

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

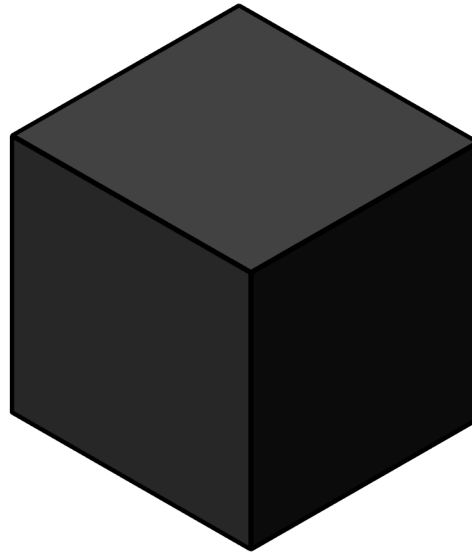
Jonathon Keeney, Ph.D.

keeneyjg@gwu.edu

Assistant Research Professor, George Washington University
Managing Director, BioCompute Executive Steering Committee

Computational Toxicology Communities of Practice
July 25th, 2019

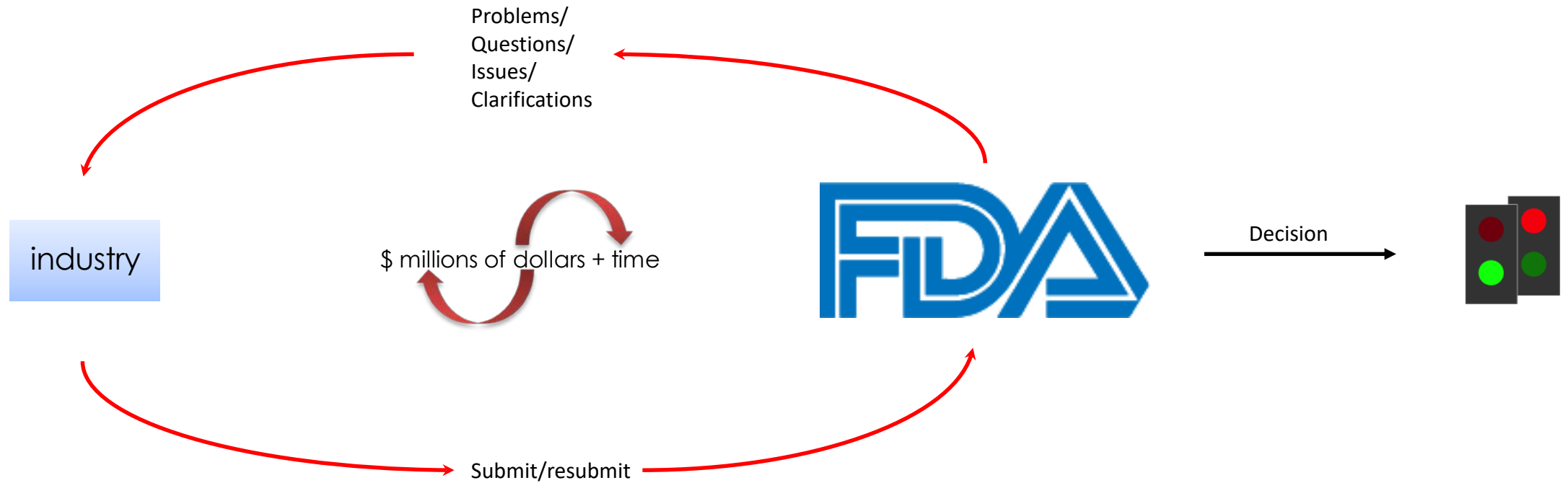
NGS Data Flows



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

```
$ fastq-dump -X 2 SRR001666 --split-3
R: $ fastq-dump -X 2 SRR001666 --split-3
W: $ fastq-dump -X 2 SRR001666 --split-3
R: $ fastq-dump -X 2 SRR001666 --split-3
W: $ fastq-dump -X 2 SRR001666 --split-3
@: $ fastq-dump -X 2 SRR001666 --split-3
+I: Read 2 spots for SRR001666
+I: Written 2 spots for SRR001666
I: @: $ head SRR001666_1.fastq SRR001666_2.fastq
G: ==> SRR001666_1.fastq <==
+I: @SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
+I: GGGTGATGGCCGCTGCCGATGGCGTCAATCCACC
+I: @SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
+I: IHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHII9IG9IC
+I: @SRR001666.2 071112_SLXA-EAS1_s_7:5:1:801:338 length=36
+I: GTTCAGGGATACGACGTTTGTATTTAAGAACTCTGA
+I: @SRR001666.2 071112_SLXA-EAS1_s_7:5:1:801:338 length=36
+I: IHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHII6IBI
+I: @SRR001666_2.fastq <==
+I: @SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
+I: AAGTTACCCCTTACAACCTTAAAGGTTTTCAAATAGA
+I: @SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
+I: IHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH>IIHHHII/
+I: @SRR001666.2 071112_SLXA-EAS1_s_7:5:1:801:338 length=36
+I: AGCAGAAGTCGATGATAATACGCGTCTTTATCAT
+I: @SRR001666.2 071112_SLXA-EAS1_s_7:5:1:801:338 length=36
+I: IHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH>IIHHH-I)8I
```

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

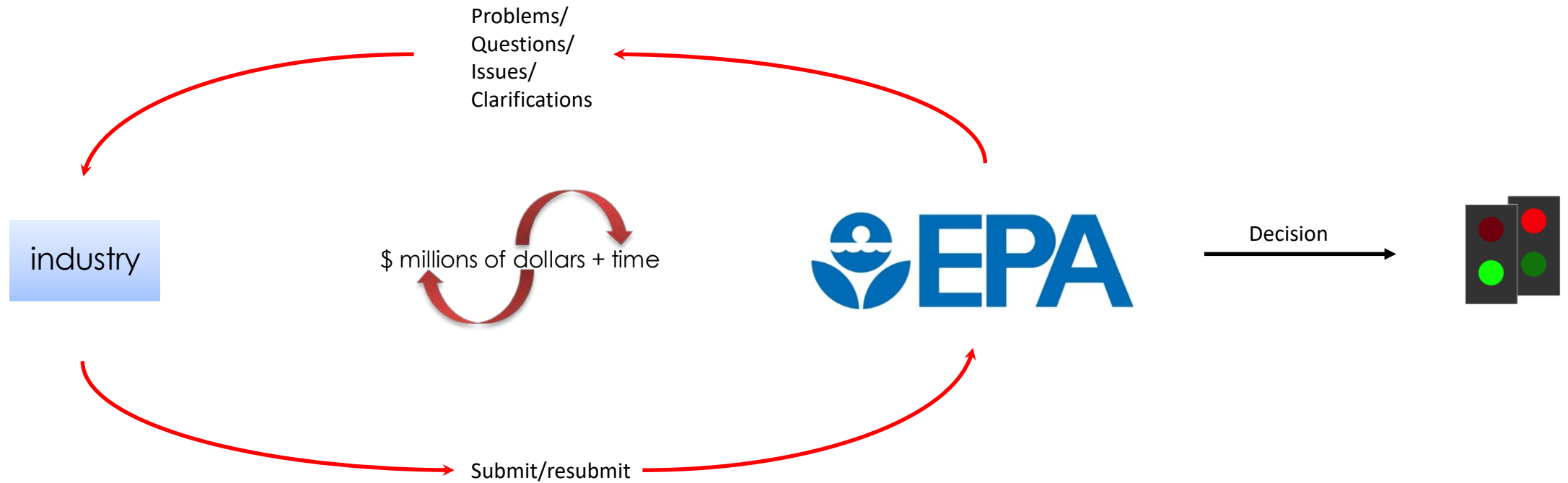
The division will accept Next Generation sequencing data generated from most standard Next Generation Sequencing (NGS) platforms provided the sponsor supplies the appropriate details for the sequencing platform, the protocols to be used for sample preparation, the raw NGS data, and the methods used to 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.

Wasted Time and Money





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

A Notice by the [Environmental Protection Agency](#) on 06/19/2015



PUBLISHED DOCUMENT



13



AGENCY:

Environmental Protection Agency (EPA).

ACTION:

Notice.

SUMMARY:

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

DOCUMENT DETAILS

Printed version:

[PDF](#)

Publication Date:

06/19/2015

Agency:

[Environmental Protection Agency](#)

Dates:

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

Comments Close:

08/18/2015

Document Type:



Assessing and Managing Chemicals under TSCA

[CONTACT US](#)

SHARE



Assessing and Managing Chemicals under TSCA
[Home](#)

[How EPA Evaluates the Safety of Existing Chemicals](#)

[Prioritizing Existing Chemicals for Risk Evaluation](#)

[Risk Evaluations for Existing Chemicals Under TSCA](#)

[Current Chemical Risk Management Activities](#)

Alternative Test Methods and Strategies to Reduce Vertebrate Animal Testing

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

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

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

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

A solution should...

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

802.11 Analogy



```
{
  "id": "obj.1277",
  "name": "Human[tax:9606] CAG repeat [SO:0000726] detection for Huntington's disease [DOID:12858]",
  "version": "1.0",
  "createdby": "hadley_king@gwmail.gwu.edu",
  "digital_signature": "QdB7vu4rLeLtMFmPHQ9S",
  "verification_status": "unreviewed",
  "publication_status": "draft",
  "usability_domain": ["Determine CAG repeats [SO:0000726] in a sequence for the diagnosis of Huntington's disease [DOID:12858] for HTT gene [HGNC:4851]: 10 - 35 CAG trinucleotide repeats: No Risk, 36 - 40 CAG trinucleotide repeats: Moderate Risk, CAG trinucleotide repeats: High Risk..."],
  "authors": [{"orcid": "0000-0002-9920-565X"}],
  "description_domain": {
    "keywords": ["disease", "human disease"...],
    "xref": ["SO:0000726", "tax:9606", "DOID:12858", "ICD:G10"...],
    "pipeline_steps": {"HIVE hexagon": {...}, "HIVE heptagon": {...}},
  }
}
```

Metadata
(not needed for computation)

```
"execution_domain": {
  "pipeline_version": "1.0",
  "platform": "HIVE",
  "driver": "hive://hive-driver",
  "script": "hive://workflows/huntingtons_disease_detection.hwf",
  "prerequisites": [
    {"name": "HIVE_hexagon", "version": "1.3"},
    {"name": "HIVE_heptagon", "version": "1.3"}],
  "env_parameters": ["HIVEv1.3"],
}
```

parametrized pipeline with all internal arguments well defined

workflow - defined in HIVE

```
"parametric_domain": {
  "divergence_threshold_percent": "10",
  "alignment_seed": "14 letters",
  "alignment_min_match_len": "50",
  "repeat_calling_min_coverage": "20",
  "repeat_calling_frequency_cutoff": "0.2",
}
```

parametric domain

Verification Kit

input domain

```
"io_domain": {
  "reference_uri_list": [
    "http://www.ncbi.nlm.nih.gov/nuccore/SRR4243395",
    "http://www.ncbi.nlm.nih.gov/nuccore/NG_009378.1"],
  "input_uri_list": ["hive://nuc-read/514683", "hive://nuc-read/514682"],
  "output_uri_list": ["hive://data/509234/repeat-profile.csv"],
  "error_domain": ["false negative discovery < 0.000001"]
}
```

output domain

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

Main Home HIVE-Portal Links Contact Help HIVE Team Logout Hadley King News

Hexagon Aligner

NNS19-1

Parameters

Progress

Results

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

Hit List Downloads

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

Histogram Alignments Stack Hit Tables Help

Alignment

```
17 AAACGATCACGTCGTTTATGCGGTCTGCTGAACTTATCAGAGACACGCTACAGAAAGACCTTAAACGCTGTTGAGCCTATCGTACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
300 AAACGATCACGTCGTTTATGCGGTCTGCTGAACTTATCAGAGACACGCTACAGAAAGACCTTAAACGCTGTTGAGCCTATCGTACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
51 GAACCTATCAGAGACACGCTACAGAAAGACCTTAAACGCTGTTGAGCCTATCGTACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
1 GAACCTATCAGAGACACGCTACAGAAAGACCTTAAACGCTGTTGAGCCTATCGTACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
95 GGTTCGAGCCTATCGTACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
124 GGTTCGAGCCTATCGTACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
95 GGTTCGAGCCTATCGTACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
1 GGTTCGAGCCTATCGTACCGCTTAAATATGAGGACAGGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
126 GGACAAGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
1 GGACAAGCACTGACAATGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
140 CAATTGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
1 CAATTGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
143 TTGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
301 TTGGCGTCCCGCCCATCTCTATGAATTCFAGAAGAAAAGTTTGTGACTTGTCTGCGCGTCCCTATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
213 ATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
1 ATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
213 ATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
1 ATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
213 ATATAAGAAATGGGGAAGGACGAGCTTATGTATGTATCTGACGAT
```

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_A1	266	8,977	1,271.3	5.7

Alignments

Start	Alignment
4258	ATTACAGaGACAGCAGAGACCTATTTGGAAAGGACCAGCCAAACTACTCTGGAAAGGT
94	ATTACAGgGACAGCAGAGACCTATTTGGAAAGGACCAGCCAAACTACTCTGGAAAGGT
4258	ATTACAGaGACAGCAGAGACCTATTTGGAAAGGACCAGCCAAACTACTCTGGAAAGGT
1	ATTACAGgGACAGCAGAGACCTATTTGGAAAGGACCAGCCAAACTACTCTGGAAAGGT

Summary

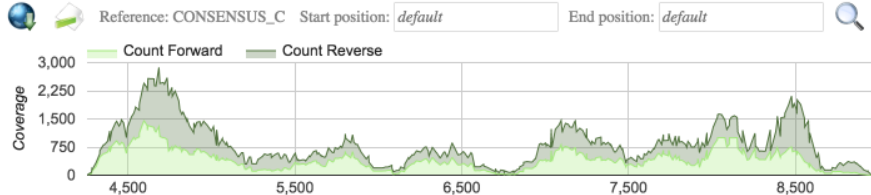
Annotation Files

- 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

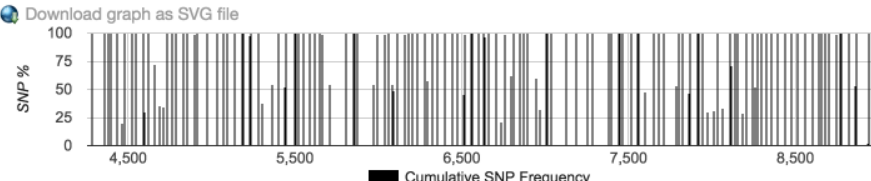
Profile Graphs

SNP Calls AA Calls Annotations

Reference: CONSENSUS_C Start position: default End position: default



Download graph as SVG file

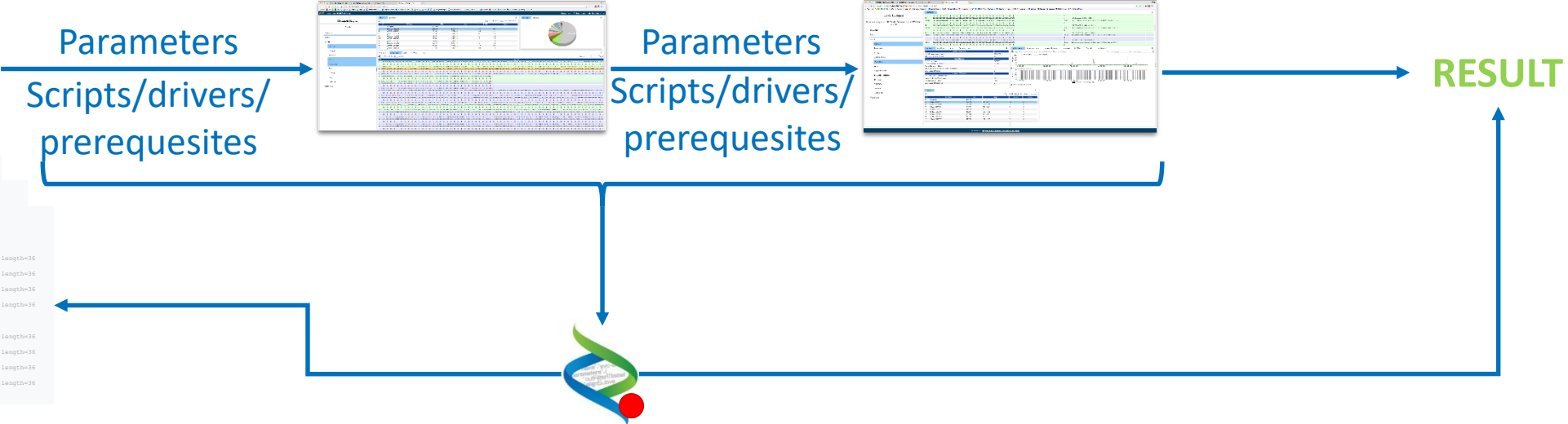


Download graph as SVG file

BCO Example: HCV-1 drug resistance



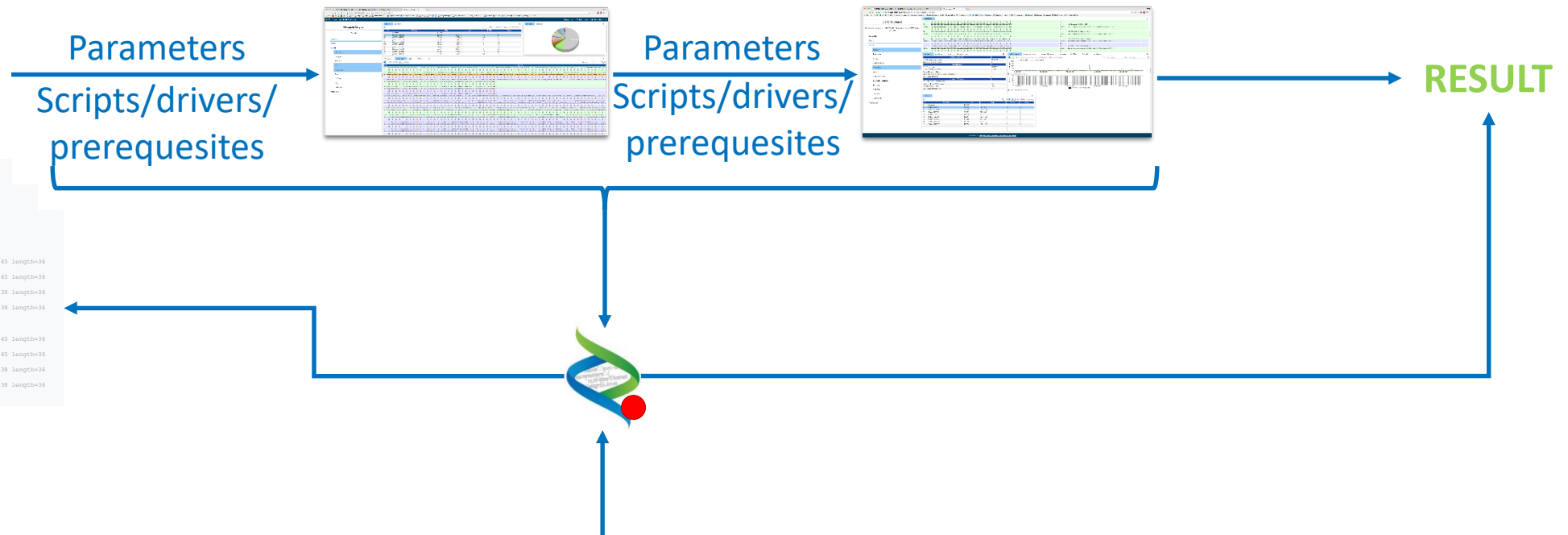
```
$ catq-dump -X 2 S88001666 --split-3
Read 2 spots for S88001666
# 1
# 2
# 3
# 4
# 5
# 6
# 7
# 8
# 9
# 10
# 11
# 12
# 13
# 14
# 15
# 16
# 17
# 18
# 19
# 20
# 21
# 22
# 23
# 24
# 25
# 26
# 27
# 28
# 29
# 30
# 31
# 32
# 33
# 34
# 35
# 36
# 37
# 38
# 39
# 40
# 41
# 42
# 43
# 44
# 45
# 46
# 47
# 48
# 49
# 50
# 51
# 52
# 53
# 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
# 1000
```



BCO Example: HCV-1 drug resistance



```
$ fastq-dump -X 2 SRR001666 --split-3
# 1
# fastq-dump -X 2 SRR001666 --split-3
# 2
# fastq-dump -X 2 SRR001666 --split-3
# 3
# fastq-dump -X 2 SRR001666 --split-3
# 4
# fastq-dump -X 2 SRR001666 --split-3
# 5
# fastq-dump -X 2 SRR001666 --split-3
# 6
# fastq-dump -X 2 SRR001666 --split-3
# 7
# fastq-dump -X 2 SRR001666 --split-3
# 8
# fastq-dump -X 2 SRR001666 --split-3
# 9
# fastq-dump -X 2 SRR001666 --split-3
# 10
# fastq-dump -X 2 SRR001666 --split-3
# 11
# fastq-dump -X 2 SRR001666 --split-3
# 12
# fastq-dump -X 2 SRR001666 --split-3
# 13
# fastq-dump -X 2 SRR001666 --split-3
# 14
# fastq-dump -X 2 SRR001666 --split-3
# 15
# fastq-dump -X 2 SRR001666 --split-3
# 16
# fastq-dump -X 2 SRR001666 --split-3
# 17
# fastq-dump -X 2 SRR001666 --split-3
# 18
# fastq-dump -X 2 SRR001666 --split-3
# 19
# fastq-dump -X 2 SRR001666 --split-3
# 20
# fastq-dump -X 2 SRR001666 --split-3
# 21
# fastq-dump -X 2 SRR001666 --split-3
# 22
# fastq-dump -X 2 SRR001666 --split-3
# 23
# fastq-dump -X 2 SRR001666 --split-3
# 24
# fastq-dump -X 2 SRR001666 --split-3
# 25
# fastq-dump -X 2 SRR001666 --split-3
# 26
# fastq-dump -X 2 SRR001666 --split-3
# 27
# fastq-dump -X 2 SRR001666 --split-3
# 28
# fastq-dump -X 2 SRR001666 --split-3
# 29
# fastq-dump -X 2 SRR001666 --split-3
# 30
# fastq-dump -X 2 SRR001666 --split-3
# 31
# fastq-dump -X 2 SRR001666 --split-3
# 32
# fastq-dump -X 2 SRR001666 --split-3
# 33
# fastq-dump -X 2 SRR001666 --split-3
# 34
# fastq-dump -X 2 SRR001666 --split-3
# 35
# fastq-dump -X 2 SRR001666 --split-3
# 36
# fastq-dump -X 2 SRR001666 --split-3
# 37
# fastq-dump -X 2 SRR001666 --split-3
# 38
# fastq-dump -X 2 SRR001666 --split-3
# 39
# fastq-dump -X 2 SRR001666 --split-3
# 40
# fastq-dump -X 2 SRR001666 --split-3
# 41
# fastq-dump -X 2 SRR001666 --split-3
# 42
# fastq-dump -X 2 SRR001666 --split-3
# 43
# fastq-dump -X 2 SRR001666 --split-3
# 44
# fastq-dump -X 2 SRR001666 --split-3
# 45
# fastq-dump -X 2 SRR001666 --split-3
# 46
# fastq-dump -X 2 SRR001666 --split-3
# 47
# fastq-dump -X 2 SRR001666 --split-3
# 48
# fastq-dump -X 2 SRR001666 --split-3
# 49
# fastq-dump -X 2 SRR001666 --split-3
# 50
# fastq-dump -X 2 SRR001666 --split-3
# 51
# fastq-dump -X 2 SRR001666 --split-3
# 52
# fastq-dump -X 2 SRR001666 --split-3
# 53
# fastq-dump -X 2 SRR001666 --split-3
# 54
# fastq-dump -X 2 SRR001666 --split-3
# 55
# fastq-dump -X 2 SRR001666 --split-3
# 56
# fastq-dump -X 2 SRR001666 --split-3
# 57
# fastq-dump -X 2 SRR001666 --split-3
# 58
# fastq-dump -X 2 SRR001666 --split-3
# 59
# fastq-dump -X 2 SRR001666 --split-3
# 60
# fastq-dump -X 2 SRR001666 --split-3
# 61
# fastq-dump -X 2 SRR001666 --split-3
# 62
# fastq-dump -X 2 SRR001666 --split-3
# 63
# fastq-dump -X 2 SRR001666 --split-3
# 64
# fastq-dump -X 2 SRR001666 --split-3
# 65
# fastq-dump -X 2 SRR001666 --split-3
# 66
# fastq-dump -X 2 SRR001666 --split-3
# 67
# fastq-dump -X 2 SRR001666 --split-3
# 68
# fastq-dump -X 2 SRR001666 --split-3
# 69
# fastq-dump -X 2 SRR001666 --split-3
# 70
# fastq-dump -X 2 SRR001666 --split-3
# 71
# fastq-dump -X 2 SRR001666 --split-3
# 72
# fastq-dump -X 2 SRR001666 --split-3
# 73
# fastq-dump -X 2 SRR001666 --split-3
# 74
# fastq-dump -X 2 SRR001666 --split-3
# 75
# fastq-dump -X 2 SRR001666 --split-3
# 76
# fastq-dump -X 2 SRR001666 --split-3
# 77
# fastq-dump -X 2 SRR001666 --split-3
# 78
# fastq-dump -X 2 SRR001666 --split-3
# 79
# fastq-dump -X 2 SRR001666 --split-3
# 80
# fastq-dump -X 2 SRR001666 --split-3
# 81
# fastq-dump -X 2 SRR001666 --split-3
# 82
# fastq-dump -X 2 SRR001666 --split-3
# 83
# fastq-dump -X 2 SRR001666 --split-3
# 84
# fastq-dump -X 2 SRR001666 --split-3
# 85
# fastq-dump -X 2 SRR001666 --split-3
# 86
# fastq-dump -X 2 SRR001666 --split-3
# 87
# fastq-dump -X 2 SRR001666 --split-3
# 88
# fastq-dump -X 2 SRR001666 --split-3
# 89
# fastq-dump -X 2 SRR001666 --split-3
# 90
# fastq-dump -X 2 SRR001666 --split-3
# 91
# fastq-dump -X 2 SRR001666 --split-3
# 92
# fastq-dump -X 2 SRR001666 --split-3
# 93
# fastq-dump -X 2 SRR001666 --split-3
# 94
# fastq-dump -X 2 SRR001666 --split-3
# 95
# fastq-dump -X 2 SRR001666 --split-3
# 96
# fastq-dump -X 2 SRR001666 --split-3
# 97
# fastq-dump -X 2 SRR001666 --split-3
# 98
# fastq-dump -X 2 SRR001666 --split-3
# 99
# fastq-dump -X 2 SRR001666 --split-3
# 100
# fastq-dump -X 2 SRR001666 --split-3
```



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



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

BCO applications for computational toxicology

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

All Chemicals in Assay Endpoint: [ACEA_ER_80hr](#)

[Excel](#)

Annotations Citations **tcpl Processing** Reagents AOPs

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

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		OUTPUTS	
file	Input 1 [REQUIRED] workflow-app-1	file	Output 1 [REQUIRED] workflow-app-1
file	Input 2 [REQUIRED] workflow-app-2	file	Output 2 [REQUIRED] workflow-app-2

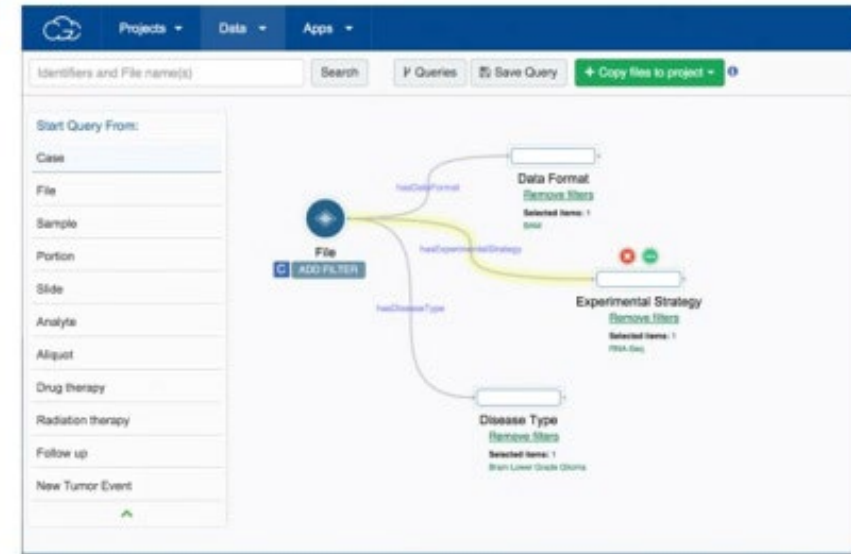


Projects | Data | Apps

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

Start Query From:

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




Galaxy Administration

Galaxy Administration | Analyze Data | Workflow | Shared Data | Admin | Help | User | Using 35.7

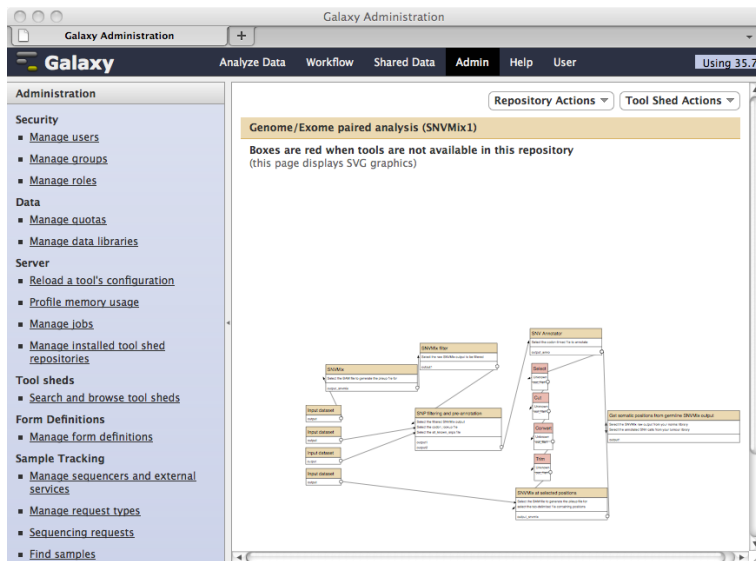
Administration

- Security
 - Manage users
 - Manage groups
 - Manage roles
- Data
 - Manage quotas
 - Manage data libraries
- Server
 - Reload a tool's configuration
 - Profile memory usage
 - Manage jobs
 - Manage installed tool shed repositories
- Tool sheds
 - Search and browse tool sheds
- Form Definitions
 - Manage form definitions
- Sample Tracking
 - Manage sequencers and external services
 - Manage request types
 - Sequencing requests
 - Find samples

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)




Main Home | HIVE Portal | Links

CensusScope

HMB25_2_R1

Parameters

Progress

Results

Taxonomy Details

Convergence

Phylogenetic Tree

Text Tree

Table

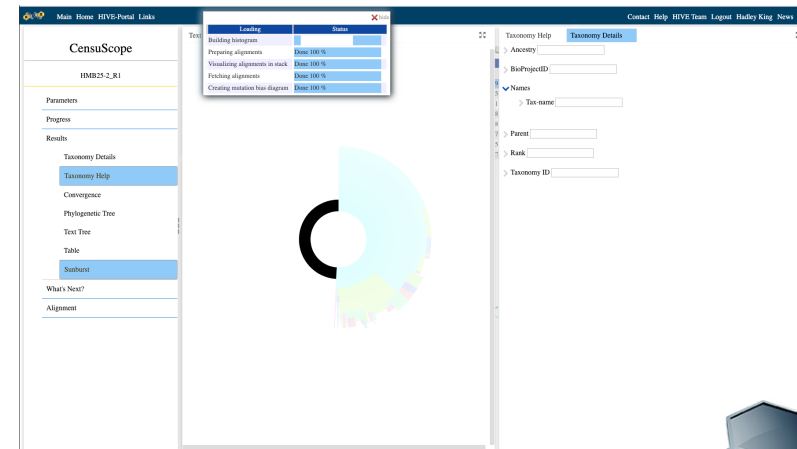
Summary

What's Next?

Alignment

Loading

Task	Progress
Building histograms	Done 100%
Preparing alignments	Done 100%
Visualizing alignments to read	Done 100%
Finding alignments	Done 100%
Creating mutation heat diagram	Done 100%



Taxonomy Help | Taxonomy Details

- Ancestry
- BioProjectID
- Names
 - Tax name
- Parent
- Rank
- Taxonomy ID



Search term

You can search by BCO ID, name or contributor.

[View Object](#)**Base type for all BioCompute Objects**

All BioCompute object types must adhere to this type in order to be compliant with BioCompute framework

bco_id

A unique identifier that should be applied to each BCO instance, generated and assigned by a BCO database engine. IDs should never be reused

checksum

A string-type, read-only value, protecting the object from internal or external alterations without proper validation generated with a SHA-256 hash function.

bco_spec_version

Version of the BCO specification used to define this document

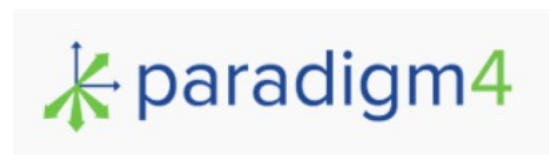
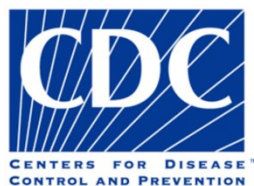
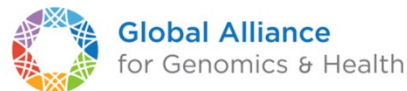
Provenance Domain

Structured field for tracking data through transformations, including contributors, reviewers, and versioning.

name

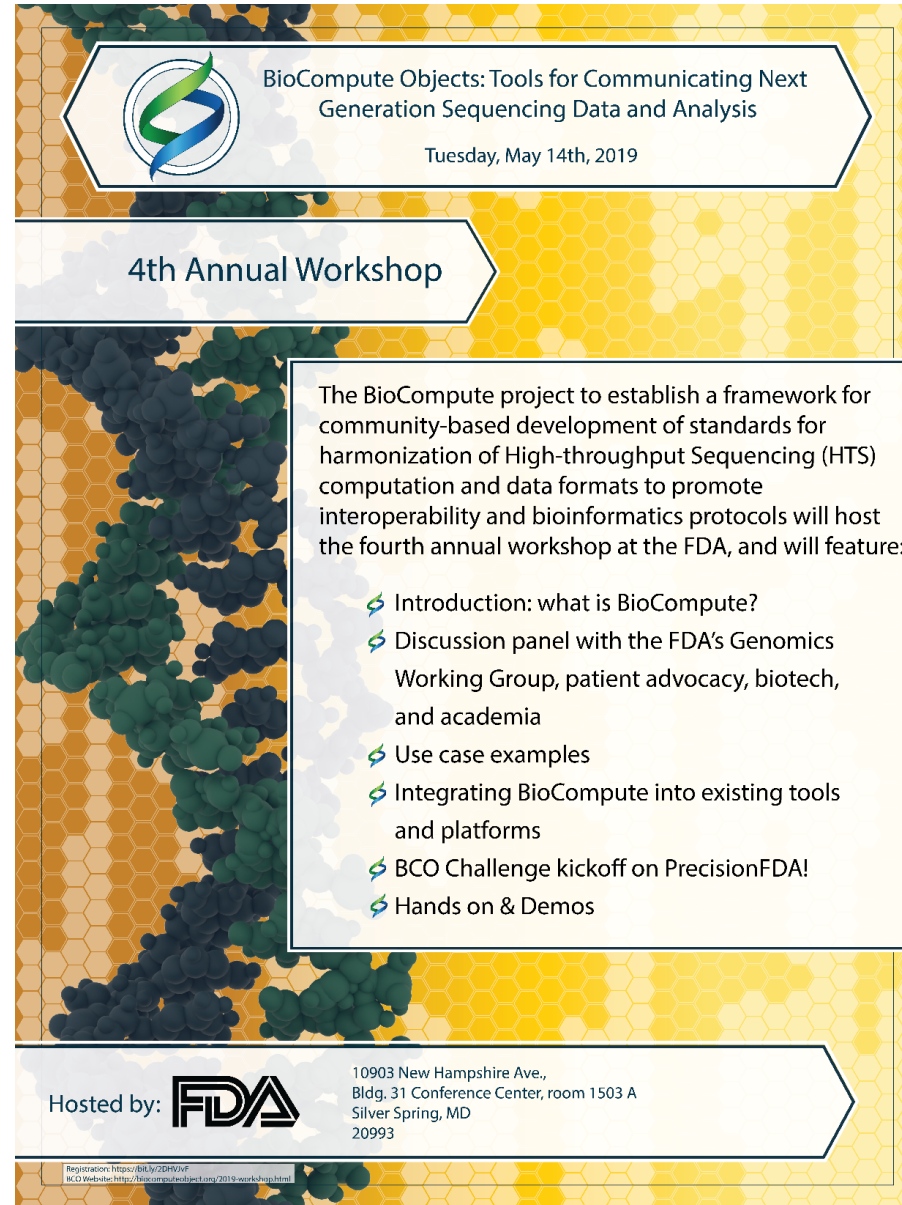
Public searchable name for BioCompute Object. This public field should take free text value using common biological research terminology supporting the terminology used in the usability_domain, external references (xref), and keywords sections.

version



2019 Workshop

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




The poster features a yellow honeycomb background with a dark green and blue DNA double helix graphic on the left side. The text is arranged in several sections: a top header box with the BioCompute logo and title, a central box for the workshop name, a large text box with details and a bulleted list of activities, and a bottom box with the host information and address.

 BioCompute Objects: Tools for Communicating Next Generation Sequencing Data and Analysis
Tuesday, May 14th, 2019

4th Annual Workshop

The BioCompute project to establish a framework for community-based development of standards for harmonization of High-throughput Sequencing (HTS) computation and data formats to promote interoperability and bioinformatics protocols will host the fourth annual workshop at the FDA, and will feature:

- ⚡ Introduction: what is BioCompute?
- ⚡ Discussion panel with the FDA's Genomics Working Group, patient advocacy, biotech, and academia
- ⚡ Use case examples
- ⚡ Integrating BioCompute into existing tools and platforms
- ⚡ BCO Challenge kickoff on PrecisionFDA!
- ⚡ Hands on & Demos

Hosted by:  10903 New Hampshire Ave.,
Bldg. 31 Conference Center, room 1503 A
Silver Spring, MD
20993

Registration: <https://bit.ly/2P9f0vF>
BCO Website: <http://biocomputeobjects.org/2019-workshop.html>

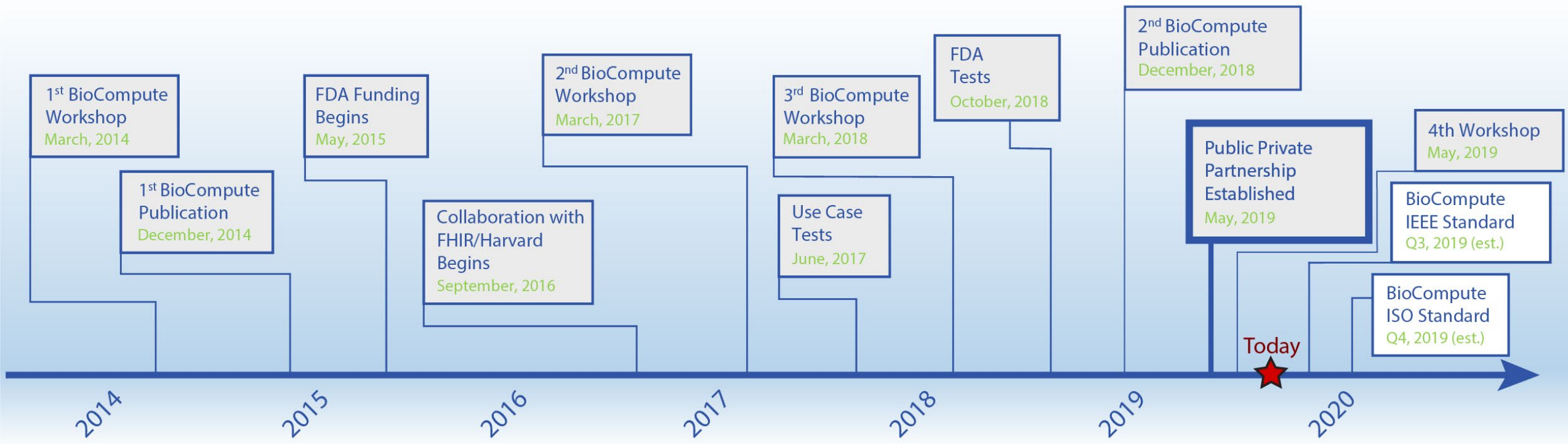
Use Case Examples

- 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

Demos

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

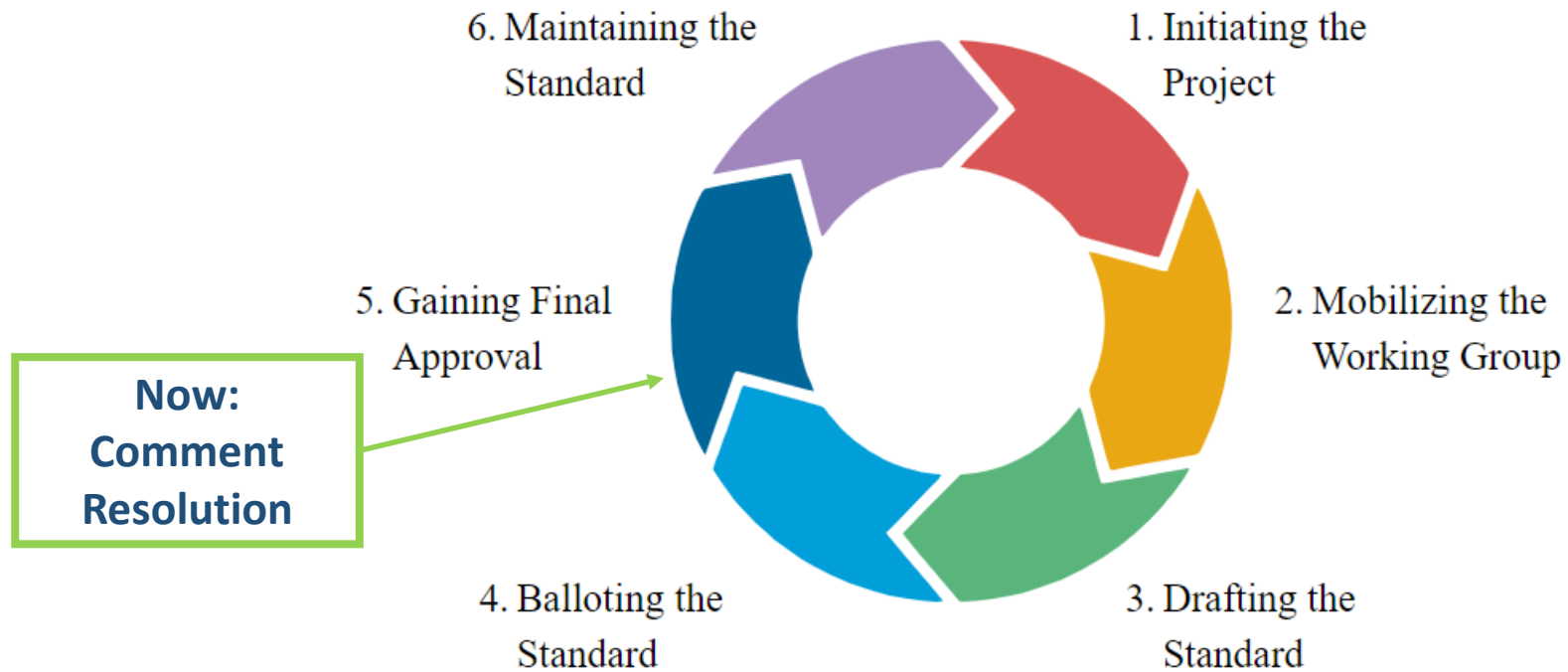
Timeline of Major Events



Future Plans

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

IEEE Standards Development Process



International Standards Organization

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



BioCompute Public-Private Partnership

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





**Raja Mazumder
(Chair)**



Dennis Dean



Carole Goble



**Jonathon Keeney
(Managing Director)**



Vahan Simonyan



Gil Alterovitz



Jeremy Goecks

**Executive
Steering
Committee**



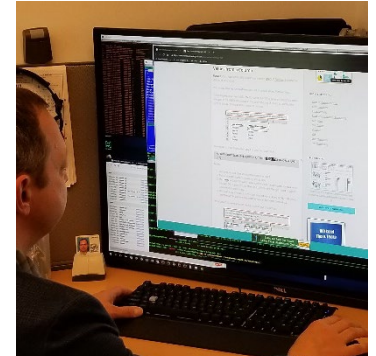
Hadley King (Chair)



Dan Taylor



Jonas Almeida



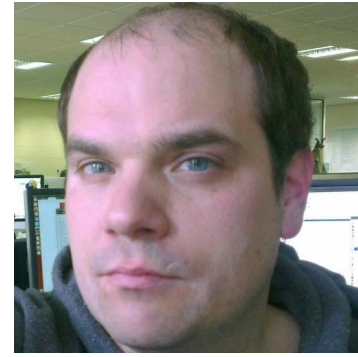
Anton Golikov



Nicola Soranzo



Elaine Thompson



Stian Soiland-Reyes



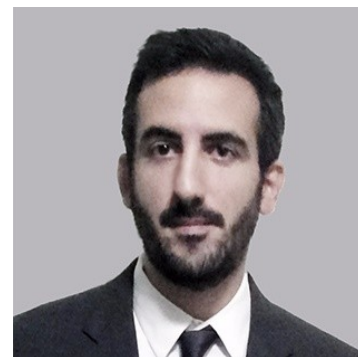
Jason Travis



Ntino Krampis



Michael R. Crusoe



Konstantinos
Karagiannis

Technical
Steering
Committee

BioCompute Object (BCO) App-a-thon



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



THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

precisionFDA 

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



App-a-thon launching May 14th 2019!

BioCompute Object (BCO) App-a-thon



App-a-thon Goals:

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



THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

precision**FDA** 

BioCompute Object (BCO) App-a-thon

May 14 through October 18

	<p>Beginner Track: Create BCOs</p>	<p>Advanced Track: Create BCO software tools</p>	<p>App-a-thon Launch: May 14, 2019 Challenge details are made available</p>
<p>BCOs and Tools</p>	 <p>Complete BCO</p>	 <p>Conformance Tool</p>  <p>Writer Tool</p>	<p>May 14 – October 18, 2019 Participants create BCOs and BCO tools</p>
<p>Conformance Check*</p>	 <p>Ability of an FDA reviewer to understand the BCO</p>	 <p>Ability to generate accurate BCO; Ability to conduct conformance check on BCOs</p>	<p>Deadline: October 18, 2019 Evaluation begins</p>
<p>Results</p>			



BioCompute Object (BCO) App-a-thon



App-a-thon Goals:

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



THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

precision**FDA** 

BioCompute Object (BCO) App-a-thon



Submissions to the Beginner Track must:

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



THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

precision**FDA** 

BioCompute Object (BCO) App-a-thon



Submissions to the Advanced Track must:

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



THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON, DC

precision**FDA** 

App-a-thon site

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

precisionFDA contact information:

Elaine Johanson, Acting Director

Office of the Chief Scientist/Office of Health Informatics

precisionfda@fda.hhs.gov