

Genomics applications in the cloud with the DNAnexus Platform



Andrey Kislyuk

BOSC 2013

The DNAnexus Platform

All

- Alignment
- Annotation
- Debugging
- Export
- Import
- Reports
- RNA-Seq

BWA
Maps reads to a reference genome using the Burrows-Wheeler Alignment (BWA) algorithm.

Reads Reference genome > Mappings

BED Importer
Converts a BED file into one or more Spans/Genes.

BED file Reference genome > Annotations

ContigSet Validator
Validate whether or not a DNAnexus object is a valid ContigSet.

Inputs > App > Outputs

1 input Reads [array] > **BWA** configured > Mappings Indexed

Reference genome

Mappings table(s) [array] > **GATK Indel Realign...** configured > recalibrated_mappings

Reference genome dbSNP

Known Indels [array]

Mappings Objects [array] > **GATK UnifiedGenot...** configured > Reference Genome

```
Processing
#
tar zxvf reference.bowt
if [ "$reads2gz" != "" ]
then
  bowtie2 -t -p 2 -x ref
else
  bowtie2 -t -p 2 -x ref
fi
samtools view -bS out
samtools sort out.b
```

Configurable cloud infrastructure for genomics

The DNAnexus Platform

Open and
comprehensive
API

X DOCS [Downloads](#) · [CLI](#) · [Developers](#) · [API](#) · [FAQ](#) · [DNAnexus Platform](#)

Projects

Projects are special containers intended for collaboration; access permissions are controlled by the users who have ADMINISTER access to the project.

Projects vs Containers

Users can create projects, either through the website or using the API method [/project/new](#). Unless they are creating a project on behalf of an organization, they will be responsible for the storage costs of the data inside the project, as well as any compute time for any jobs run from the project. Projects have various metadata (name, description, tags, properties). The project also has two project-wide boolean flags:

- **PROTECTED**: If set to true, only members with "ADMINISTER" permission level can delete objects. Otherwise, members with "CONTRIBUTE" permission level can also delete objects from the project. To aid in the reproducibility of results from running apps in the project, [project caches](#) are not persistent and are recreated for each time an app is run.
- **RESTRICTED**: If set to true, the objects of this project cannot be cloned, and any running applet that reads from this project cannot write results to any other project (see later discussion for what this means, too).

In contrast, users cannot create generic data containers directly. Containers are generated automatically for different purposes involved in running an analysis, sharing data, etc. and will always be associated with a project or user/organization that will be responsible for its storage costs.

Project Metadata

Projects can have names, descriptions, properties, and tags. Properties may be automatically set by the website front-end and are used for facilitating user experience.

List of API Methods

Project API Methods

The following methods are specified on this page:

- [/project/new](#)
- [/project-xxxx/addTags](#)

DNAnexus API Specification v1.0.0

- **Introduction**
 - Entity IDs
 - Protocols
 - Authentication
- **Users**
- **Data Containers**
 - Folders and Deletion
 - Cloning
 - Projects
 - Project Permissions and Sharing
- **Data Object Lifecycle**
 - Types
 - Details and Links
 - Visibility
- **Data Object Metadata**
 - Name
 - Properties
 - Tags
- **Data Object Classes**
 - Records
 - Files
 - Genomic Tables
- **Running Analyses**
 - I/O and Run Specifications
 - Job Input and Output
 - Applets and Entry Points
 - Apps
 - Containers for Execution
- **Search**
- **Directory of API Methods**

Developer Portal

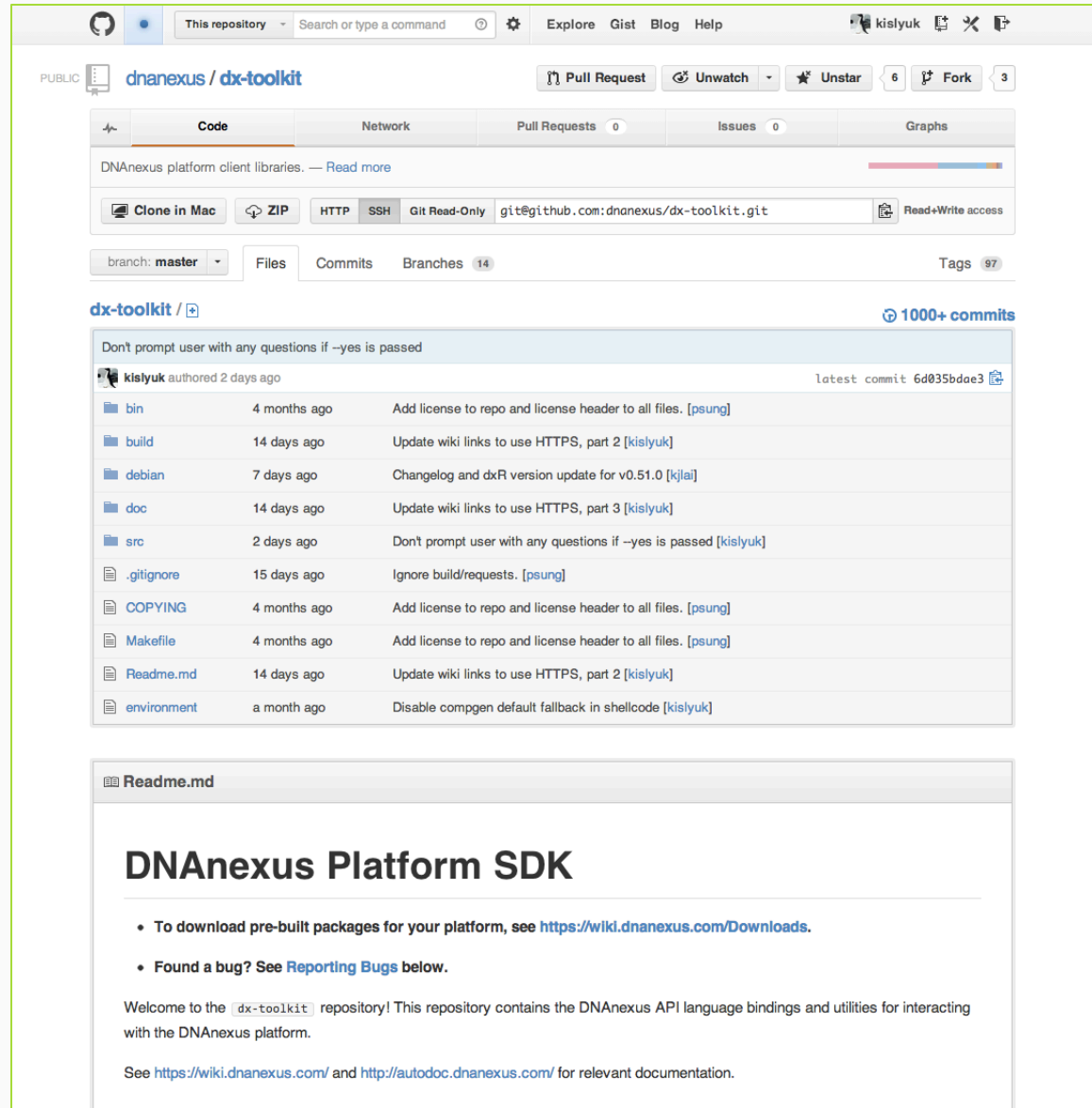
The DNAnexus Platform

Open and
comprehensive
SDK

Linux
OS X

Windows

Upload agents
(Dropbox-like)



The screenshot displays the GitHub interface for the `dnanexus/dx-toolkit` repository. At the top, there are navigation links for 'Explore', 'Gist', 'Blog', and 'Help'. The repository name is prominently displayed, along with options to 'Pull Request', 'Unwatch', 'Unstar', and 'Fork'. Below this, there are tabs for 'Code', 'Network', 'Pull Requests', 'Issues', and 'Graphs'. The main content area shows a list of files and folders, each with its name, the time since the last commit, and a brief description of the changes. The files listed include `bin`, `build`, `debian`, `doc`, `src`, `.gitignore`, `COPYING`, `Makefile`, `Readme.md`, and `environment`. The README section is partially visible, showing the title 'DNAnexus Platform SDK' and a list of instructions for downloading pre-built packages and reporting bugs.

File/Folder	Time since commit	Description
bin	4 months ago	Add license to repo and license header to all files. [psung]
build	14 days ago	Update wiki links to use HTTPS, part 2 [kislyuk]
debian	7 days ago	Changelog and dxR version update for v0.51.0 [kjlai]
doc	14 days ago	Update wiki links to use HTTPS, part 3 [kislyuk]
src	2 days ago	Don't prompt user with any questions if --yes is passed [kislyuk]
.gitignore	15 days ago	Ignore build/requests. [psung]
COPYING	4 months ago	Add license to repo and license header to all files. [psung]
Makefile	4 months ago	Add license to repo and license header to all files. [psung]
Readme.md	14 days ago	Update wiki links to use HTTPS, part 2 [kislyuk]
environment	a month ago	Disable compgen default fallback in shellcode [kislyuk]

DNAnexus Platform SDK

- To download pre-built packages for your platform, see <https://wiki.dnanexus.com/Downloads>.
- Found a bug? See [Reporting Bugs](#) below.

Welcome to the `dx-toolkit` repository! This repository contains the DNAnexus API language bindings and utilities for interacting with the DNAnexus platform.

See <https://wiki.dnanexus.com/> and <http://autodoc.dnanexus.com/> for relevant documentation.

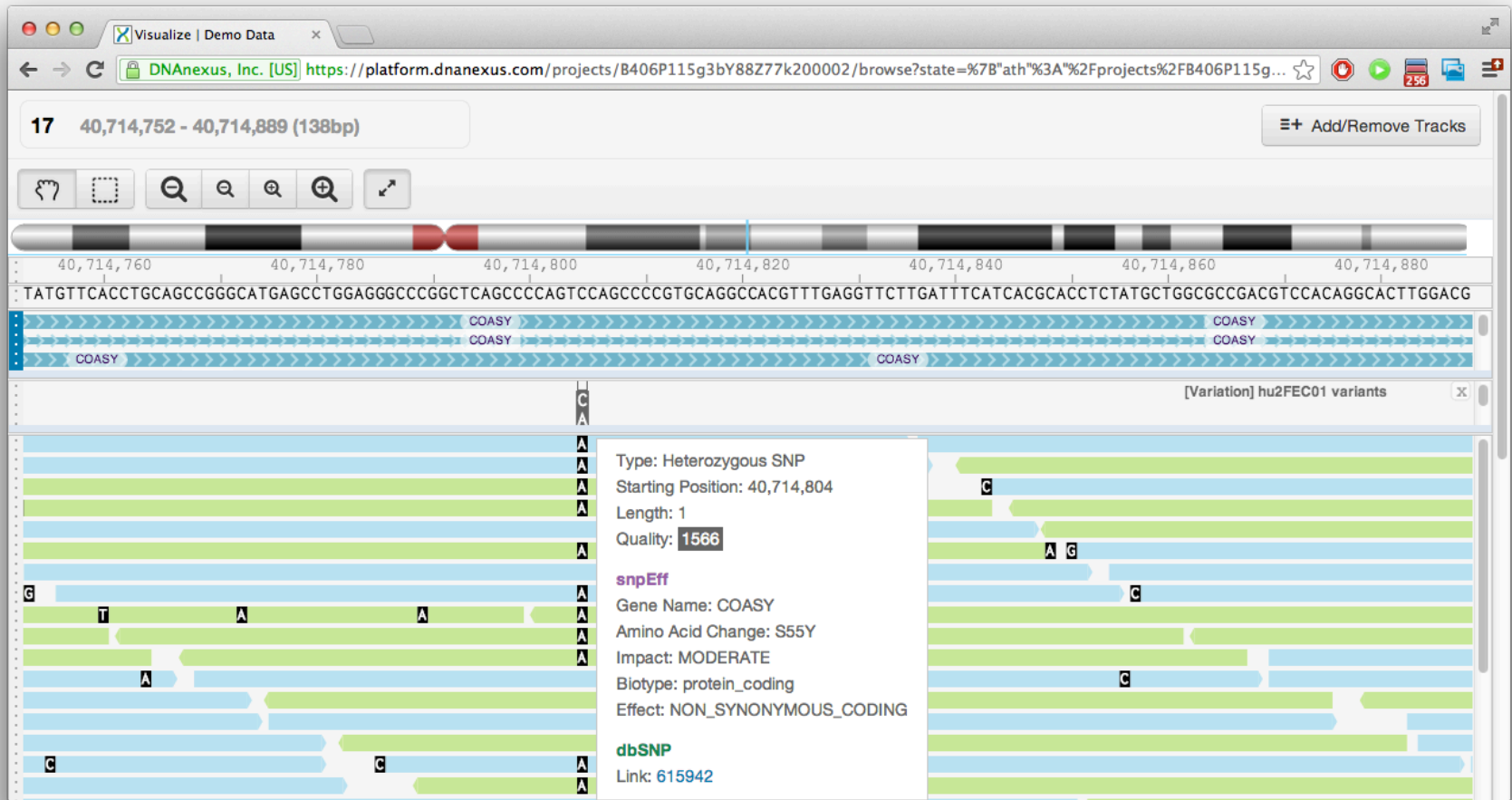
Batteries are included

The DNAnexus Platform



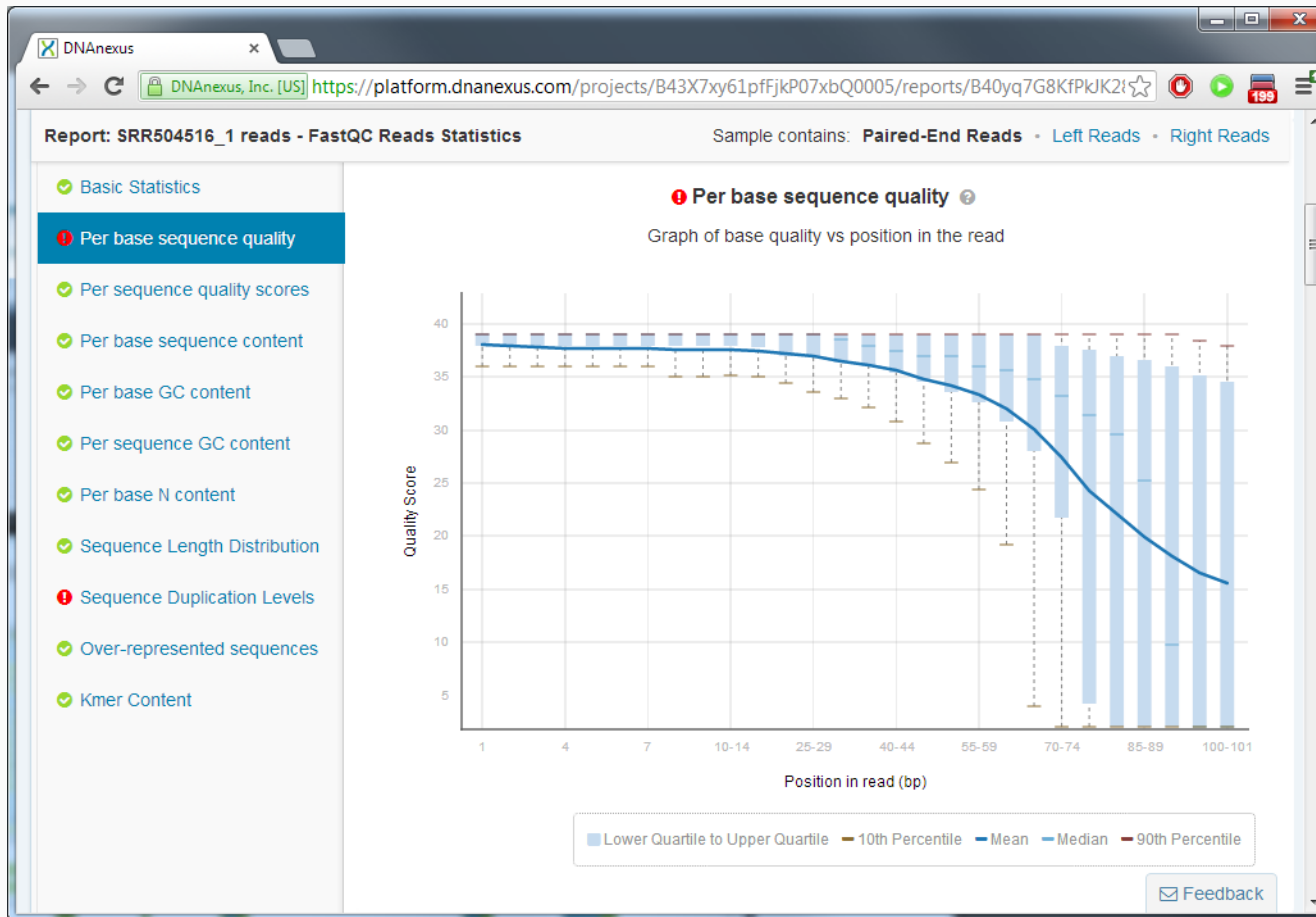
Powerful visualization tools built in

The DNAnexus Platform



Powerful visualization tools built in

The DNAnexus Platform



Quality control reports

Visualize | Vibrio cholerae x

DNAnexus, Inc. [US] https://platform.dnanexus.com/projects/B4G58y8jbfpf41KvV9k0005J/browse

Projects Apps Help ▾ 259 Andrej Kislyuk ▾

Vibrio cholerae ChIPseq 7.56 kb Admin & Private Share

Select Tracks to Visualize

Reference Genome Special tracks: Reference Ruler

All projects ▸ Demo Data

Your visualization engine here
(there's an API for that)

- Demo Data
 - 23andMe Exome Variati...
 - Developer Quicksta...
 - ERP001228 (whole-gen...
 - Illumina BodyMap2 hea...
 - SRR100022

SUGGESTIONS

- Reference Genomes
- Vibrio cholerae ChIPseq

<input type="checkbox"/>	b37 Ensembl Genes r70 /Illumina BodyMap2 heart tissue ChIP-Seq	Jan 30, 2013 6:
<input type="checkbox"/>	ERP001228 mappings /ERP001228 (whole-genome)	Jan 24, 2013 9:
<input type="checkbox"/>	ERR000894 RNA-seq mappings /Illumina BodyMap2 heart tissue RNA-Seq	Jan 29, 2013 3:
<input type="checkbox"/>	SRR100022 mappings to b37 (realigned and recalibrated) /SRR100022	Jan 24, 2013 11:
<input type="checkbox"/>	SRR100022 variants by GATK (annotated) /SRR100022	Jan 25, 2013 1:
<input type="checkbox"/>	hu2FEC01 variants /23andMe Exome Variation	Jan 7, 2013 12:
<input type="checkbox"/>	ERP001228 mappings coverage /ERP001228 (whole-genome)	Feb 27, 2013 5:

Ok

DNAnexus Platform · Downloads · Docs · CLI · Developers · API · Answers

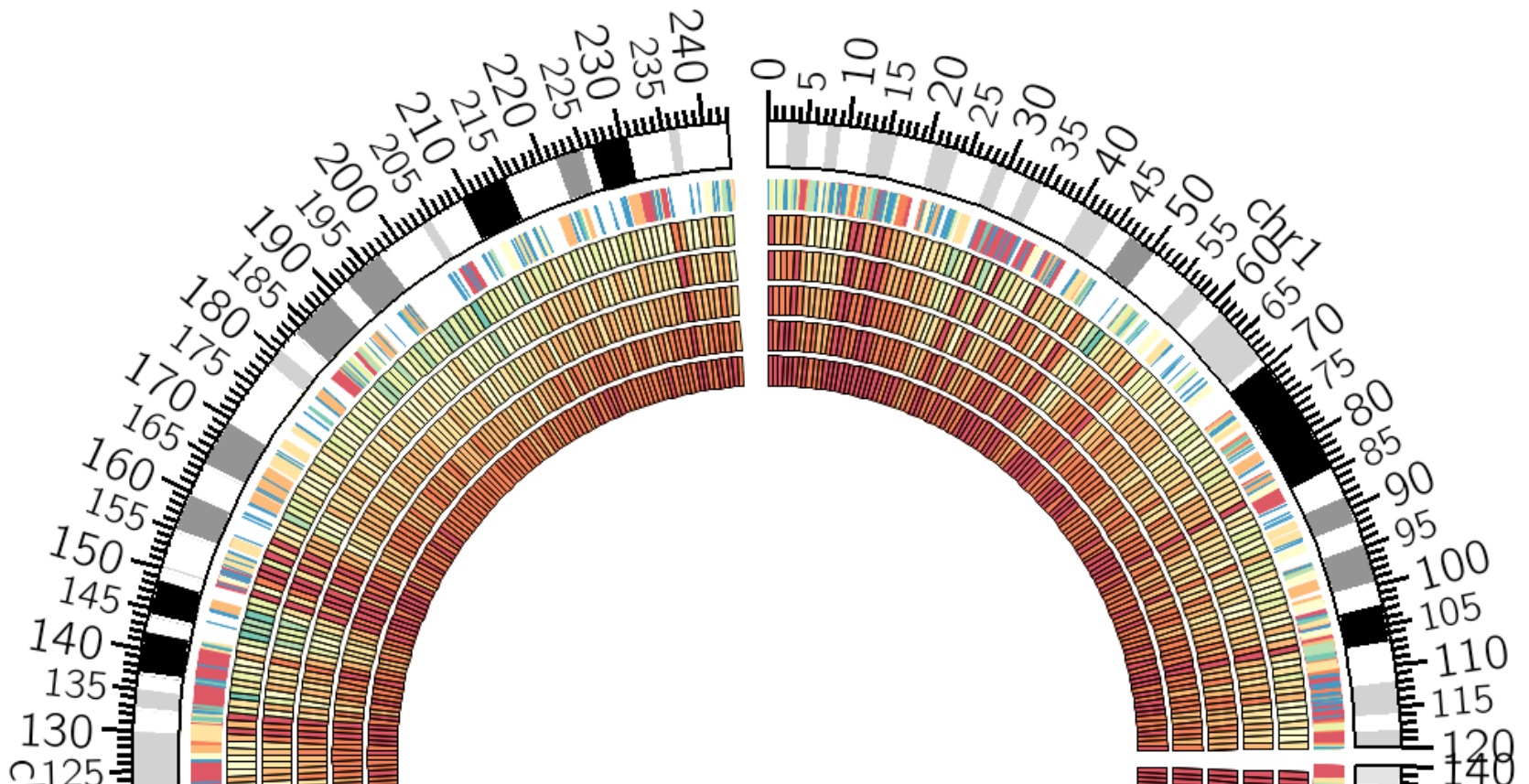
Feedback

Yes, we have Circos

[← Circos](#) 0.10 GB data usage Admin your access Private access policy Share 2 Members

[Manage](#) [Monitor](#) [Visualize](#) ⚙️

Report: Custom Circos Report



**DNAneXus is ready
for clinical data**

Secure, compliant, audited

Data compliance in accordance with:

HIPAA

GCP

CLIA

21 CFR
Part 11

21 CFR
Part 58

21 CFR
Part 493

European Data Privacy laws and regulations

Those are not just acronyms...

- All data encrypted with full-disk AES-256 at rest, SSL on the move
- Production access controls
- Third-party security audits
- Optional 2-Factor Auth
- LXC (Linux Containers) hypervisor
- Auditable by user

Your data is yours

We will **never** hold your data hostage

- Always exportable
- Always downloadable
- We're not allowed to look at it

The DNAnexus Platform

Reliability

- Triple data redundancy
- Geographically distributed
- Job-level hardware fault tolerance
- Reproducible and auditable results for 6+ years

DNAnexus is ready for clinical data

Blazing fast development

Open-source stack

App wizard

walks you through app creation

Learn by example

fork our repos

Collaborate

deploy apps from GitHub

The image shows a composite of two screenshots. The top screenshot is the 'DNAnexus App Wizard, API v1.0.0' interface. It has a dark background and contains the following text: 'Basic Metadata', 'Please enter basic metadata fields that will be used to describe your app. Optional fields are denoted by options with square brackets. At the end of this wizard, the files necessary for building your app will be generated from the answers you provide.', 'The name of your app must be unique on the DNAnexus platform. After creating your app for the first time, you will be able to publish new versions using the same app name. App names are restricted to alphanumeric characters (a-z, A-Z, 0-9), and the characters ".", "-", and "_'.', and 'App Name'. The bottom screenshot is a GitHub page for the 'dnanexus' organization. It features the GitHub logo, a search bar, and a list of repositories. The repositories listed are: 'vcf_importer' (Python, 1 star, 0 forks, 0 mirrors, last updated 30 minutes ago), 'samtools_mpileup' (C, 1 star, 0 forks, 0 mirrors, last updated 33 minutes ago), 'gtf_importer' (Python, 0 stars, 0 forks, 0 mirrors, last updated 39 minutes ago), 'gff_importer' (Python, 0 stars, 0 forks, 0 mirrors, last updated 41 minutes ago), 'dx-toolkit' (C++, 6 stars, 0 forks, 0 mirrors, last updated 42 minutes ago), 'gatk_variant_annotator' (C, 0 stars, 0 forks, 0 mirrors, last updated an hour ago), and 'gatk_unifiedgenotyper' (C, 1 star, 0 forks, 0 mirrors, last updated an hour ago). The page also shows 34 public repositories and 8 members for the dnanexus organization.

Debug quickly

```
2. home:3 [bash] (bash)
home:3 [bash] (bash) 961
kislyuk@aurora:~>dx watch job-B6K14F05fyY6XPPfv4g0000P
Watching job job-B6K14F05fyY6XPPfv4g0000P. Press Ctrl+C to stop.
* GATK Pipeline - BWA (re-run) (gatk_pipeline:main) (done) job-B6K14F05fyY6XPPfv4g0000P
  vince 2013-05-21 17:32:44 (runtime 0:02:11)
2013-05-21 17:33:02 GATK Pipeline - BWA (re-run) INFO Logging initialized (priority)
2013-05-21 17:33:02 GATK Pipeline - BWA (re-run) INFO Logging initialized (bulk)
2013-05-21 17:33:09 GATK Pipeline - BWA (re-run) STDOUT Installing apt packages openjdk-6-jre-headless tabix
  pypy dx-toolkit-beta
2013-05-21 17:33:39 GATK Pipeline - BWA (re-run) STDOUT >>> Unpacking resources.tar.gz to /
2013-05-21 17:33:41 GATK Pipeline - BWA (re-run) STDERR python running (job ID job-B6K14F05fyY6XPPfv4g0000P)
2013-05-21 17:33:41 GATK Pipeline - BWA (re-run) STDOUT Recalibrated Table: gtable-B6K14k85fyY9x7fFx13Q00FK
2013-05-21 17:33:42 GATK Pipeline - BWA (re-run) STDERR * Starting dx-contigset-to-fasta...
2013-05-21 17:33:42 GATK Pipeline - BWA (re-run) STDERR - Getting details for ContigSet record-B6Jx5vQ5fyYBz
  8G4X9b00054...
2013-05-21 17:33:42 GATK Pipeline - BWA (re-run) STDERR {"contigs":{"names":["1"],"offsets":[0],"sizes":[249
  250621]},"flat_sequence_file":{"$dnanexus_link":"file-B6Jx5Q85fyY9QK07zj10001f"}}
2013-05-21 17:33:42 GATK Pipeline - BWA (re-run) STDERR - Downloading sequence for chromosome 1 (offset = 0,
  length = 249250621)...
2013-05-21 17:33:54 GATK Pipeline - BWA (re-run) STDERR - writing FASTA...
2013-05-21 17:34:21 GATK Pipeline - BWA (re-run) STDERR * Finished dx-contigset-to-fasta.
2013-05-21 17:35:06 GATK Pipeline - BWA (re-run) STDOUT splits: 1
2013-05-21 17:35:06 GATK Pipeline - BWA (re-run) STDOUT [0]
2013-05-21 17:35:06 GATK Pipeline - BWA (re-run) STDOUT -L 1:10000000-11000000
2013-05-21 17:35:06 GATK Pipeline - BWA (re-run) STDOUT [u' -L 1:10000000-11000000']
2013-05-21 17:35:06 GATK Pipeline - BWA (re-run) STDOUT [u' -L 1:10000000-11000000']
* GATK Pipeline - BWA (re-run) (gatk_pipeline:main) (done) job-B6K14F05fyY6XPPfv4g0000P
  vince 2013-05-21 17:32:44 (runtime 0:02:11)
  Output: recalibrated_mappings = gtable-B6K14k85fyY9x7fFx13Q00FK
  variants = gtable-B6K1Fbj5fyY0VgyZ8vjQ00Xf
kislyuk@aurora:~>
```

Type Directory - DNAnexus

https://wiki.dnanexus.com/Type-Directory

Type Directory

The following Types have been defined to facilitate interoperability between applications in the DNAnexus Platform. For more info about using Types, see one of the following:

- [Data Type Conventions](#) for an introduction to types.
- [Types in the API Specification](#) for information about how to use types via the API.
- [Conventions for language conventions](#) that are used in the type specifications below.

Foundations for

Types

3. python2.7

python2.7 961

Row	chr	lo	hi	name	span_id	type	strand	score	is_coding	parent_id	frame	desc
524288	15	65297191	65297274	ENST00000220058.8	568429	CDS	-	-2147483648	True	53758	-1	
524289	15	65297191	65297274	ENST00000560717.7	568454	3' UTR	-	-2147483648	True	53760	-1	
524290	15	65297191	65297274	ENST00000558460.8	568441	CDS	-	-2147483648	True	53759	-1	
524291	15	65298450	65298529	ENST00000560717.6	568453	3' UTR	-	-2147483648	True	53760	-1	
524292	15	65298450	65298529	ENST00000558460.7	568440	CDS	-	-2147483648	True	53759	-1	
524293	15	65298450	65298529	ENST00000220058.7	568428	CDS	-	-2147483648	True	53758	-1	
524294	15	65308170	65308865	ENST00000543678.4	568462	3' UTR	-	-2147483648	True	53761	-1	
524295	15	65308170	65321977	MTFMT	53761	transcript	-	-2147483648	True	53757	-1	
524296	15	65308592	65308865	ENST00000558614.5	568467	exon	-	-2147483648	False	53762	-1	
524297	15	65308592	65321912	MTFMT	53762	transcript	-	-2147483648	True	53757	-1	

1212362 more rows

(END)

Report: A specification for how to generate a report that can be viewed in the platform. It may also be one of the following subtype. Feedback

Massive on-demand compute



Instant-on supercomputer
at your fingertips...

...only when you need it

Spin up thousands of instances

Specify instance types

Pay by the second



Community Collaboration

DNAnexus Answers

The screenshot displays the DNAnexus Answers forum interface. At the top, there is a search bar and navigation links for Posts, Tags, Users, Badges, and Log In. Below the navigation, there are tabs for different content types: Show All, Questions¹³, Unanswered⁵, Forum, Tips & Tutorials, Feature & App Requests, and News¹. The main content area shows a list of questions with their respective statistics (votes, answers, views) and titles. The questions are:

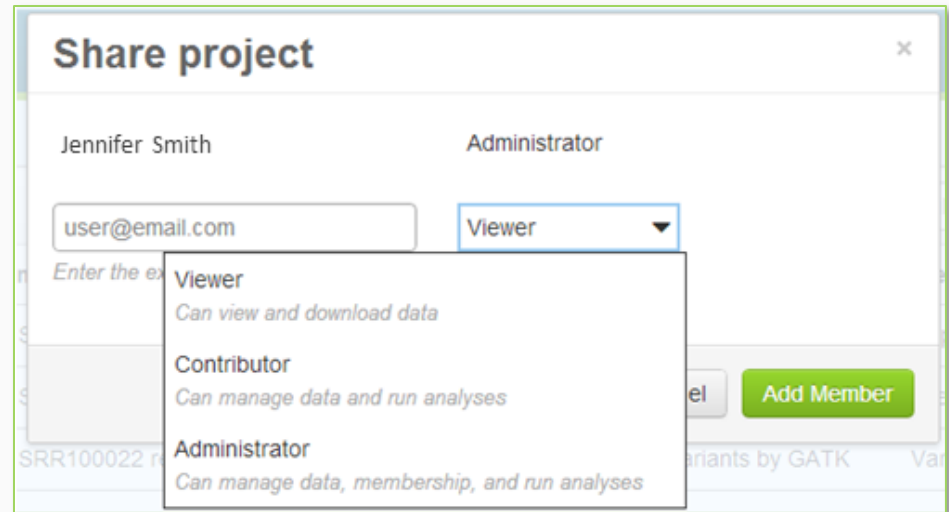
- How do I connect to an FTP server from a worker?** (0 votes, 1 answer, 10 views)
- I can't log-in to the new platform with my classic password.** (0 votes, 1 answer, 6 views)
- How do I package a Linux executable into an app?** (0 votes, 1 answer, 10 views)
- What's the difference between an app and an applet?** (0 votes, 3 answers, 25 views)
- How do I install software requirements for my app?** (0 votes, 1 answer, 7 views)
- How do I write my app in my favorite programming language?** (0 votes, 1 answer, 17 views)
- What are the default user limits for processes running inside the Linux execution environment?** (1 vote, 1 answer, 10 views)
- How do I request more memory/CPU for my app?** (0 votes, 1 answer, 9 views)

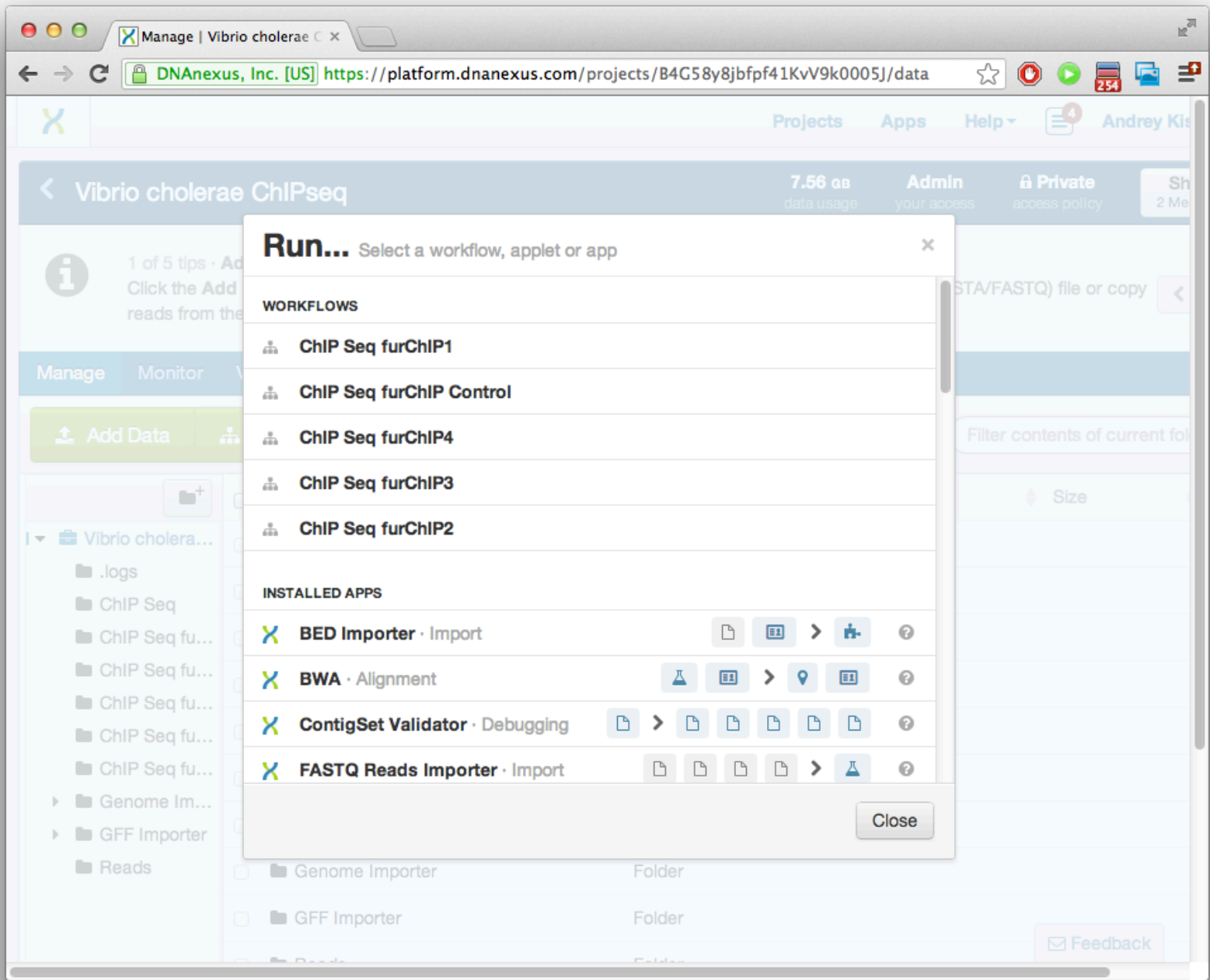
On the right side, there are sections for News (dx-toolkit 0.48.2 released, dx-toolkit 0.39.1 released [closed]), Recent Tags (fasta, fastq, bash, array, json, dx-jobutil-parse-link, linux, developers, apps, errors, data-files, data-upload, log-in, sign-up), and Important Tags (dx-toolkit, dxr).

A community working together

Instant collaboration

- **Eliminate** data transfer headaches
- **Collaborate** on data, tools, workflows in one environment
- **Enable** bioinformatics experts to deliver tools to biologists





Vibrio cholerae CHIPseq

7.56 GB data usage Admin your access Private access policy

1 of 5 tips · Add Click the Add reads from the

Manage Monitor

Add Data

- Vibrio cholerae
 - .logs
 - ChIP Seq
 - ChIP Seq fu...
 - ChIP Seq fu...
 - ChIP Seq fu...
 - ChIP Seq fu...
 - ChIP Seq fu...
 - ChIP Seq fu...
 - Genome Im...
 - GFF Importer
 - Reads

Run... Select a workflow, applet or app

WORKFLOWS

- ChIP Seq furChIP1
- ChIP Seq furChIP Control
- ChIP Seq furChIP4
- ChIP Seq furChIP3
- ChIP Seq furChIP2

INSTALLED APPS

- BED Importer · Import
- BWA · Alignment
- ContigSet Validator · Debugging
- FASTQ Reads Importer · Import

Close

ST/FASTQ) file or copy

Filter contents of current fol

Size

Feedback

Workflow | Trio Exome

DNAexus, Inc. [US] <https://platform.dnanexus.com/projects/B5Xv0vx96fk1vP8Jj0V0008X/workflows/B...>

Projects Apps Help 4 Andrey Kislyuk

Andrey's demos 39.20 GB data usage Admin your access Private access policy Share 1 Member

Manage Monitor Visualize

Trio Exome Saved 2 apps configured Run

Inputs	App	Outputs
<ul style="list-style-type: none">1 input Reads [array]b37 (indexed for BWA) Reference genome	BWA configured	<ul style="list-style-type: none">MappingsIndexed reference genome
<ul style="list-style-type: none">1 input Mappings table(s) [array]b37 Reference genomeSRR100022_1.filt.fastq.gz dbSNPKnown Indels [array]	GATK Indel Realig... configured	<ul style="list-style-type: none">recalibrated_mappings

Add a Step

Feedback

So how do I add my app?

1. Run App Wizard

```
kislyuk@aurora:~>dx-app-wizard
DNAnexus App Wizard, API v1.0.0
```

Basic Metadata

Please enter basic metadata fields that will be used to describe your app. Optional fields are denoted by options with square brackets. At the end of this wizard, the files necessary for building your app will be generated from the answers you provide.

The **name** of your app must be unique on the DNAnexus platform. After creating your app for the first time, you will be able to publish new versions using the same app name. App names are restricted to alphanumeric characters (a-z, A-Z, 0-9), and the characters ".", "_", and "-".

App Name: spades

The **title**, if provided, is what is shown as the name of your app on the website. It can be any valid UTF-8 string.

Title : SPAdes

The **summary** of your app is a short phrase or one-line description of what your app does. It can be any UTF-8 human-readable string.

Summary : SPAdes assembler

The **description** of your app is a longer piece of text describing your app. It can be any UTF-8 human-readable string, and it will be interpreted using Markdown (see <http://daringfireball.net/projects/markdown/syntax/> for more details).

Description :

app spec

```
1 {
2   "name": "spades2",
3   "title": "Spades",
4   "dxapi": "1.0.0",
5   "version": "0.0.1",
6   "inputSpec": [
7     {
8       "name": "forward_reads",
9       "class": "file",
10      "optional": false
11    },
12    {
13      "name": "reverse_reads",
14      "class": "file",
15      "optional": false
16    }
17  ],
18  "outputSpec": [
19    {
20      "name": "assembly",
21      "class": "file"
22    }
23  ],
24  "runSpec": {
25    "interpreter": "bash",
26    "file": "src/spades2.sh"
27  }
28 }
29
```

Line 1, Column 1

Spaces: 2

JSON

2. Add entry point code

```
spades2.sh — spades
OPEN FILES
FOLDERS
  ▼ spades
    ► resources
    ▼ src
      spades2.sh
    ► test
      dxapp.json
      Readme.developer.md
      Readme.md
  spades2.sh x
1  #!/bin/bash
2  # spades2 0.0.1
3  # Generated by dx-app-wizard.
4  #
5  # Basic execution pattern: Your app will run on a single machine from
6  # beginning to end.
7  #
8  # Your job's input variables (if any) will be loaded as environment
9  # variables before this script runs. Any array inputs will be loaded
10 # as bash arrays.
11 #
12 # Any code outside of main() (or any entry point you may add) is
13 # ALWAYS executed, followed by running the entry point itself.
14 #
15 # See https://wiki.dnanexus.com/Developer-Portal for tutorials on how
16 # to modify this file.
17
18 main() {
19
20     echo "Value of forward_reads: '$forward_reads'"
21     echo "Value of reverse_reads: '$reverse_reads'"
22
23     # The following line(s) use the dx command-line tool to download your file
24     # inputs to the local file system using variable names for the filenames. To
25     # recover the original filenames, you can use the output of "dx describe
26     # "$variable" --name".
27
28     dx download "$forward_reads" -o forward_reads
29     dx download "$reverse_reads" -o reverse_reads
30
31     # Fill in your application code here.
32     #
33     # To report any recognized errors in the correct format in
34     # $HOME/job_error.json and exit this script, you can use the
```

Line 1, Column 1

Spaces: 4 Shell Script (Bash)

```
spades2.sh — spades
OPEN FILES
FOLDERS
▼ spades
  ► resources
  ▼ src
    spades2.sh
  ► test
    dxapp.json
    Readme.developer.md
    Readme.md
spades2.sh
31 # Fill in your application code here.
32 #
33 # To report any recognized errors in the correct format in
34 # $HOME/job_error.json and exit this script, you can use the
35 # dx-jobutil-report-error utility as follows:
36 #
37 # dx-jobutil-report-error "My error message"
38 #
39 # Note however that this entire bash script is executed with -e
40 # when running in the cloud, so any line which returns a nonzero
41 # exit code will prematurely exit the script; if no error was
42 # reported in the job_error.json file, then the failure reason
43 # will be AppInternalError with a generic error message.
44
45 ... spades.py -1 forward_reads -2 reverse_reads -o assembly
46
47 tar -cf assembly.tar.gz assembly
48
49 # The following line(s) use the dx command-line tool to upload your file
50 # outputs after you have created them on the local file system. It assumes
51 # that you have used the output field name for the filename for each output,
52 # but you can change that behavior to suit your needs. Run "dx upload -h"
53 # to see more options to set metadata.
54
55 assembly=$(dx upload assembly.tar.gz --brief)
56
57 # The following line(s) use the utility dx-jobutil-add-output to format and
58 # add output variables to your job's output as appropriate for the output
59 # class. Run "dx-jobutil-add-output -h" for more information on what it
60 # does.
61
62 dx-jobutil-add-output assembly "$assembly" --class=file
63 }
64
```

2 lines, 60 characters selected

Spaces: 4 Shell Script (Bash)



3. Build, test, publish

```
kislyuk@aurora:~/Desktop/projects>dx-build-app spades
WARNING:dxpy:app is missing a summary, please add one in the "summary" field of dxapp.json
Created temporary project project-B6YGgkxJVY8XbX507GYQ00k7 to build in
DEBUG:dxpy:Building in /Users/kislyuk/Desktop/projects/spades
DEBUG:dxpy:Uploading in spades
Created applet applet-B6YGjJVJY8XbX507GYQ00kK successfully
Will create app with spec: {u'name': u'spades', u'title': u'Spades', u'outputSpec': [{u'name': u'assembly',
  u'class': u'file'}], u'runSpec': {u'interpreter': u'bash', u'file': u'src/spades2.sh'}, u'version': u'0.0.1
', u'inputSpec': [{u'optional': False, u'name': u'forward_reads', u'class': u'file'}, {u'optional': False, u
'name': u'reverse_reads', u'class': u'file'}], u'dxapi': u'1.0.0'}
Attempting to create version 0.0.1...
App spades/0.0.1 does not yet exist
Created app app-B6YGjJb333PXbX507GYQ00kQ
Uploaded app spades/0.0.1 (app-B6YGjJb333PXbX507GYQ00kQ) successfully
You can publish this app with:
  dx api app-spades/0.0.1 publish "{\"makeDefault\": true}"
kislyuk@aurora:~/Desktop/projects>
```

done

App | SPAdes

← → ↻ <https://staging.dnanexus.com/app/spades> ☆ 🔒 🎮 259 📄 ☰

 Projects Apps Help ▾  5 Andrey Kislyuk ▾

SPAdes

Added by Andrey Kislyuk Developer

Info Versions Developer notes

SPAdes Genome Assembler

Inputs

- **forward_reads** :
- **reverse_reads** :

Outputs

- **assembly** :

Run this app from the command line

```
$ dx run spades
# For help specifying
inputs:
$ dx run spades -h
```

To get dx, download the [Platform SDK](#).

Pricing

Compute cost (variable)


Permissions


No special permissions

Latest Update

Version
0.0.1+build.20130530.0703 May 30, 2013

Added by

 Andrey Kislyuk

 Feedback

SPAdes | Workflow

https://staging.dnanexus.com/projects/B407bKFZQ06jzxvyYJKQ002J/data/SRR100022

Projects Apps Help 5 Andrey Kislyuk

Run Analysis for SPAdes View job progress in your project's Monitor tab.

SPAdes 1 app unconfigured **Run**

Inputs > App > Outputs

forward_reads reverse_reads > SPAdes set Inputs > assembly

Add a Step

Close

<input type="checkbox"/>	📍 SRR100022 reads mapped to b37	Mappings	Ready	180,042,082 rows	Feb :
<input type="checkbox"/>	📁 SRR100022 reads	Reads	Ready	90,021,041 rows	Feb :
<input type="checkbox"/>	📄 SRR100022_1.fastq.gz	File	Ready	5.54 GB	Feb :
<input type="checkbox"/>	📄 SRR100022_2.fastq.gz	File	Ready	5.56 GB	Feb :

Feedback

App | BWA

DNAnexus, Inc. [US] <https://platform.dnanexus.com/app/bwa/info>

Projects Apps Help ▾ 4 Andrey Kislyuk ▾

BWA

Added by DNAnexus

Developer

Info Versions Developer notes

Maps reads to a reference genome using the Burrows-Wheeler Aligner

Burrows-Wheeler Aligner (BWA) is an efficient program that aligns relatively short nucleotide sequences against a long reference sequence such as the human genome. This app runs BWA to map **letterspace** reads to a reference genome and produce mappings.

Inputs:

Reads: An array of Reads table objects that will be mapped to the reference genome. If more than one Reads object is provided, the results are combined into a single Mappings output.

Output name: The name of the resulting Mappings table object (optional; if not provided, the name will be based on the Reads name).

Reference genome: The genome that the reads will be mapped against. BWA requires a special processing on the genome, called indexing; this processing can take several hours for long genomes. If the genome given in the input is not indexed for BWA, the app will automatically index it and include an indexed version in the output, for future use. When possible, please run this app with an indexed genome to avoid re-indexing. DNAnexus provides several pre-indexed genomes in the 'Reference Genomes' public project.

Mapping algorithm: BWA implements two different algorithms, both based on Burrows-Wheeler Transform (BWT). The first algorithm, called 'aln' is designed for short queries up to ~200bp with low error rate (<3%). It does gapped global alignment w.r.t. queries, supports paired-end reads, and is one of the fastest short read alignment algorithms to date while also visiting suboptimal hits. The second algorithm, called 'bwasw', is designed for long reads with more errors. It performs heuristic Smith-Waterman-like alignment to find high-scoring local hits (and thus chimera). On low-error short queries, 'bwasw' is slower and less accurate than the first algorithm, but on long queries, it is better. Using a value of 'auto' will automatically choose a suitable algorithm based on the length of the reads in the

Run this app from the command line

```
$ dx run bwa
# For help specifying
inputs:
$ dx run bwa -h
```

To get dx, download the [Platform SDK](#).

Pricing

Compute cost (variable)

Permissions ⓘ

No special permissions


Latest Update


Version 1.0.4 May 23, 2013

Upstream Project

Version: 0.6.2-r126
License: GPLv3

Added by

 George Asimenos

 Andrew Carroll

[Feedback](#)

less accurate than the first algorithm, but on long queries, it is better. Using a value of 'auto' will automatically choose a suitable algorithm based on the length of the reads in the input (however, this is currently not implemented, and using 'auto' will lead to the 'aln' algorithm always being chosen).

Reads per chunk: This app parallelizes itself by dividing the input into chunks of a certain size (a certain number of reads), and mapping each chunk individually. Lower chunk sizes lead to higher levels of parallelization, reducing the wall-clock time that one has to wait for the app to finish. However, lower chunks sizes may also increase the cost of running the app in the cloud, as they lead to a higher number of chunks, each of which adds a constant processing overhead. The default value is 25 million reads per chunk; DNAnexus suggests caution when experimenting with this parameter.

Discard unmapped rows?: If selected, unmapped reads will not be included in the Mappings output.

Low-level parameters: Users familiar with the BWA executable can directly manipulate the parameters that are used for the `bwa aln`, `bwa samse`, `bwa sampe` and `bwa bwasw` calls. These parameters are: `aln_n`, `aln_o`, `aln_e`, `aln_i`, `aln_d`, `aln_l`, `aln_k`, `aln_m`, `aln_M`, `aln_O`, `aln_E`, `aln_R`, `aln_q`, `aln_N`, `sampe_a`, `sampe_o`, `sampe_n`, `sampe_N`, `sampe_c`, `sampe_s`, `sampe_n`, `sw_a`, `sw_b`, `sw_q`, `sw_r`, `sw_w`, `sw_m`, `sw_T`, `sw_c`, `sw_z`, `sw_s`, `sw_N`. Each one of these parameters directly correspond to the respective command-line argument, e.g. `aln_o` corresponds to the `-o` option of `bwa aln` (maximum number of gap opens). Certain options, such as the `-t` option of `bwa aln`, are not exposed to users because they are set by the app, based on the kind of cloud environment that the app runs on.

Outputs:

Mappings: The mappings produced by BWA are output in Mappings table object. (Developers can look at <http://wiki.dnanexus.com/Types/Mappings> for more information). Mappings objects can be then used as inputs to certain variation calling apps, mappings QC apps, etc.; they can also be visualized in the genome browser.

Indexed reference genome: An indexed version of the reference genome, for future use as input to this app. As mentioned earlier, if the app is given a reference genome that is not indexed for BWA, it will index it. This output contains the indexed version so that you can provide it as input to future invocations of the app.

CATEGORIES [Read Mapping](#)

SOURCE [Browse source code](#)

Fork this app: `git clone git@github.com:dnanexus/bwa.git`

CITATIONS

- Li, H., Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, doi:10.1093/bioinformatics/btp324
- Li, H., Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, doi:10.1093/bioinformatics/btp698

License: GPLv3

Added by

- Vince Ramey
- Andrew Carroll
- George Asimenos
- Joseph Dale
- Phil `<script>alert('uh-oh')`
- Jenkins Staging

[Support](#)

Reproducibility

Ever try to reproduce results from a bioinformatics paper?

How about **CLIA compliance**?

All objects are versioned

Analysis I/O is read-only

Jobs enter into project's *permanent record*

Publishing

Authors who publish their software

- Don't worry about supporting diverse installs
 - You installed my package on WHAT?
- Leverage all Platform features
 - Accessible UI
- Compose with other apps
 - It's an ecosystem
 - Publish your workflows as apps, too

Recognition

Users are encouraged to cite app authors

One-click bibliographies coming soon

The screenshot shows a workflow interface for a project named "Vibrio cholerae ChIPseq". The top navigation bar includes "Projects", "Apps", "Help", and a user profile for "Andrey Kislyuk". The project details show "7.56 GB data usage", "Admin your access", "Private access policy", and a "Share" button for "2 Members". Below the project name, there are tabs for "Manage", "Monitor", and "Visualize". The main workflow area is titled "ChIP Seq furCHIP4" and is marked as "Saved". It features a "Run" button and a "3 apps done" indicator. The workflow is organized into three horizontal tracks, each with "Inputs", "App", and "Outputs" sections. The first track shows "1 input Reads [array]" and "Reference genome" as inputs, leading to a "BWA" app (job done), which produces "Mappings" and "Indexed reference genome" outputs. The second track shows "via BWA Mappings" and "Reference Genome" as inputs, leading to a "Coverage Histogram..." app (job done), which produces a "Coverage Histogram" output. The third track shows "via BWA Mappings" and "Reference Genome" as inputs, leading to a "ChIPseq Analysis" app (job done), which produces "Peak Table" and "Enrichment Track" outputs. A "GFF of Reference Genome" input is also shown at the bottom of the third track. An "Add a Step" button is visible at the bottom of the interface.

**We care about
developers**

DNAexus developer program

- \$1000 credit
- App Bounties
- Featured Apps

Vision

DNAnexus roadmap

- AWS Glacier – *cheap data archival*
- Projects as publications
Supplementary materials!

And of course...

- More apps

DNAnexus roadmap

Always improving...

...And so are our cloud providers

We pass on the savings to you

DNAnexus is the platform for
publishing your algorithms

DNAnexus is the platform for
delivering genomics results to users

Acknowledgments

DNAAnexus is
Andreas Sundquist
Arend Sidow
Serafim Batzoglou

Major Investors



Thank you

dnanexus.com