



Research Summit 2024



An introduction to the Research Environment: live training session

Emily Perry

Research Engagement Manager

9 July 2024



Agenda

1

Genomic data from the 100,000 Genomes Project and NHS GMS

2

Clinical and phenotype data

3

Tools in the Genomics England Research Environment

4

Command line and programmatic access

5

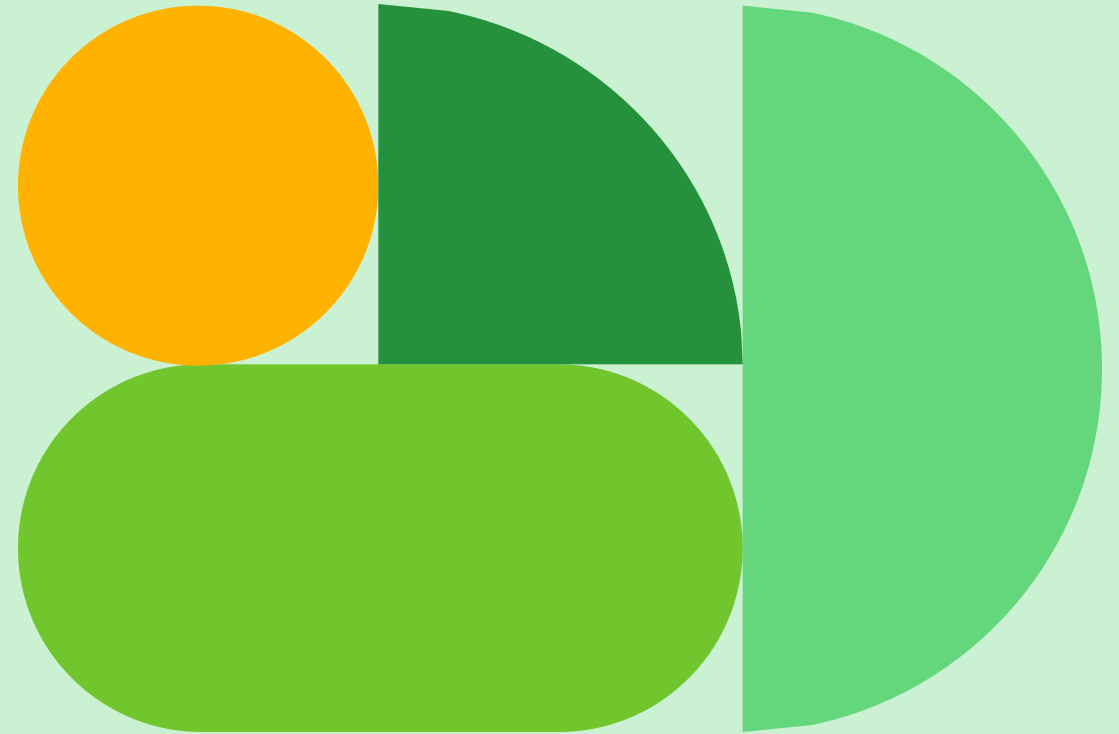
Restricted export

6

Getting access to the Genomics England Research Environment

7

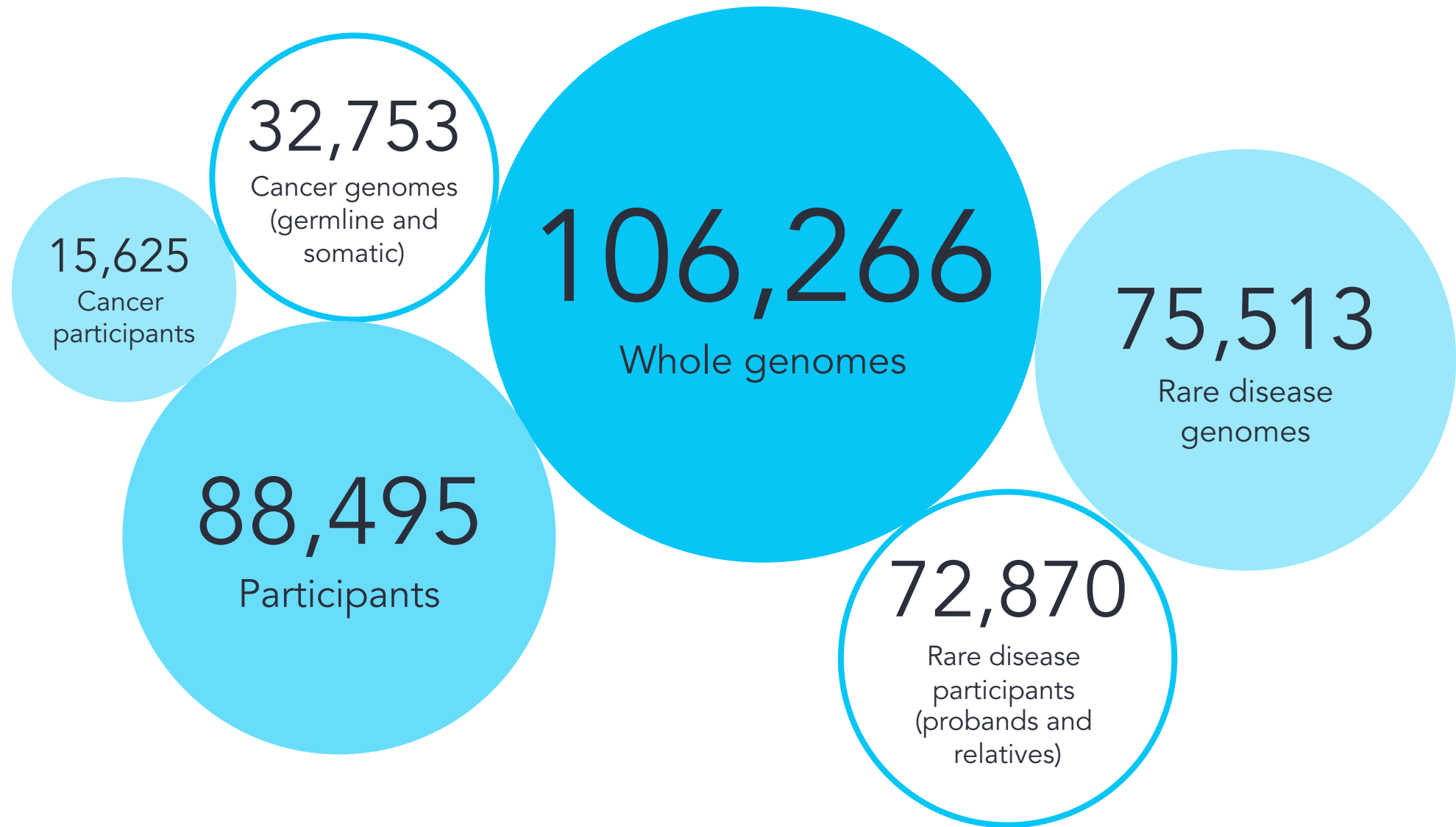
Help and questions



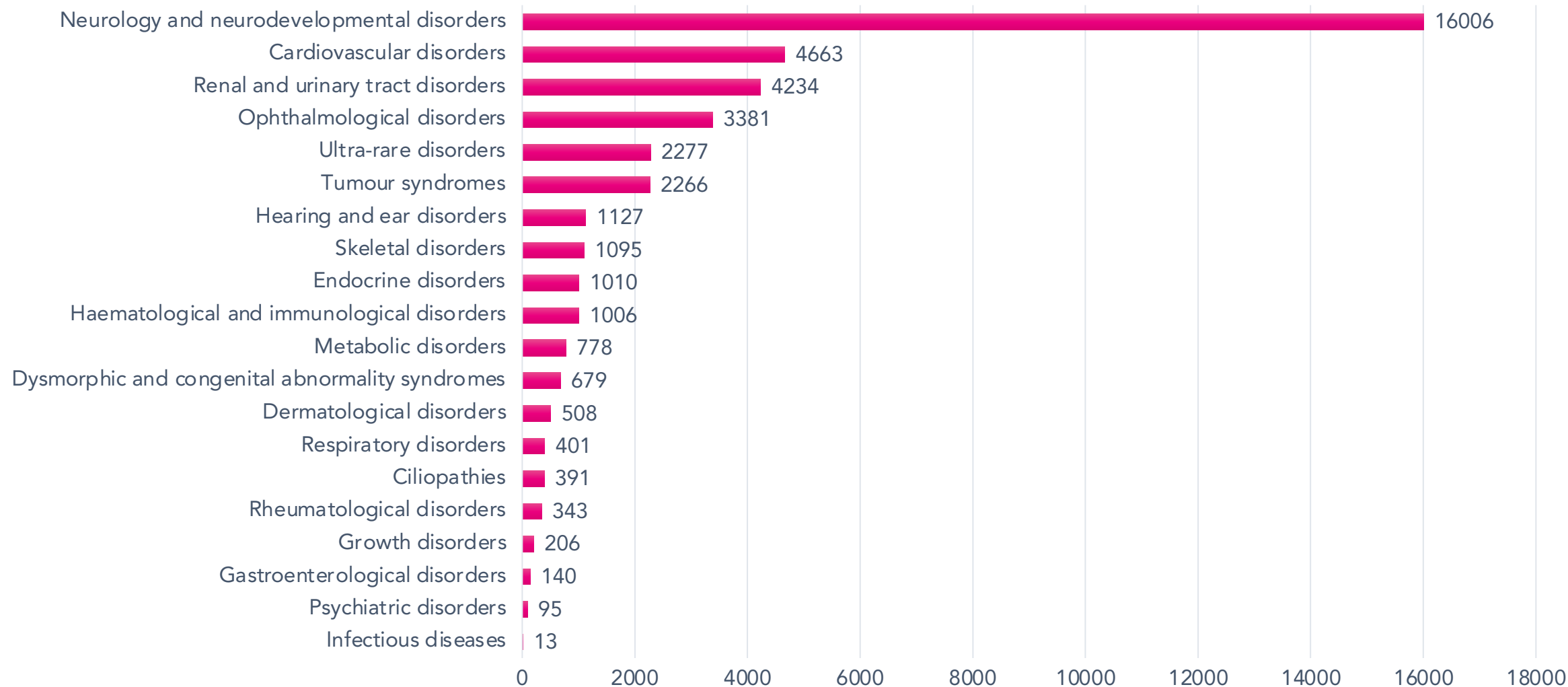
Genomic data from the 100,000 Genomes Project and NHS GMS



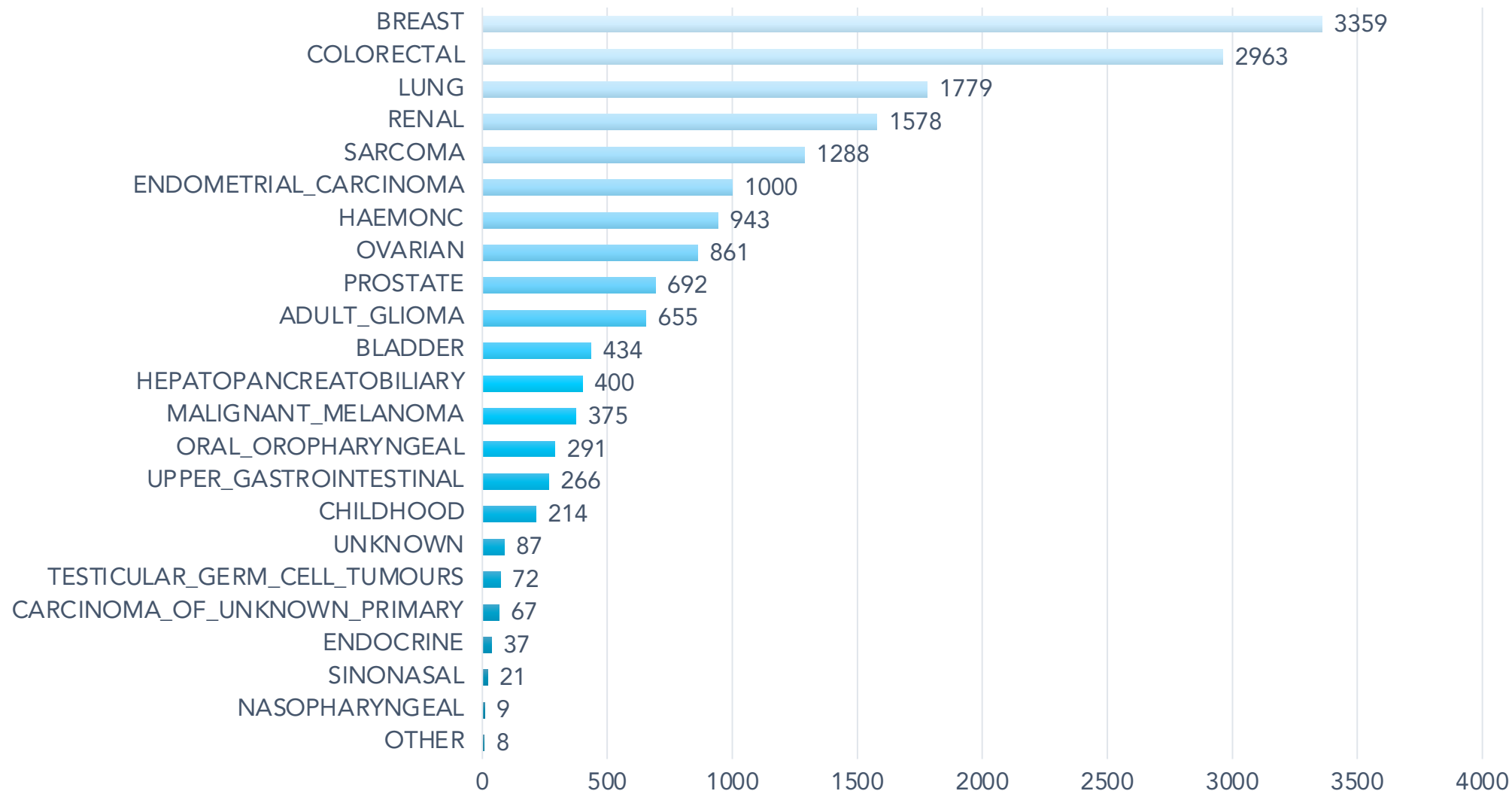
100,000 Genomes Project



100,000 Genomes rare disease



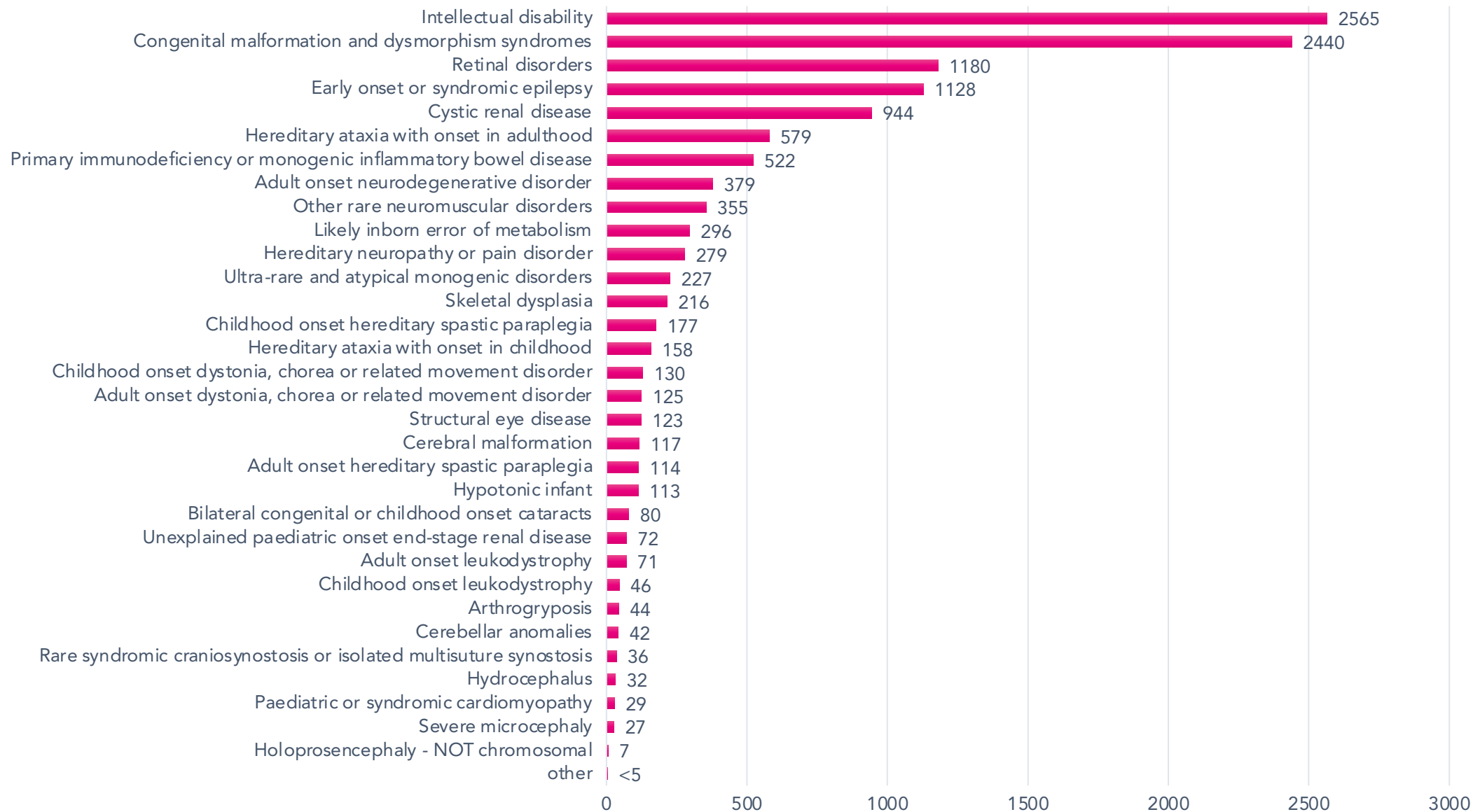
100,000 Genomes cancer



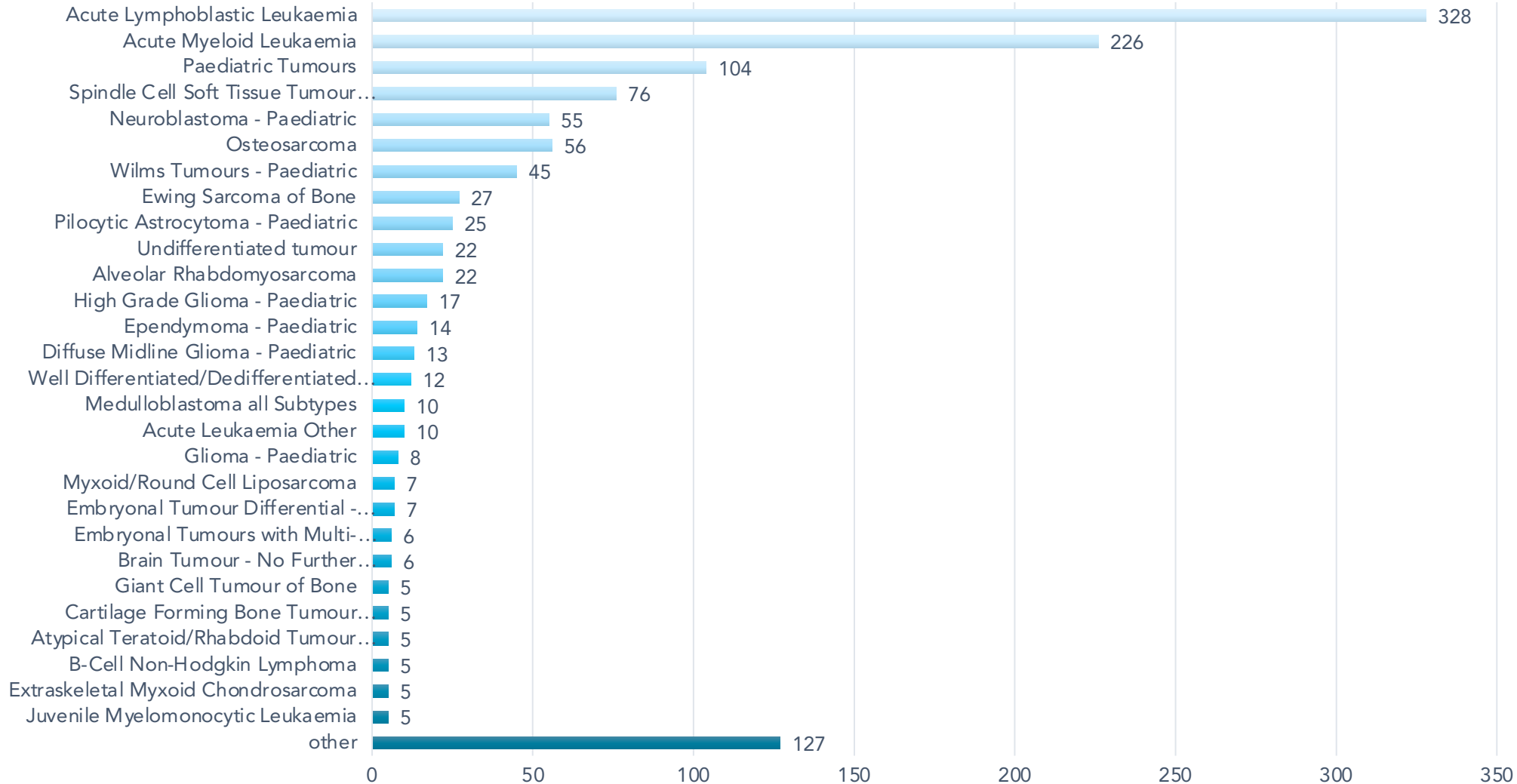
NHS GMS



NHS GMS rare disease



NHS GMS cancer



For **EVERY** genome

Alignment
as BAM or CRAM
files

Variant calls
as VCFs, including
gVCF, repeats VCFs,
structural variant
VCFs

Analysis
Variant tiering and
tumour mutational
signatures

Clinical and phenotype data

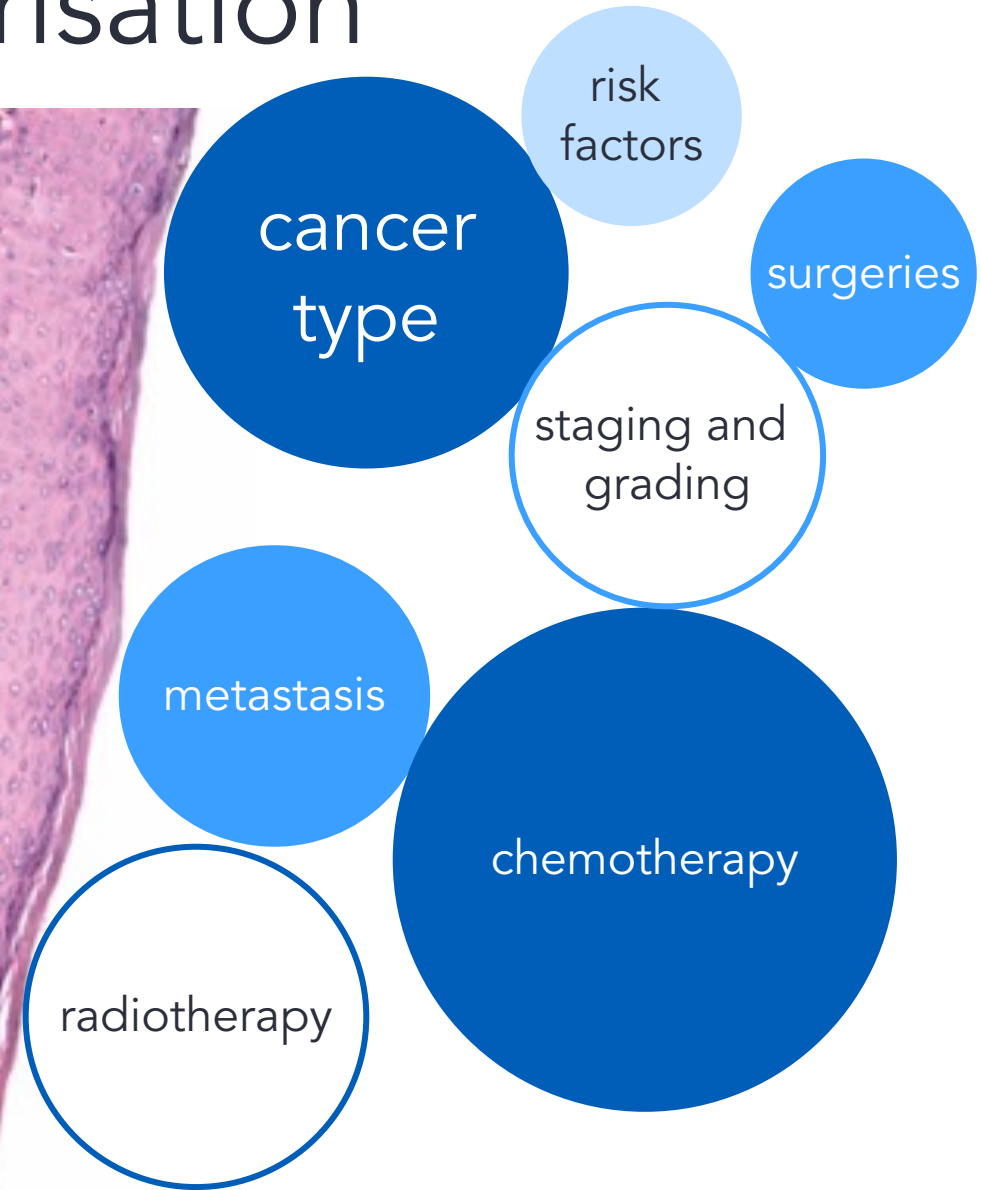
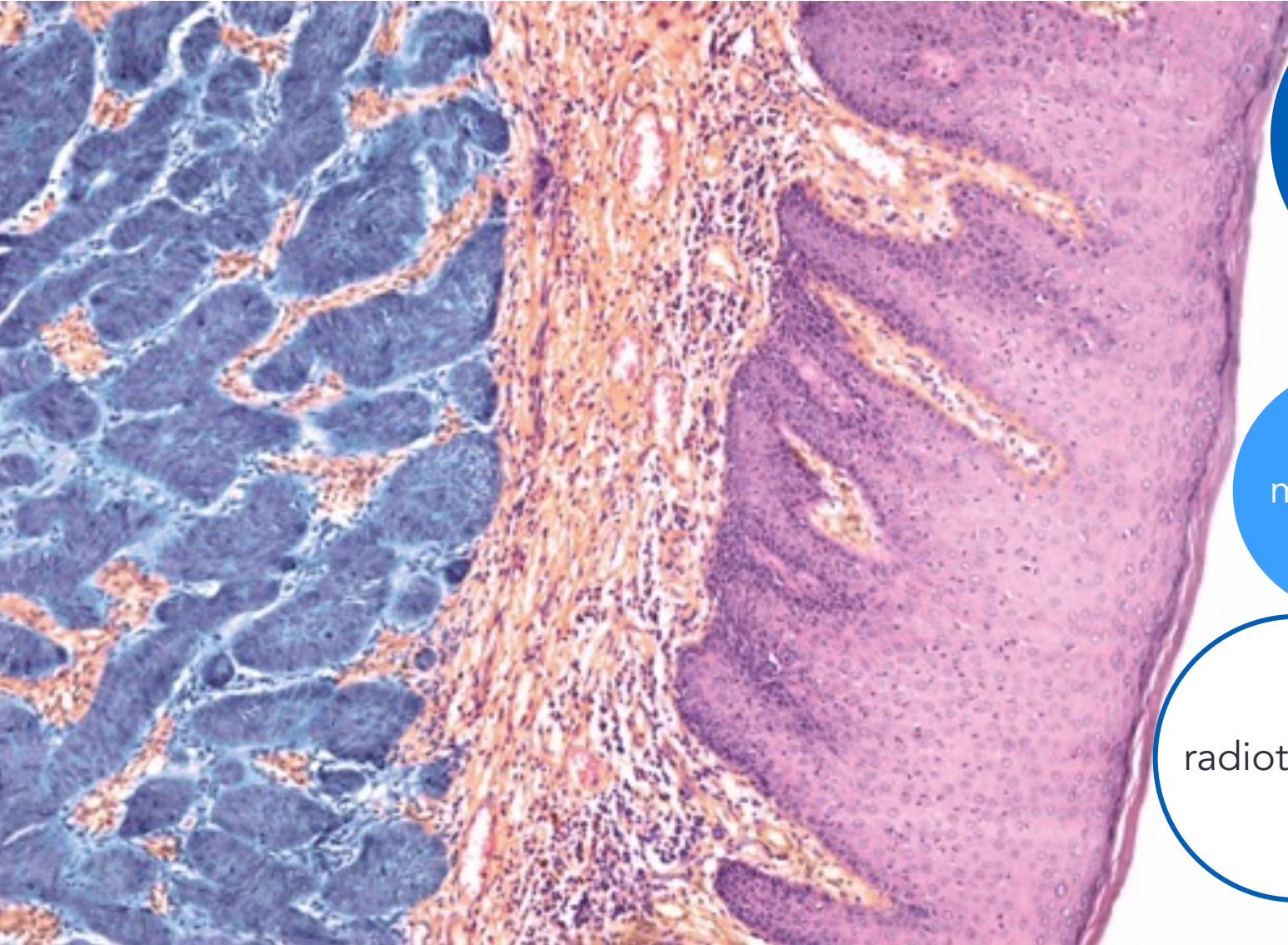


Rare disease phenotyping

- Disease classification
- HPO terms present/absent
- Measurements and observations (not universal)
 - general measurements
 - early childhood observations
 - details of imaging (but not results)
 - genetic tests
 - lab tests

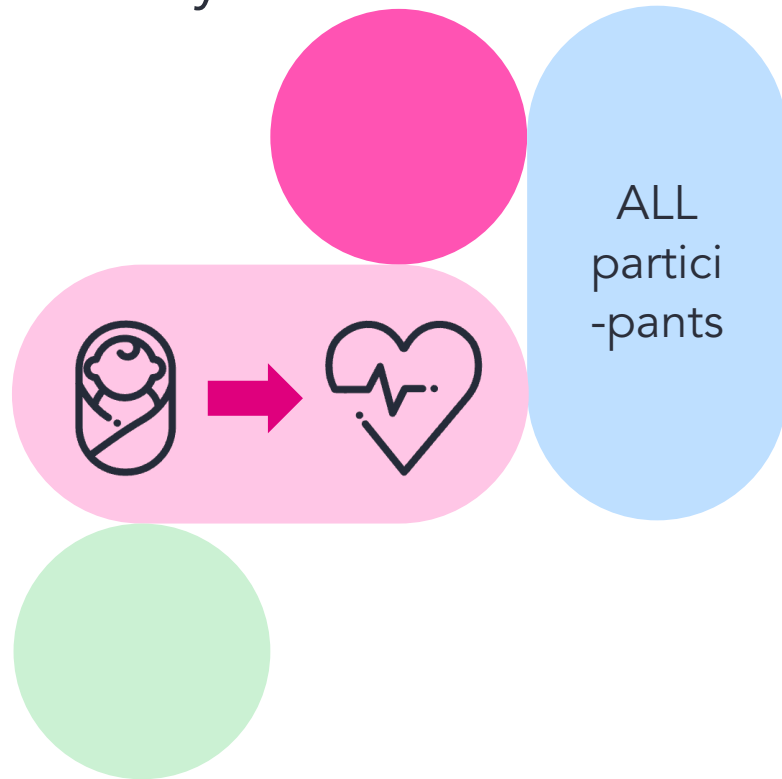


Cancer tumour characterisation

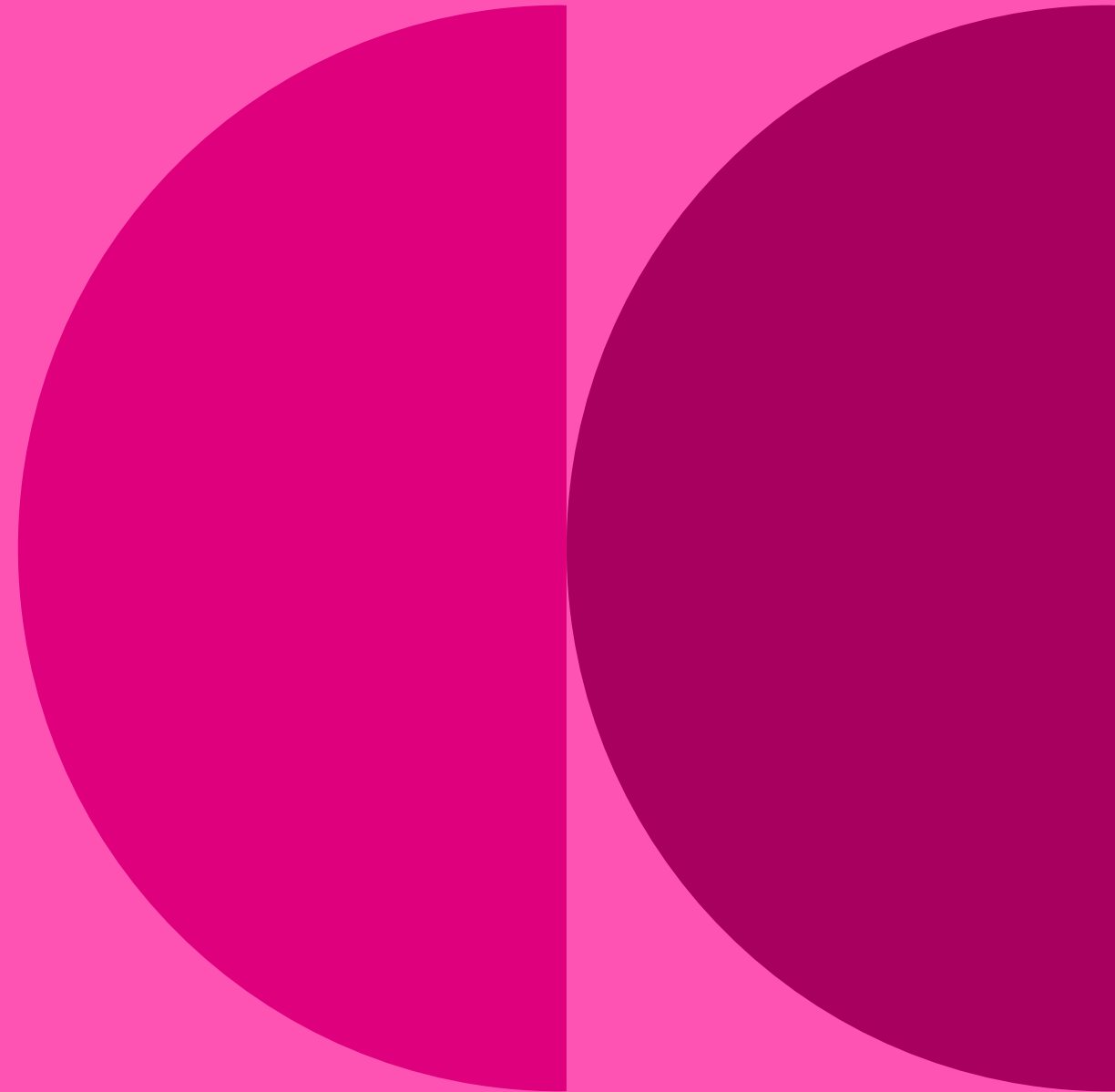


Medical history – 100k

- NHSE hospital episode statistics
- Mental health data
- Mortality



Tools in the Genomics England Research Environment



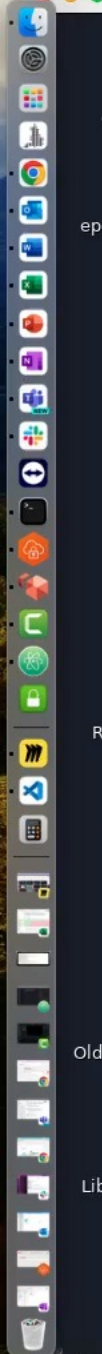
LabKey

- Central database of:
 - Clinical data
 - Results of bioinformatics analysis
 - Locations of genomic files
- Point and click interface
- API

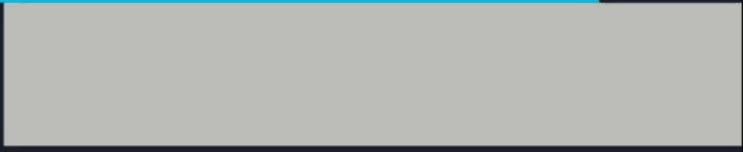
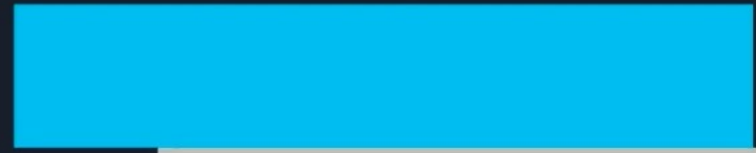


LabKey®

LabKey demo



- Computer
- Text Editor
- Airlock
- Research Environment Documentation
- Welcome Pack
- eperry's Home
- Data Discovery
- Participant Explorer
- report.tsv
- Trash
- Firefox
- Visual Studio Code
- Ensembl
- 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
- CloudOS



Participant Explorer

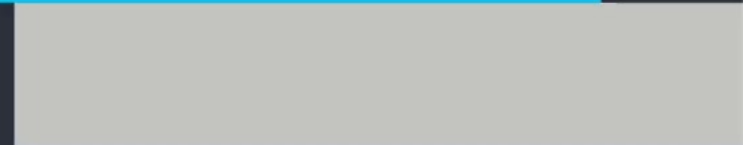
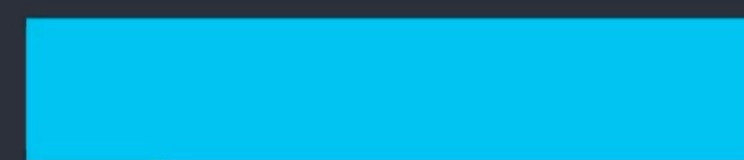
- Search for participants by:
 - IDs
 - Clinical concepts
 - Personal details
- View/compare medical histories



Participant Explorer demo



- Computer
- Text Editor
- Airlock
- Research Environment Documentation
- Welcome Pack
- eperry's Home
- Data Discovery
- Participant Explorer
- report.tsv
- Trash
- Firefox
- Visual Studio Code
- Ensembl
- 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
- CloudOS



Interactive variant analysis (IVA)

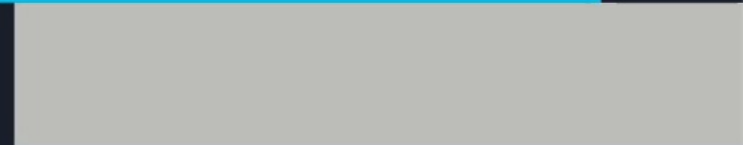
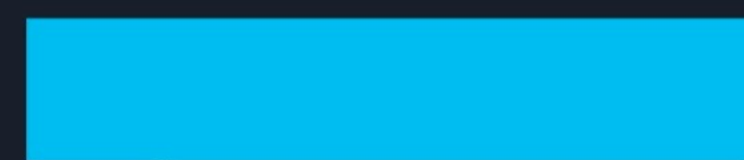
- Find participants with genetic variants
- Filter variants in a participant by family genotypes
- Filter on genome features



IVA demo

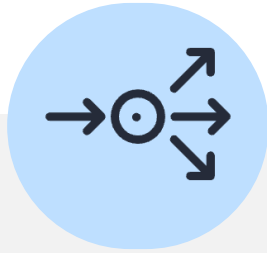


- Computer
- Text Editor
- Airlock
- Research Environment Documentation
- Welcome Pack
- eperry's Home
- Data Discovery
- Participant Explorer
- report.tsv
- Trash
- Firefox
- Visual Studio Code
- Ensembl
- Document Viewer
- IGV Browser
- RE Messages
- IVA 2.0
- R
- Terminal Emulator
- Rocket Chat
- LibreOffice
- Panel App
- Research Registry
- GVim
- Labkey
- Git GUI
- RStudio
- Old Firefox Data
- Open Targets
- Emacs
- LibreOffice 7.6
- CloudOS

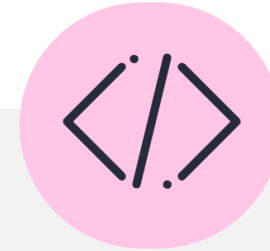


Command line and programmatic access

LabKey API



Combine queries between tables



Work in a variety of programming languages (support for Python and R) using SQL queries



Replicate queries between releases and analyses

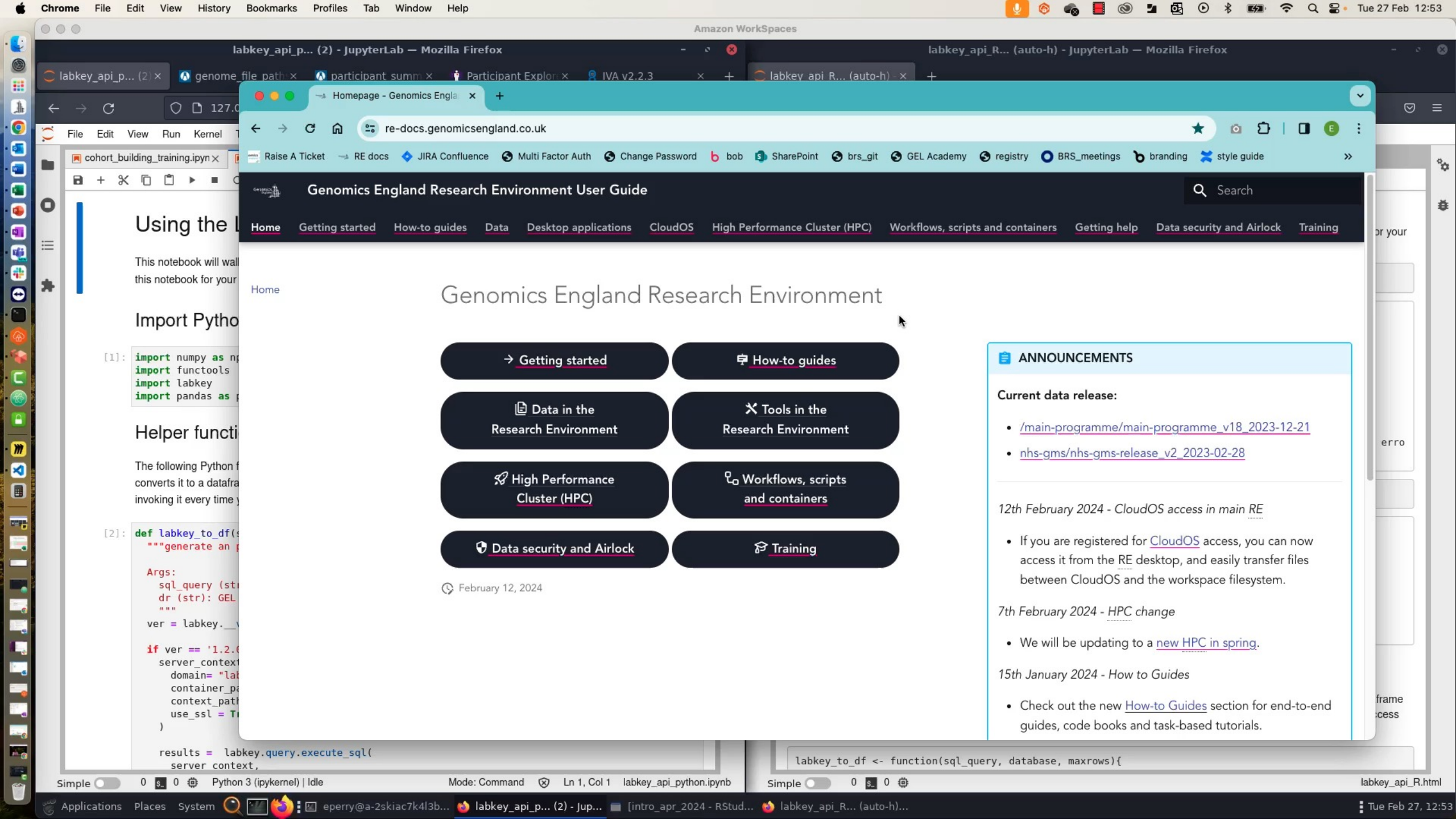


Work locally and on the HPC

Programming tools in the RE



LabKey demo



Genomics England Research Environment

- [Getting started](#)
- 📄 [How-to guides](#)
- 📄 [Data in the Research Environment](#)
- ⚙️ [Tools in the Research Environment](#)
- 🚀 [High Performance Cluster \(HPC\)](#)
- 🔗 [Workflows, scripts and containers](#)
- 🔒 [Data security and Airlock](#)
- 🎓 [Training](#)

🕒 February 12, 2024

ANNOUNCEMENTS

Current data release:

- [/main-programme/main-programme_v18_2023-12-21](#)
- [nhs-gms/nhs-gms-release_v2_2023-02-28](#)

12th February 2024 - [CloudOS access in main RE](#)

- If you are registered for [CloudOS](#) access, you can now access it from the [RE](#) desktop, and easily transfer files between CloudOS and the workspace filesystem.

7th February 2024 - [HPC change](#)

- We will be updating to a [new HPC in spring](#).

15th January 2024 - [How to Guides](#)

- Check out the new [How-to Guides](#) section for end-to-end guides, code books and task-based tutorials.

```
[1]: import numpy as np
import functools
import labkey
import pandas as pd
```

```
[2]: def labkey_to_df(sql_query, database, maxrows):
    """generate an dataframe from labkey"""
    ver = labkey.__version__

    if ver == '1.2.0':
        server_context = labkey.get_server_context(
            domain="labkey",
            container_path="labkey",
            context_path="labkey",
            use_ssl=True
        )
```

```
labkey_to_df <- function(sql_query, database, maxrows){
```

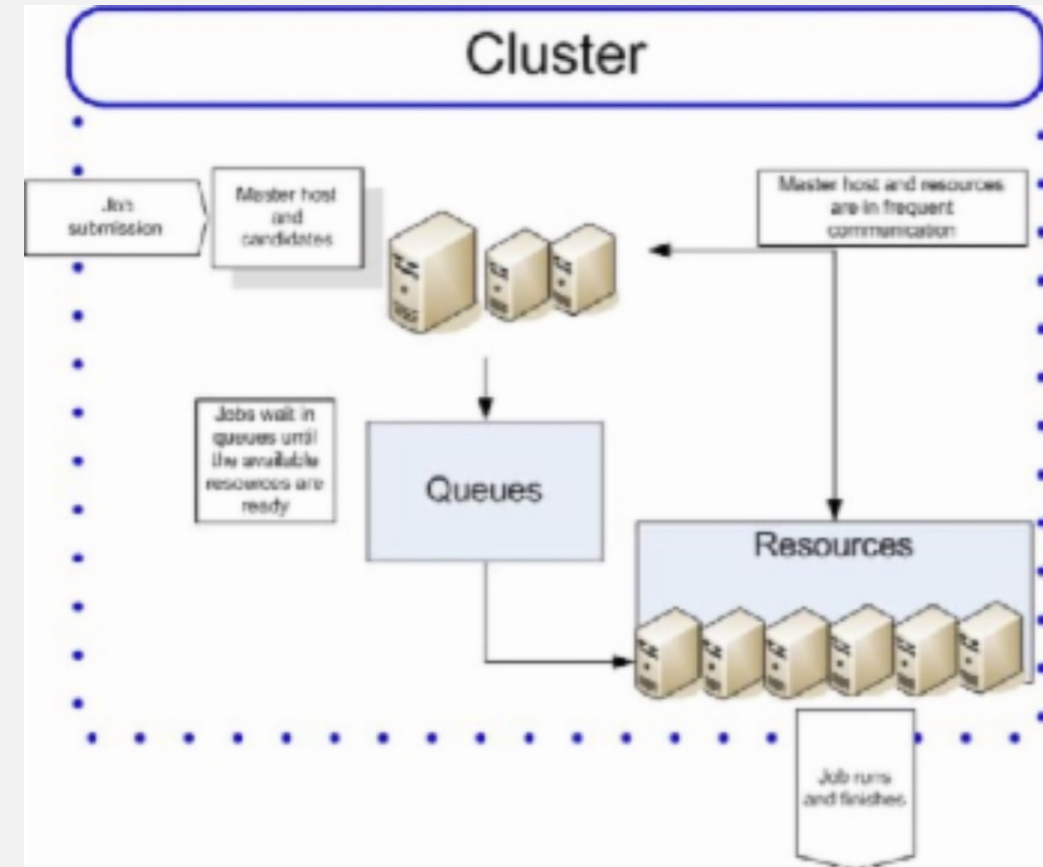
High Performance Cluster

Programming languages

Bioinformatics tools

Pipelining tools

Singularity



HPC demo



Computer, Text Editor, Airlock, Research Environment Documentation, Welcome Pack

eperry's Home, Data Discovery, Participant Explorer, report.tsv

Trash, Firefox, Ensemble, Document Viewer, IGV Browser, RE Messages, IVA 2.0, R, Rocket Chat, LibreOffice, Panel App, Gvim, Labkey, Git GUI, Old Firefox Data, Open Targets, Emacs, LibreOffice 7.6, CloudOS

```

eperry@a-2skiac7k413bq:~
File Edit View Search Terminal Tabs Help
eperry@a-2skiac... x eperry@a-2skiac... x eperry@a-2skiac... x eperry@a-2skiac... x
[eperry@a-2skiac7k413bq ~]$

```



Restricted export

Our contract with participants

"...although researchers can look at your data and ask questions about it, they can only take away the answers to their questions (their results). They can't copy or take away any of your individual data."



A Trusted Research Environment

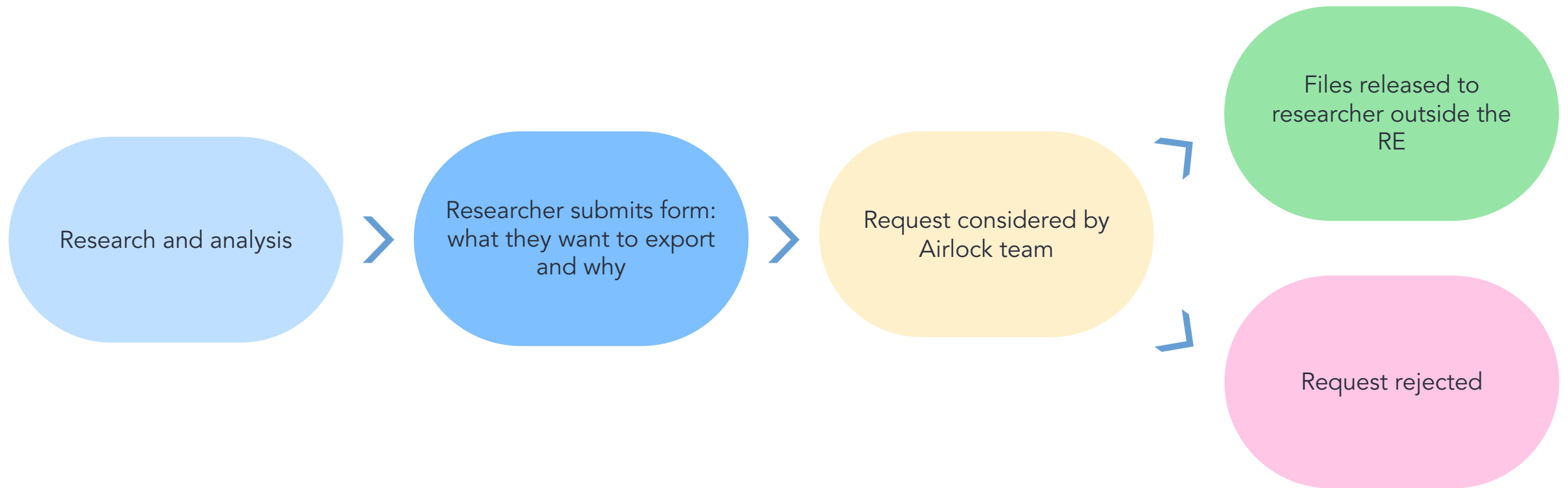
Researchers can...

- Work with the data in the RE
- Copy/paste in
- Bring in Containers
- Access whitelisted websites
- Request to export the results of their analysis

Researchers cannot...

- Share folders between their computer and the VM
- Copy/paste out
- Export files
- Access most of the internet

The Airlock



Airlock demo

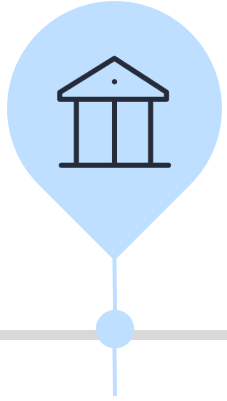
The desktop environment features a dark theme. The dock on the left contains icons for applications such as Computer, eperry's Home, Trash, Ensembl, Rocket Chat, Gvim, Old Firefox Data, LibreOffice 7.6, Text Editor, Data Discovery, Firefox, Document Viewer, IGV Browser, IVA 2.0, R, Rocket Chat, LibreOffice, Panel App, Gvim, Labkey, Git GUI, Open Targets, CloudOS, Airlock, Research Environment Documentation, Participant Explorer, report.tsv, Visual Studio Code, RE Messages, Terminal Emulator, Research Registry, and RStudio. The 'Genomics england' logo is prominently displayed on the right side. The taskbar at the bottom shows several terminal windows with titles like [eperry@a-2skiac7k4]3..., [labkey_api_p... (2) - Ju..., [intro_apr_2024 - RStud..., and [labkey_api_R... (auto-h... [eperry].



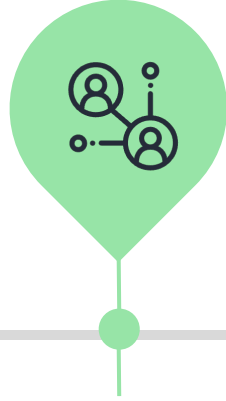
Getting access to the Genomics England Research Environment



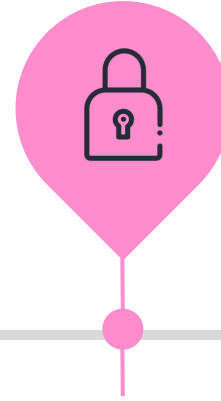
Applying for access - Academia



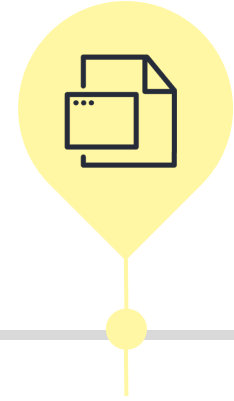
University/Institute
signs agreement



Researcher applies to
join Research Network



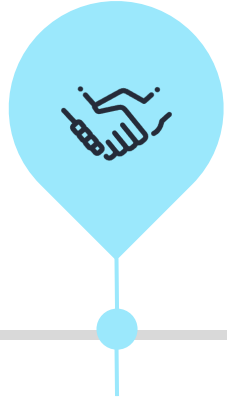
Researcher
completes
information
governance training



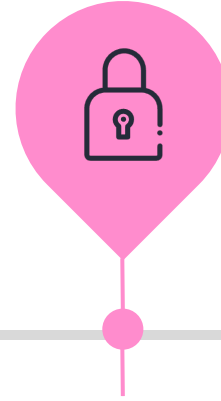
Access the RE



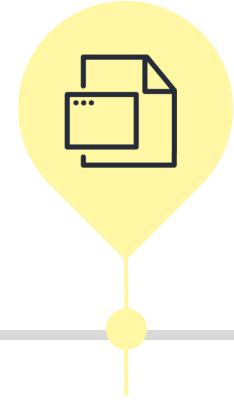
Applying for access - Industry



Company joins
Discovery Forum

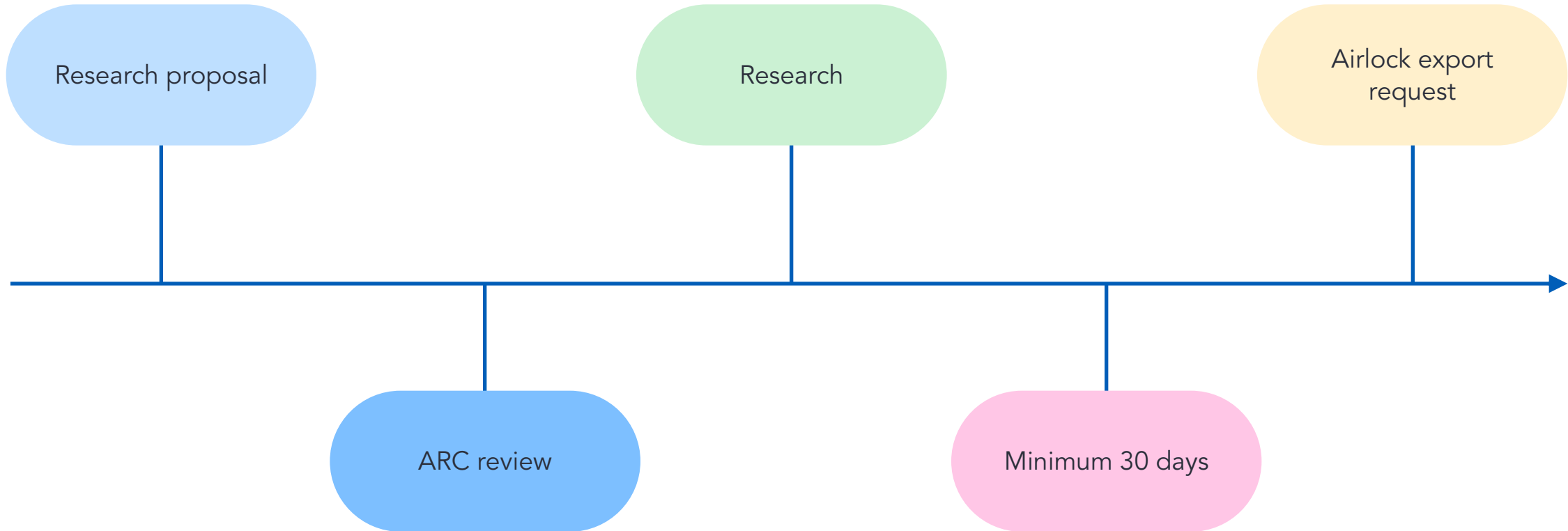


Researcher
completes
information
governance training



Access the RE

Project proposal



Help and questions

Getting help



Come and find us at the booth today



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>

Who to talk to today...

Industry partnerships



James



Kate

Academic research



Pela



Ailin

Bioinformatics analysis with the RE



Roel



Emily



Loukas



Asier



Chris



Christian



Miruna



Magda



Elena



Ella



Xiangrui



Training sessions

16/7

Getting medical records for participants

10/9

Using GEL data for publications and reports

8/10

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

12/11

Running workflows on the HPC and Cloud

10/12

Introduction to the RE

Free monthly
virtual
training

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



Feedback



Thank you

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