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
```
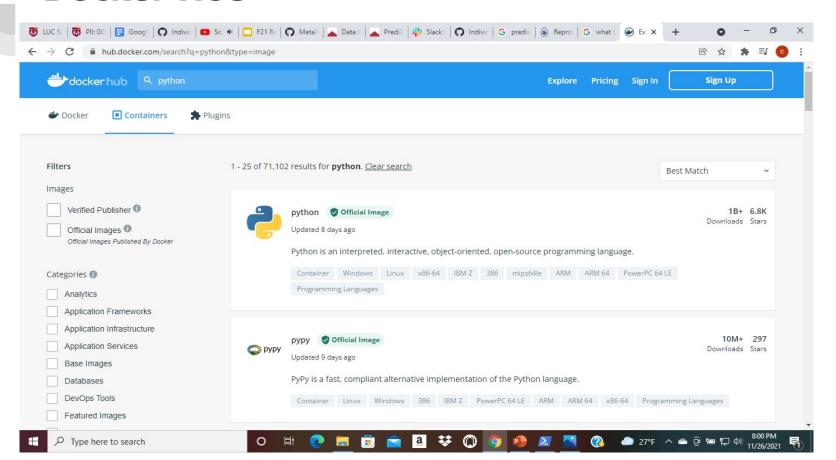
Dockerfile example

```
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"]
```
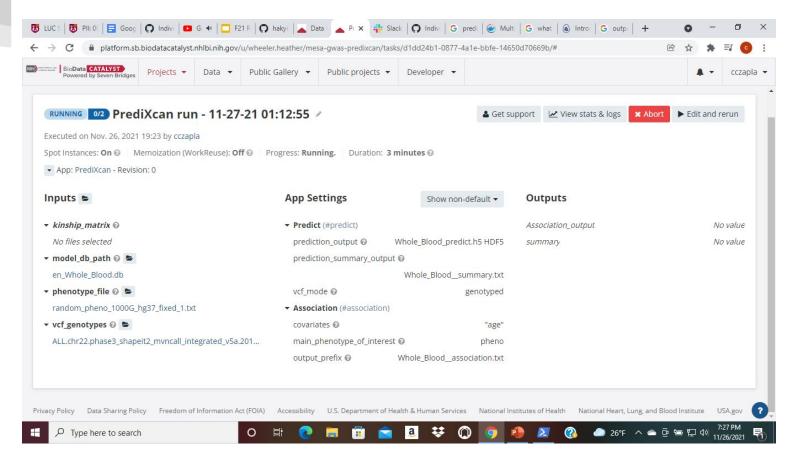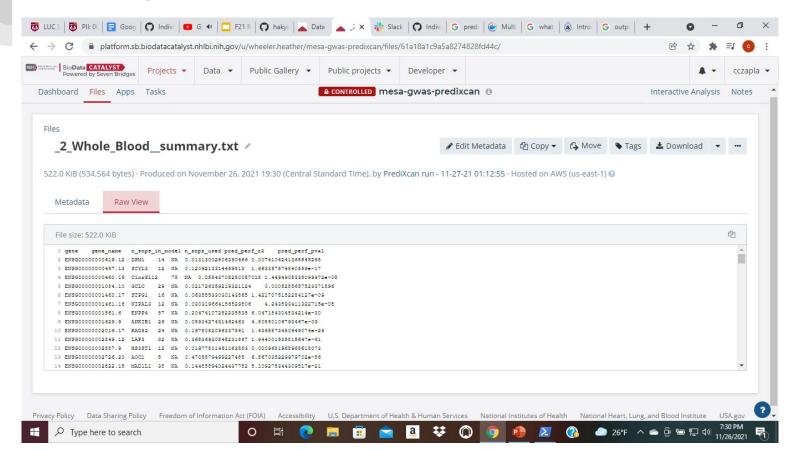
Docker Hub

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

Results

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