

Sep 01, 2024

ICA Data Ingestion Protocol

DOI

dx.doi.org/10.17504/protocols.io.n92ld8188v5b/v1

Gordon Qian¹, Ryan Davis¹, Jennifer Johnston²

¹University of Sydney; ²NysnoBio

ASAP Collaborative Rese...

Team Kirik



courtney.wright Wright

University of Sydney

OPEN  ACCESS



DOI: **dx.doi.org/10.17504/protocols.io.n92ld8188v5b/v1**

Protocol Citation: Gordon Qian, Ryan Davis, Jennifer Johnston 2024. ICA Data Ingestion Protocol. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.n92ld8188v5b/v1>

License: This is an open access protocol distributed under the terms of the **[Creative Commons Attribution License](#)**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: August 30, 2024

Last Modified: September 01, 2024

Protocol Integer ID: 106735

Keywords: ASAPCRN, Neuronal Genome Atlas for Parkinsons (NGAP), Illumina Connected Analytics (ICA), Whole Genome Sequencing, Variants, iPSC

Funders Acknowledgement:

Michael J Fox Foundation

Grant ID: ASAP-000497



Disclaimer

User account required to access ICA

Abstract

Step-by-step protocol for uploading genome sequencing data into the Neuronal Genome Atlas for Parkinsons (NGAP). This Atlas resource is hosted on Illumina Connected Analytics (ICA). The Ingestion Protocol outlines main data upload using a command line interface (CLI), and offers alternative data ingestion methods to provide broad access across the ASAP CRN. Alternative methods provided include: the Amazon Web Services (AWS), the Illumina Service Connector, or Illumina BaseSpace.

Materials

- Cell line genome sequencing data

Before start

User account required to access ICA.

This protocol details the use of a cloud based platform for analysis and storage of human patient cell line sequencing data, including iPSC and fibroblast whole genome sequence, using the Illumina suite of analysis software and assets. This uses DRAGEN and other DNA sequence analysis tools. Host capability for custom pipeline design, dockerisation of tools, and various other analyses.



Data ingestion methods

- 1 Data ingestion can be performed via different methods where one may be preferred over the other due to data storage location or the user's operating system.

- 2 **Primary Upload method**

Illumina's **Command Line Interface (CLI)** which can be executed locally or from high-performance computing (HPC) servers where large sequencing data is typically located. We provide detailed instructions below for data upload using this method into **Illumina Connected Analytics (ICA)** Database Platform

- 3 **Other Upload Methods**

Data Upload from AWS

- If your data is located on Amazon Web Services (AWS), you may directly upload the data onto ICA using the AWS CLI.
- Instructions for this can be found here: <https://help.ica.illumina.com/tutorials/datatransfer#aws-cli>

Data Upload via Service Connector

- Illumina's Service Connector allows users to sync data between the platform's cloud-hosted data store and a user's local machine, which can be your computer or server.
- Instructions for this can be found here: <https://help.ica.illumina.com/project/p-connectivity/service-connector>

Data Upload from BaseSpace

- If your data is located on BaseSpace, you may copy your data onto ICA through Illumina's "BaseSpace to ICA Data Copy" application.

Command Line Interface (CLI)

- 4 **CLI Installation**

- The latest CLI installation file can be found here: <https://help.ica.illumina.com/command-line-interface/cli-releasehistory>
- Depending on the user's operating system installation instructions will differ.

Mac/Linux Instructions:

1. Place the CLI in a folder that is included in your \$PATH environment variable list of paths, typically /usr/local/bin.
2. You will also need to make the file executable so that the CLI can run:



```
sudo chmod a+x /usr/local/bin/icav2
```

Windows Instructions:

1. Place the CLI in a folder that is included in your \$PATH environment variable list of paths. In Windows this is typically in the "C:\Program Files" folder.
2. If you do not have write access to that folder, open a CMD window in administrator mode and type the following commands:

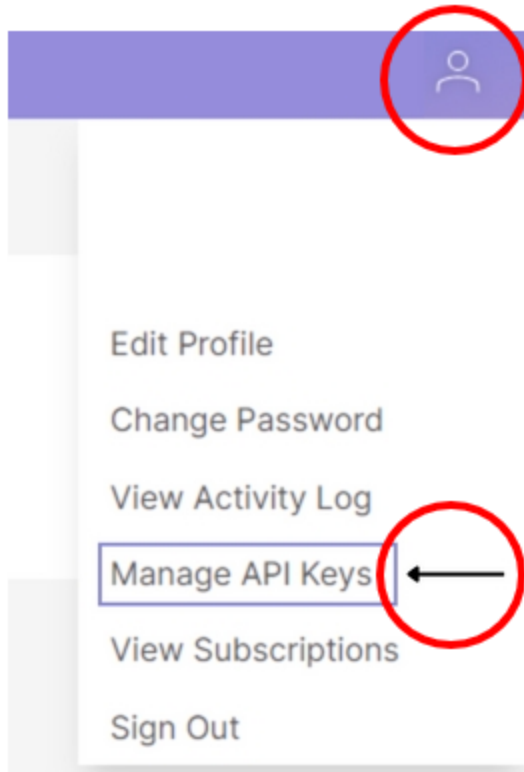
```
mkdir "C:\Program Files\Illumina"  
copy icav2.exe "C:\Program Files\Illumina"
```

Additional information can be found here: <https://help.ica.illumina.com/command-line-interface/cli-installation>

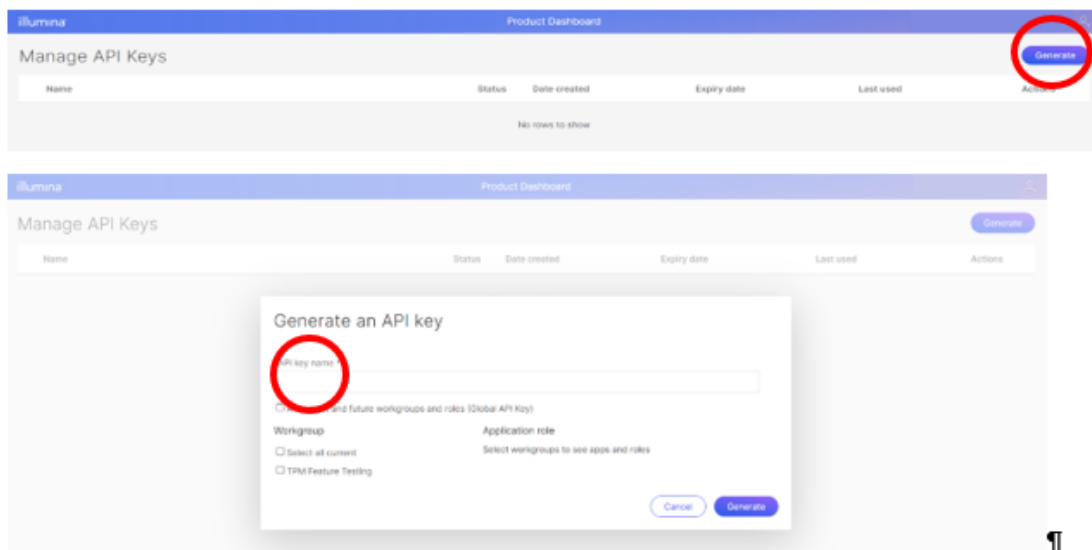
Authentication

- 5 The usage of Illumina's CLI requires an API key associated with the user's Illumina Connected Analytics (ICA) account access. This can be acquired through the following steps:

1. Locate the "Manage API Keys" menu from Illumina's platform home

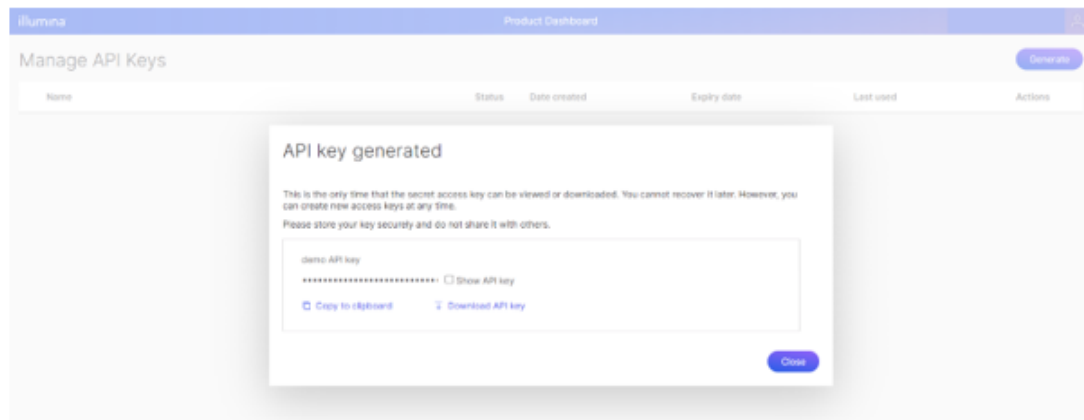


2. Click the button to generate a new API Key. Provide a name for the API Key. Then choose to either include all workgroups or select the workgroups to be included. Selected workgroups will be accessible with the API Key.





3. Click to generate the API Key. The API Key is then presented (hidden) with a button to show the key to be copied and a link to download to a file to be stored securely for future reference. Once the window is closed, the key contents will not be accessible through the domain login page, so **be sure to store it securely for future reference** if needed.



4. Authenticate the usage of the CLI using the following commands:

```
icav2 config set
Creating /Users/johngnome/.icav2/config.yaml
Initialize configuration settings [default]
server-url [ica.illumina.com]:
x-api-key : EXAMPLE_API_KEY
output-format (allowed values table,yaml,json defaults to table) :
colormode (allowed values none,dark,light defaults to none) :
```

Input the API Key generated from the product dashboard when prompted for x-api-key.

Data Upload

- 6
- In order to upload data onto ICA, you must first retrieve the relevant project ID you wish to upload the data to. This can be found by listing all projects using the command:

```
icav2 projects list
```

- The first column of the output will show the project ID. Save this as it will be used for data upload.



- To upload a file called “Sample-1_S1_L001_R1_001.fastq.gz” to the project, copy the project id and use the command syntax below:

```
icav2 projectdata list --project-id <project-id>
```

- To verify the file has uploaded, run the following to get a list of all files stored within the specified project:

```
icav2 projectdata list --project-id <project-id>
```

- The general upload syntax is the follow: `icav2 projectdata upload <localFileFolder> <remote-path> --project-id <project-id>`
- You are limited to only uploading 1 file or folder at a time, so it is recommended to store all files in 1 folder to upload with one command line input.