlining BCO generation with platform features

BCO Workflow Use Cases

- Demonstrating ease of use with multiple workflows
- Highlighting the need for system support and ongoing maintenance
- Elucidating new challenges with neoantigen and machine learning workflows

Towards Best Practices and Regulatory Submission

- Proposing use of characteristic – risk – mitigation framework

Acknowledgements



Christian Frech, PhD
Celsius Therapeutics



Kaushik Ghose, PhD
Google



Soner Koc, MS
Senior Data Scientist



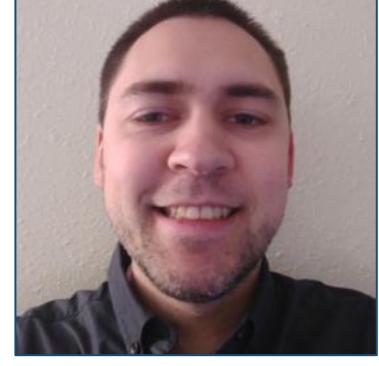
Phillip Webster, PhD

Genomic Scientist



Saul Acevedo, PhD

Data Scientist



Jeffrey Grover, PhD
Entrada Therapeutics

George Washington University

Raja Mazumder, PhD

Jonathan Keeney, PhD

Amanda Bell

Charles King

Janisha Patel

Seven Bridges

Philip Brooks, PhD

Alison Leaf, PhD

Michele Mattoni, PhD

Manisha Ray, PhD

Isheeta Sethi, PhD



Nan Xiao, PhD
Merck

Questions and Discussion

Happy to Discuss and Demo



Dennis A. Dean, II, PhD

dennis.dean@sevenbridges.com

<https://www.linkedin.com/in/dennis-a-dean-ii-ph-d-4b97b1b/>

Phillip Webster, PhD

phil.webster@sevenbridges.com

<https://www.linkedin.com/in/phillipjwebster/>

Extra Slides

BCO in Support of Modern Workflow (CWL) Development

