What tools and workflows should I use to fulfil an overall goal?

Emily Perry, Eleni Christodoulou

8th October 2024

Data security 

- This training session will include data from the GEL Research Environment
- As part of your IG training you have agreed to not distribute these data in any way
- You are not allowed to:
    - Invite colleagues to watch this training with you
    - Take any screenshots or videos of the training
    - Share your webinar link (we will remove anyone who is here twice)
- We will record this training and distribute the censored video afterwards

Presenters

Emily Perry
Research Engagement Manager

Eleni Christodoulou
Solutions Architect at Lifebit

Questions

All your microphones are muted

Use the Zoom Q&A to ask questions

Upvote your favourite questions: if we are short on time we will prioritise those with the most votes

Questions

Miruna Carmen Barbu
Bioinformatician - Research Services

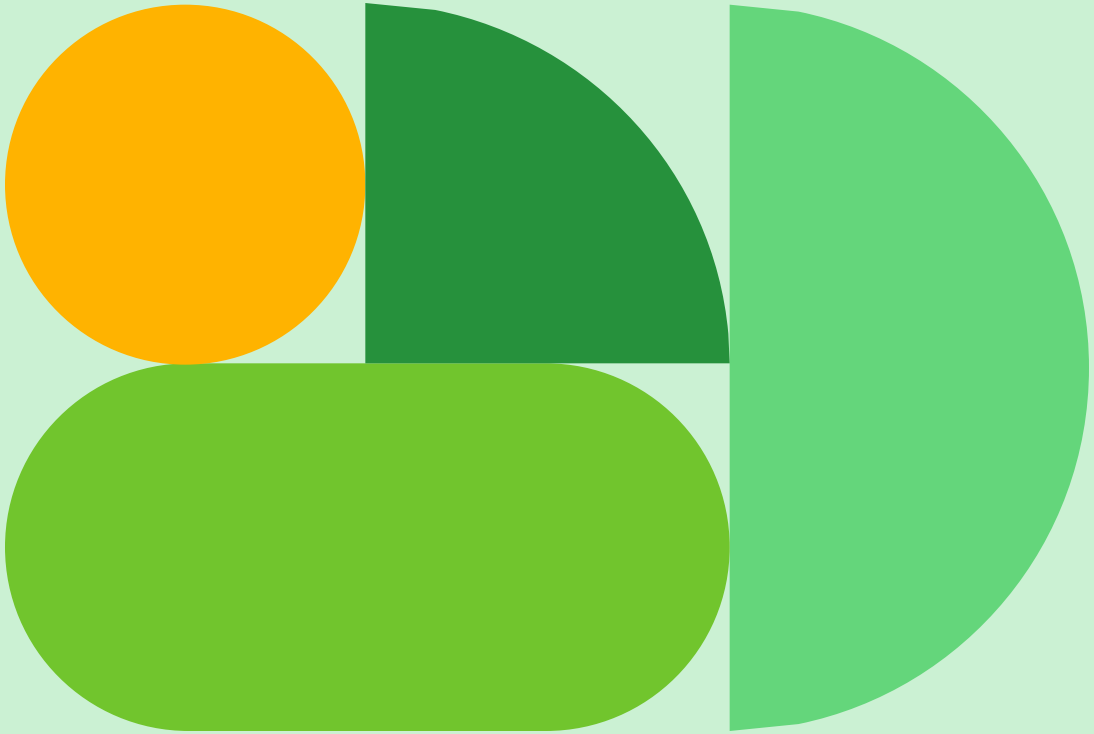
Elena Bernabeu
Bioinformatician - Research Services

Alex Ho
Bioinformatician - Research Services

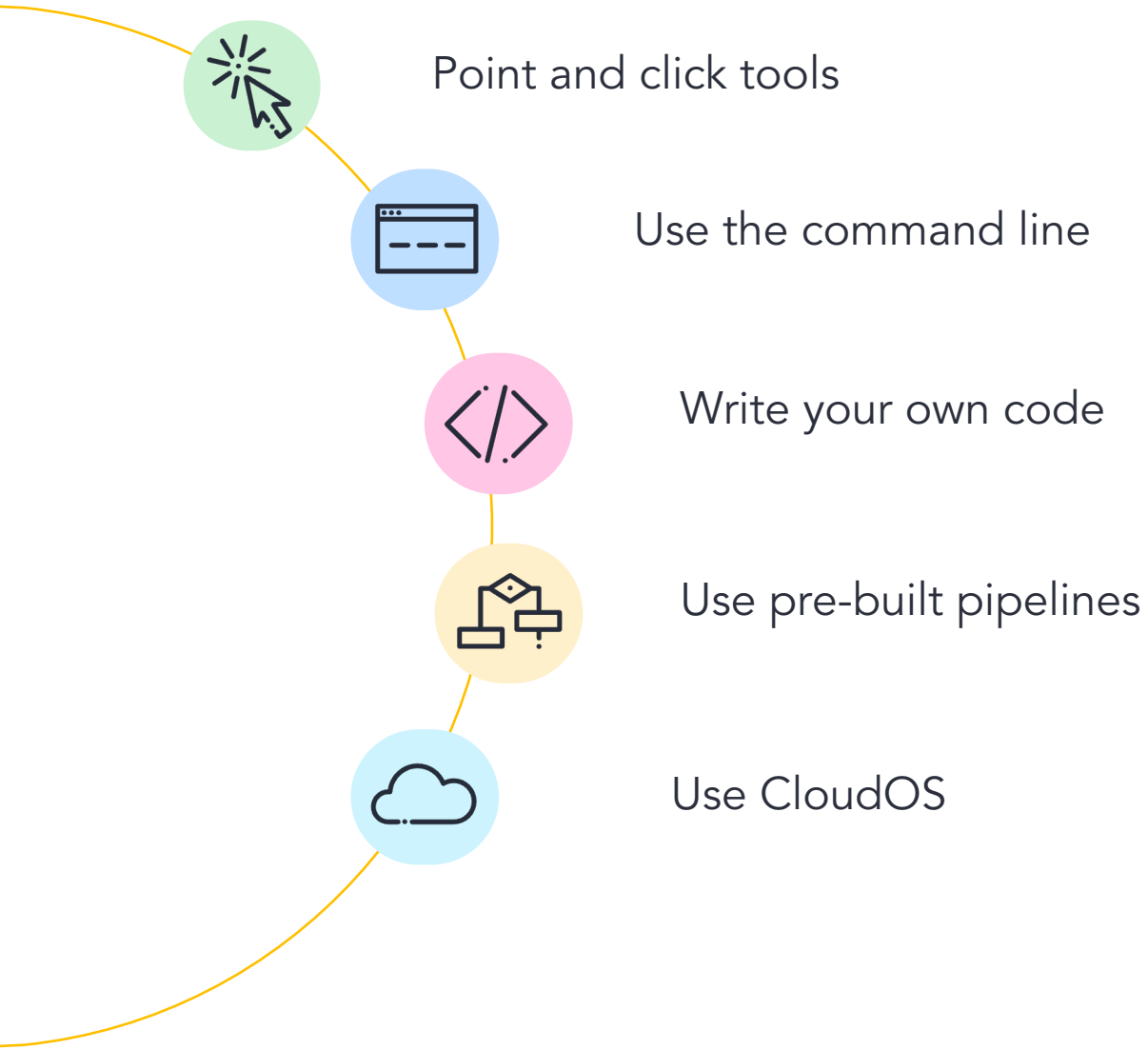
Charlotte Maughan
Client Manager - LifeBit

Agenda

1. Introduction and admin
2. Identifying variants associated with a phenotype
3. Identifying phenotypes associated with a gene
4. Finding diagnoses for patients who didn’t get one through primary clinical interpretation
5. Help and questions



This session



Point and click tools

Use the command line

Write your own code

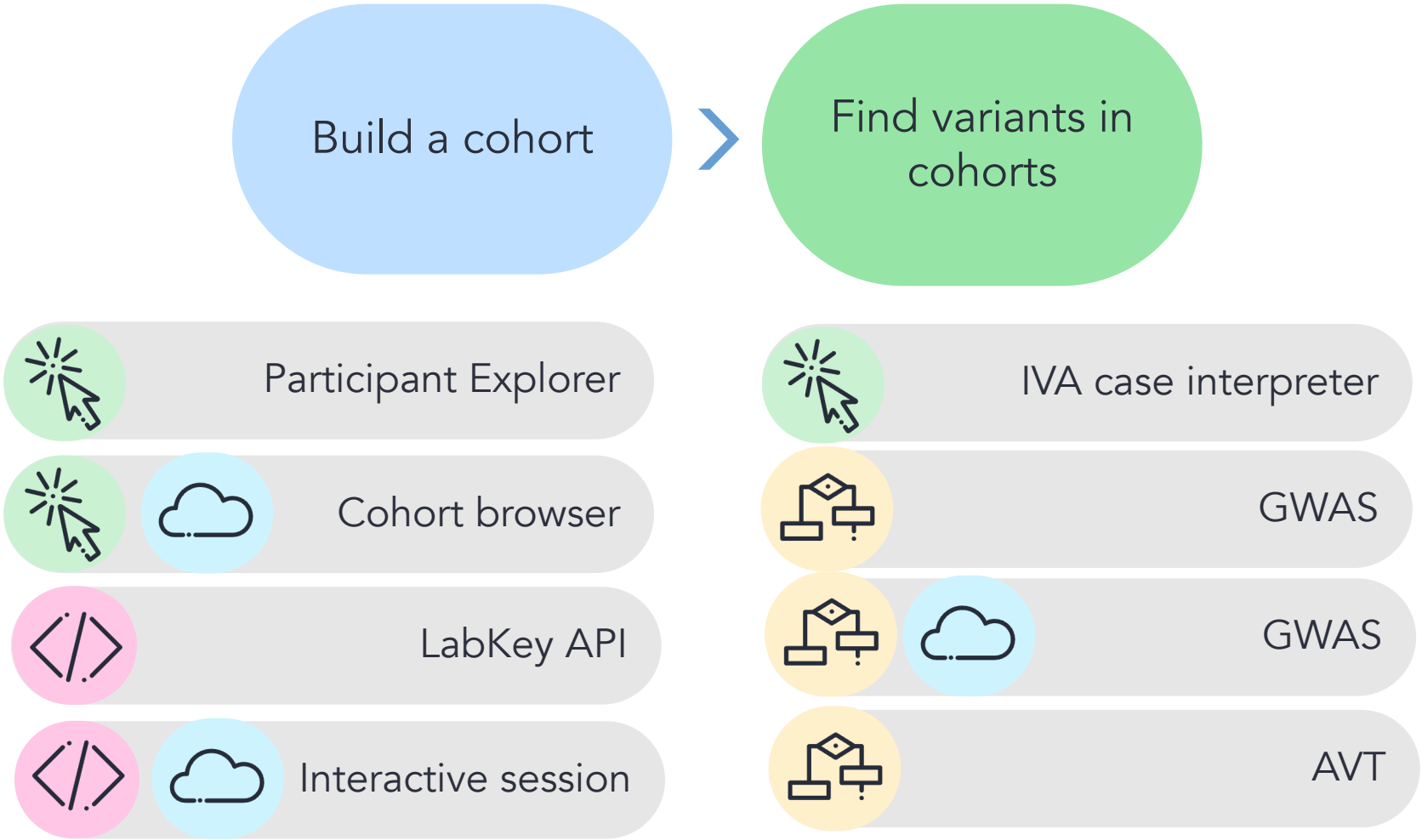
Use pre-built pipelines

Use CloudOS

Links to more detailed online tutorials – live sessions often also available

2. Identifying variants associated with a phenotype

Identifying variants associated with a phenotype



https://re-docs.genomicsengland.co.uk/phenotype_to_variants/

Build a cohort – Participant Explorer

- Filter participants by
  - Phenotypes, diagnoses
  - Personal details
- Reverse the search for a control cohort
- Download tables of search results

https://re-docs.genomicsengland.co.uk/pxa_cohorts/

Build a cohort – Cohort Browser

Search for participants by

Standardised OMOP tables

Structured clinical data tables

Reverse the search for a control cohort

Create tables within CloudOS and download to VDI

https://re-docs.genomicsengland.co.uk/cb_cohorts/

Build a cohort – LabKey API

- Use SQL queries
- Tutorials in Python and R
- Cancer tutorial
- Rare disease tutorial

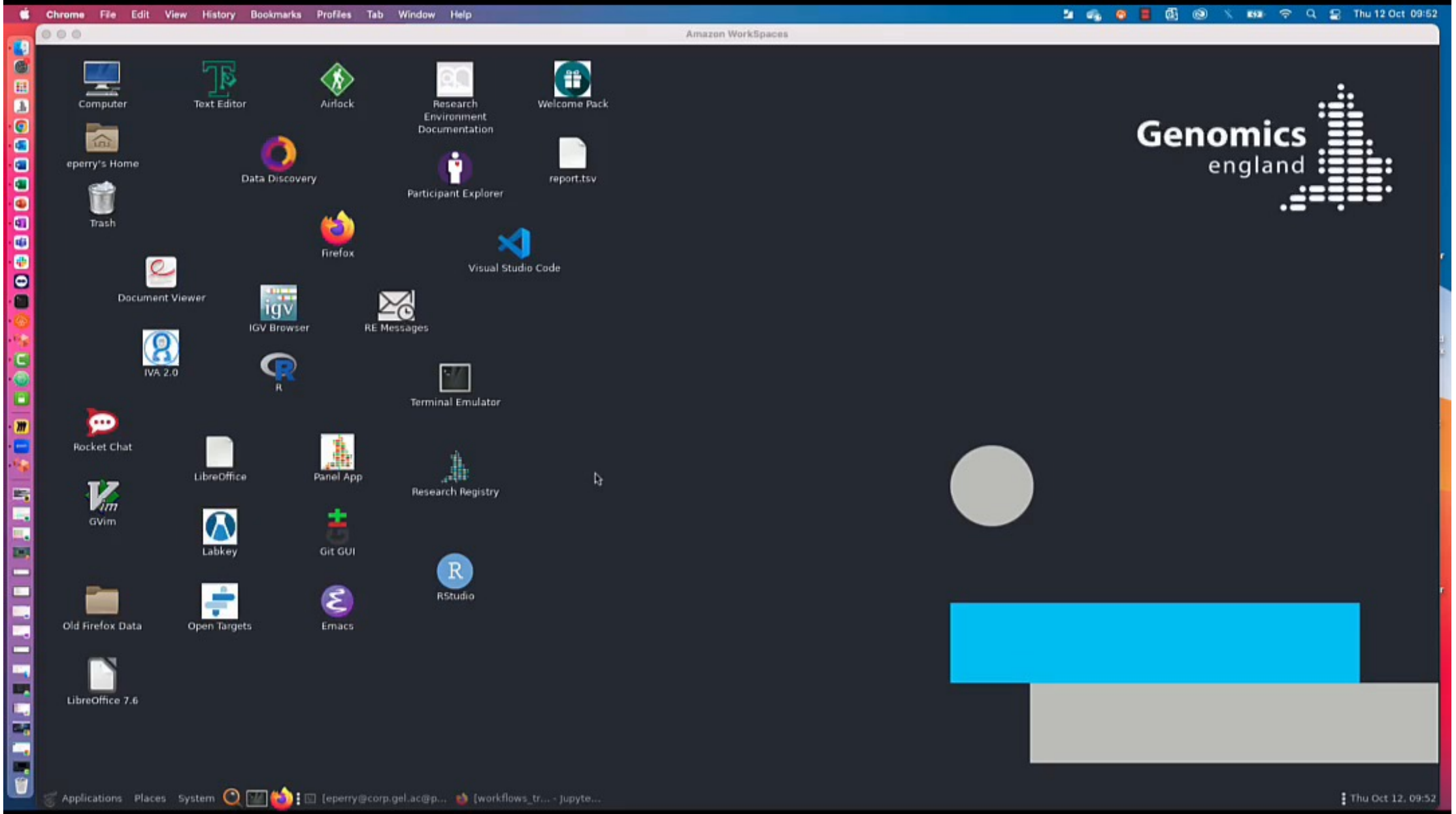
https://re-docs.genomicsengland.co.uk/cancer_cohorts/
https://re-docs.genomicsengland.co.uk/rd_cohorts/

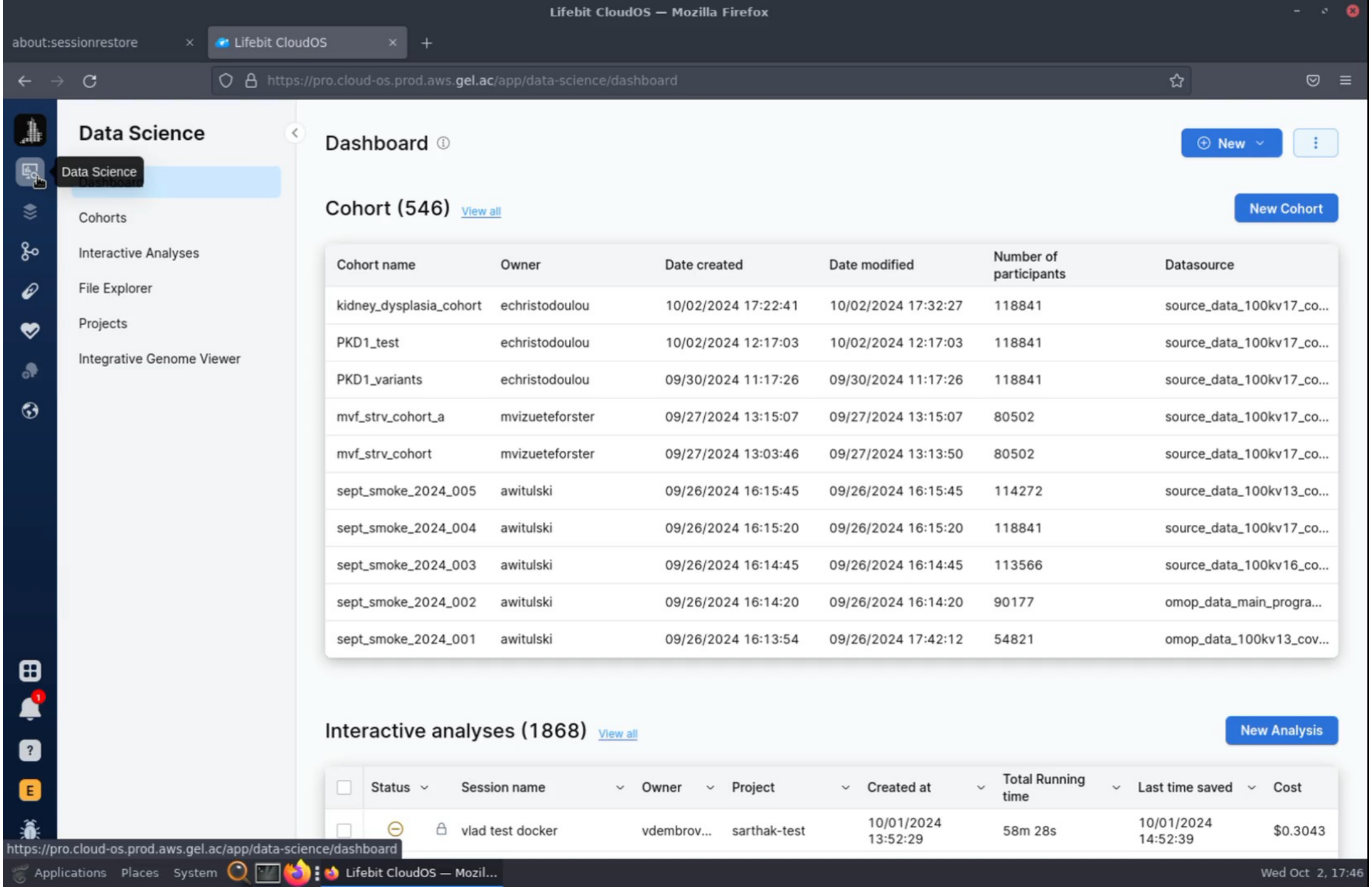
Build a cohort – Interactive session

- Use SQL queries
- Tutorial in R
- Cancer tutorial
- Rare disease tutorial

https://re-docs.genomicsengland.co.uk/cancer_cohorts/
https://re-docs.genomicsengland.co.uk/rd_cohorts/

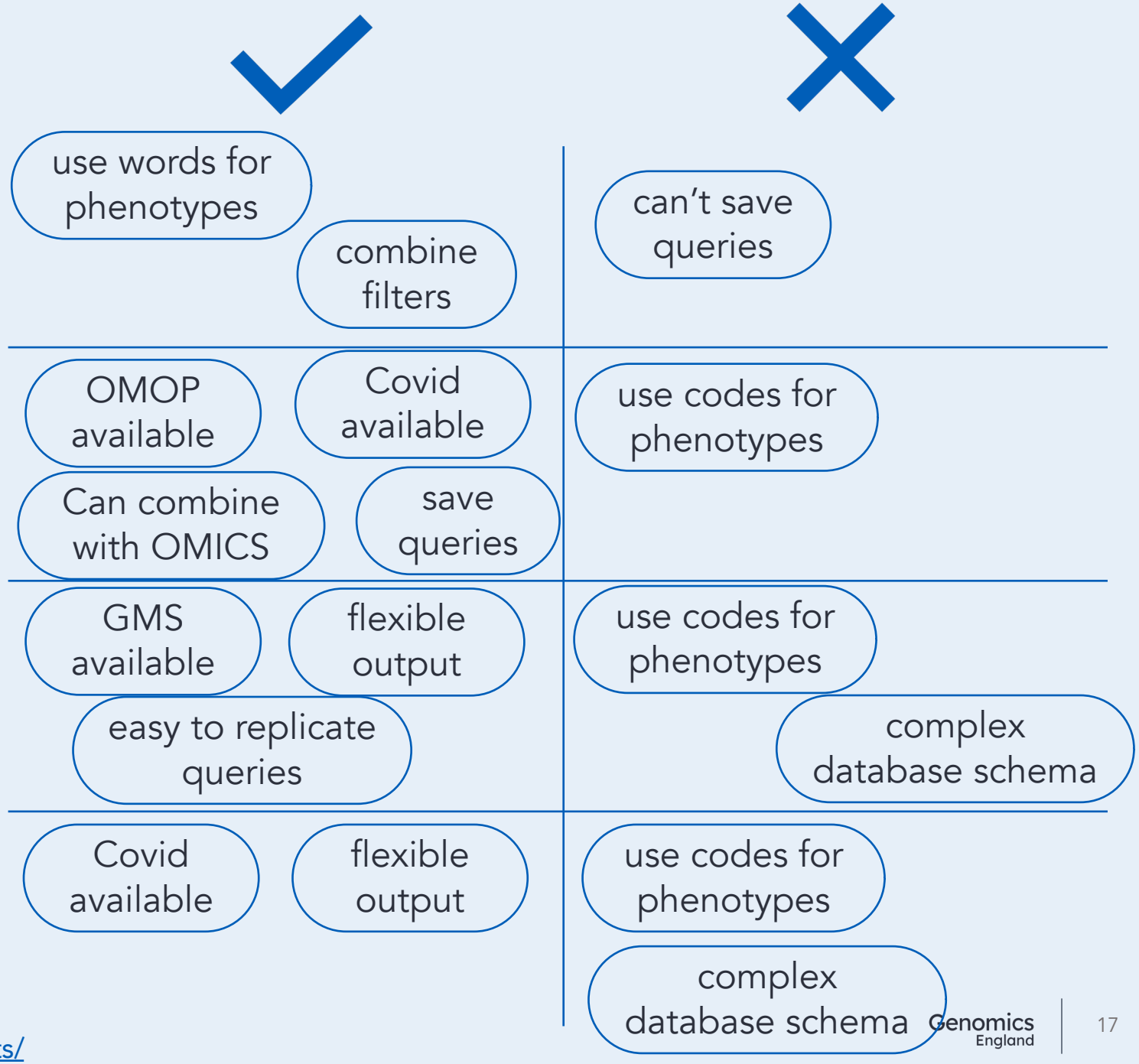
Build a cohort demo


Chrome File Edit View History Bookmarks Profiles Tab Window Help
Thu 12 Oct 09:52
Amazon WorkSpaces
Computer
Text Editor
Airlock
Research Environment Documentation
Welcome Pack
eperry's Home
Data Discovery
Participant Explorer
report.tsv
Trash
Firefox
Visual Studio Code
Document Viewer
IGV Browser
RE Messages
IVA 2.0
R
Terminal Emulator
Rocket Chat
LibreOffice
Panel App
Research Registry
GVim
Labkey
Git GUI
RStudio
Old Firefox Data
Open Targets
Emacs
LibreOffice 7.6
Genomics england
Applications Places System
Thu Oct 12, 09:52


Lifebit CloudOS — Mozilla Firefox
about:sessionrestore
Lifebit CloudOS
https://pro.cloud-os.prod.aws.gel.ac/app/data-science/dashboard
Data Science
Data Science
Cohorts
Interactive Analyses
File Explorer
Projects
Integrative Genome Viewer
Dashboard
New
Cohort (546) View all
New Cohort
Cohort name | Owner | Date created | Date modified | Number of participants | Datasource
kidney_dysplasia_cohort | echristodoulou | 10/02/2024 17:22:41 | 10/02/2024 17:32:27 | 118841 | source_data_100kv17_co...
PKD1_test | echristodoulou | 10/02/2024 12:17:03 | 10/02/2024 12:17:03 | 118841 | source_data_100kv17_co...
PKD1_variants | echristodoulou | 09/30/2024 11:17:26 | 09/30/2024 11:17:26 | 118841 | source_data_100kv17_co...
mvf_strv_cohort_a | mvizueteforster | 09/27/2024 13:15:07 | 09/27/2024 13:15:07 | 80502 | source_data_100kv17_co...
mvf_strv_cohort | mvizueteforster | 09/27/2024 13:03:46 | 09/27/2024 13:13:50 | 80502 | source_data_100kv17_co...
sept_smoke_2024_005 | awitulski | 09/26/2024 16:15:45 | 09/26/2024 16:15:45 | 114272 | source_data_100kv13_co...
sept_smoke_2024_004 | awitulski | 09/26/2024 16:15:20 | 09/26/2024 16:15:20 | 118841 | source_data_100kv17_co...
sept_smoke_2024_003 | awitulski | 09/26/2024 16:14:45 | 09/26/2024 16:14:45 | 113566 | source_data_100kv16_co...
sept_smoke_2024_002 | awitulski | 09/26/2024 16:14:20 | 09/26/2024 16:14:20 | 90177 | omop_data_main_progra...
sept_smoke_2024_001 | awitulski | 09/26/2024 16:13:54 | 09/26/2024 17:42:12 | 54821 | omop_data_100kv13_cov...
Interactive analyses (1868) View all
New Analysis
Status | Session name | Owner | Project | Created at | Total Running time | Last time saved | Cost
vlad test docker | vdembrov... | sarthak-test | 10/01/2024 13:52:29 | 58m 28s | 10/01/2024 14:52:39 | $0.3043
https://pro.cloud-os.prod.aws.gel.ac/app/data-science/dashboard
Applications Places System
Lifebit CloudOS — Mozil...
Wed Oct 2, 17:46

Build a cohort

- Participant Explorer
- Cohort browser
- LabKey API
- Interactive session



https://re-docs.genomicsengland.co.uk/phenotype_to_variants/

Find variants in cohort – IVA case interpreter

- See all variants in a participant
- Filter variants by
  - Genes, consequences, variant frequency etc
  - Family genotypes

https://re-docs.genomicsengland.co.uk/iva_case/

Find variants in cohort - GWAS

- Run on the HPC
- Run as a CloudOS batch job
- GWAS summaries
- Manhattan plots

https://re-docs.genomicsengland.co.uk/gwas/
https://lifebit.atlassian.net/wiki/spaces/CD/pages/813040563/GWAS+System+Tools

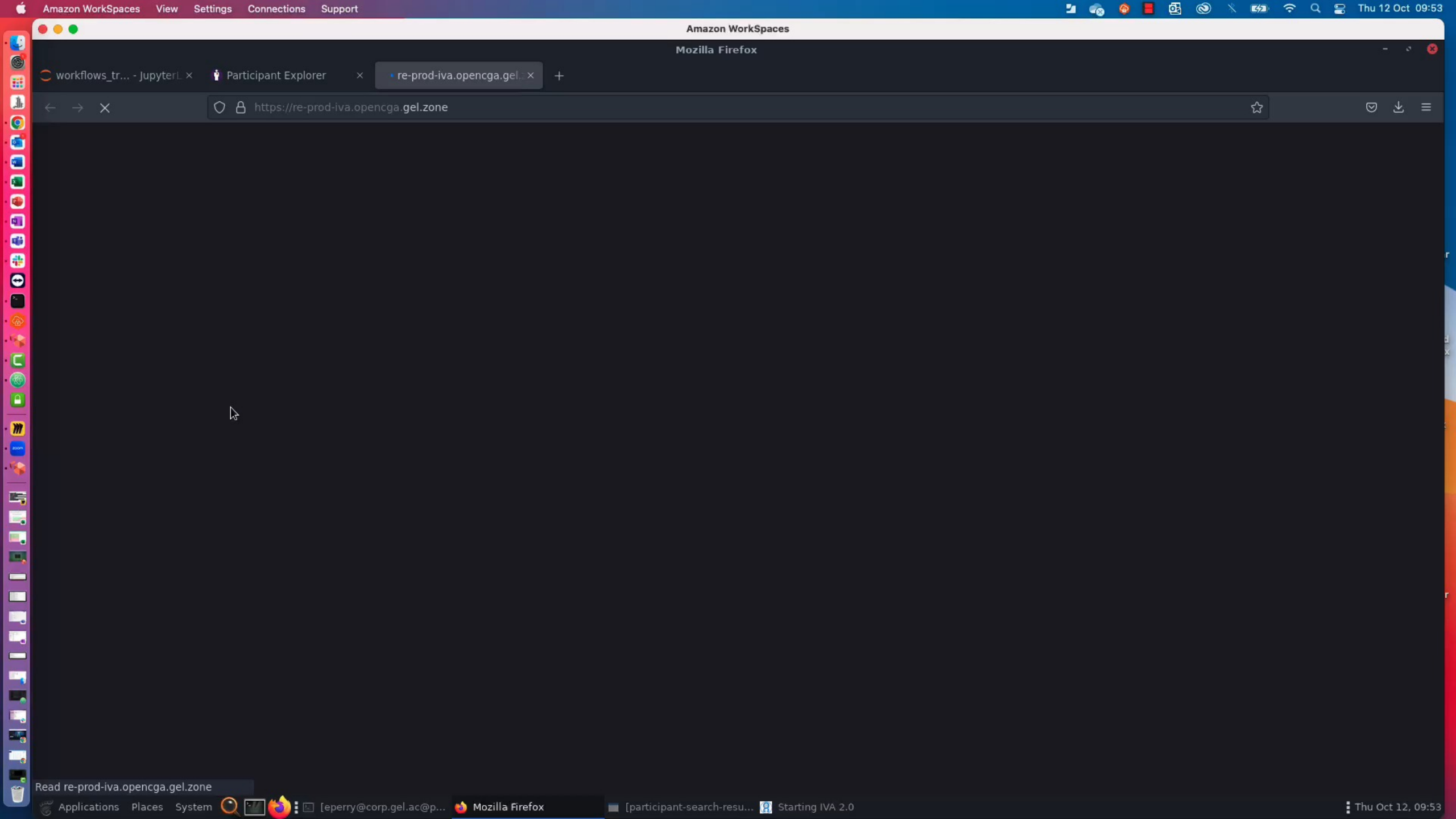
Find variants in cohort - AVT

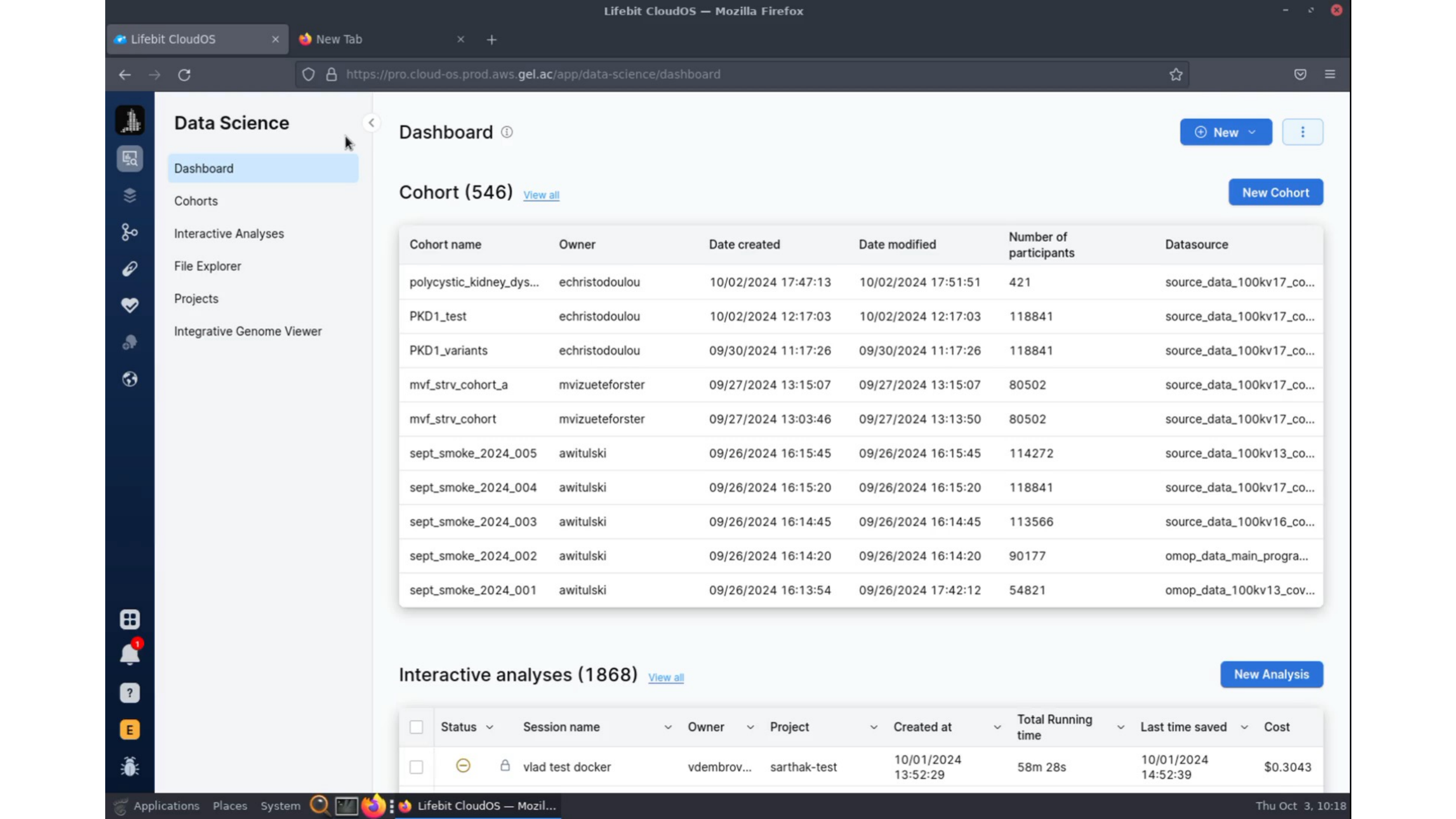
Run on the HPC

Output depends on the tools used

https://re-docs.genomicsengland.co.uk/avt/

Find variants in cohort demo


Amazon WorkSpaces View Settings Connections Support
Thu 12 Oct 09:53
Amazon WorkSpaces
Mozilla Firefox
workflows_tr... - JupyterL
Participant Explorer
re-prod-iva.opencga.gel.
https://re-prod-iva.opencga.gel.zone
Read re-prod-iva.opencga.gel.zone
Applications Places System
[eperry@corp.gel.ac@p...
Mozilla Firefox
[participant-search-resu...
Starting IVA 2.0
Thu Oct 12, 09:53


Lifebit CloudOS — Mozilla Firefox
Lifebit CloudOS
New Tab
https://pro.cloud-os.prod.aws.gel.ac/app/data-science/dashboard
Data Science
Dashboard
Cohorts
Interactive Analyses
File Explorer
Projects
Integrative Genome Viewer
Dashboard
New
Cohort (546) View all
New Cohort
Cohort name | Owner | Date created | Date modified | Number of participants | Datasource
polycystic_kidney_dys... | echristodoulou | 10/02/2024 17:47:13 | 10/02/2024 17:51:51 | 421 | source_data_100kv17_co...
PKD1_test | echristodoulou | 10/02/2024 12:17:03 | 10/02/2024 12:17:03 | 118841 | source_data_100kv17_co...
PKD1_variants | echristodoulou | 09/30/2024 11:17:26 | 09/30/2024 11:17:26 | 118841 | source_data_100kv17_co...
mvf_strv_cohort_a | mvizueteforster | 09/27/2024 13:15:07 | 09/27/2024 13:15:07 | 80502 | source_data_100kv17_co...
mvf_strv_cohort | mvizueteforster | 09/27/2024 13:03:46 | 09/27/2024 13:13:50 | 80502 | source_data_100kv17_co...
sept_smoke_2024_005 | awitulski | 09/26/2024 16:15:45 | 09/26/2024 16:15:45 | 114272 | source_data_100kv13_co...
sept_smoke_2024_004 | awitulski | 09/26/2024 16:15:20 | 09/26/2024 16:15:20 | 118841 | source_data_100kv17_co...
sept_smoke_2024_003 | awitulski | 09/26/2024 16:14:45 | 09/26/2024 16:14:45 | 113566 | source_data_100kv16_co...
sept_smoke_2024_002 | awitulski | 09/26/2024 16:14:20 | 09/26/2024 16:14:20 | 90177 | omop_data_main_progra...
sept_smoke_2024_001 | awitulski | 09/26/2024 16:13:54 | 09/26/2024 17:42:12 | 54821 | omop_data_100kv13_cov...
Interactive analyses (1868) View all
New Analysis
Status | Session name | Owner | Project | Created at | Total Running time | Last time saved | Cost
vlad test docker | vdembrov... | sarthak-test | 10/01/2024 13:52:29 | 58m 28s | 10/01/2024 14:52:39 | $0.3043
Applications Places System
Lifebit CloudOS — Mozil...
Thu Oct 3, 10:18

Find variants in cohort

IVA case interpreter

GWAS

GWAS

AVT

easy to search and interpret

one participant at a time

pre-built

common variant impact

flexible options

complex to run

many input files needed

pre-built

run with point and click

common variant impact

flexible options

many input files needed

pre-built

flexible options

rare variant impact

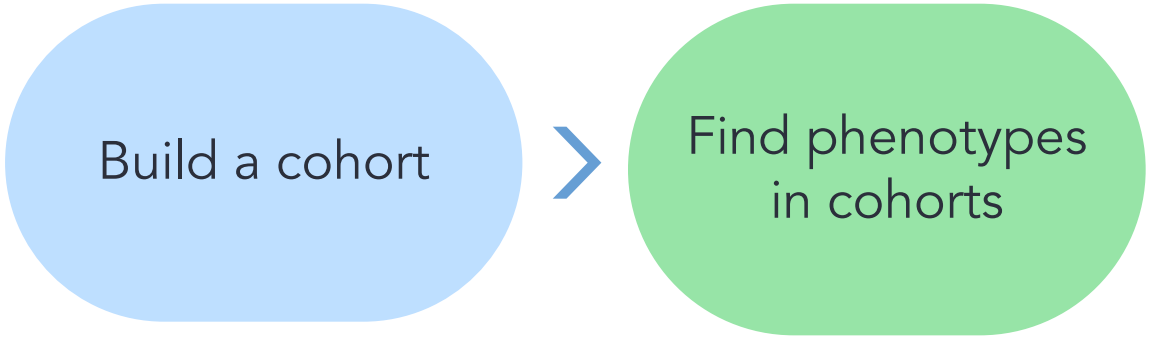
many input files

complex to interpret results

https://re-docs.genomicsengland.co.uk/phenotype_to_variants/

3. Identifying phenotypes associated with a gene

Identifying phenotypes associated with a gene



- IVA
- Cohort browser
- Small/structural variant
- Small/structural variant
- AggV2

- Participant Explorer
- Cohort browser
- LabKey API
- Interactive session

https://re-docs.genomicsengland.co.uk/gene_to_phenotype/

Build a cohort – IVA variant browser

See all variants

Filter variants by

Genes, consequences, variant frequency etc

See all participants with variant

https://re-docs.genomicsengland.co.uk/iva_variant/

Build a cohort – Cohort browser

- See all variants
- Filter variants by
- Genes, consequences, variant frequency etc
- See all participants with variant

https://lifebit.atlassian.net/wiki/spaces/CD/pages/813107122/Filter+cohort+by+genomic+data

Build a cohort – Small/structural variant workflows

- Run on the HPC
- Run as a CloudOS batch job
- Input list of genes
- List of variants in gene(s)
- List of participants with variants

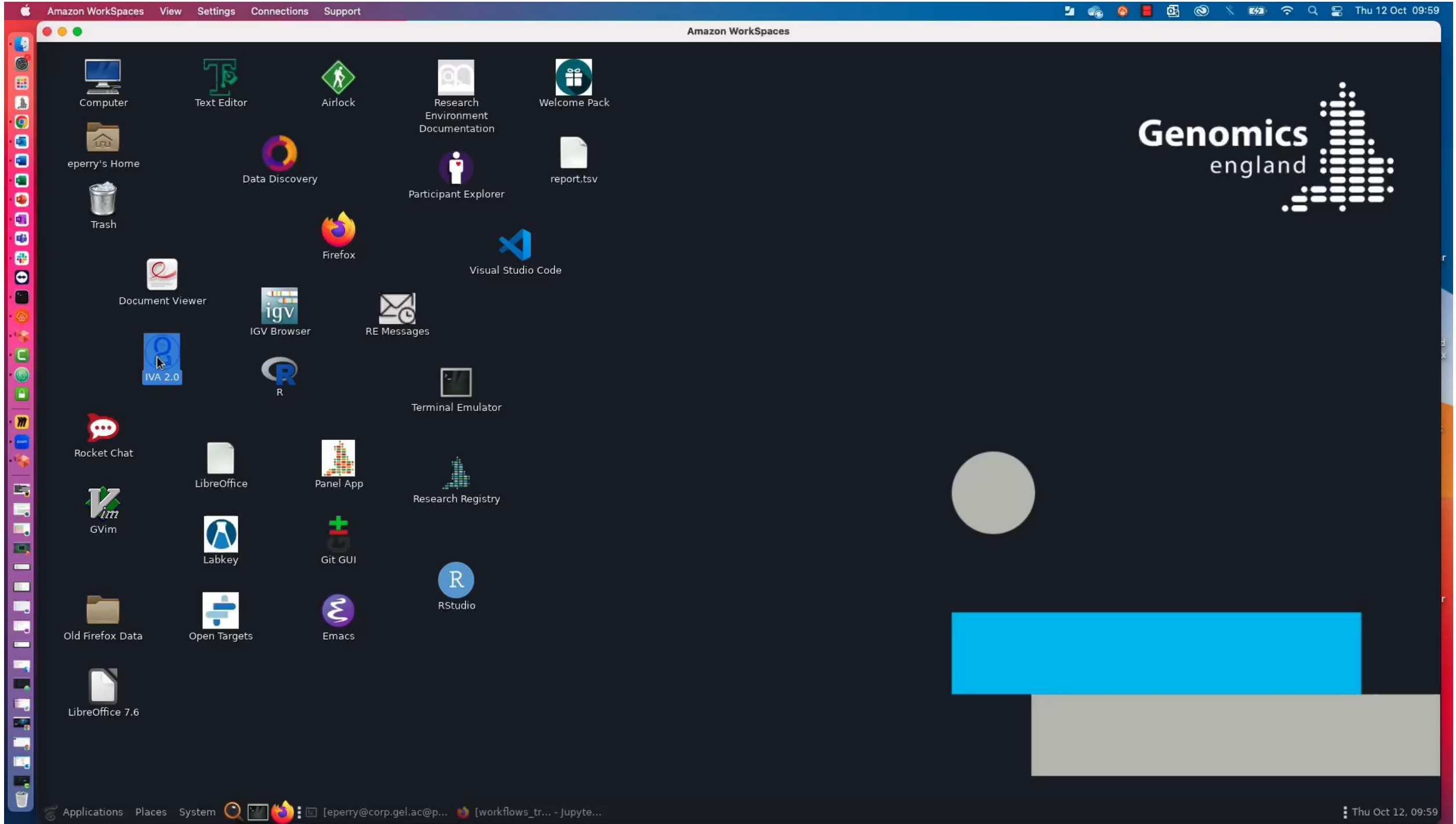
https://re-docs.genomicsengland.co.uk/structural_variant/
https://re-docs.genomicsengland.co.uk/small_variant/

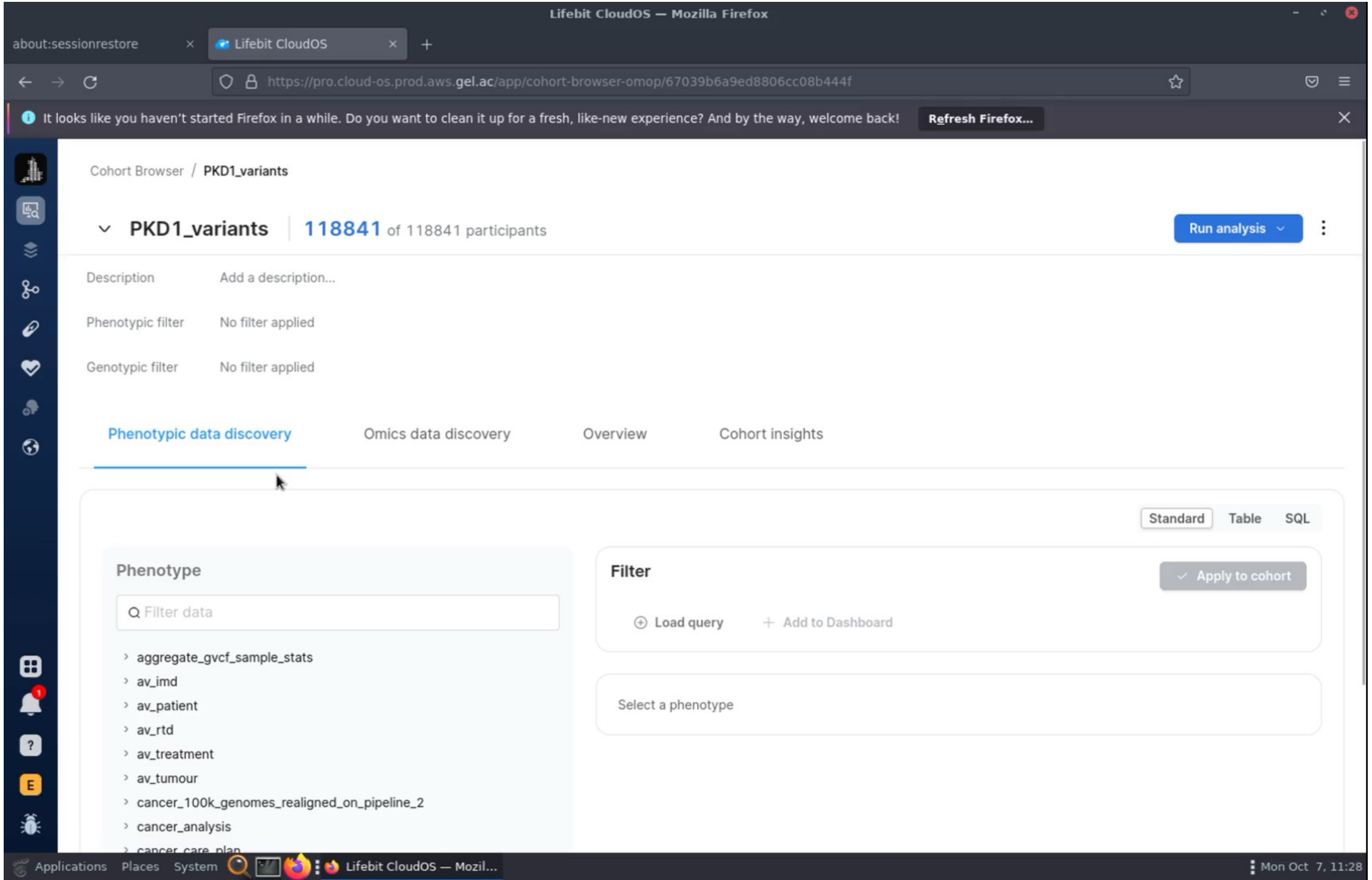
Build a cohort – AggV2

- Aggregated VCFs of all variants
- Participant genotypes
- bcftools to pull out genotypes

https://re-docs.genomicsengland.co.uk/aggv2

Find participants with variants demo


Amazon WorkSpaces View Settings Connections Support
Thu 12 Oct 09:59
Amazon WorkSpaces
Computer
Text Editor
Airlock
Research Environment Documentation
Welcome Pack
eperry's Home
Data Discovery
Participant Explorer
report.tsv
Trash
Firefox
Visual Studio Code
Document Viewer
IGV Browser
RE Messages
IVA 2.0
R
Terminal Emulator
Rocket Chat
LibreOffice
Panel App
Research Registry
GVim
Labkey
Git GUI
RStudio
Old Firefox Data
Open Targets
Emacs
LibreOffice 7.6
Genomics england
Applications Places System
[eperry@corp.gel.ac@p...
[workflows_tr... - Jupyte...
Thu Oct 12, 09:59


Lifebit CloudOS — Mozilla Firefox
about:sessionrestore
Lifebit CloudOS
https://pro.cloud-os.prod.aws.gel.ac/app/cohort-browser-omop/67039b6a9ed8806cc08b444f
It looks like you haven't started Firefox in a while. Do you want to clean it up for a fresh, like-new experience? And by the way, welcome back!
Refresh Firefox...
Cohort Browser / PKD1_variants
PKD1_variants
118841 of 118841 participants
Run analysis
Description
Add a description...
Phenotypic filter
No filter applied
Genotypic filter
No filter applied
Phenotypic data discovery
Omics data discovery
Overview
Cohort insights
Standard
Table
SQL
Phenotype
Filter data
aggregate_gvcf_sample_stats
av_imd
av_patient
av_rtd
av_treatment
av_tumour
cancer_100k_genomes_realigned_on_pipeline_2
cancer_analysis
Filter
Apply to cohort
Load query
Add to Dashboard
Select a phenotype
Applications
Places
System
Lifebit CloudOS — Mozil...
Mon Oct 7, 11:28

Find participant with variants


IVA


Cohort browser


Small/structural variant


Small/structural variant


AggV2

easy to search and interpret

output participants one variant at a time

combine phenotype and omics

save queries

pre-built

CloudOS and HPC

extra input files for CloudOS

GRCh37 and GRCh38

CloudOS user friendly

use BCFtools

command line familiarity

subset participants

https://re-docs.genomicsengland.co.uk/phenotype_to_variants/

Find phenotypes in cohorts – Participant Explorer

Search for participant(s) by ID(s)

See medical history for a participant

Compare medical history between participants

https://re-docs.genomicsengland.co.uk/medhist_pxa/

Find phenotypes in cohorts – Cohort browser

- OMOP standardised format
- Source 100kGP clinical tables
- Plots of phenotype frequencies

https://lifebit.atlassian.net/wiki/spaces/CD/pages/813107011/Get+insights+on+your+cohort

Find phenotypes in cohorts – LabKey API

- Use SQL queries
- Tutorials in Python and R
- Detailed medical history tables

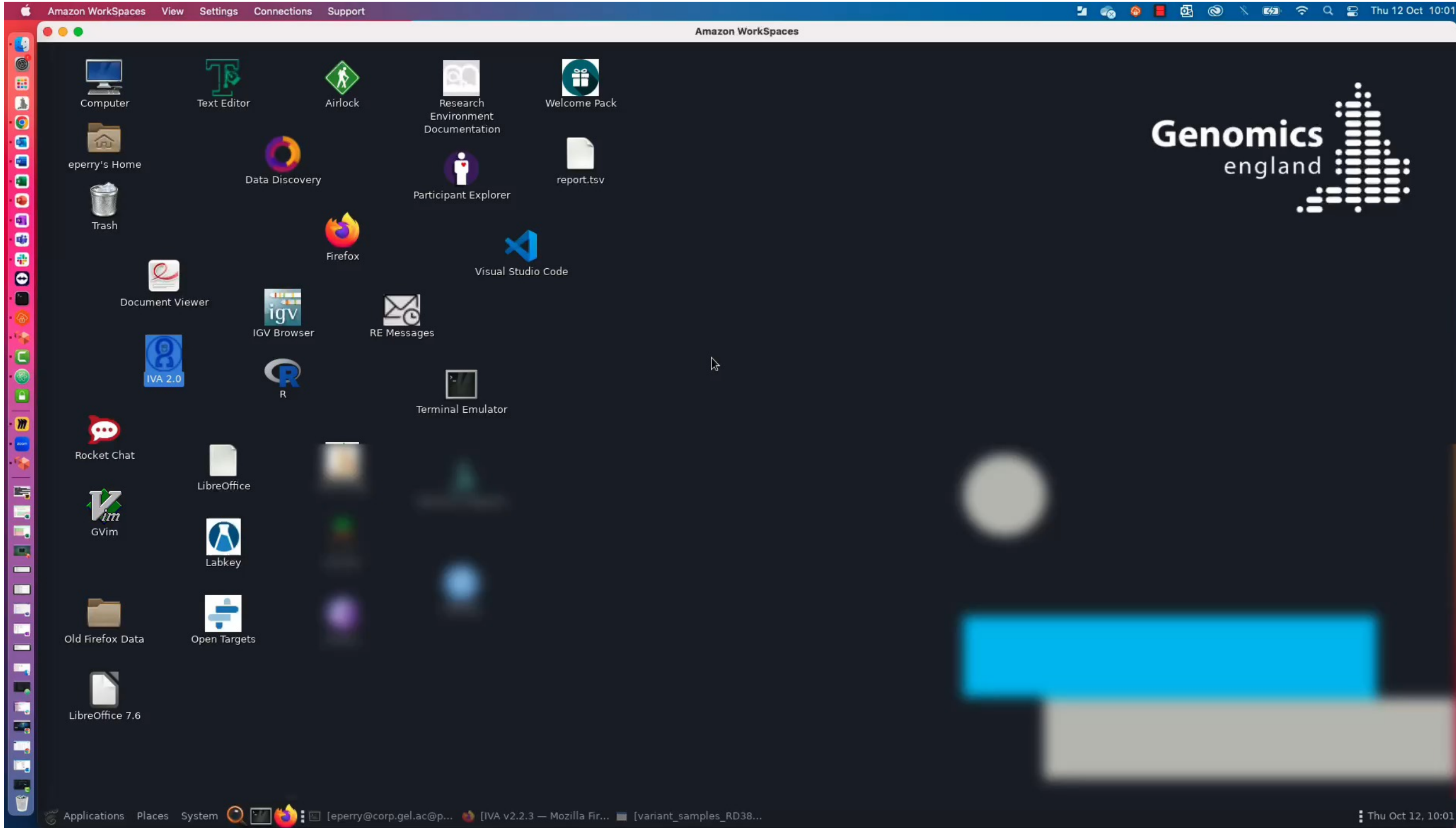
https://re-docs.genomicsengland.co.uk/medhist_api/

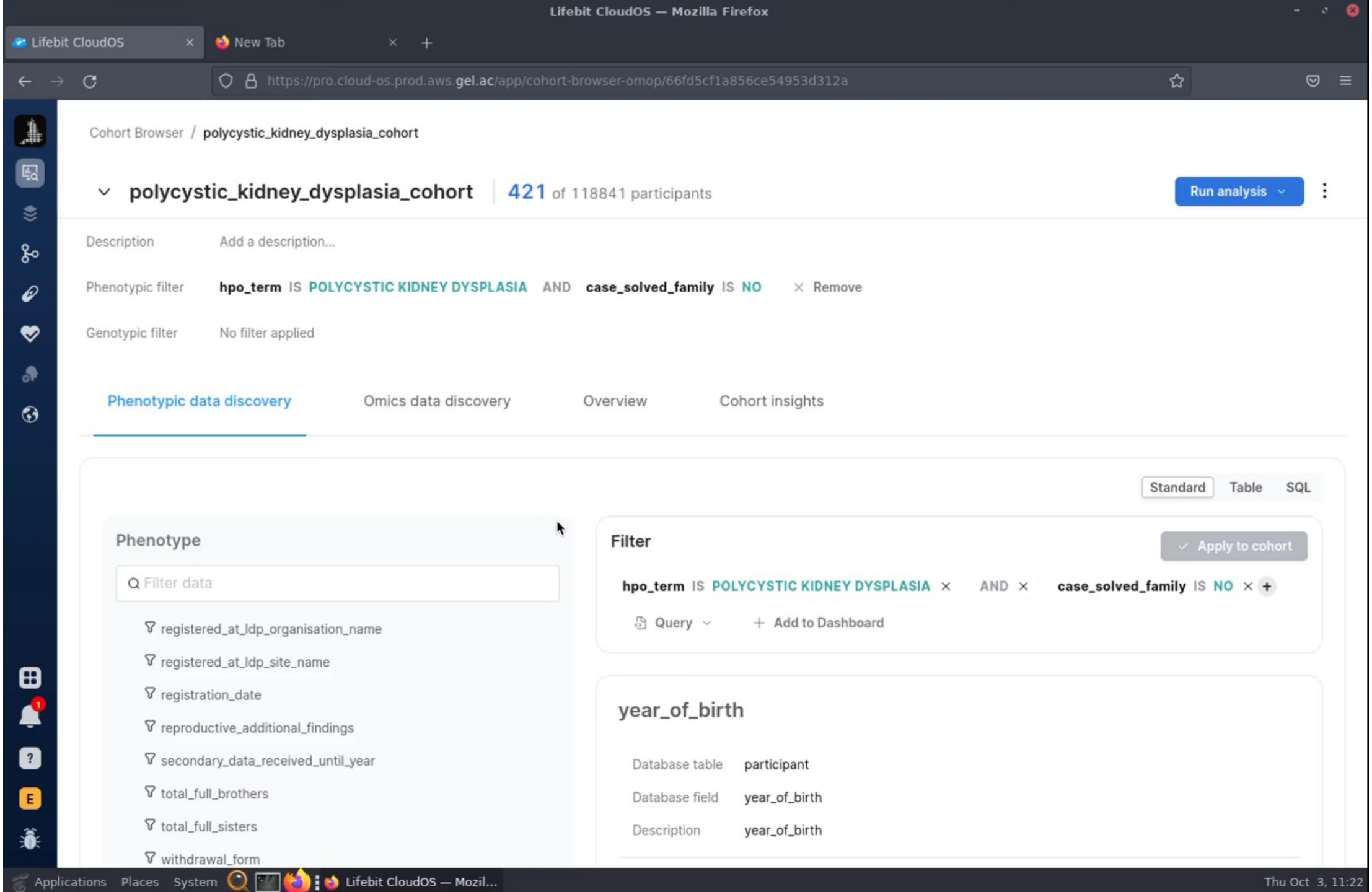
Find phenotypes in cohorts – Interactive session

- Use SQL queries
- Tutorial in R
- Detailed medical history tables

https://re-docs.genomicsengland.co.uk/medhist_api/

Find phenotypes in cohorts demo


Amazon WorkSpaces View Settings Connections Support
Thu 12 Oct 10:01
Amazon WorkSpaces
Computer
Text Editor
Airlock
Research Environment Documentation
Welcome Pack
eperry's Home
Data Discovery
Participant Explorer
report.tsv
Trash
Firefox
Visual Studio Code
Document Viewer
IGV Browser
RE Messages
IVA 2.0
R
Terminal Emulator
Rocket Chat
LibreOffice
GVim
Labkey
Old Firefox Data
Open Targets
LibreOffice 7.6
Genomics england
Applications Places System
[eperry@corp.gel.ac@p...
[IVA v2.2.3 — Mozilla Fir...
[variant_samples_RD38...
Thu Oct 12, 10:01


Lifebit CloudOS — Mozilla Firefox
Lifebit CloudOS
New Tab
https://pro.cloud-os.prod.aws.gel.ac/app/cohort-browser-omop/66fd5cf1a856ce54953d312a
Cohort Browser / polycystic_kidney_dysplasia_cohort
polycystic_kidney_dysplasia_cohort
421 of 118841 participants
Run analysis
Description
Add a description...
Phenotypic filter
hpo_term IS POLYCYSTIC KIDNEY DYSPLASIA AND case_solved_family IS NO
Remove
Genotypic filter
No filter applied
Phenotypic data discovery
Omics data discovery
Overview
Cohort insights
Standard
Table
SQL
Phenotype
Filter data
registered_at_ldp_organisation_name
registered_at_ldp_site_name
registration_date
reproductive_additional_findings
secondary_data_received_until_year
total_full_brothers
total_full_sisters
withdrawal_form
Filter
Apply to cohort
hpo_term IS POLYCYSTIC KIDNEY DYSPLASIA
AND
case_solved_family IS NO
Query
Add to Dashboard
year_of_birth
Database table
participant
Database field
year_of_birth
Description
year_of_birth
Applications
Places
System
Lifebit CloudOS — Mozil...
Thu Oct 3, 11:22

Find phenotypes in cohorts

Participant Explorer

Cohort browser

LabKey API

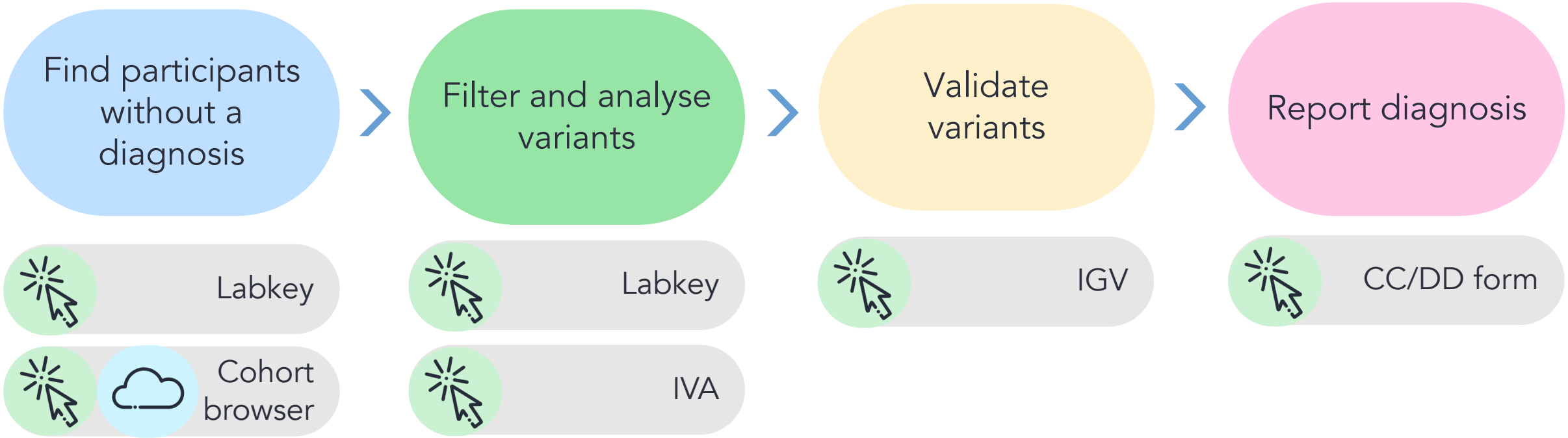
Interactive session

✓	✗
easy to search and interpret phenotypes as words	limited scale can't save queries
graphical output of phenotypes save queries build and analyse in one tool	phenotypes as codes
flexible output	phenotypes as codes
flexible output	phenotypes as codes

https://re-docs.genomicsengland.co.uk/phenotype_to_variants/

4. Finding diagnoses for patients who didn’t get one through primary clinical interpretation

Finding diagnoses for patients who didn't get one through primary clinical interpretation



https://re-docs.genomicsengland.co.uk/patient_to_diagnosis/

Find participants without a diagnosis - Labkey

Search for participants without solved cases

See unapproved diagnoses

https://re-docs.genomicsengland.co.uk/exit_questionnaire/

Find participants without a diagnosis – Cohort Browser

Search for participants without solved cases

See unapproved diagnoses

https://re-docs.genomicsengland.co.uk/exit_questionnaire/

Filter and analyse variants - Labkey

See variants prioritised by tiering and exomiser

Look at panels used

https://re-docs.genomicsengland.co.uk/tiering/
https://re-docs.genomicsengland.co.uk/exomiser/

Filter and analyse variants – IVA case interpreter

- See all variants in a participant
- Filter variants by
  - Genes, consequences, variant frequency etc
  - Family genotypes

https://re-docs.genomicsengland.co.uk/iva_case/

Validate variants – IGV

Identify family BAM file locations

Visualise in IGV

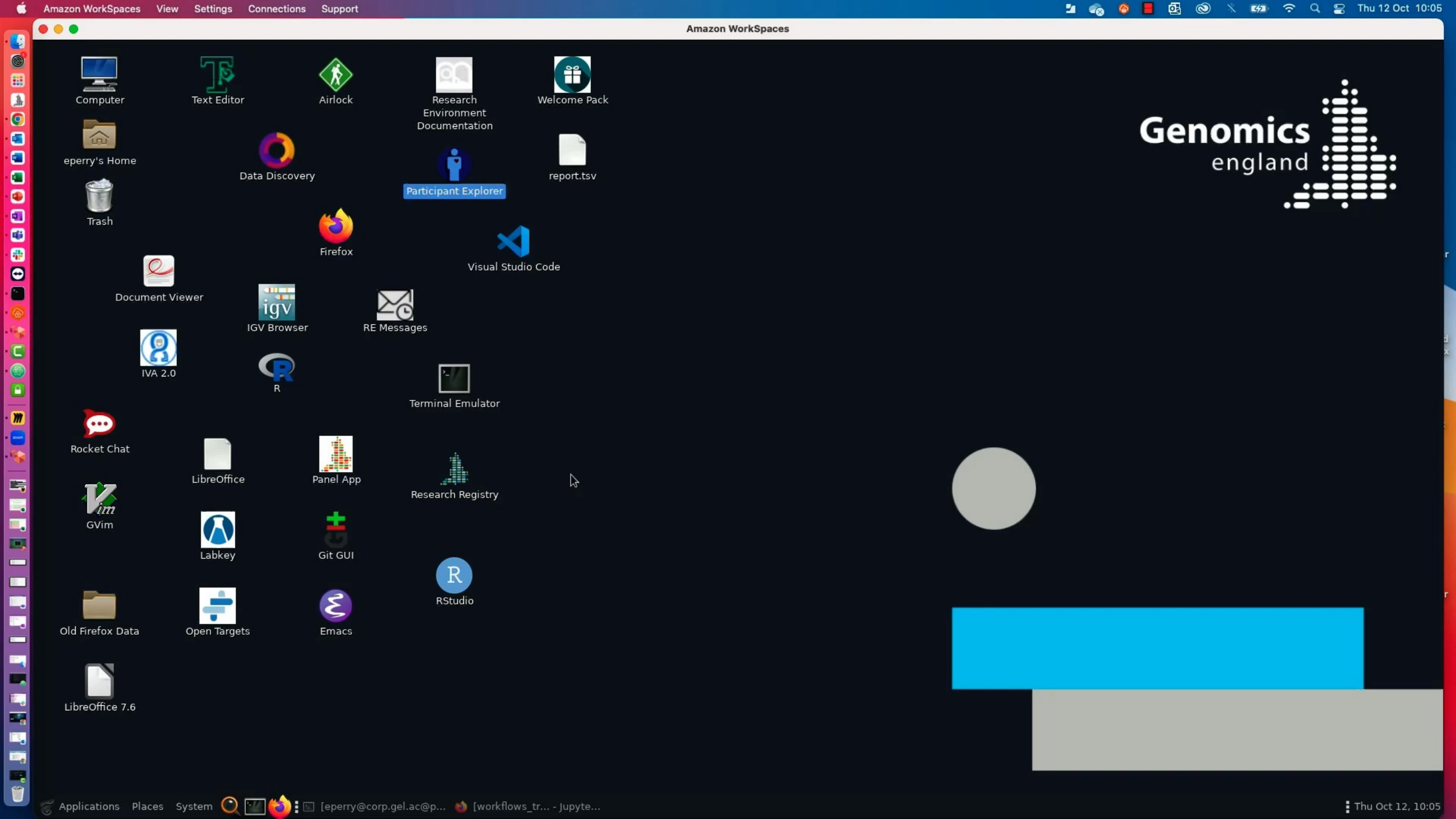
https://re-docs.genomicsengland.co.uk/igv/

Report diagnosis – CI/DD form

- Request to contact clinician for:
  - Consent to publish
  - More samples or phenotype data
- Report diagnosis to GEL

https://re-docs.genomicsengland.co.uk/cri/

Diagnostic discovery demo


Amazon WorkSpaces View Settings Connections Support
Thu 12 Oct 10:05
Amazon WorkSpaces
Computer
Text Editor
Airlock
Research Environment Documentation
Welcome Pack
eperry's Home
Data Discovery
Participant Explorer
report.tsv
Trash
Firefox
Visual Studio Code
Document Viewer
IGV Browser
RE Messages
IVA 2.0
R
Terminal Emulator
Rocket Chat
LibreOffice
Panel App
Research Registry
GVim
Labkey
Git GUI
RStudio
Old Firefox Data
Open Targets
Emacs
LibreOffice 7.6
Genomics england
Applications Places System
[eperry@corp.gel.ac@p...
[workflows_tr... - Jupyte...
Thu Oct 12, 10:05

5. Help and questions

Getting help

Check our documentation:
https://re-docs.genomicsengland.co.uk/
Click on the documentation icon in the environment

Contact our Service Desk:
https://jiraservicedesk.extge.co.uk/plugins/servlet/desk

Questions

All your microphones are muted

Use the Zoom Q&A to ask questions

Upvote your favourite questions: if we are short on time we will prioritise those with the most votes

Training sessions 2024

12/11	Running workflows on the HPC and Cloud
10/12	Introduction to the RE

Materials from past training all online

https://re-docs.genomicsengland.co.uk/upcoming/

Training sessions 2025

3rd Tuesday every month

Introduction to the RE

21/1 18/2 18/3

15/4 20/5 17/6

https://re-docs.genomicsengland.co.uk/upcoming/

Materials from past training all online

Training sessions 2025

14/1	Using the Research Environment for clinical diagnostic discovery
11/2	Importing data and tools to use in the RE
11/3	Working with R in the RE
8/4	Working with python in the RE
13/5	Building cancer cohorts and survival analysis
10/6	Building rare disease cohorts with matching controls

https://re-docs.genomicsengland.co.uk/upcoming/

In-person training day

- For any registered RE users
- 20th November 2024
- Held in our offices at Canary Wharf
- Hybrid attendance
- Hands-on practicals and exercises

https://www.genomicsengland.co.uk/events/research-environment-training-event-november-2024

Feedback

Thank you

Visit: https://re-docs.genomicsengland.co.uk/