

Docker and Running PrediXcan on BioDataCatalyst

Camilla Czapla and Wheeler Lab





Docker

- Software for building “containers”
- Container image:
 - Separate filesystem
 - Dependencies
 - Code
 - Created by a Dockerfile: set of commands
- Container:
 - Runs image
 - Isolated environment



Dockerfile outline

```
FROM <existing image name>
```

```
RUN <installation commands>
```

```
RUN <installation commands>
```

```
CMD <default commands or arguments>
```

```
Other commands: COPY, WORKDIR, ENTRYPOINT
```



Dockerfile example

```
FROM alpine
```

```
COPY sum.py /home
```

```
RUN apk add --update python3 py3-pip python3-dev
```

```
#allow people to enter arguments for sum.py
```

```
ENTRYPOINT ["python3", "/home/sum.py"]
```

```
#default arguments
```

```
CMD ["10", "11"]
```

Docker Hub

Browser tabs: LUC S, PIL: 00, Google, Indivi, Sc, F21 R, Meta, Data, Predi, Slack, Indivi, predi, Repro, what t, Ex x

Address bar: hub.docker.com/search?q=python&type=image

Navigation: Docker Containers Plugins

Search: python

Buttons: Explore Pricing Sign In Sign Up

Filters

Images


- ☐ Verified Publisher
- ☐ Official Images
Official Images Published By Docker

Categories

- ☐ Analytics
- ☐ Application Frameworks
- ☐ Application Infrastructure
- ☐ Application Services
- ☐ Base Images
- ☐ Databases
- ☐ DevOps Tools
- ☐ Featured Images

1 - 25 of 71,102 results for **python**. [Clear search](#)

Best Match



python Official Image


Updated 8 days ago

Python is an interpreted, interactive, object-oriented, open-source programming language.

Container Windows Linux x86-64 IBM Z 386 mips64le ARM ARM 64 PowerPC 64 LE

Programming Languages

18+ 6.8K
Downloads Stars



pypy Official Image

Updated 9 days ago

PyPy is a fast, compliant alternative implementation of the Python language.

Container Linux Windows 386 IBM Z PowerPC 64 LE ARM ARM 64 x86-64

Programming Languages

10M+ 297
Downloads Stars

Windows taskbar: Type here to search, 27°F, 8:00 PM 11/26/2021



BioDataCatalyst

- Cloud platform for biological data science
- Public and controlled data
- GWAS applications among others
- Data tools allowing the download or upload of files through the command line
- Interactive Analysis through RStudio available of data within your projects

PrediXcan on BioDataCatalyst

The screenshot displays the BioDataCatalyst web interface for a PrediXcan run. The browser address bar shows the URL: `platform.sb.biodatacatalyst.nhlbi.nih.gov/u/wheeler.heather/mesa-gwas-predixcan/tasks/d1dd24b1-0877-4a1e-bbfe-14650d70669b/#`. The page header includes navigation links: Projects, Data, Public Gallery, Public projects, and Developer. The user is logged in as 'cczapla'.

The main content area shows a 'RUNNING' status for 'PrediXcan run - 11-27-21 01:12:55'. It indicates the run was executed on Nov. 26, 2021, 19:23 by 'cczapla'. The progress is 'Running' and the duration is '3 minutes'. The app version is 'PrediXcan - Revision: 0'.

The interface is divided into three main sections: Inputs, App Settings, and Outputs.

- Inputs:**
 - kinship_matrix:** No files selected.
 - model_db_path:** en_Whole_Blood.db
 - phenotype_file:** random_pheno_1000G_hg37_fixed_1.txt
 - vcf_genotypes:** ALL.chr22.phase3_shapeit2_mvncall_integrated_v5a.201...
- App Settings:**
 - Predict (#predict):**
 - prediction_output:** Whole_Blood_predict.h5 HDF5
 - prediction_summary_output:** Whole_Blood_summary.txt
 - vcf_mode:** genotyped
 - Association (#association):**
 - covariates:** "age"
 - main_phenotype_of_interest:** pheno
 - output_prefix:** Whole_Blood_association.txt
- Outputs:**
 - Association_output:** No value
 - summary:** No value

The footer contains links for Privacy Policy, Data Sharing Policy, Freedom of Information Act (FOIA), Accessibility, U.S. Department of Health & Human Services, National Institutes of Health, National Heart, Lung, and Blood Institute, and USA.gov. The system clock shows 7:27 PM on 11/26/2021.

A screenshot of a Windows 10 taskbar. On the left is the Start button (Windows logo). Next to it is a search bar with the text "Type here to search". To the right of the search bar are several application icons: a white circle, a document icon, the Microsoft Edge browser icon, the File Explorer icon, the Microsoft Store icon, the Mail app icon, the Amazon app icon, the OneDrive icon, and the Google Chrome icon. Further right are icons for a presentation (PowerPoint), a code editor (Visual Studio), and a help icon (question mark). The system tray on the right shows the weather (26°F), network status, volume, and a clock displaying 7:30 PM on 11/26/2021. A notification icon with a number 1 is also visible.



Future Plans

- Run PrediXcan workflow successfully on BioDataCatalyst
- Make GitHub repository outlining steps for using Docker, BioDataCatalyst CLI, and launching a Dockstore Workflow on a Seven Bridges platform (BioDataCatalyst)
- Run PrediXcan and S-PrediXcan on BioDataCatalyst
- Write Dockerfile and workflow compatible with Seven Bridges



References

Workflow GitHub: <https://github.com/erikaesquinca>

PrediXcan/MetaXcan GitHub: <https://github.com/hakyimlab>

Docker tutorial GitHub:

<https://github.com/carpentries-incubator/docker-introduction>

Authors of docker-introduction GitHub: Jeremy Cohen, David Eyers, Christina Koch, Sarah Stevens, Andy Turner, Dave Welch

Wheeler Lab

Heather Wheeler

Daniel Araújo

Isabelle Gregga

Henry Wittich

Rhea Prag

Chris Nguyen

Kayla Bozeman



Thank You