Working with the Genomics England Research Environment

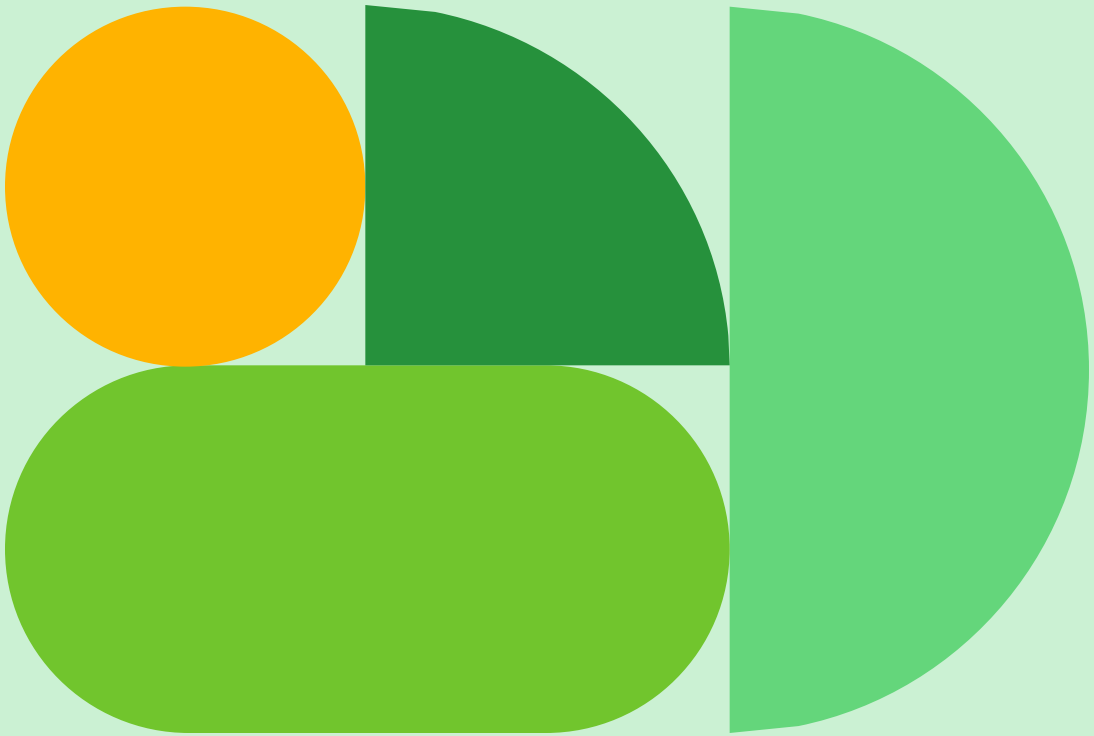
Emily Perry

Research Engagement Manager

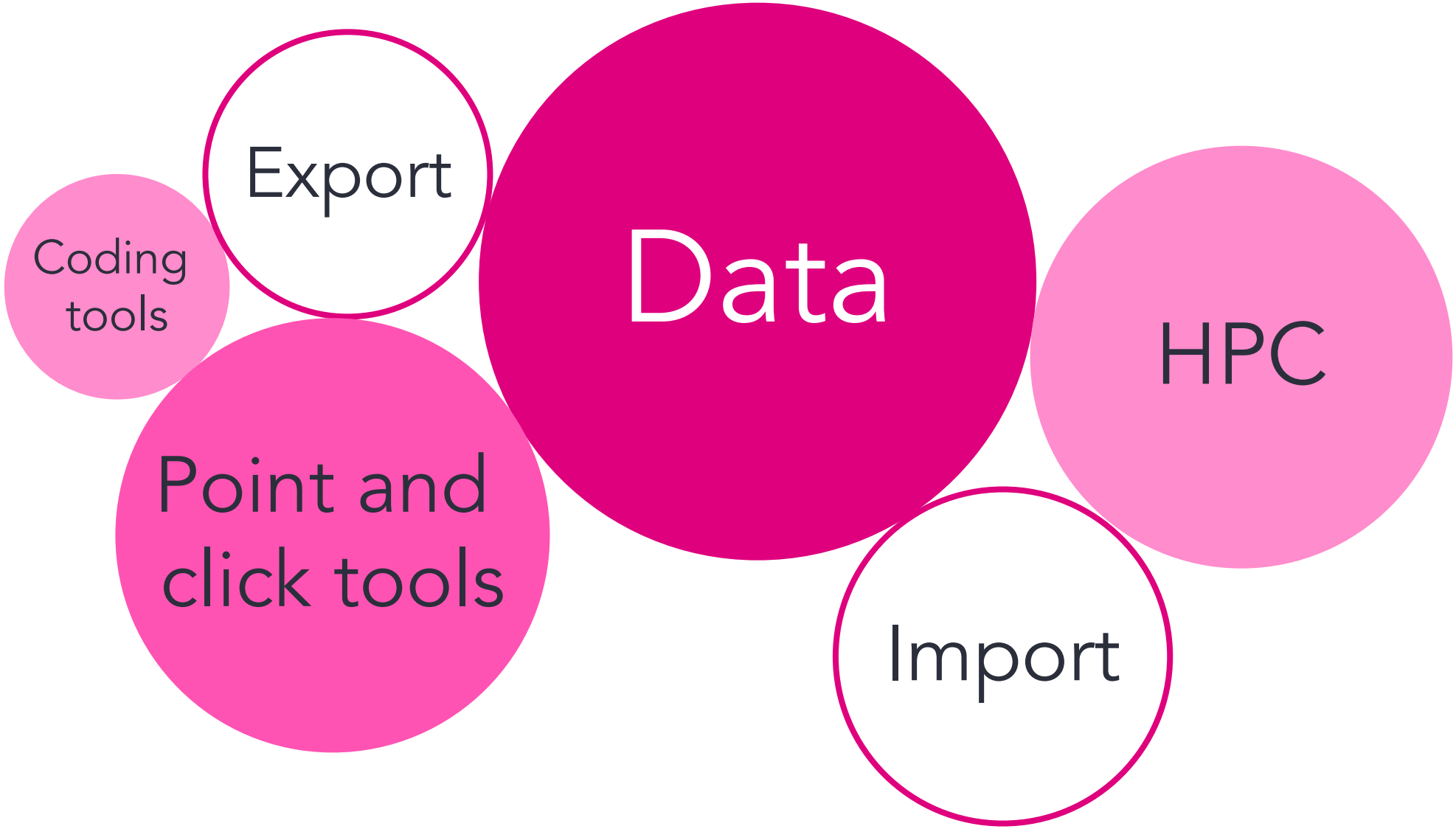
20th November 2024

Agenda

9.00-9.10	Introduction and admin
9.10-10.15	Teaching
10.15-10.45	Break
10.45-12.30	Teaching
12.30-13.30	Lunch
13.30-15.00	Teaching
15.00-15.30	Break
15.30-16.30	Teaching
16.30-17.00	Open Q&A/clinic



We will cover…



Questions

No such thing as a stupid question

Helpers

Roel Bevers
Senior Bioinformatician - Research Services

Christian Bouwens
Bioinformatician - Research Services

Alex Ho
Bioinformatician BRSC

Matthieu Vizuete-Forster
Bioinformatician - Research Services

Canary Wharf
L21 Fire Exits Stairs

Key

- Stairs
- Lifts
- Door
- Emergency fire exit
- Hub — Kitchen, social hub
- 1 person phone room
- 2 person phone room
- Bookable meeting space/room accessible to everyone
- Not bookable
- Disabled toilets
- Womens toilets
- Mens toilets
- Showers
- Refresh, recycle and print
- Lockers
- 2 person booths
- Bookable desk accessible to everyone
- Accessible to everyone

X Fire Exits stairs for safe escape to LG

Genomics England

REF-FAC-011
Canary Wharf Level 21
Fire Exits
V1.0

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
- If you are in person, you are not allowed to take any photos or videos
- If you are joining virtually, 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)

What is the RE?

A virtual machine

Genomic data

Phenotype data

Point and click tools

HPC and command line tools

A Trusted Research Environment

You can…

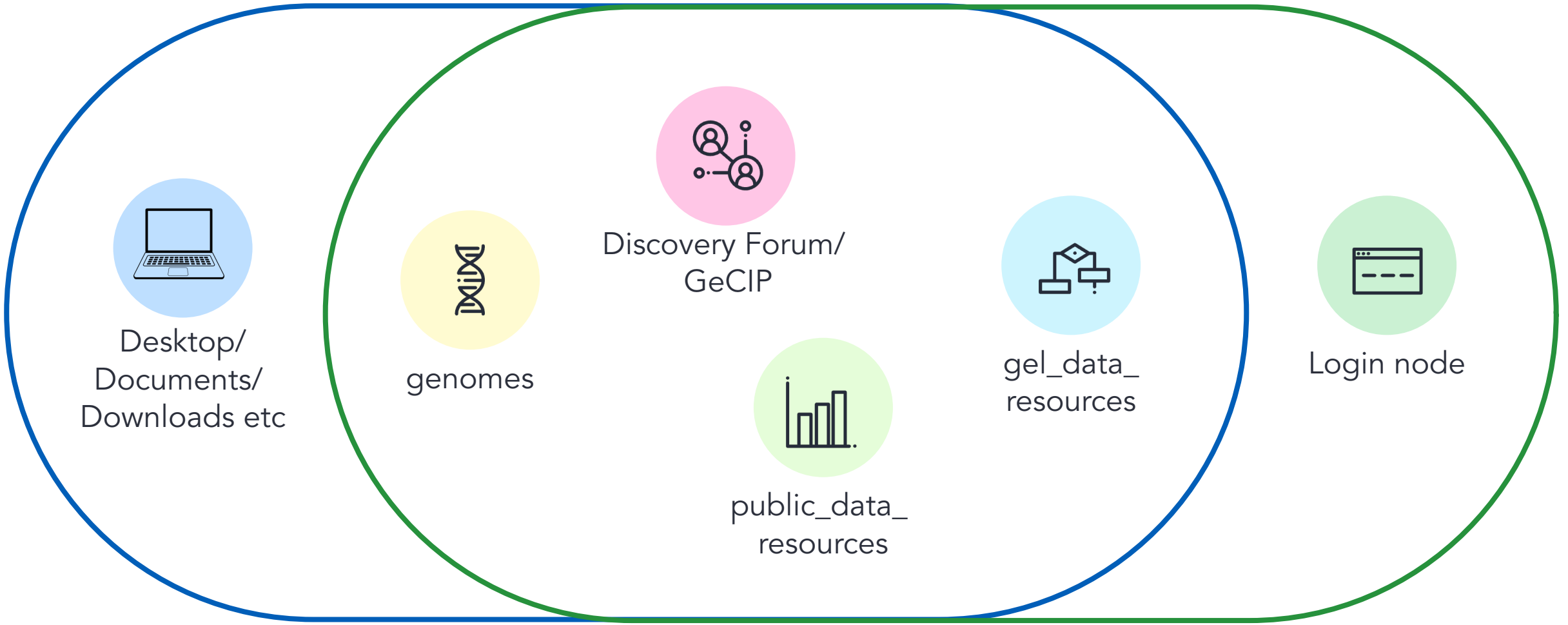
- Work with the data in the RE
- Copy/paste in
- Bring in Containers
- Access whitelisted websites
- Request to export the results of your analysis

You cannot…

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

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

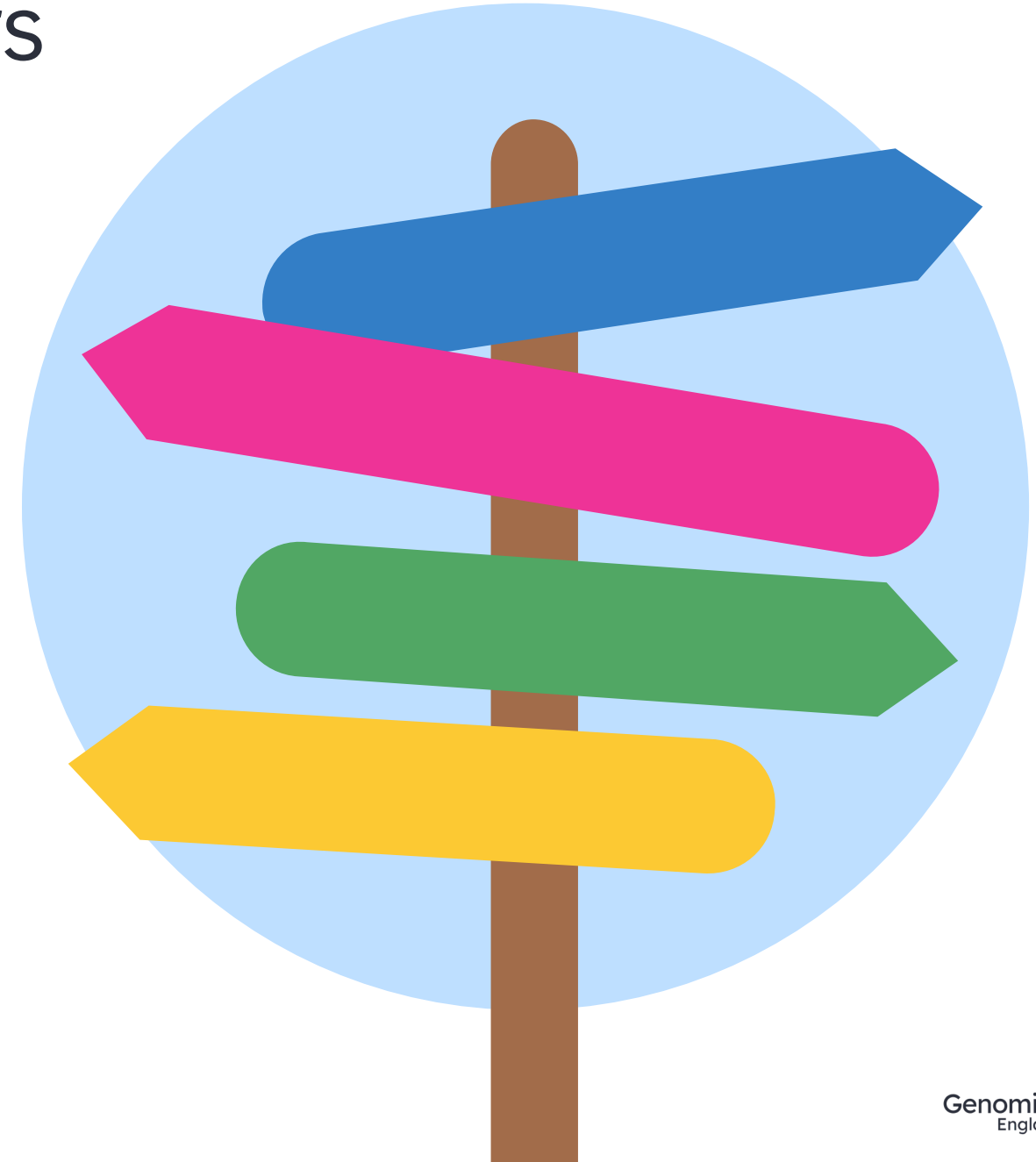
Files in the RE



What is the RE? demo

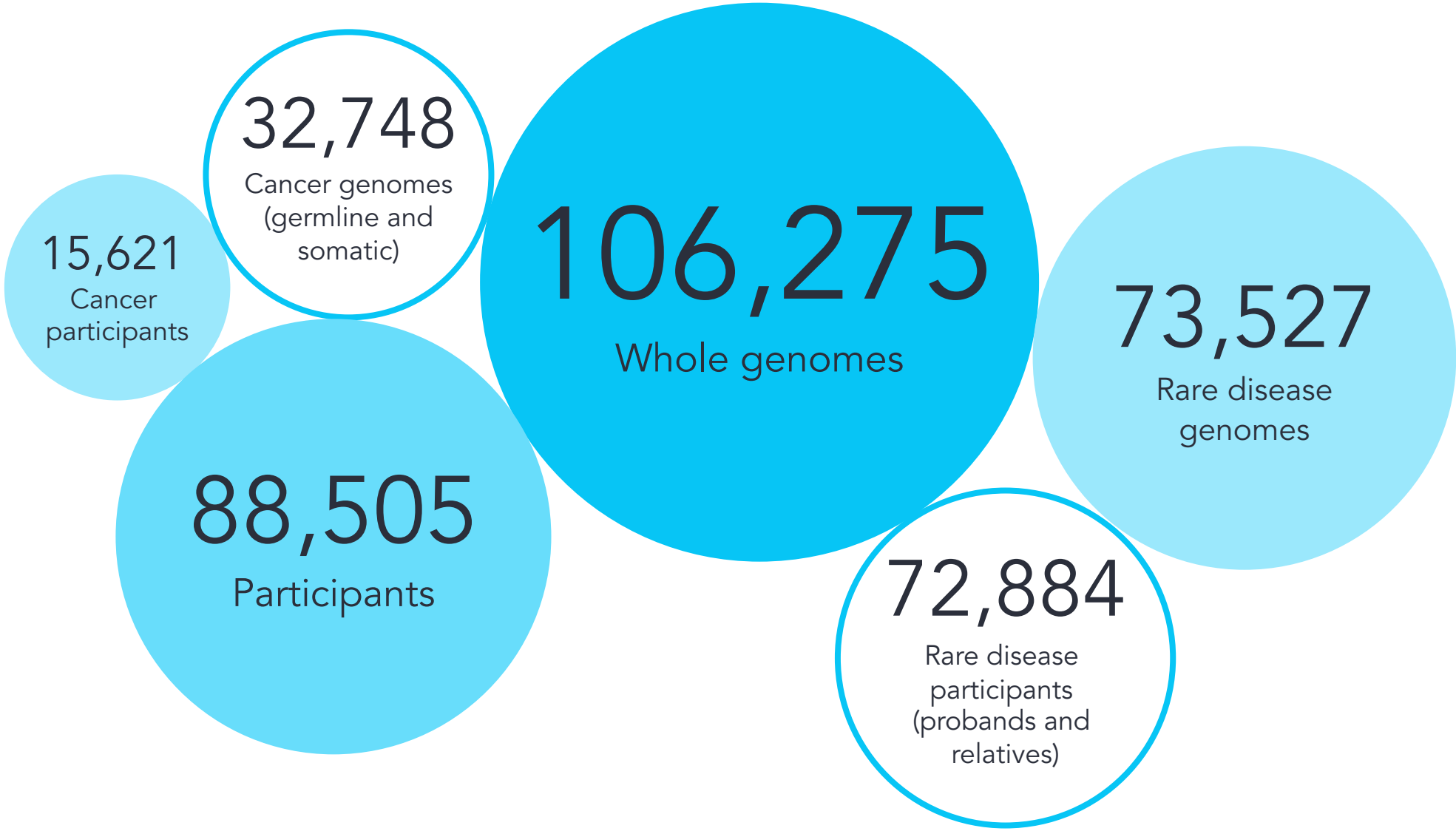
Task – working folders

1. Login to the RE
2. Identify your re_gecip or discovery_forum folder
3. Create a working directory in that folder for today
4. Copy the materials for today from gel_data_resources/example_scripts/workshop_scripts/all_day_training_Nov_2024 to that working directory



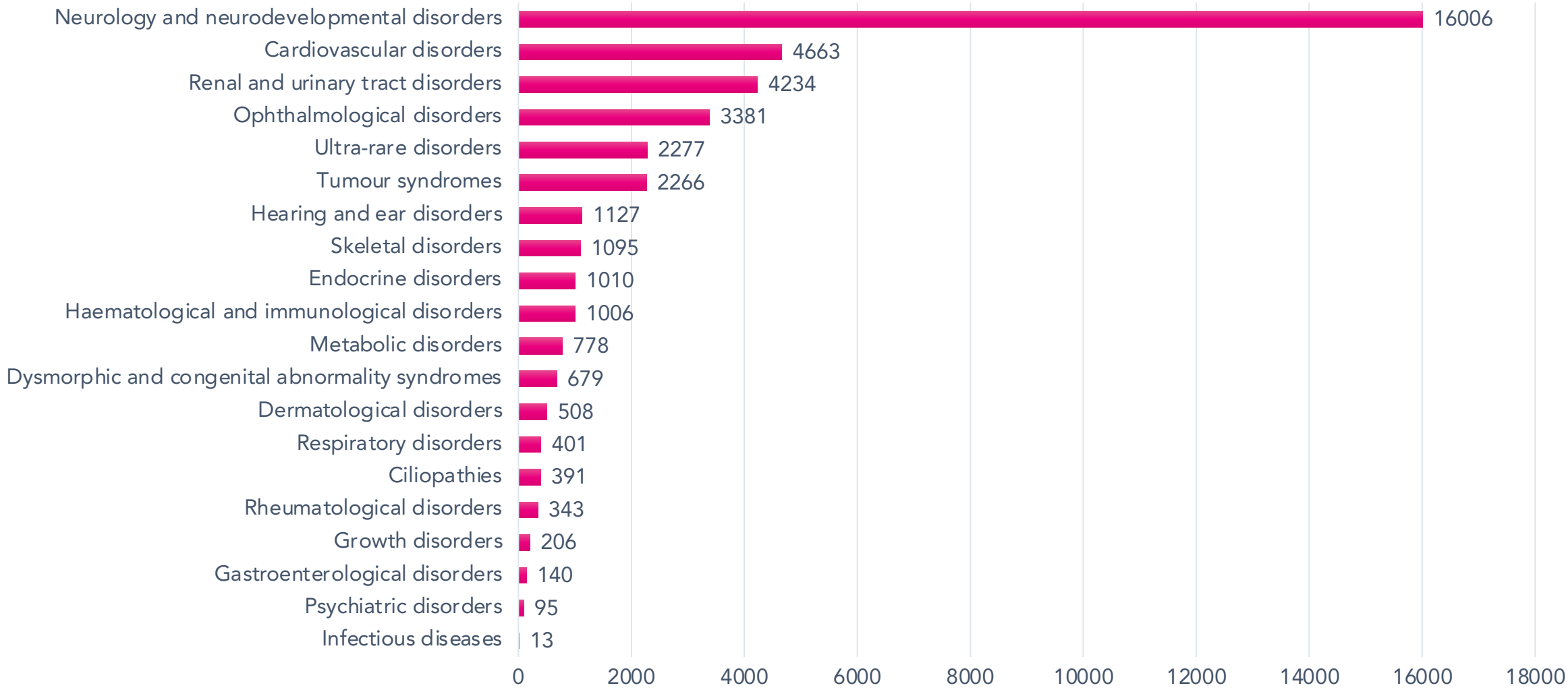
Genomic data from the 100,000 Genomes Project and NHS GMS

100,000 Genomes Project

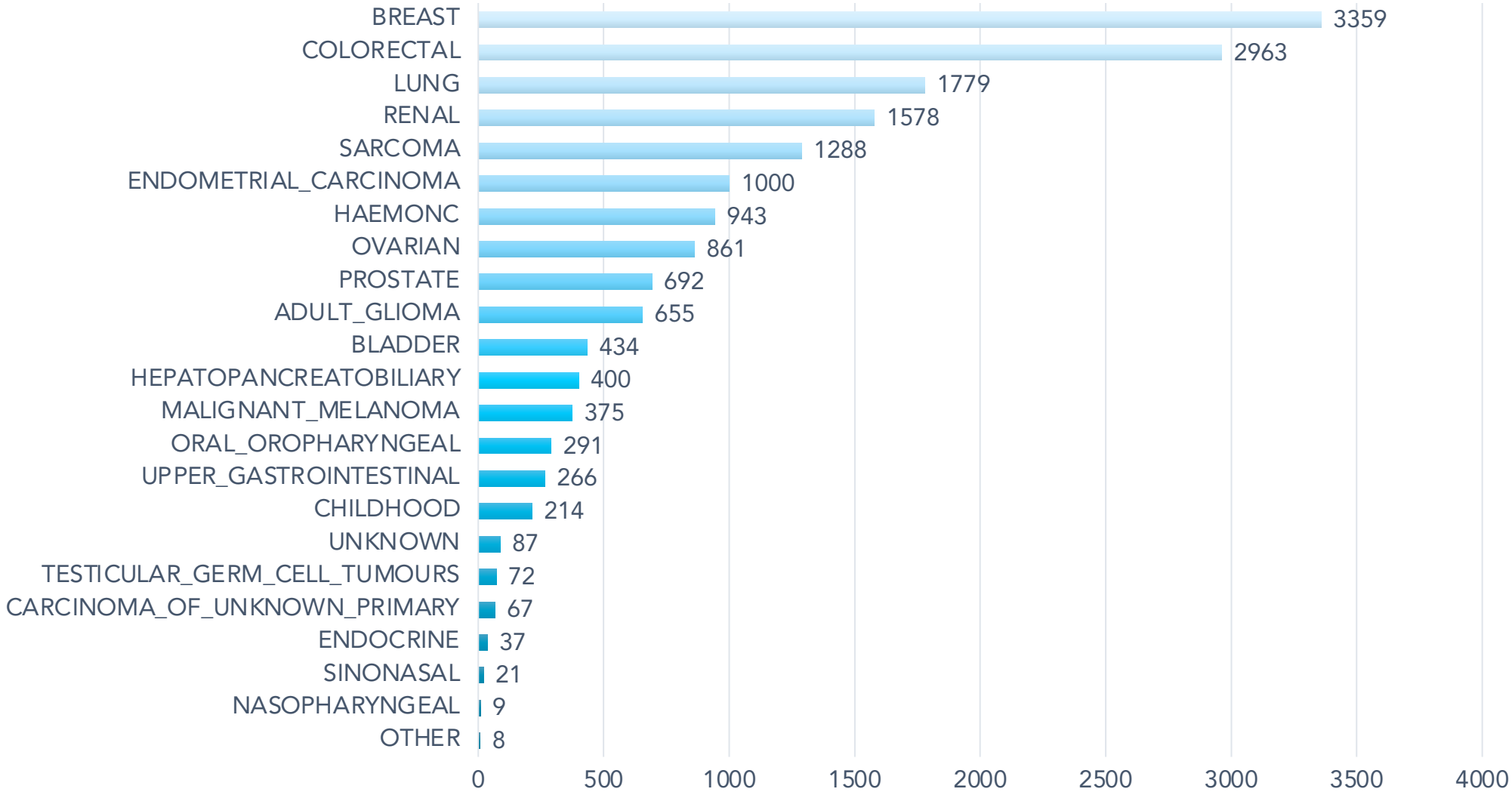


https://re-docs.genomicsengland.co.uk/release19/

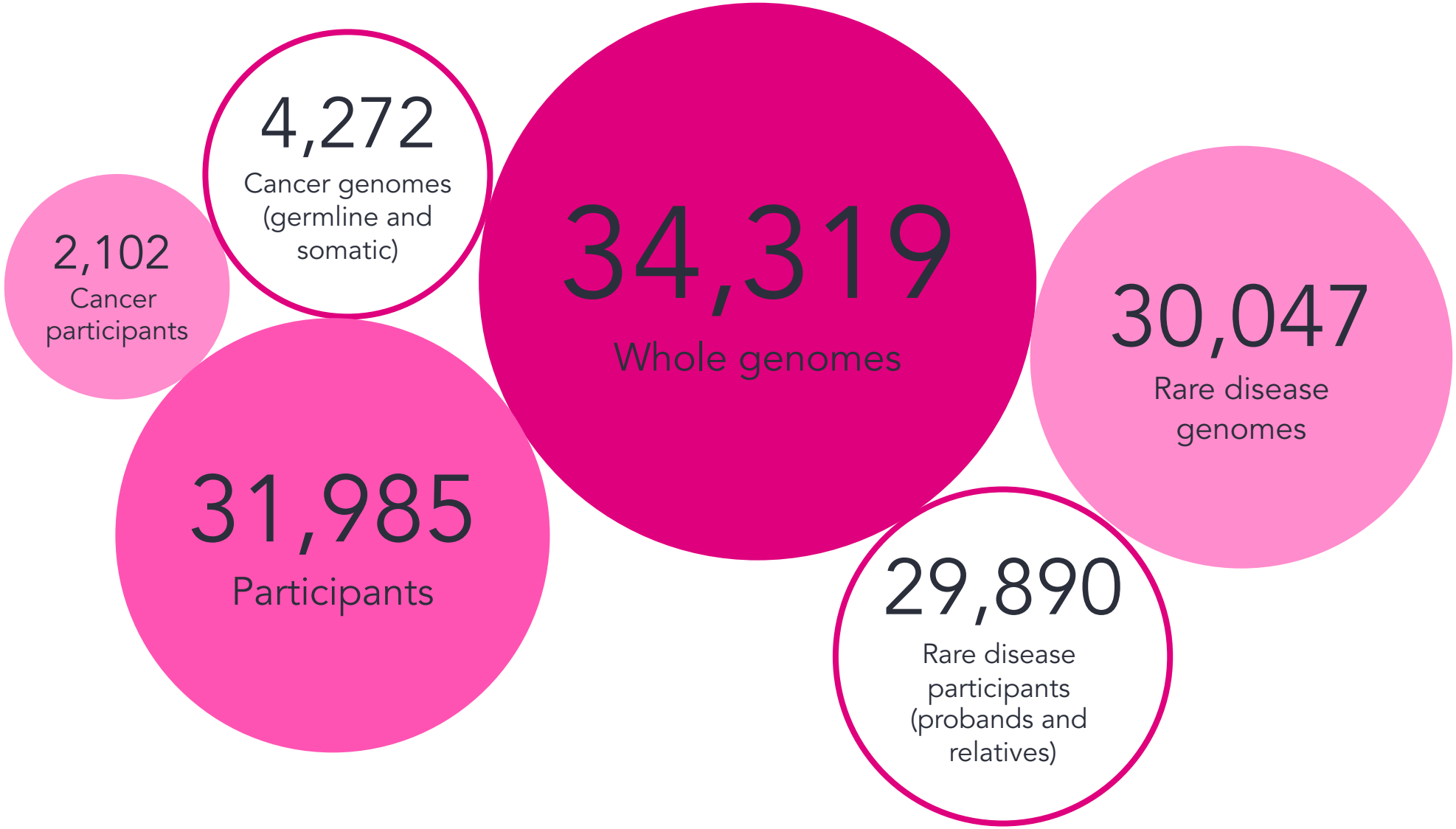
100,000 Genomes rare disease



100,000 Genomes cancer

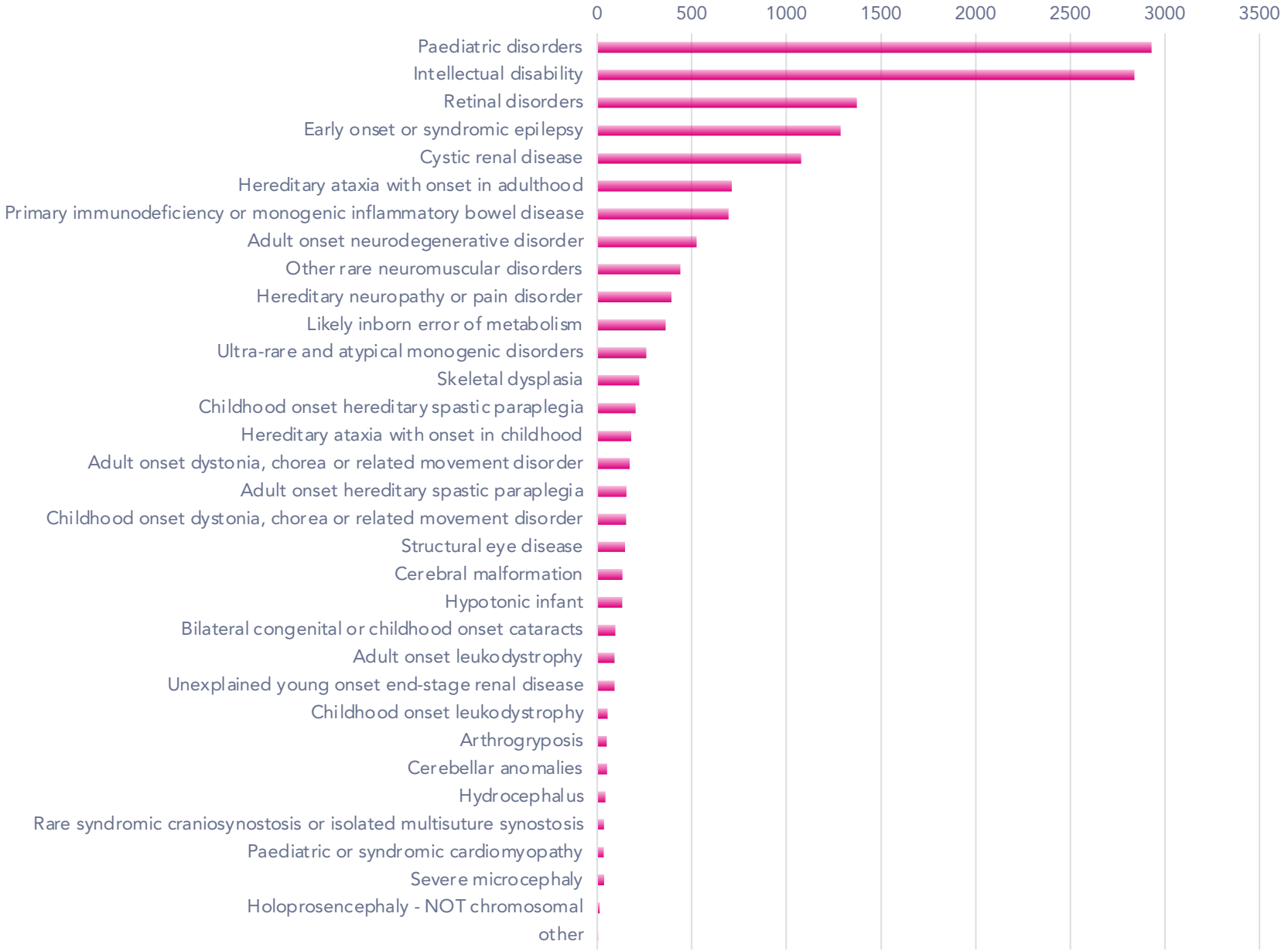


NHS GMS

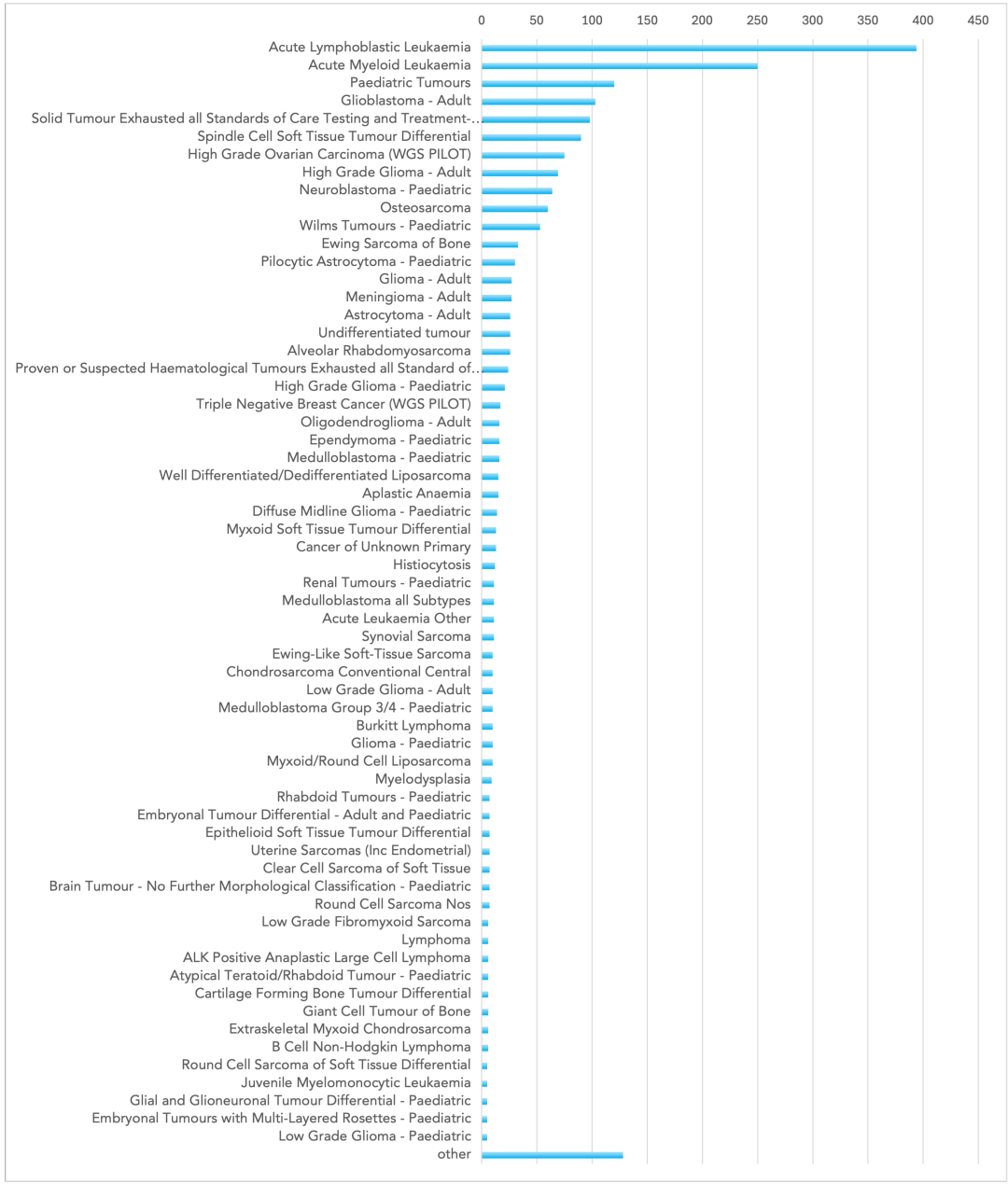


https://re-docs.genomicsengland.co.uk/gms_release4/

NHS GMS rare disease



NHS GMS cancer



For EVERY genome

Alignment

as BAM or CRAM files

Variant calls

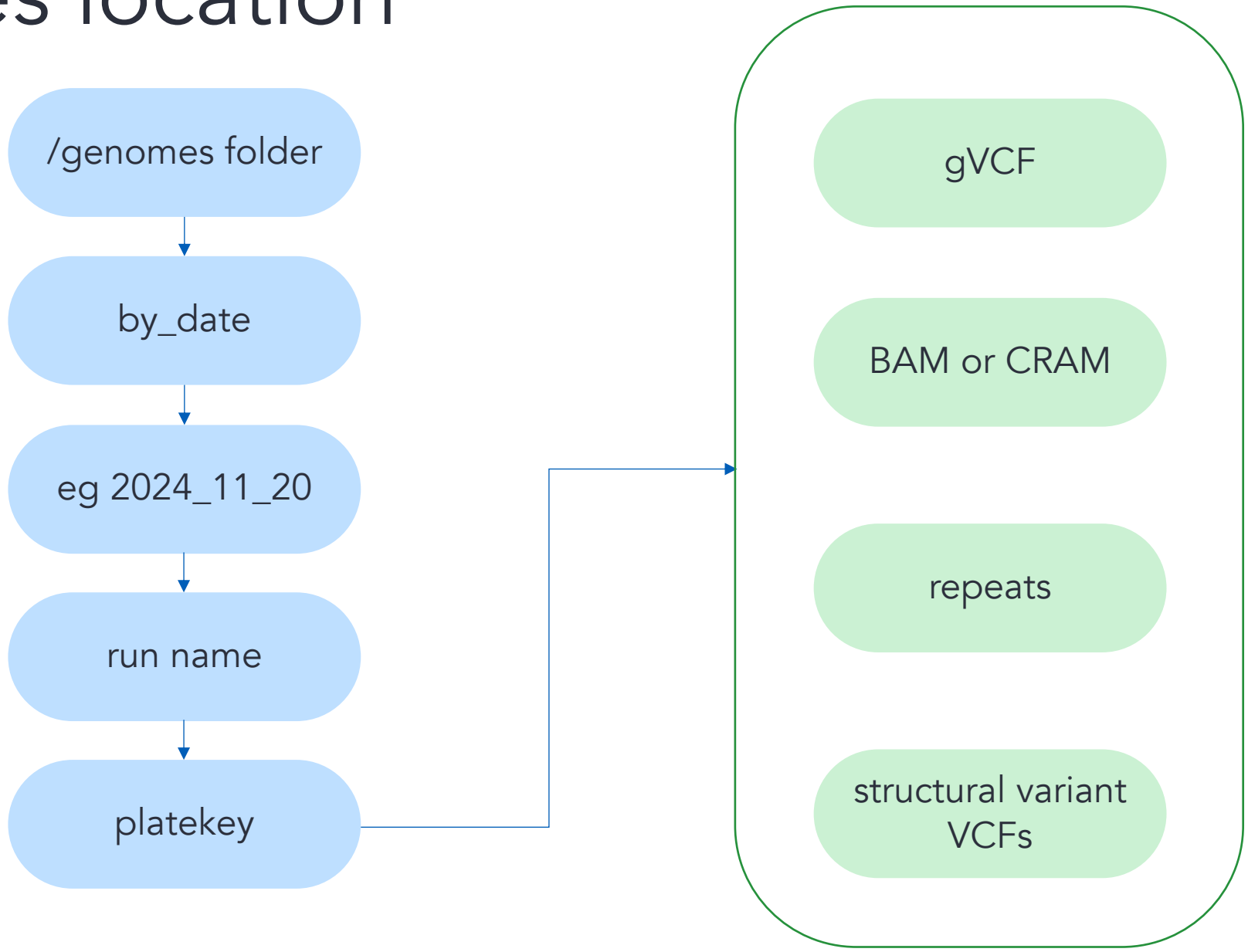
as VCFs, including gVCF, repeats VCFs, structural variant VCFs

Analysis

Variant tiering and tumour mutational signatures

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

Genomes location



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

Genomic data demo

Interactive variant analysis (IVA)

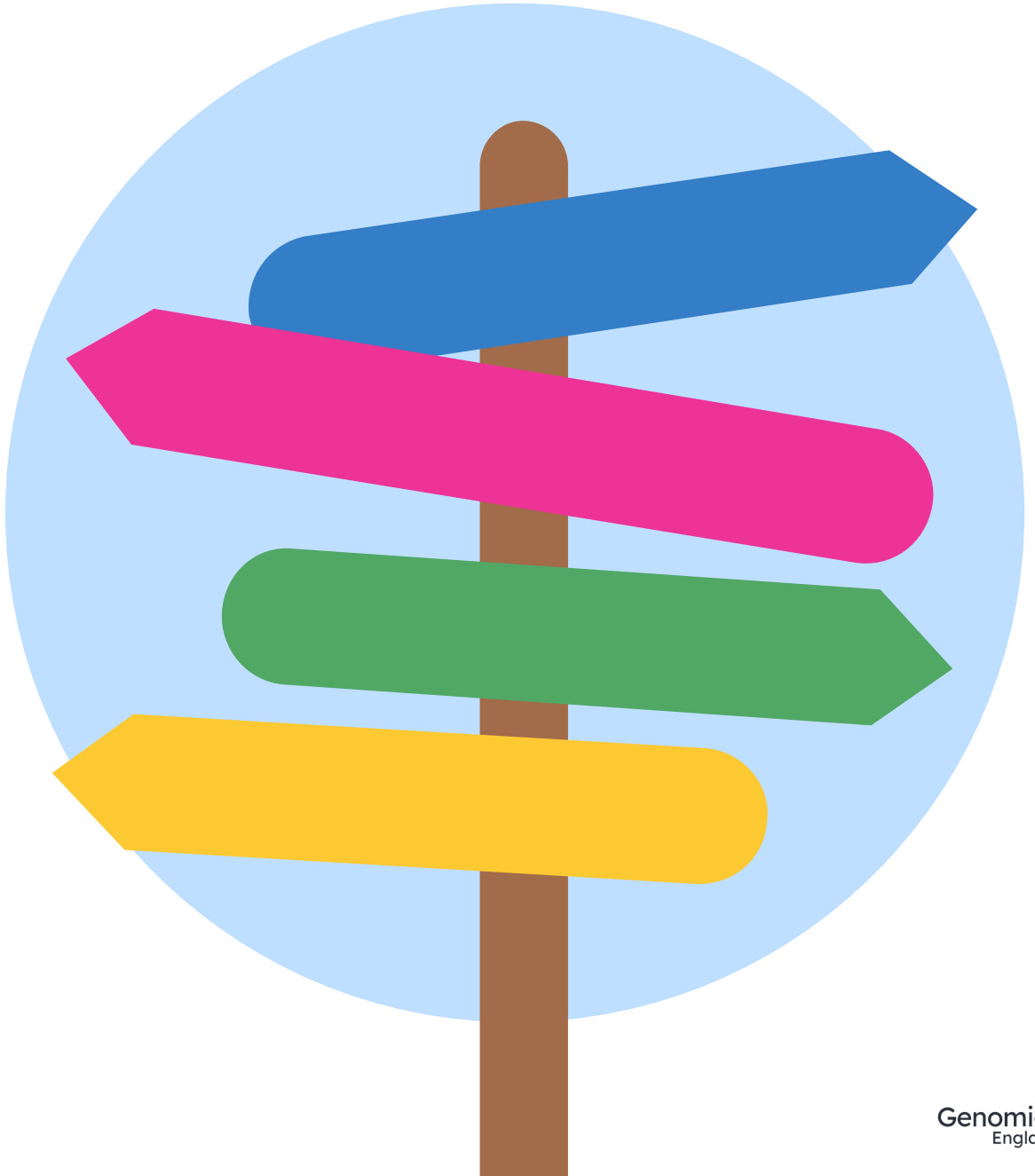
- Find participants with genetic variants
- Filter variants in a participant by family genotypes
- Filter on genome features

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

IVA demo

Task – IVA

1. Go to the Variant Browser for cancer germline GRCh38
2. Filter to find all rare LoF SNVs in *BRCA2*.
3. Choose a variant and find all participants with that variant
4. Go to the case interpreter for 122006585 in the rare disease GRCh38 programme
5. Filter to find all LoF *de novo* variants within genes on the disorders of sex development panel.



Aggregate VCFs

Combined gVCFs

Quality metrics

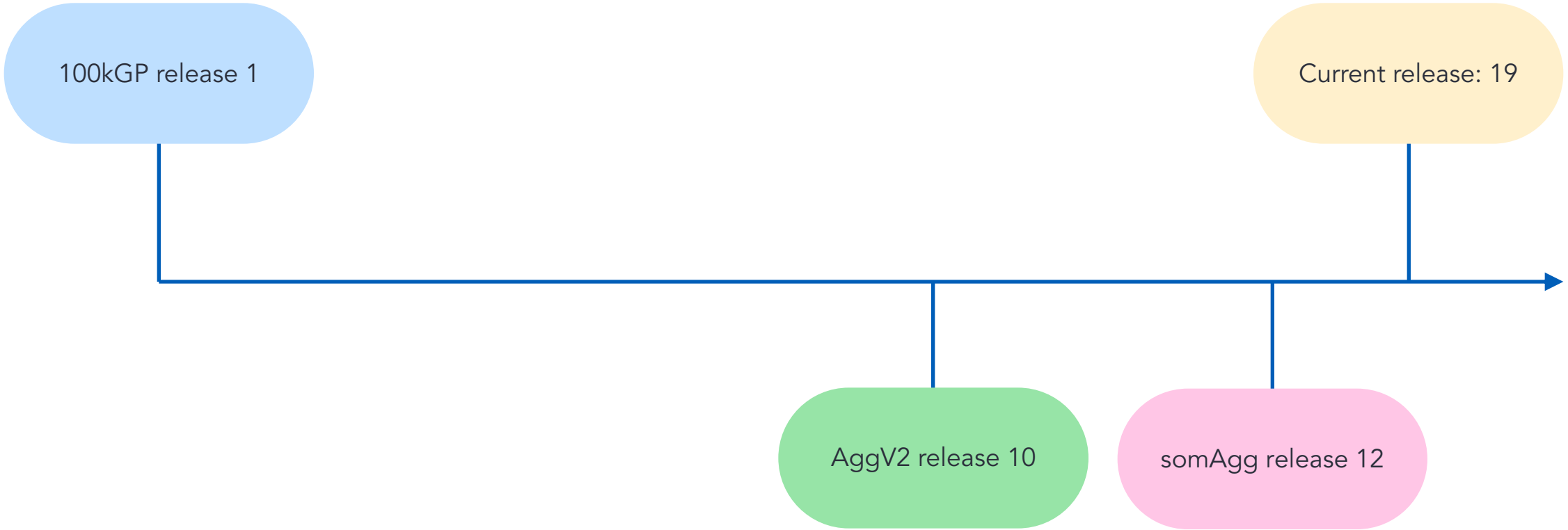
Functional annotation files

Genotype files – platekeys

1371 chunks

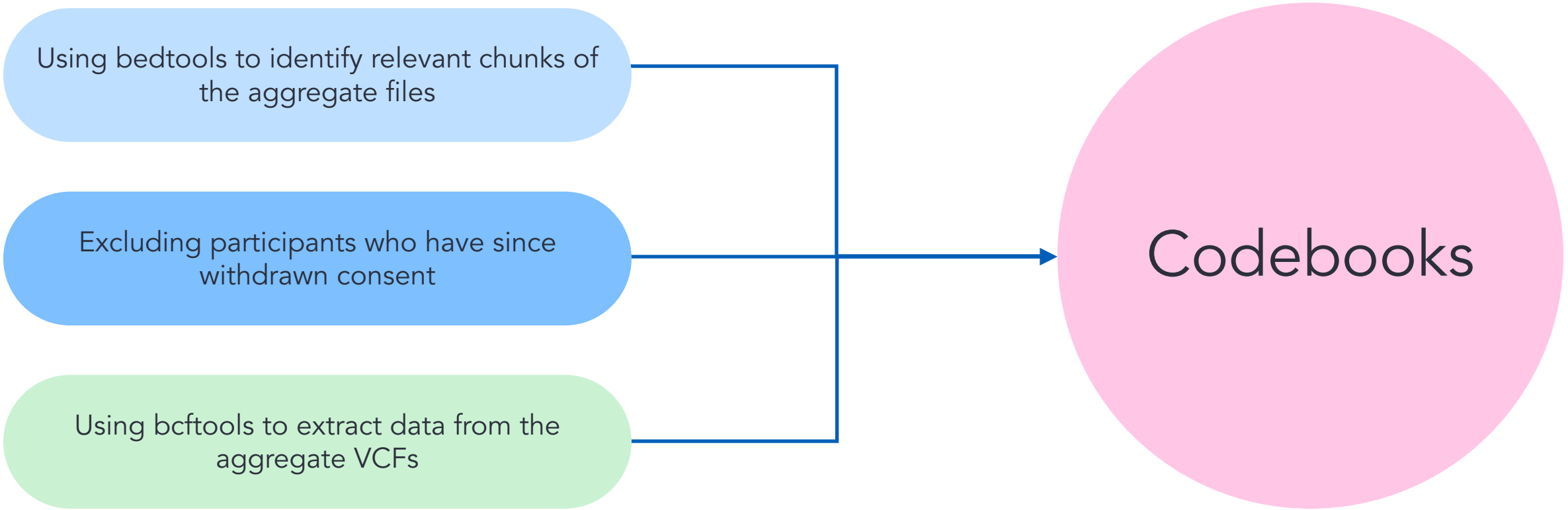
https://re-docs.genomicsengland.co.uk/aggv2/ https://re-docs.genomicsengland.co.uk/somAgg/

Aggregate VCFs



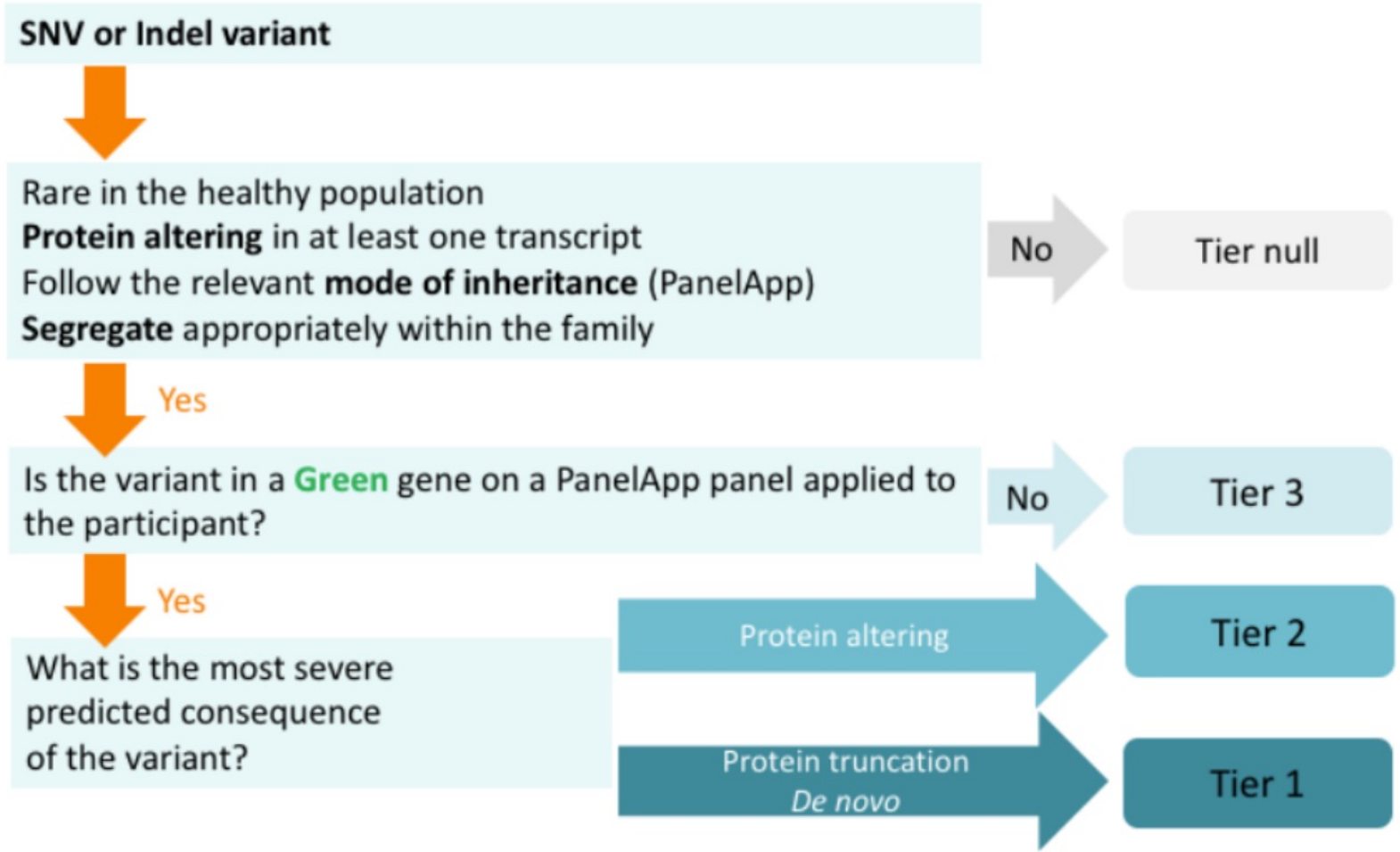
https://re-docs.genomicsengland.co.uk/aggv2/ https://re-docs.genomicsengland.co.uk/somAgg/

Aggregate VCFs



https://re-docs.genomicsengland.co.uk/aggv2_code_book/ https://re-docs.genomicsengland.co.uk/somAgg_code_book/

Rare disease tiering



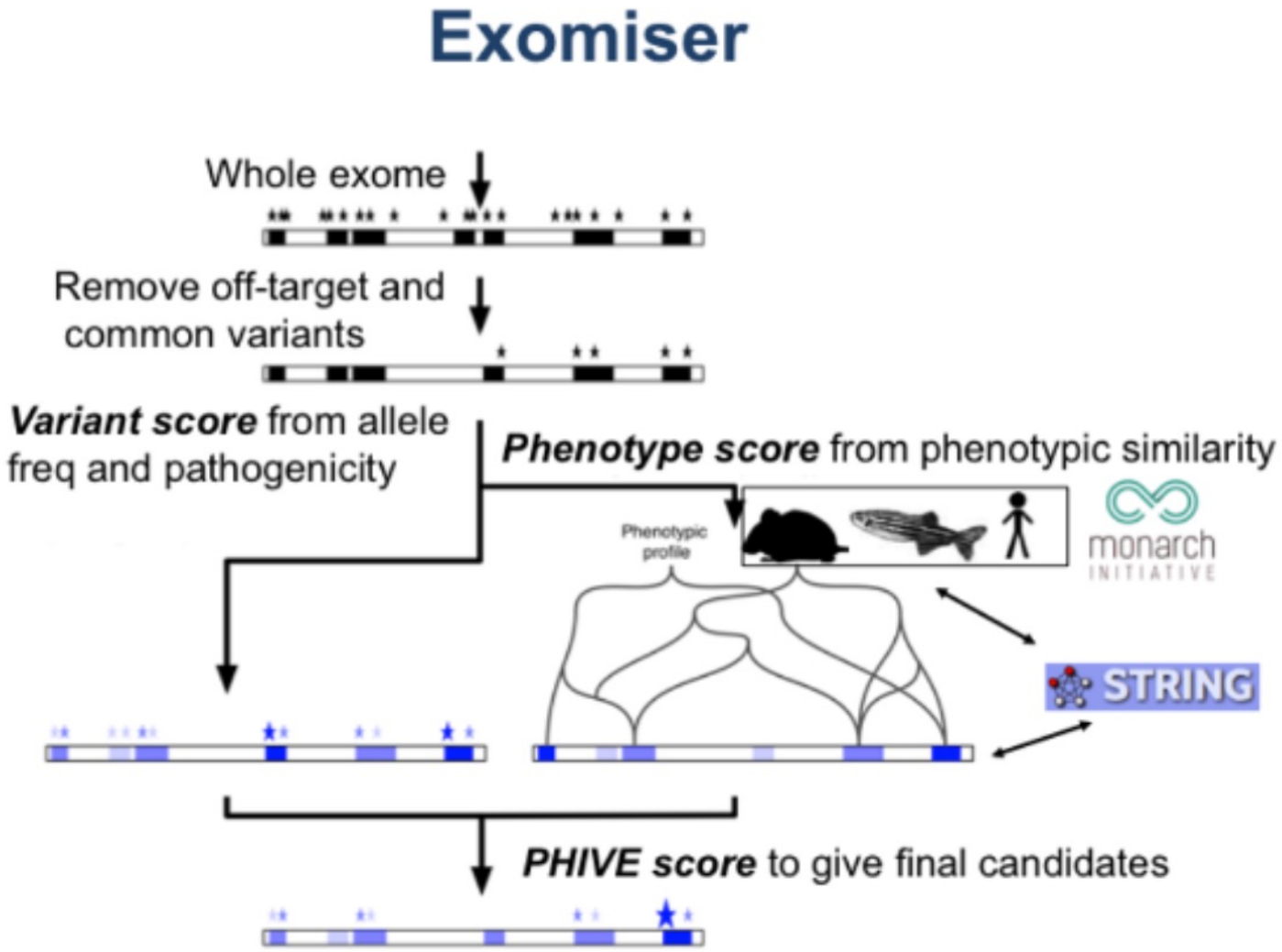
https://re-docs.genomicsengland.co.uk/tiering/

Rare disease tiering based on PanelApp genes

List ↑	Entity	Reviews	Mode of inheritance	Details
Filter Entities				8 Entities
Green	ATP1A3	1 review 1 green	MONOALLELIC, autosomal or pseudoautosomal, NOT imprinted	**Sources** • Expert Review • Expert Review Green **Phenotypes** • 601338 • 614820 **Tags**
Green	DFNB59	2 reviews 1 green	BIALLELIC, autosomal or pseudoautosomal	**Sources** • Expert Review • Expert Review Green **Phenotypes** • 610219 **Tags** new-gene-name
Green	OPA1	2 reviews 1 green	MONOALLELIC, autosomal or pseudoautosomal, NOT imprinted	**Sources** • Eligibility statement prior genetic testing • Expert Review Green **Phenotypes** • Optic atrophy 1, OMIM:165500 • Optic atrophy plus syndrome, OMIM:125250 **Tags**
Green	OTOF	1 review 1 green	BIALLELIC, autosomal or pseudoautosomal	**Sources** • Expert Review Green • Radboud University Medical Center, Nijmegen **Phenotypes** • 601071 **Tags**
Amber	DIAPH3	3 reviews 1 red	BOTH monoallelic and biallelic, autosomal or pseudoautosomal	**Sources** • Expert Review Amber • Radboud University Medical Center, Nijmegen **Phenotypes** • Auditory neuropathy, autosomal dominant, 1, 609129 **Tags**

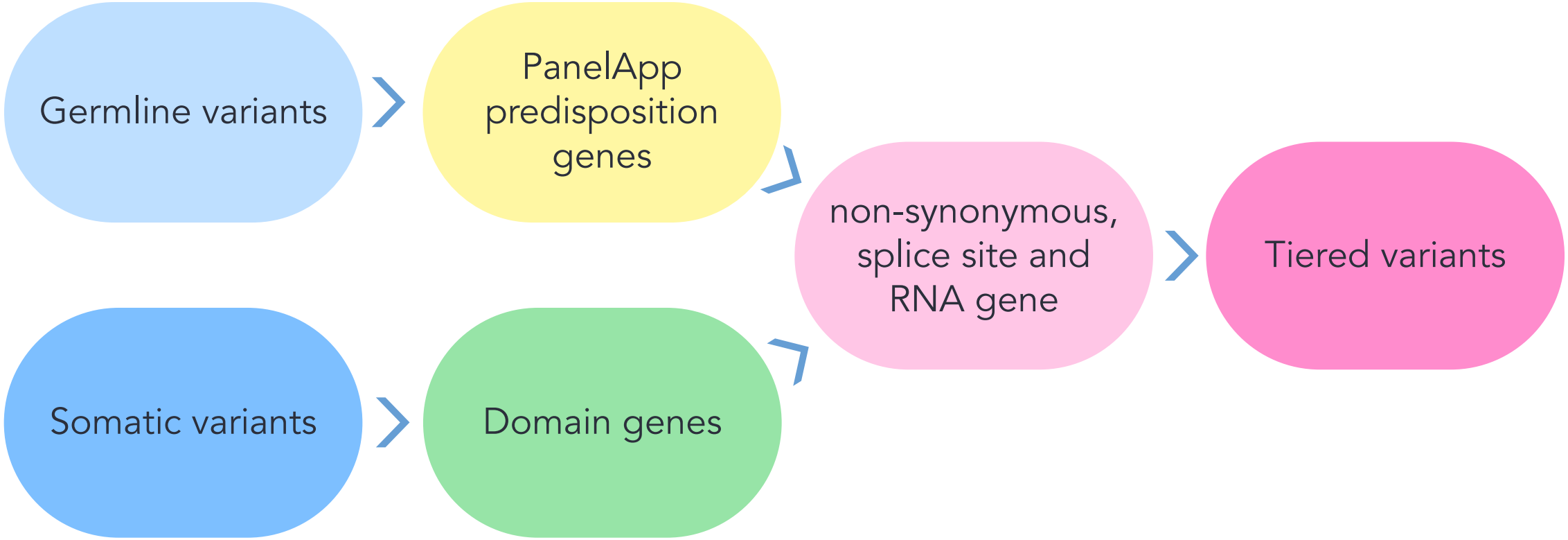
https://re-docs.genomicsengland.co.uk/panelapp/

Rare disease Exomiser



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

Cancer tiering



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

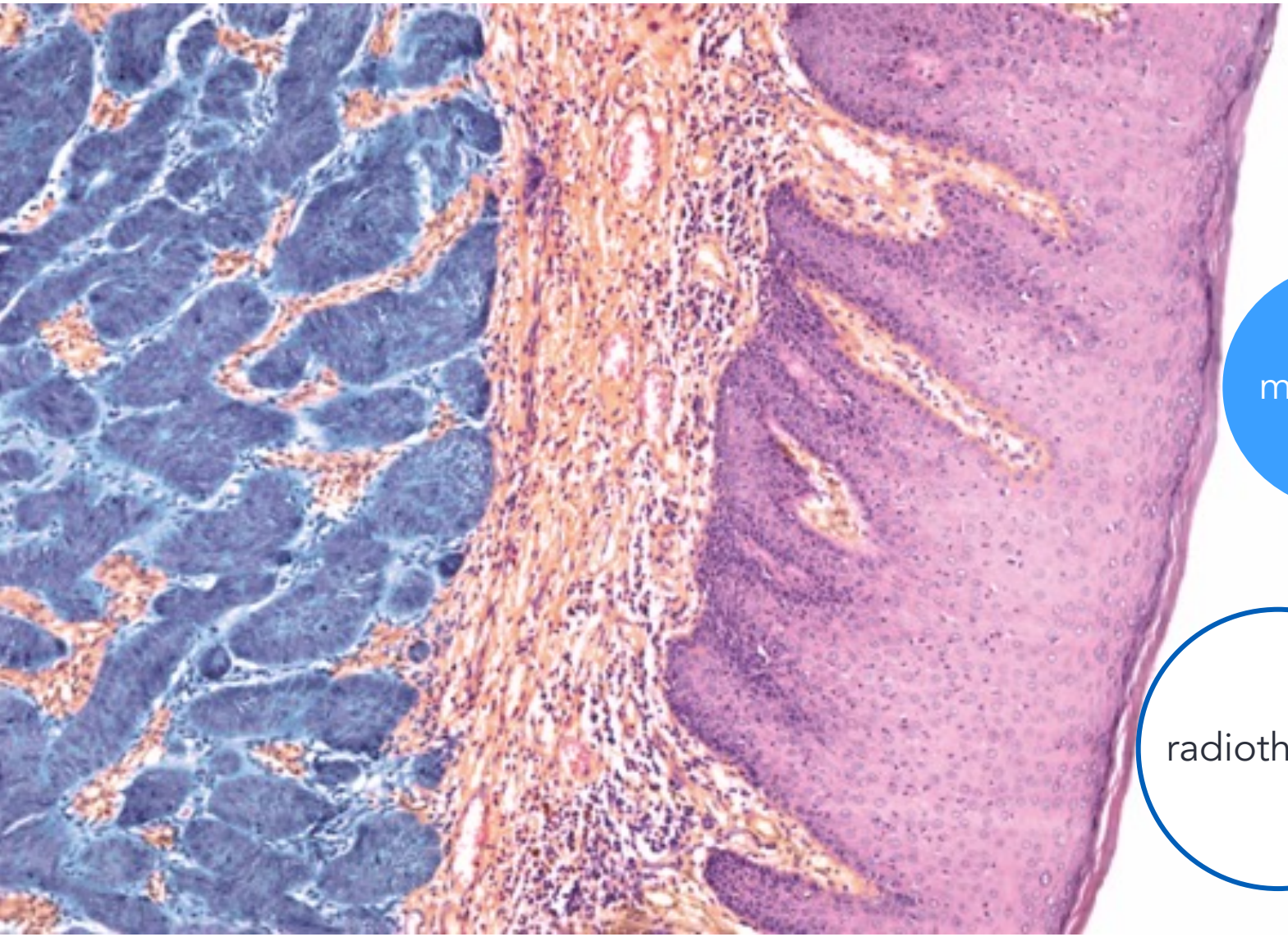
Clinical and phenotype data

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

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

Cancer tumour characterisation



risk factors

cancer type

surgeries

staging and grading

metastasis

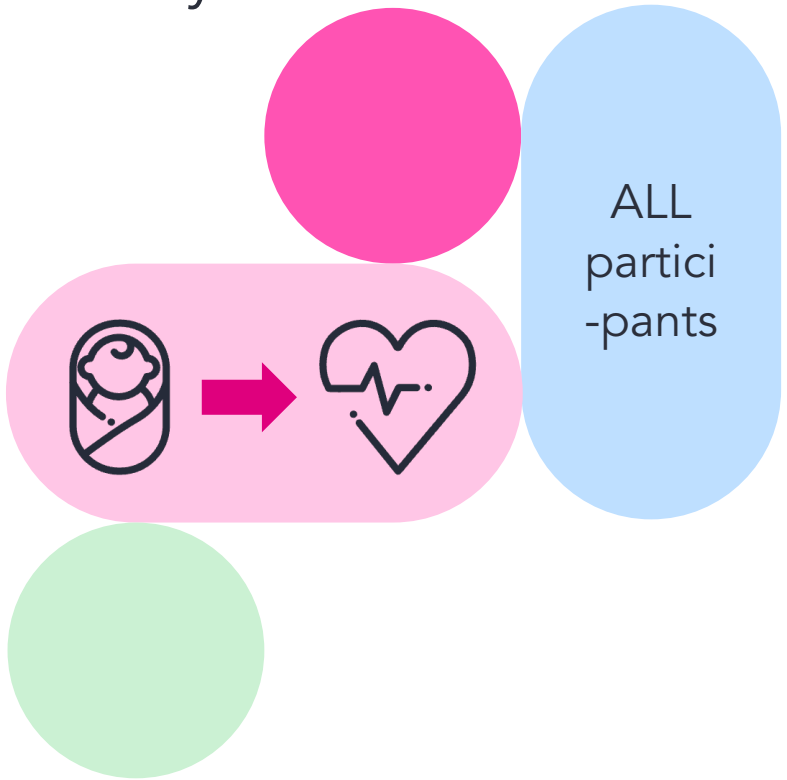
chemotherapy

radiotherapy

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

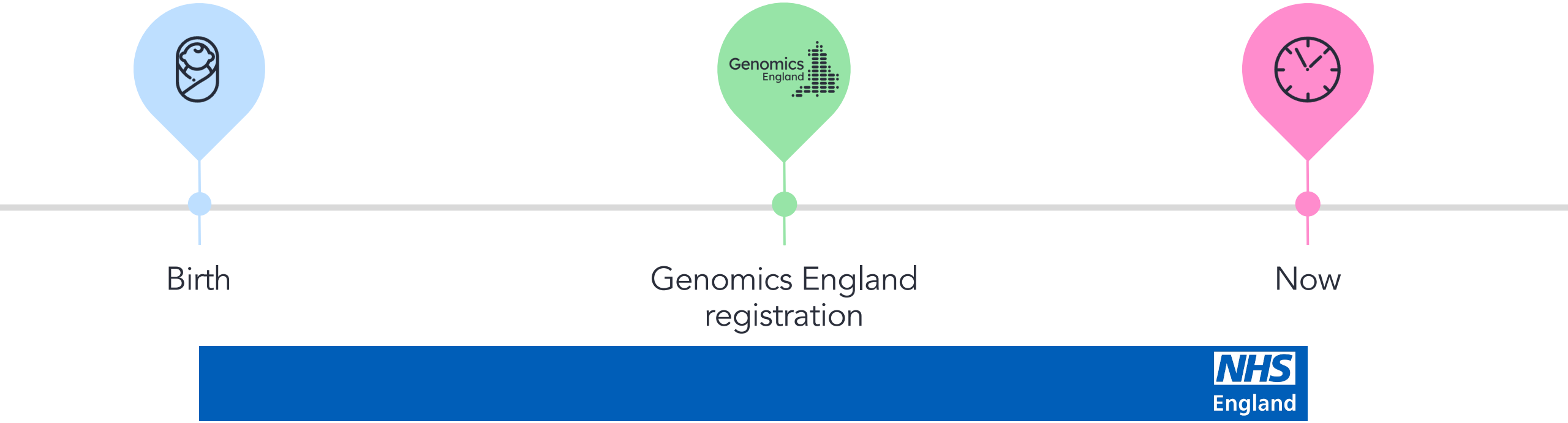
Medical history – 100k (GMS coming soon)

- NHSE hospital episode statistics
- Mental health data
- Mortality



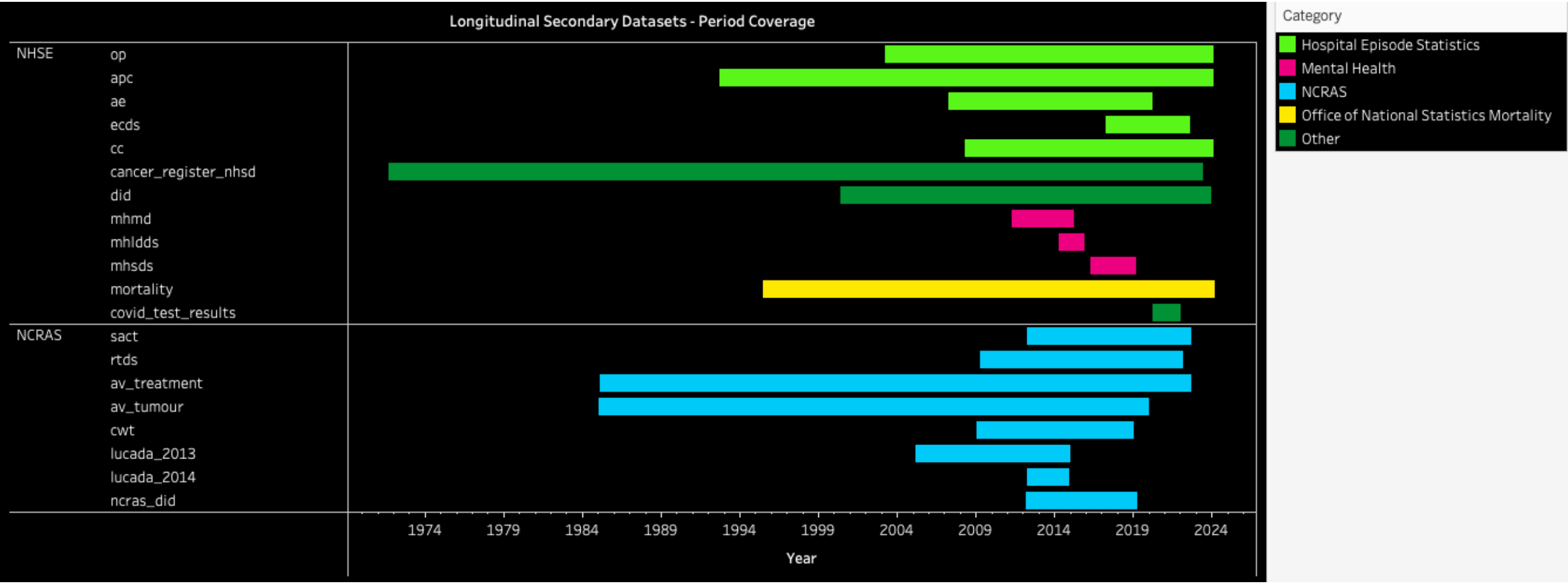
https://re-docs.genomicsengland.co.uk/general_clinical/

Genomics England data ingestion



https://re-docs.genomicsengland.co.uk/general_clinical/#hospital-episodes-statistics-from-nhse

Medical history coverage over time

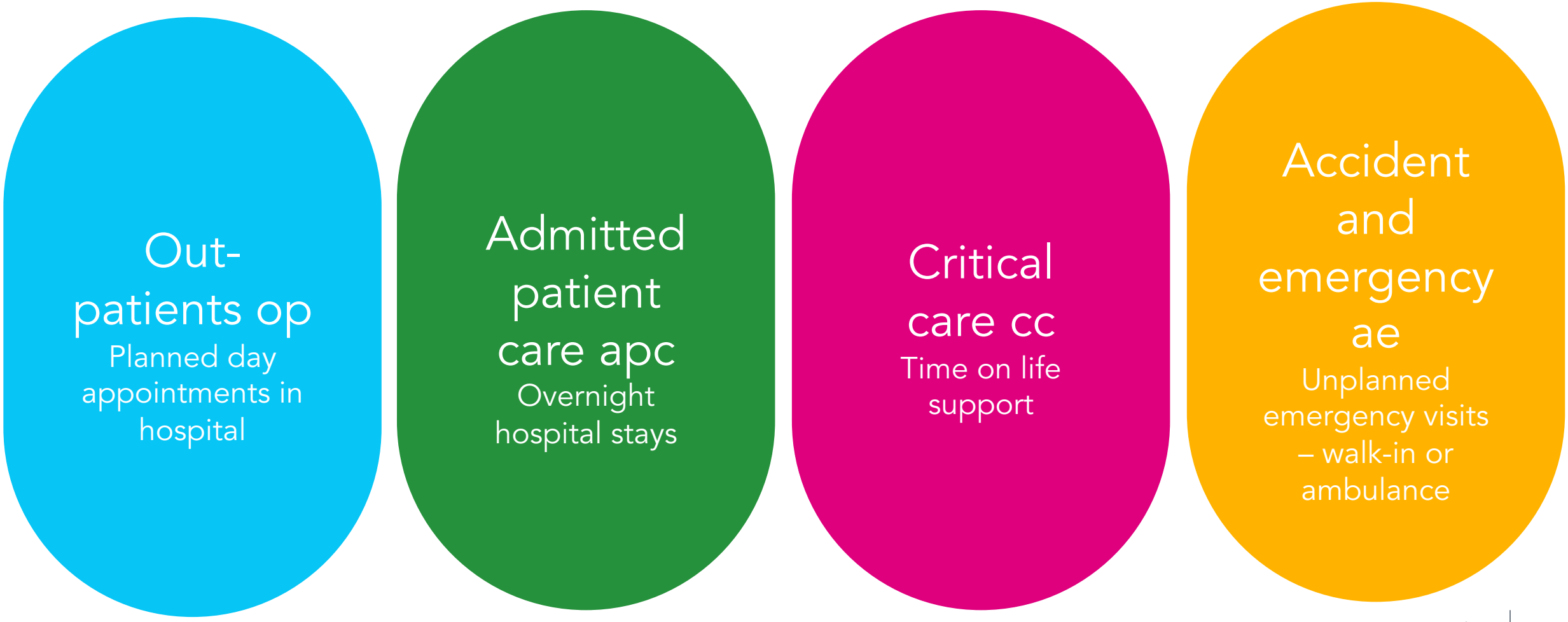


https://re-docs.genomicsengland.co.uk/general_clinical/#hospital-episodes-statistics-from-nhse

Genomics England data ingestion

https://re-docs.genomicsengland.co.uk/general_clinical/#hospital-episodes-statistics-from-nhse

Hospital episode statistics

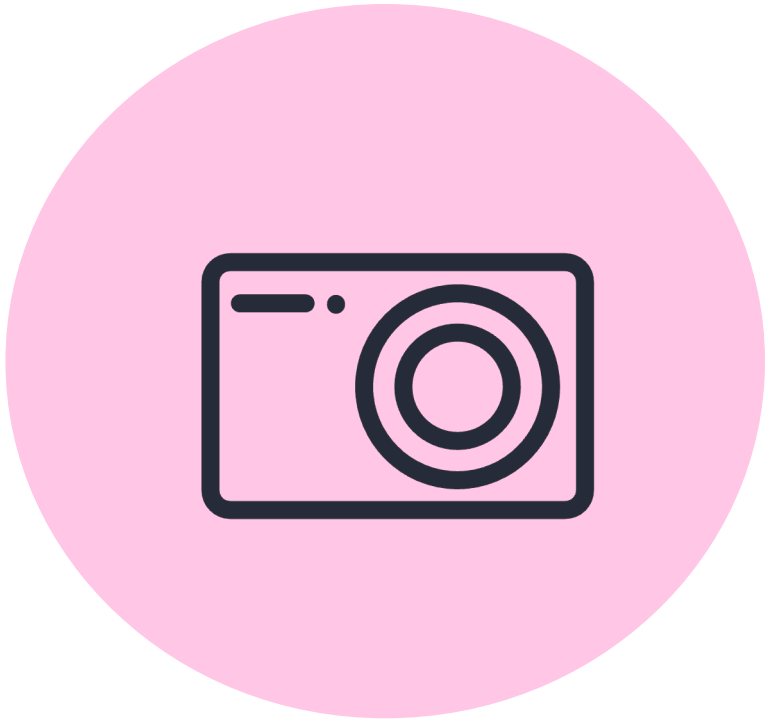


https://re-docs.genomicsengland.co.uk/general_clinical/#hospital-episodes-statistics-from-nhse

We don’t have…

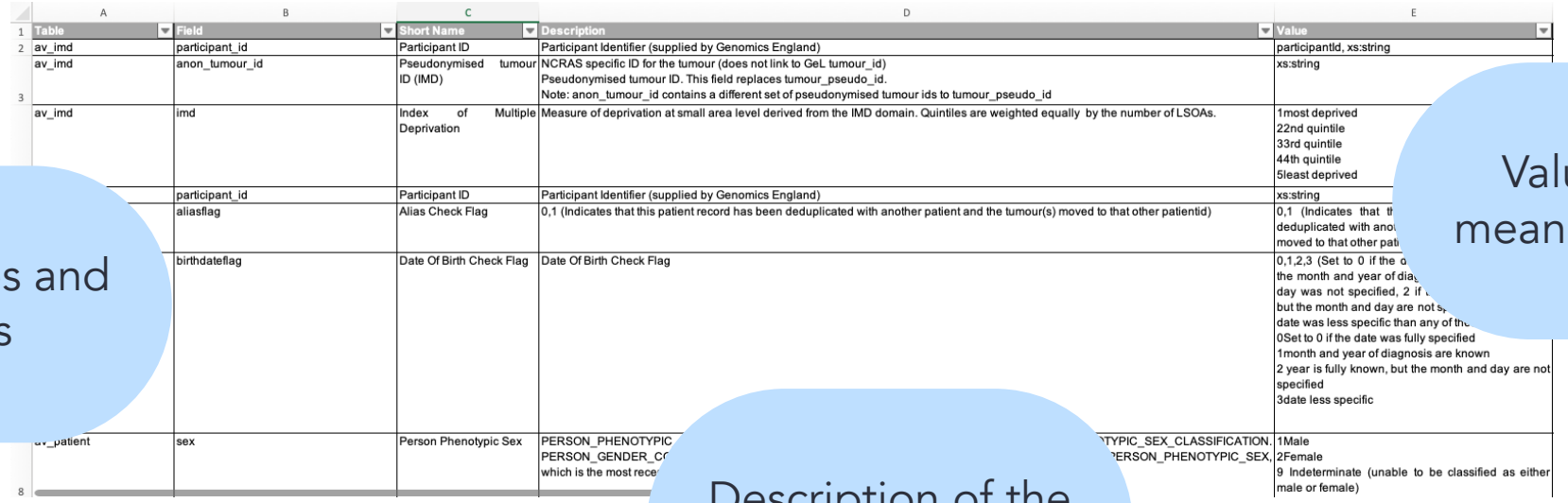
Free text

GP data



diagnostic imaging

Data dictionary



	A	B	C	D	E
1	Table	Field	Short Name	Description	Value
2	av_imd	participant_id	Participant ID	Participant Identifier (supplied by Genomics England)	participantId, xs:string
3	av_imd	anon_tumour_id	Pseudonymised tumour ID (IMD)	NCRAS specific ID for the tumour (does not link to GeL tumour_id) Pseudonymised tumour ID. This field replaces tumour_pseudo_id. Note: anon_tumour_id contains a different set of pseudonymised tumour ids to tumour_pseudo_id	xs:string
	av_imd	imd	Index of Multiple Deprivation	Measure of deprivation at small area level derived from the IMD domain. Quintiles are weighted equally by the number of LSOAs.	1most deprived 22nd quintile 33rd quintile 44th quintile 5least deprived
		participant_id	Participant ID	Participant Identifier (supplied by Genomics England)	xs:string
		aliasflag	Alias Check Flag	0,1 (Indicates that this patient record has been deduplicated with another patient and the tumour(s) moved to that other patientid)	0,1 (Indicates that th deduplicated with ano moved to that other pati
		birthdateflag	Date Of Birth Check Flag	Date Of Birth Check Flag	0,1,2,3 (Set to 0 if the d the month and year of dia day was not specified, 2 if but the month and day are not s date was less specific than any of th 0Set to 0 if the date was fully specified 1month and year of diagnosis are known 2 year is fully known, but the month and day are not specified 3date less specific
8	av_patient	sex	Person Phenotypic Sex	PERSON_PHENOTYPIC PERSON_GENDER_C which is the most rece	TYPIC_SEX_CLASSIFICATION. ERSON_PHENOTYPIC_SEX, 1Male 2Female 9 Indeterminate (unable to be classified as either male or female)

Lists of tables and columns

Value type or meaning of codes

Description of the data

https://re-docs.genomicsengland.co.uk/100k_data_dictionary https://re-docs.genomicsengland.co.uk/GMS_data_dictionary

LabKey

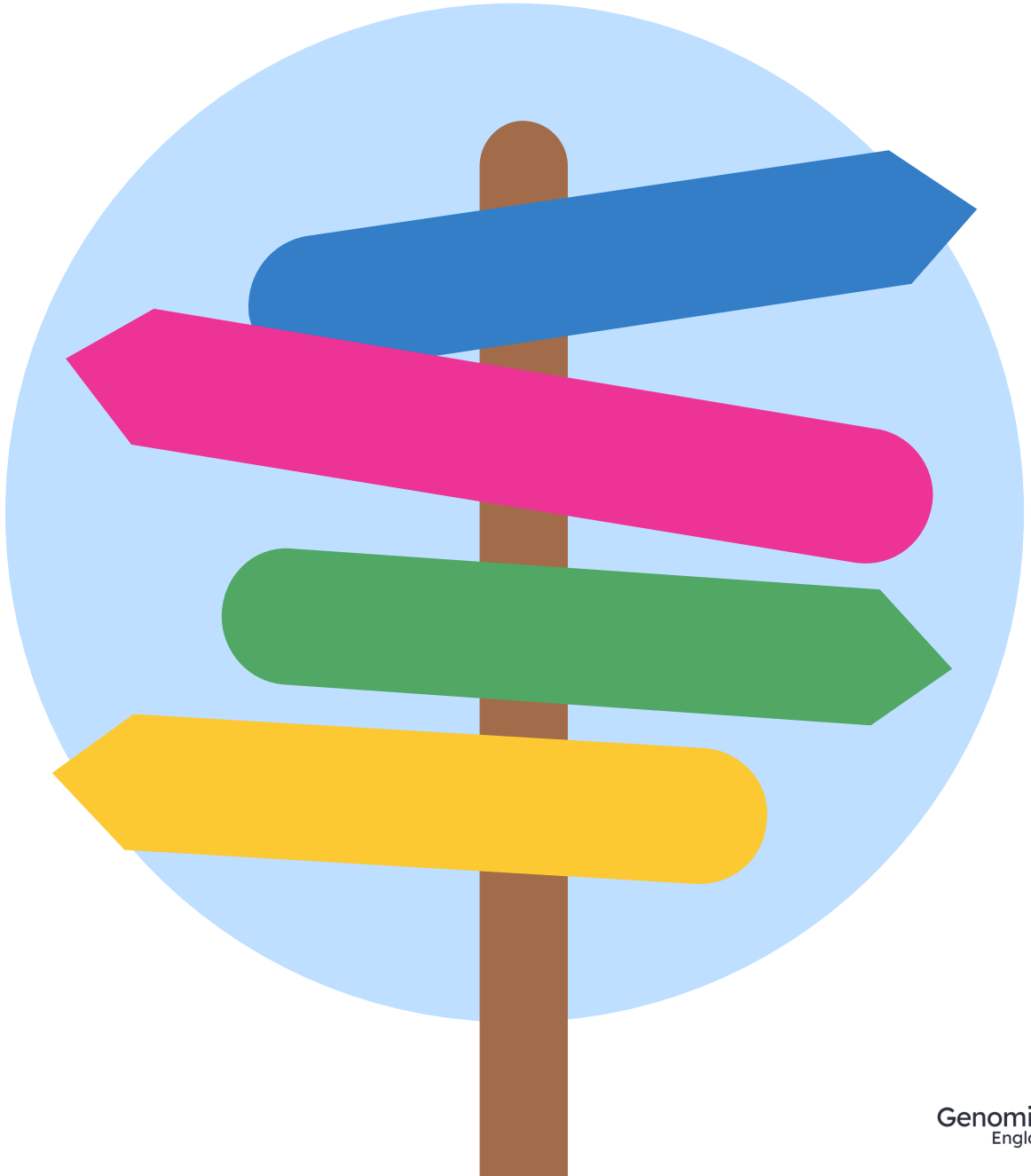
- Central database of:
  - Clinical data
  - Results of bioinformatics analysis
  - Locations of genomic files
- Point and click interface
- API

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

Clinical data in LabKey demo

Task – LabKey

1. Open the did table for the 100kGP.
2. Use the data dictionary to find what field in the table is the date of imaging.
3. Filter the table to see details of imaging done after 1st January 2020.



Participant Explorer

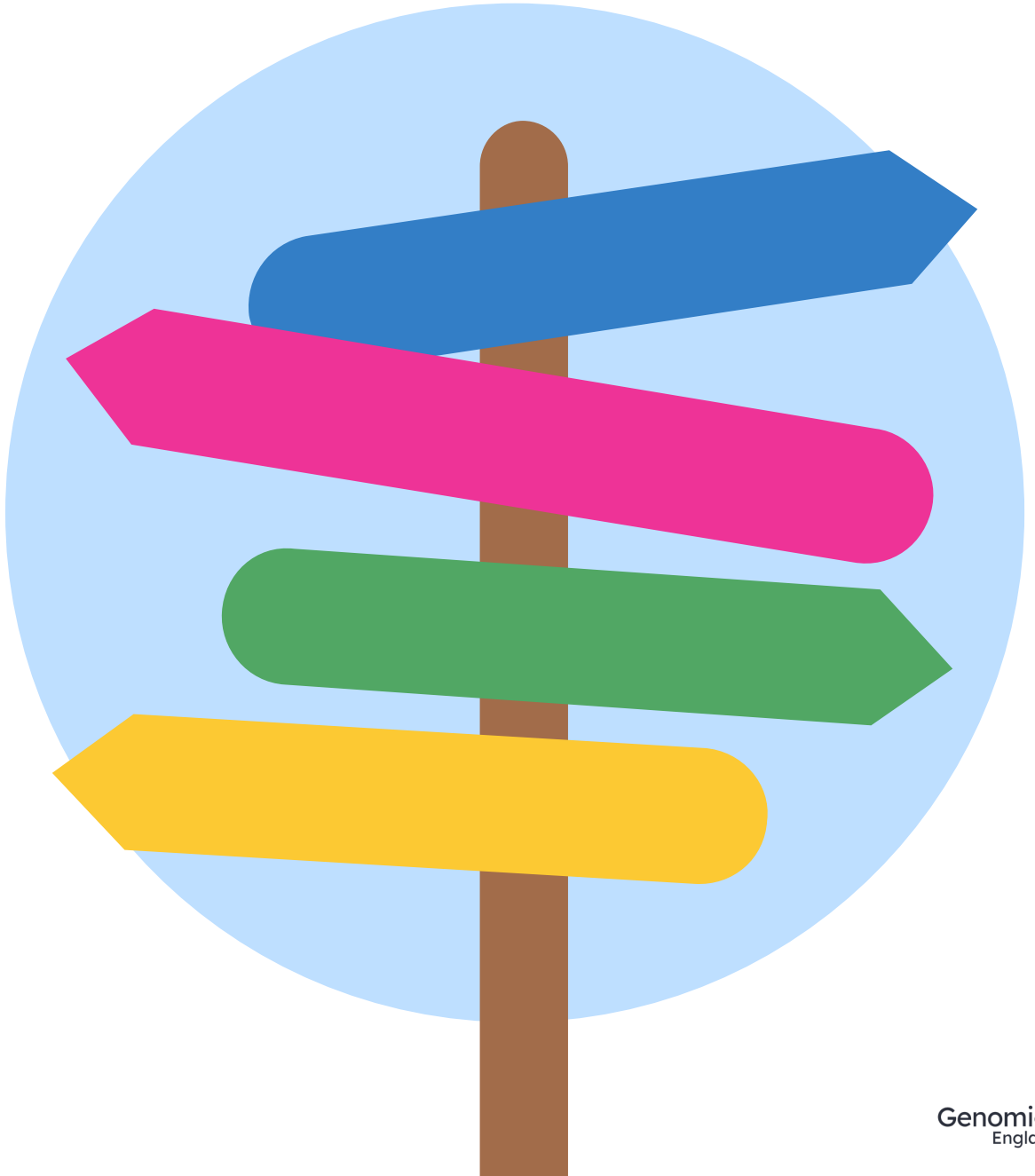
- Search for participants by:
  - IDs
  - Clinical concepts
  - Personal details
- View/compare medical histories

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

Participant Explorer demo

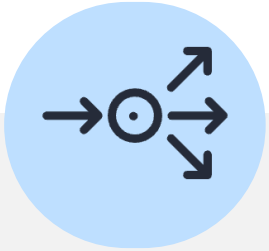
Task – PX

1. Search for female participants with hip dysplasia born since 2000.
2. Download the genome path details for all participants who match your search
3. Look at the medical history for one of the participants in your search.
4. Compare the medical history for the first four participants in your results, find clinical concepts common to all.

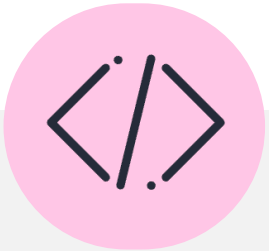


Programmatic access to clinical data

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

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

LabKey .netrc

- You can access the same data via the LabKey API as you can through other means
- You will need to configure access to the LabKey API with your username and password
    - In your home directory
    - On the HPC
- You do this by editing a file called .netrc

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

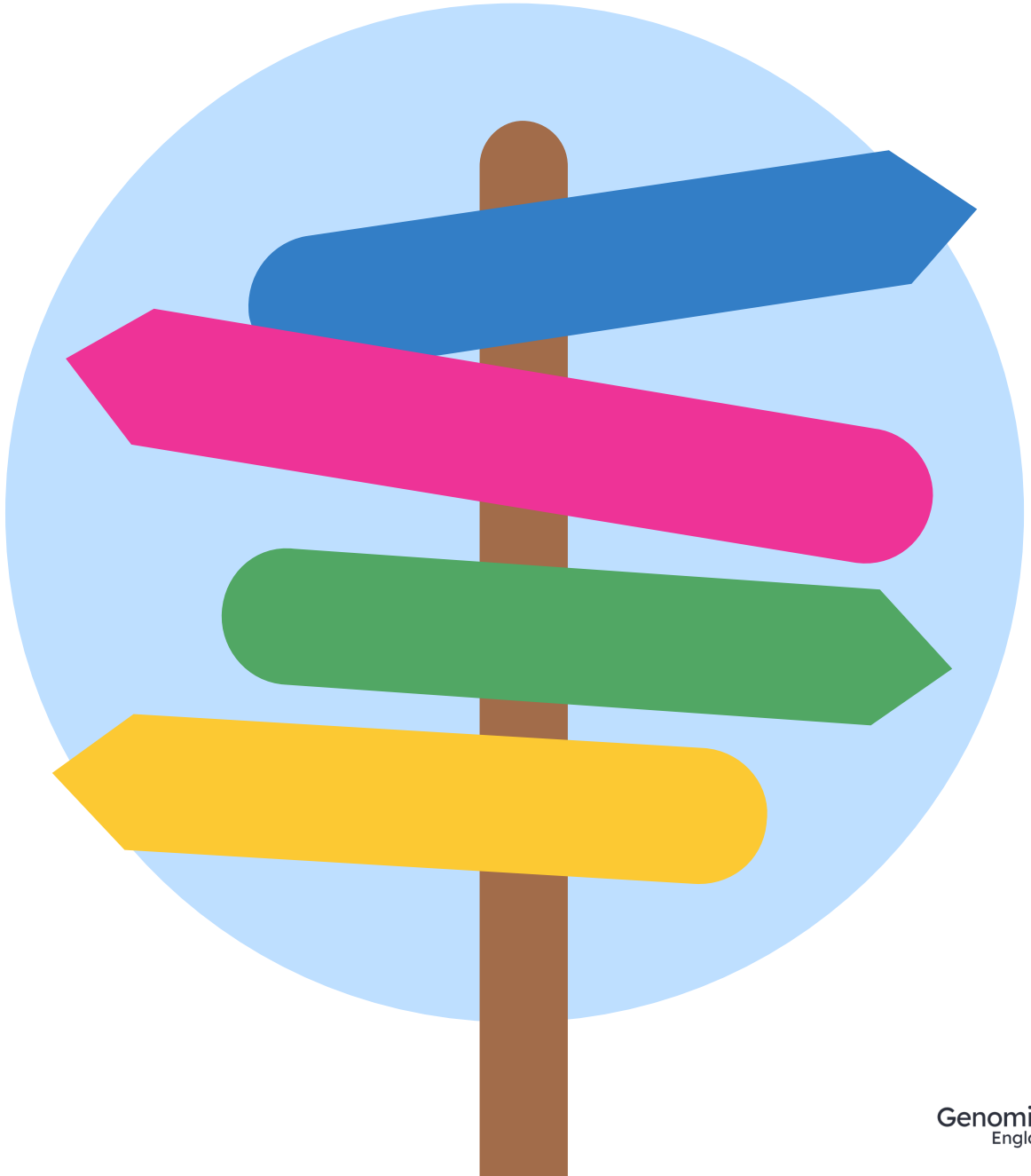
Programming tools in the RE

https://re-docs.genomicsengland.co.uk/hpc_jupyter/ https://re-docs.genomicsengland.co.uk/enable_rstudio/

LabKey API demo

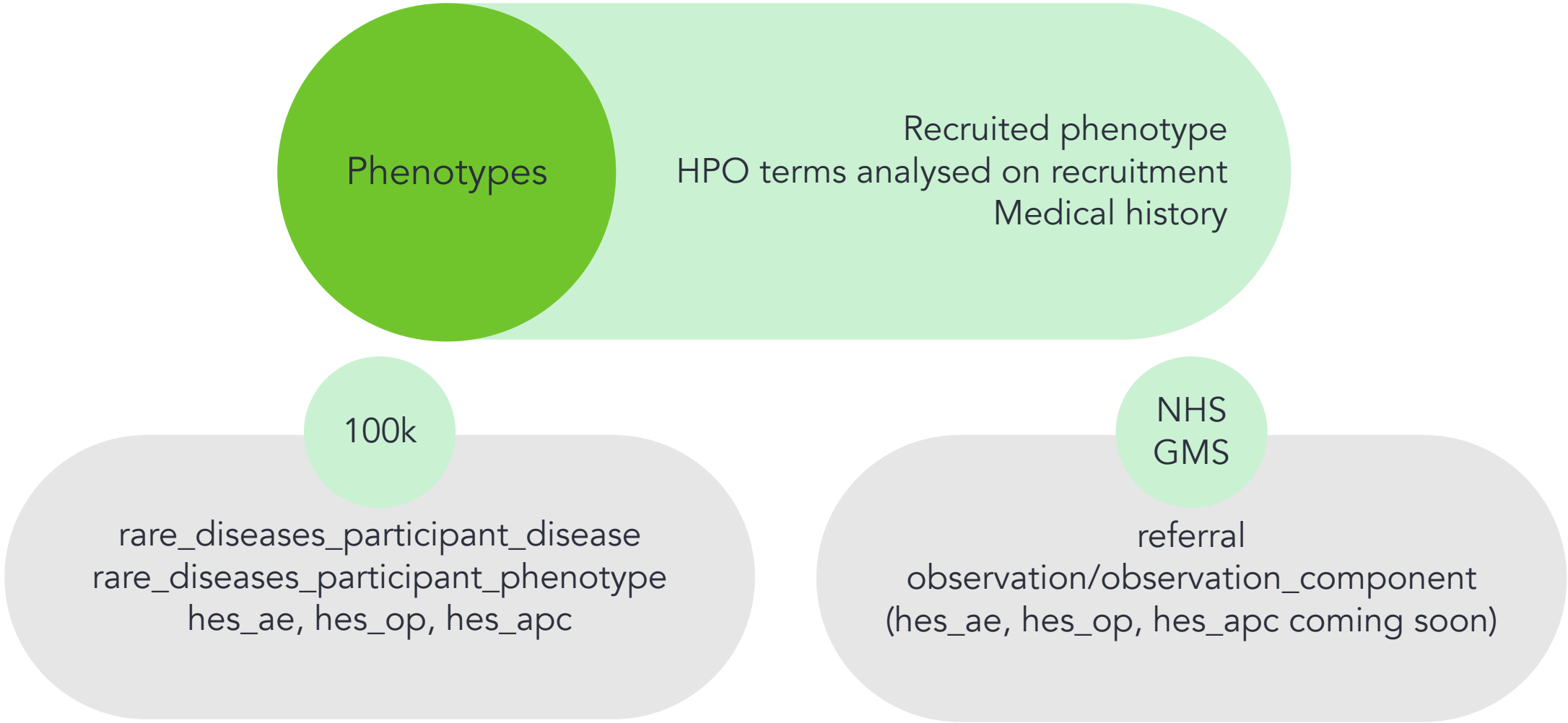
Task – LabKey API

1. Open the course Jupyter or Rstudio notebook
2. Create an SQL query to find solved cases in the NHS GMS cohort and pull out the recruited disease for each of them
3. Use the labkey_to_dataframe function to query LabKey
4. Visualise the dataframe

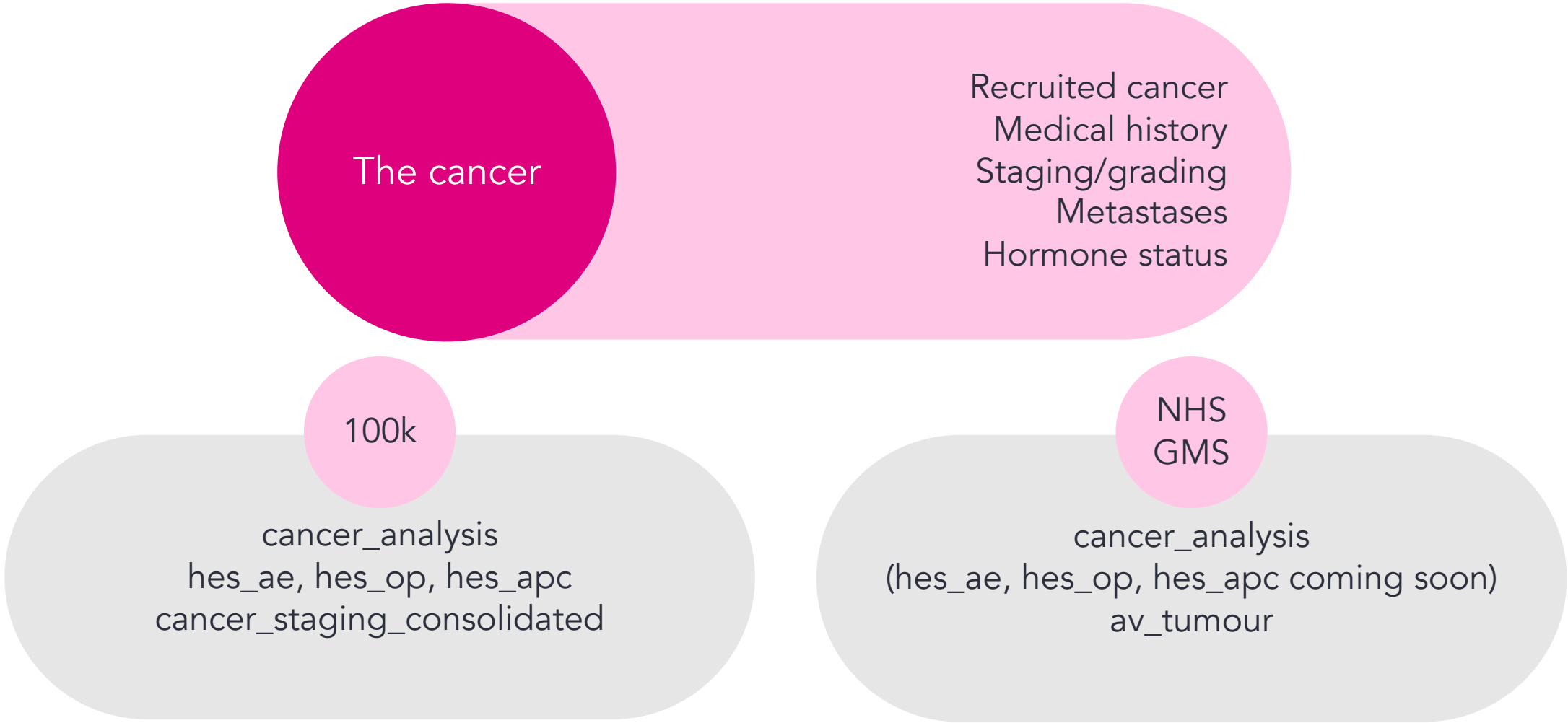


Cohort building

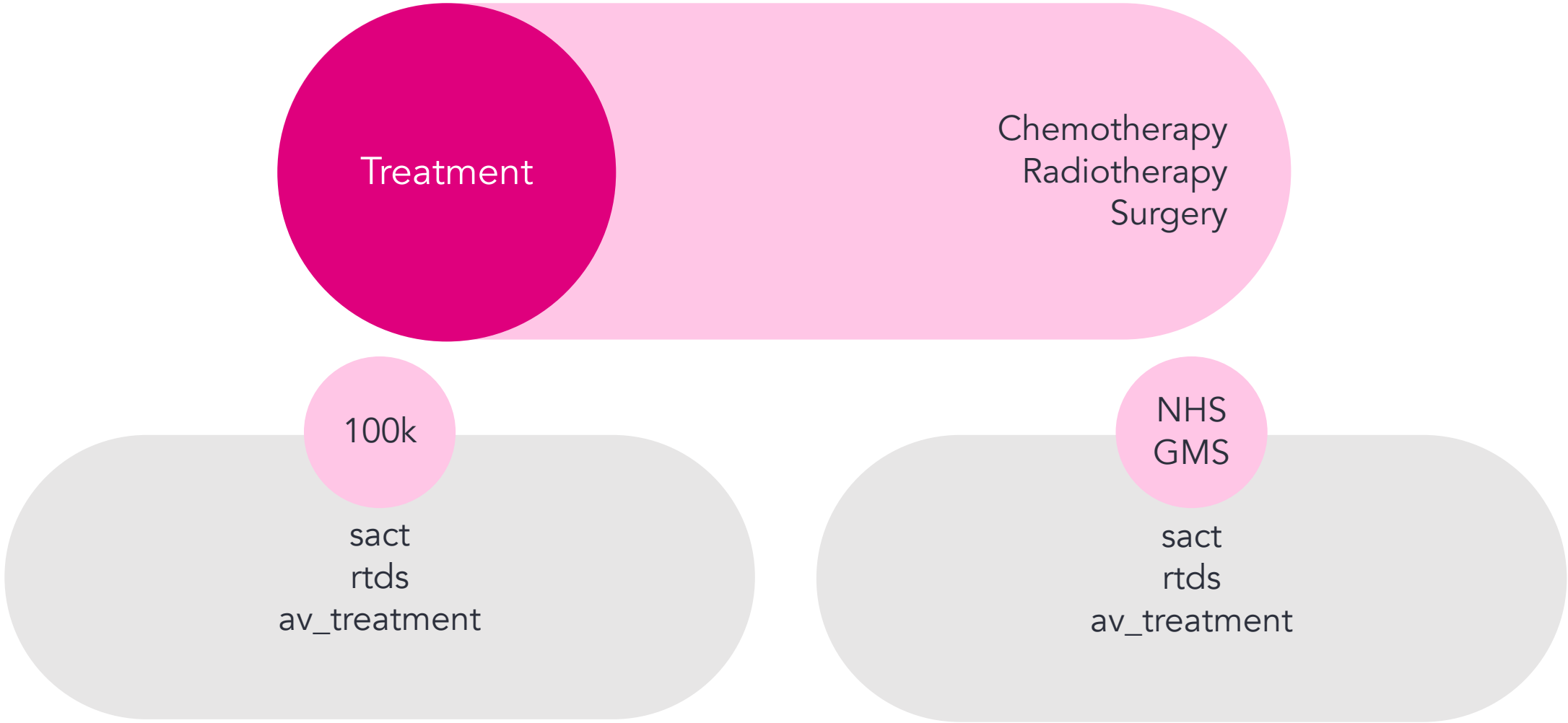
Rare disease cohort parameters



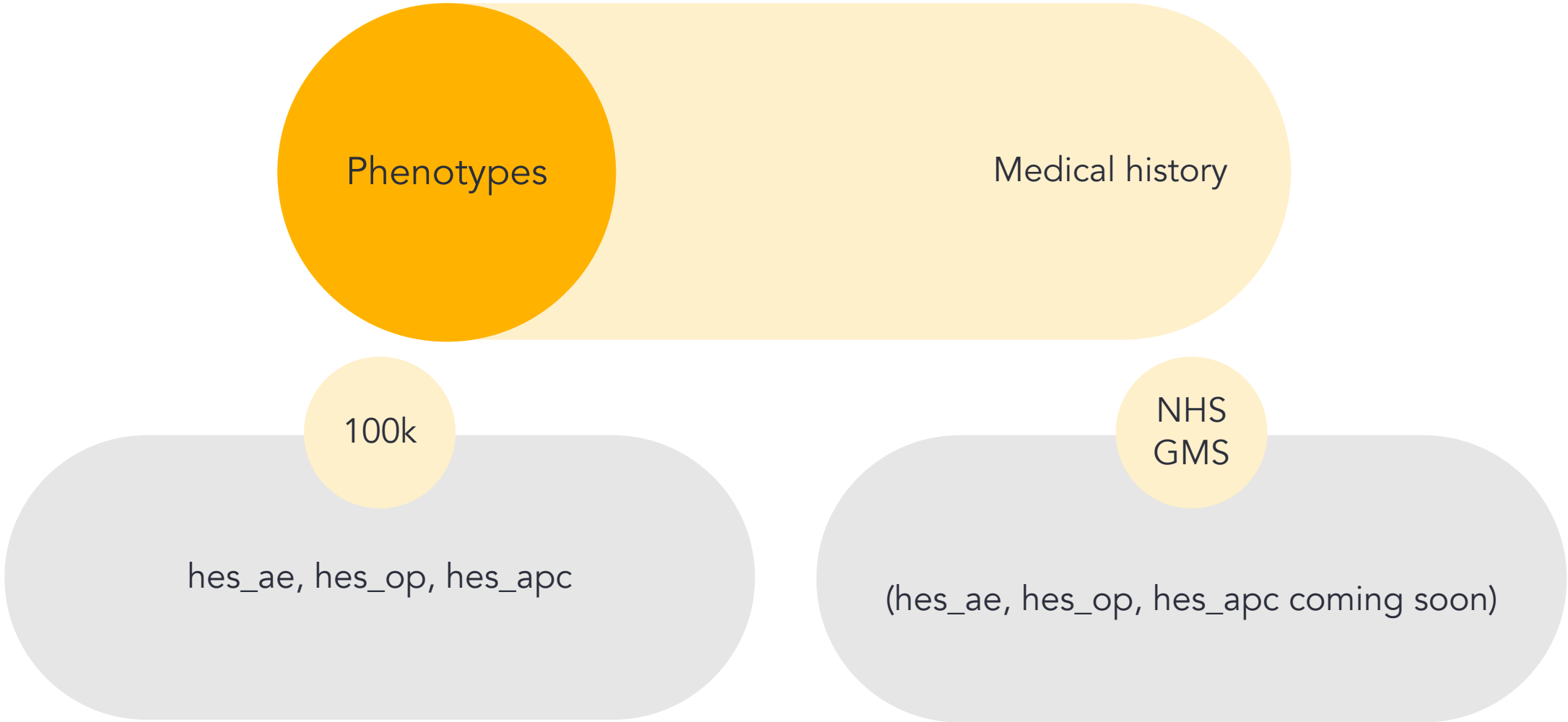
Cancer cohort parameters



Cancer cohort parameters



Common disease cohort parameters



General inclusion criteria

karyotypic = phenotypic sex
No aneuploidy

(for 100k)
aligned to GRCh38
in AggV2

assess kinship

blood
EDTA extraction
PCR-free

include MZ twins?

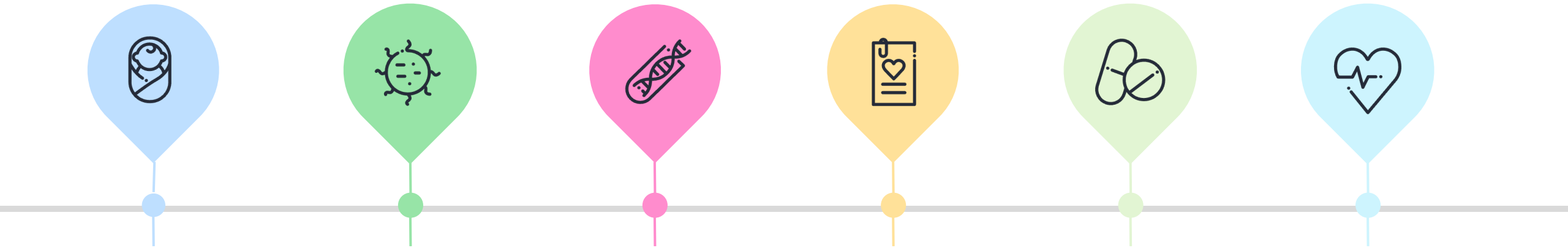
Additional QC

Match case/control

work with a single ethnicity

match sex ratios and age distribution

Derive age from dates



Year of birth

Diagnosis

Sampling

Phenotyping

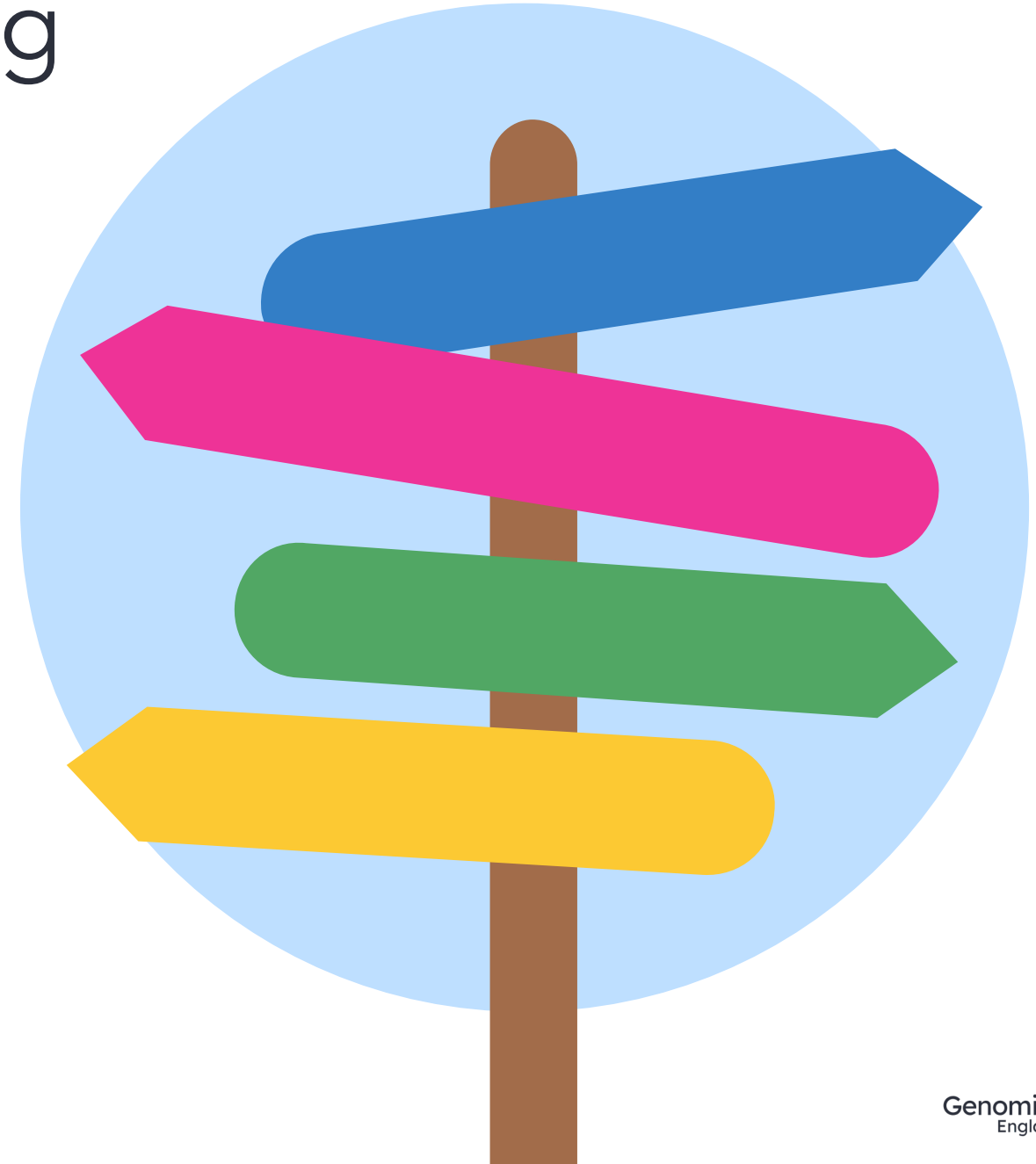
Treatment

Death

Cohort building demo

Task – Cohort building

1. Use the LabKey API to identify all participants recruited for Motor Neurone Disease/Amyotrophic Lateral Sclerosis.
2. Expand the search to add in all participants with the HPO term HP:0007354 or the ICD10 code G12.2 in the medical history.
3. Get the filepaths of the gVCF files for all participants.



Use tools on the HPC

What is an HPC?

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

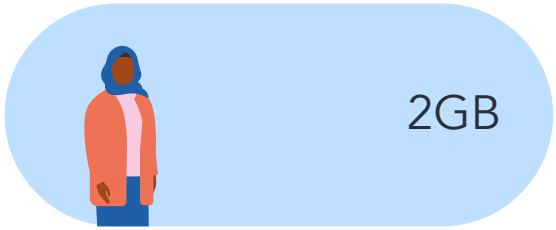
What is an HPC?

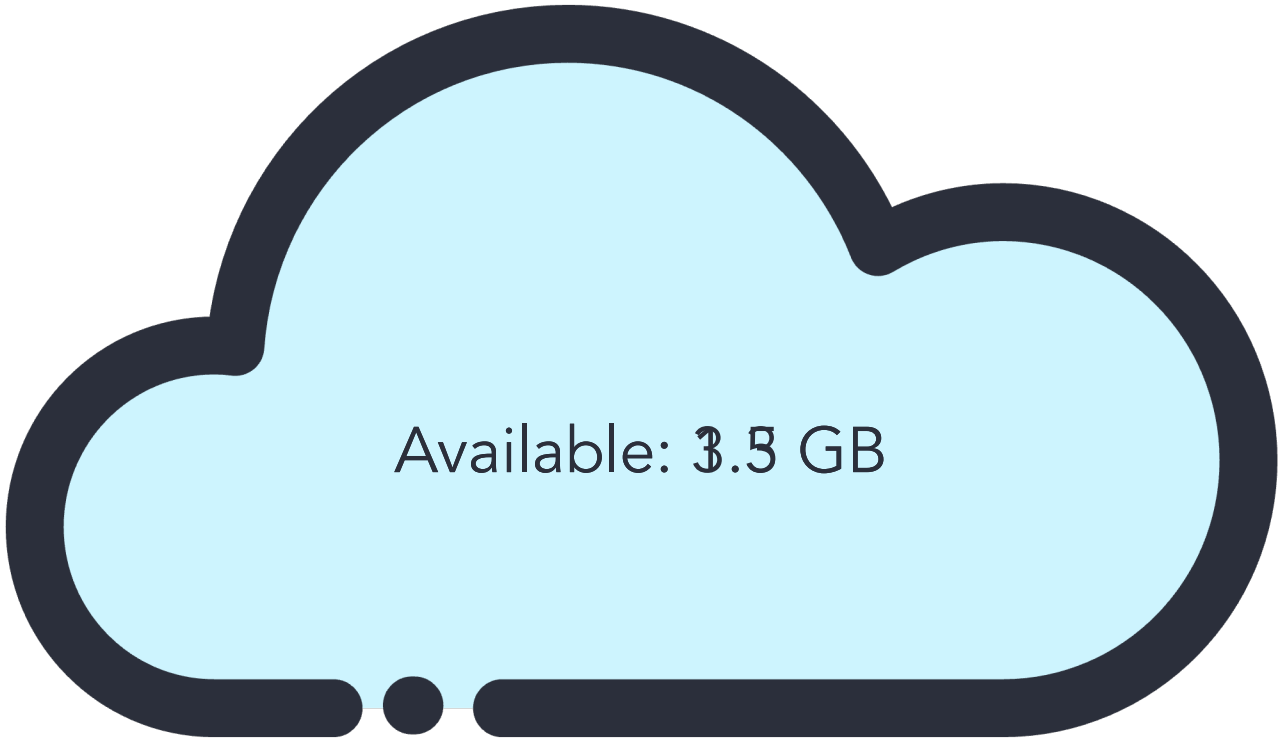
Lots of compute power

Shared with other researchers

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

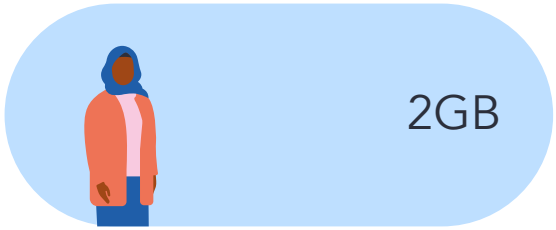
How do we share resources?





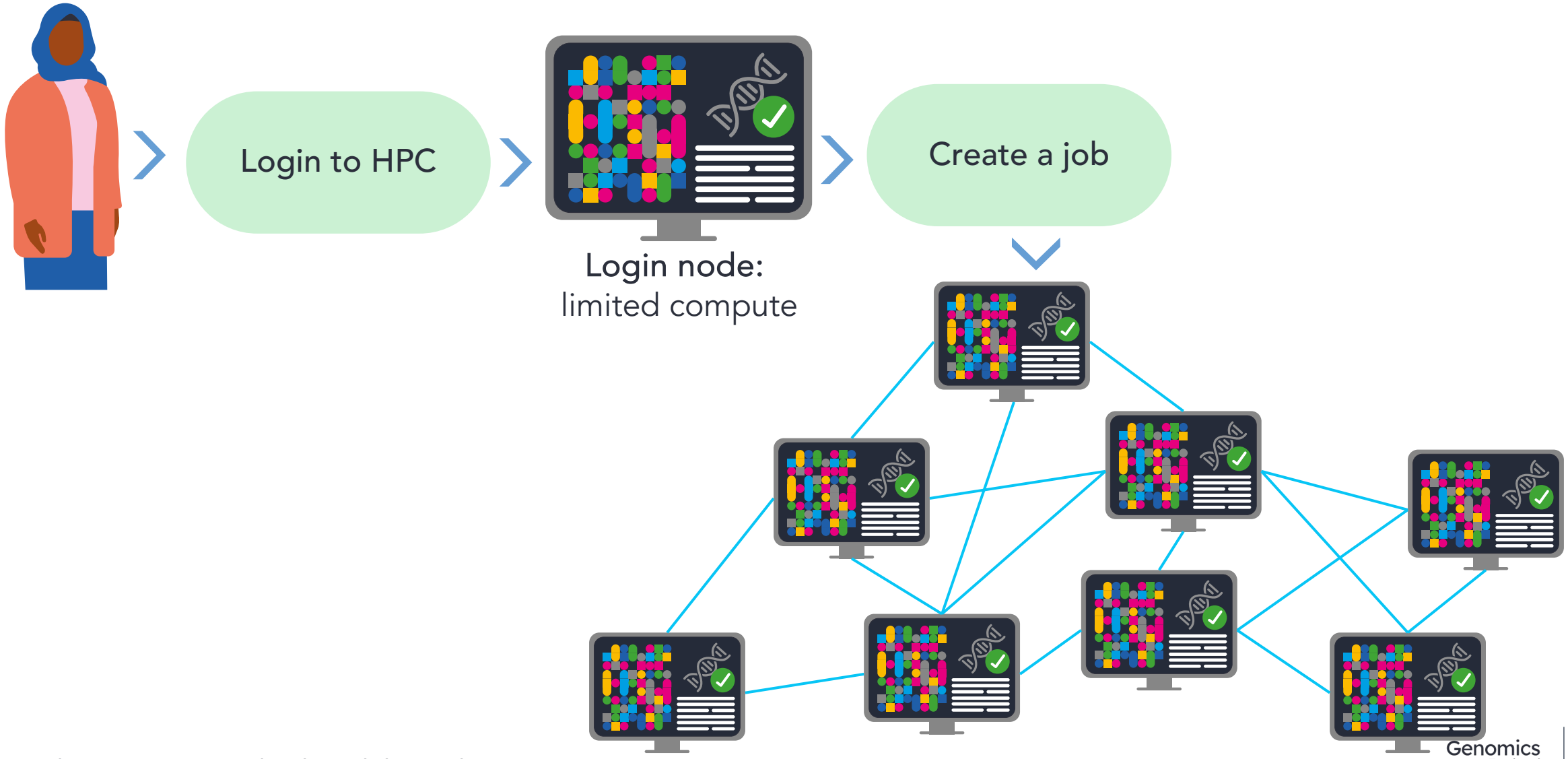
https://re-docs.genomicsengland.co.uk/hpc_about/

How do we share resources?



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

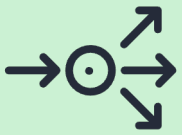
What is a "job"?



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

Two types of job

Interactive jobs



Batch jobs

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

Interactive jobs – the inter queue

require regular input

make decisions based on the results of the previous command

developing tools that you'll scale up later

exploratory analysis

interactive coding tools

GUI tools

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

A job on the inter queue

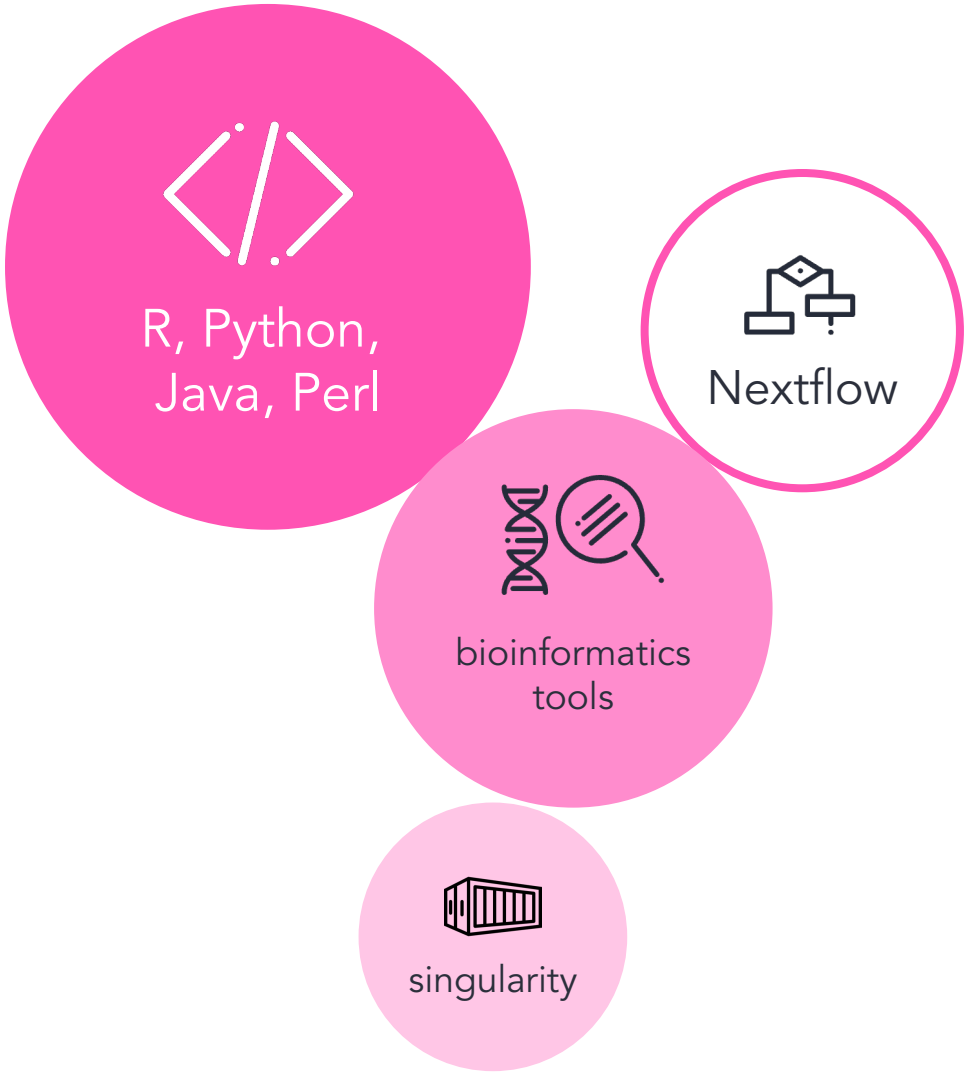
```
bsub -q inter -P <your_project_code> -R
rusage[mem=1000] -M 1000 -n 1 -Is /bin/bash
```

Do some work

```
bkill <job number>
```

https://re-docs.genomicsengland.co.uk/hpc_queues/#the-inter-queue

Software on the HPC



delly/1.2.6
denovoGear/1.1.1
discover/0.9.5
dotnet/2.0.0
dotnet/8.0.1 (D)
drop/1.2.4
eigen/3.3.9
exomiser/13.3.0
exonerate/2.2.0
fastqc/0.12.1
fetk/1.9.3
ffmpeg/6.0
fribidi/1.0.12
gatk/4.5.0.0
gauchian/1.0.2
gcc/10.4.0
gcta/1.94
gdal/3.7.0
geos/3.12.1
gistic/2.0.23
gmp/6.2.1
gnu-parallel/20190222
gnu/4.4
gradle/8.5
guppy/3.4.5
gvcfgenotyper/2019.02.26
haplocheck/1.3.3
hipstr/0.7
hisat2/2.2.1
hla-la/1.0.3
hmftools/2024-02-06
homer/4.11
htslib/1.18
igv/2.17.1
imagemagick/7.1.0
java/1.8
java/11.0.2
java/17.0.2
java/19.0.2 (D)
jq/1.7.1
kallisto/0.50.1
king/2.3.2
kraken/1.1.1
kraken2/2.1.3
lapack/3.12.0
ldsc/1.0.1
singularity/3.8.3
pindel/0.2.5b8
platypus/0.8.1
plink/1.9
plink/2.00a3.3LM
plink/2.0 (D)
plink_seq/0.10
popdel/1.5.0
proj/8.2.1
prsice-2/2.3.5
pycircos/1.0.2
pysam/0.22.0
python/3.8
python/3.8.1
python/3.11 (D)
readline/8.0
regenie/3.4.1
repeatDetector/1.0
rtg-tools/3.12.1
rvtests/2.1.0
saige/1.0.9
salmon/1.10.0
samtools/1.16.1
shapeit4/4.2.2
sniffles/1.0.11
somalier/0.2.19
sqlite3/3.40.0
squirls/2.0.1
stack/2.15.7
star/2.7.2a
star/2.7.11a (D)
strelka/2.9.10
superSTR/1.0.1
svanna/1.0.4
tabix/1.18
trimmomatic/0.39
udunits/2.2.28
vcf2maf/1.6.21
vcfanno/0.3.4
vcflib/1.0.9
vcftools/0.1.16
verifyBamID/2.0.1
vt/0.57721
xz/5.4.7
zlib/1.3
zulu/21.0.1
aws-cli/2.15
singularity/4.1.1 (D)

AdapterRemoval/2.3.3
AutoDock_Vina/1.2.5
BWA/0.7.17
BerkeleyDB/3.01
Bio-DB-HTS/3.01
CADD/1.6
CNView/1.0
CNVnator/0.4.1
CaVEMan/1.15.3
ExpansionHunter/3.2.2
ExpansionHunter/4.0.2 (D)
ExpansionHunterDenovo/0.9.0
GSL/2.7
MEDICC2/1.0.2
MPFR/4.2.0
R/3.6.3
R/4.2.1
R/4.3.3 (D)
REViewer/0.2.7
aliview/1.28
ampliconArchitect/1.3.r7
ampliconClassifier/1.1.1
annotSV/3.3.7
annovar/2019Nov
annovar/2024-03-14 (D)
ant/1.9.16
apbs/3.4.1
asmc-asmc/2024-02-26
automake/1.15
bamtools/2.5.2
bcftools/1.16
beagle/5.4
bedops/2.4.41
bedtools/2.30.0
bedtools/2.31.0 (D)
blast+/2.15
blat/1.0
bolt-lmm/2.4.1
boost/1.83
bowtie2/2.5.2
canvas/1.40.0.1613
circos/0.69-9
clang/16.0.6
cmake/3.24.3
cpan/1.7047
cromwell/v65
curl/7.81.0
cython/3.0.8
cytoscape/3.10.1
ldstore/2.0
libdeflate/1.20
libgit2/1.6.2
libtiff/3.4
libtiff/4.3.0
libtiff/4.5.0 (D)
libunwind/1.8.0
liftover/1.0
linasm/1.13
llvm/16.0.6
locuszoom/1.4
lollipop/0.3.0
lumpy/0.3.1
mafft/7.520
magma/1.10
manta/1.6.0
matlab/8.1
matlab/24.1 (D)
maven/3.9.6
meme/5.5.5
metal/1.0
miniconda3/23.11.0
miniforge3/23.11.0-0
minimap2/2.26
mosaicHunter/2024-02-14
mplayer/1.5
msisensor-pro/1.2.0
msisensor/0.6
multiqc/1.19
music2/0.2
mutserve/2.0.0-rc15
mutsig2cv/3.11
ncurses/6.4
new_fugue/2010-06-02
nextflow/22.10.5
nextflow/23.04
nextflow/23.10-with-plugins
nextflow/23.10
nextflow/24.04.2-with-plugins (D)
nf-core/0.3.1
nf-test/0.7.3
nf-test/0.8.2
nf-test/0.9.0 (D)
nodejs/16.9.0
openrefine/3.7.4
openssl/1.1.1o
pandoc/3.3
perl/5.38.2
picard/3.1.

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

Loading software

```
module avail myfavouritesoftware
```

list of available software including the string: "myfavouritesoftware"

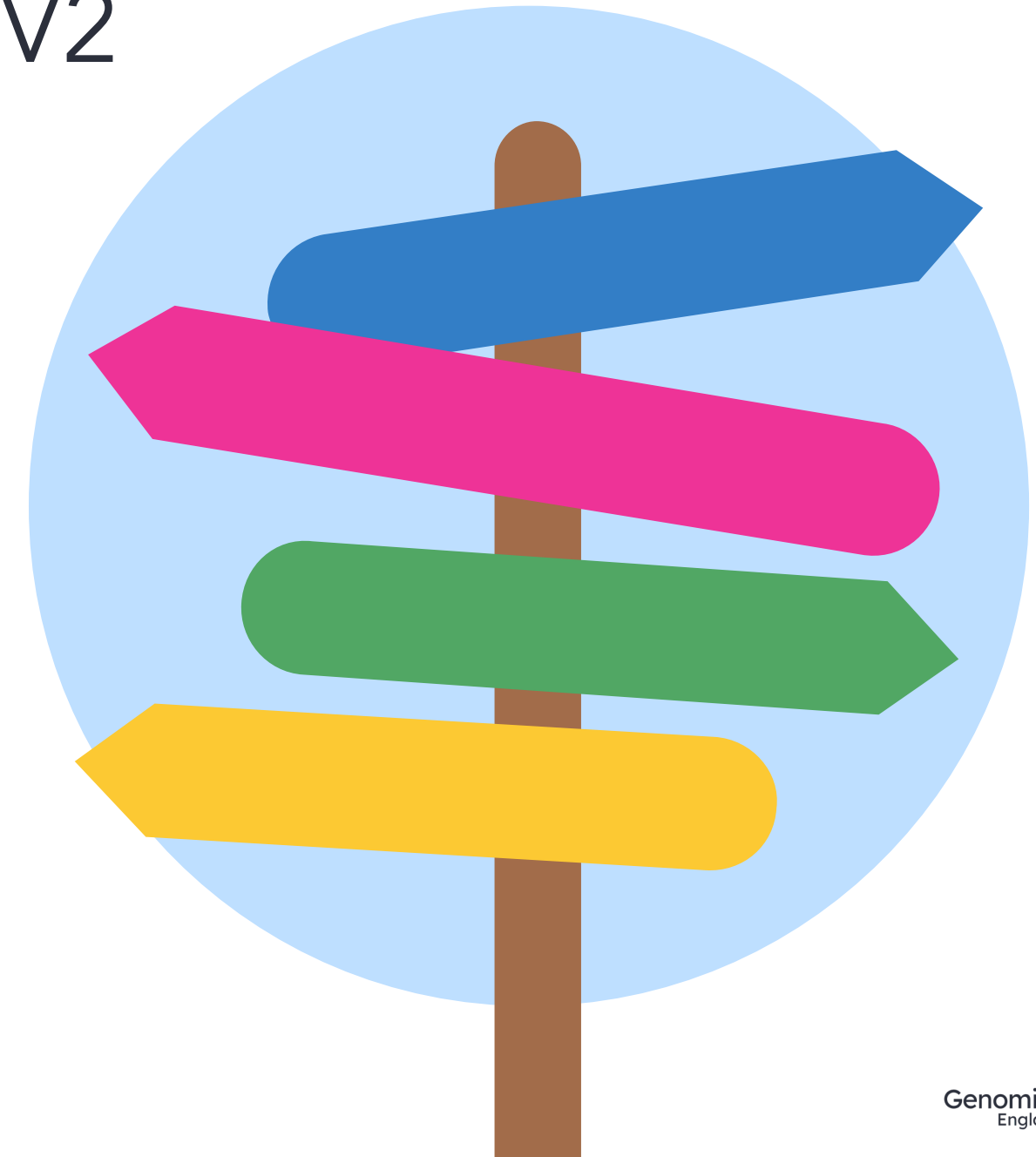
```
module load myfavouritesoftware/3.2
```

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

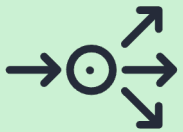
HPC interactive job and using tools demo

Task – Querying AggV2

1. Start an interactive job on the HPC
2. Load bedtools and use it to find the AggV2 chunk that contains the locus 2:135851076
3. Find all participants with an alternative allele at the locus 2:135851076
4. Kill your interactive job



Batch jobs



pipelines

next-step triggered in pipeline

Queues

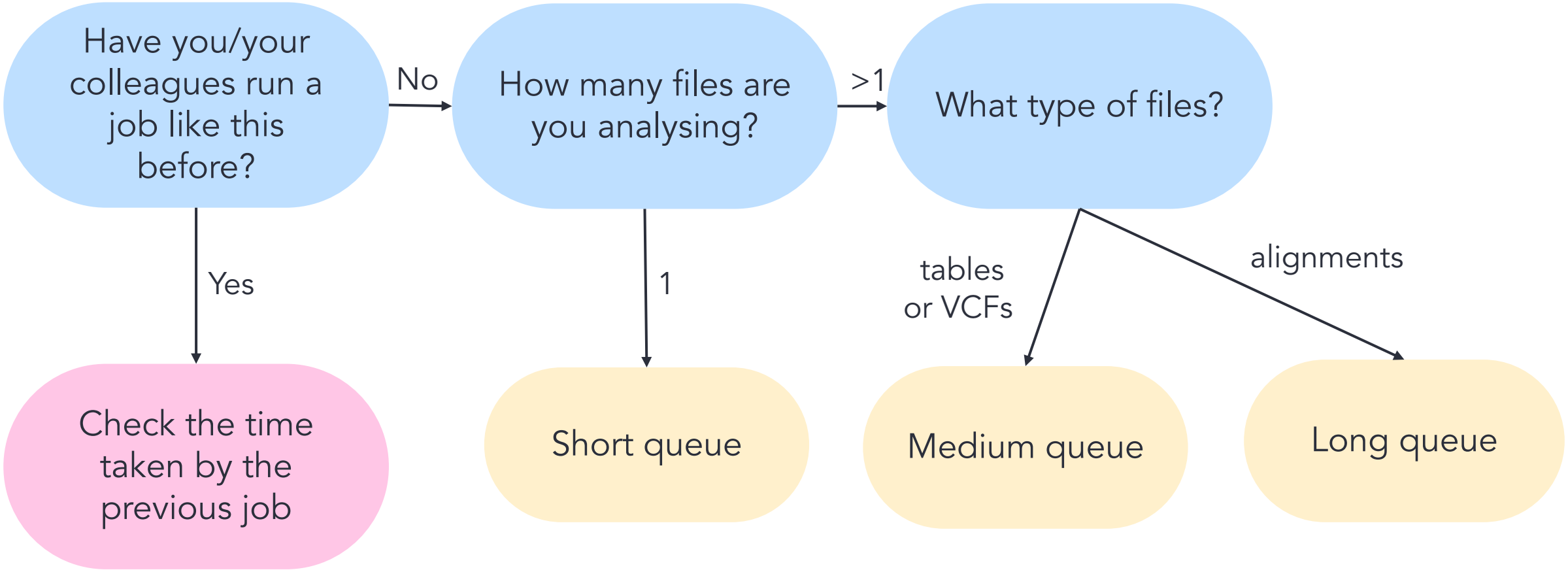
Short <4 hours

Medium 4-24 hours

Long 24 hours +

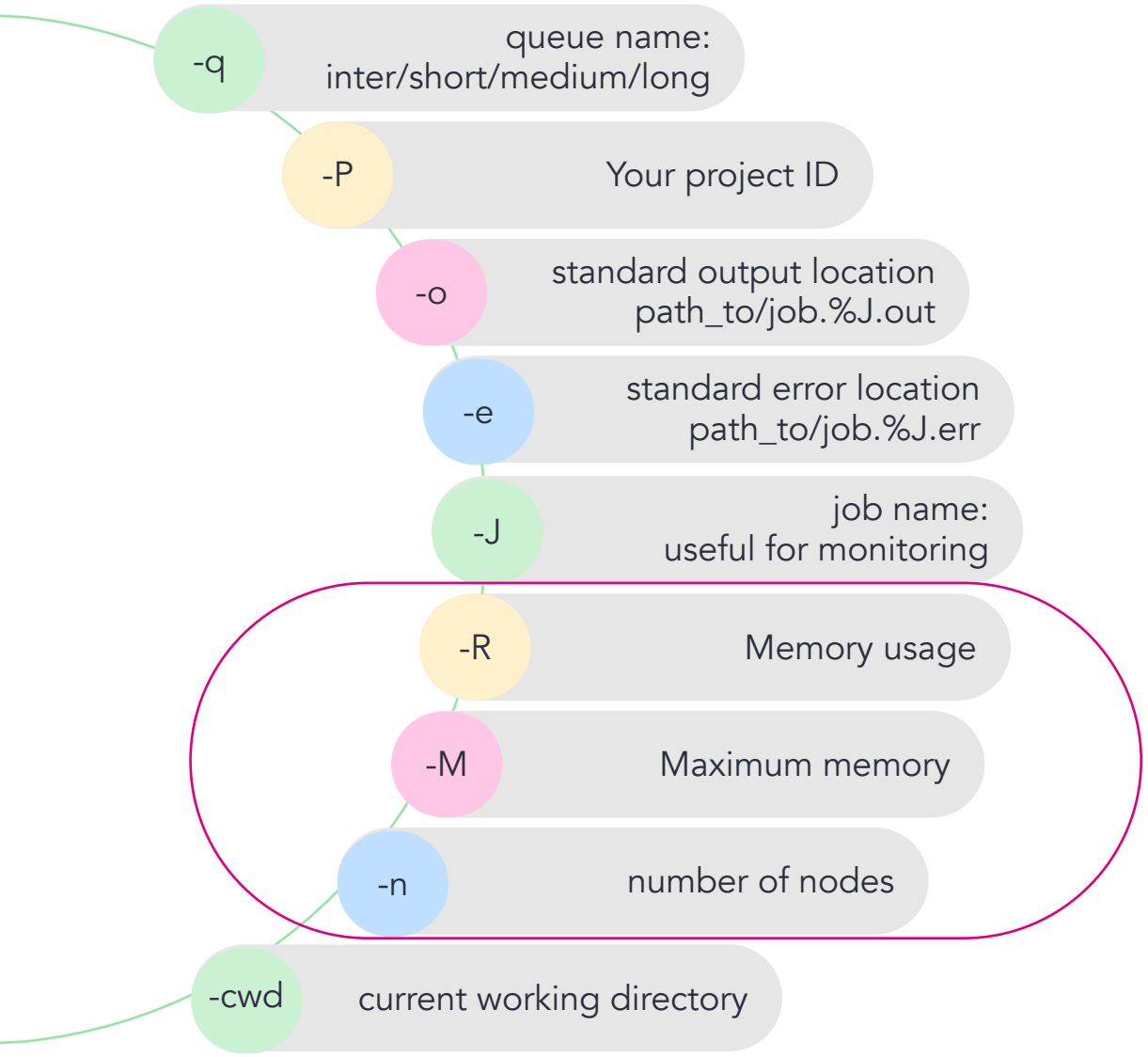
https://re-docs.genomicsengland.co.uk/hpc_queues/

Choosing a batch queue



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

Creating a job - parameters



bsub

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

Memory usage

-R Memory usage

- How much memory you're requesting in MB
- rusage[mem=1000]
- There needs to be this much memory available on the queue for your job to start

-M Maximum memory

- Maximum memory in MB
- 1000
- Your job will terminate if you exceed this

-n number of nodes

- How many nodes
- 1

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

Memory requested

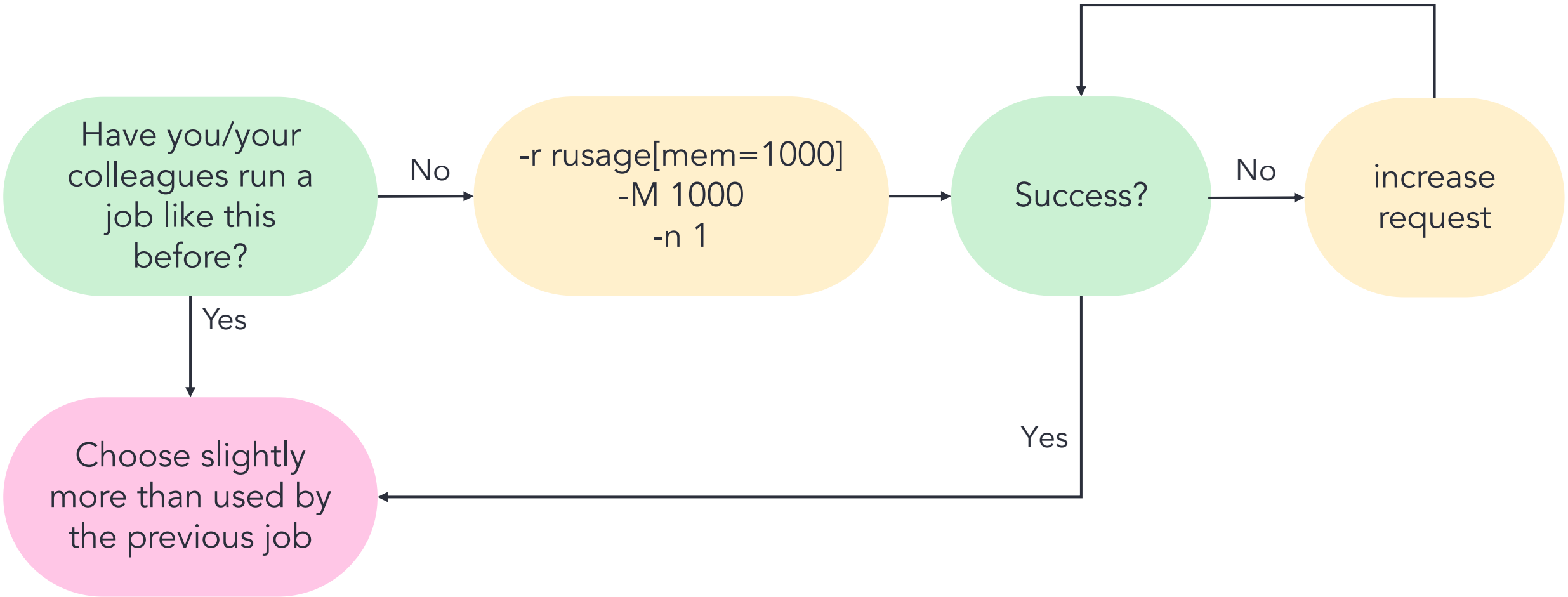
`Job killed after reaching LSF memory usage limit`

`PENDING`

Memory requested

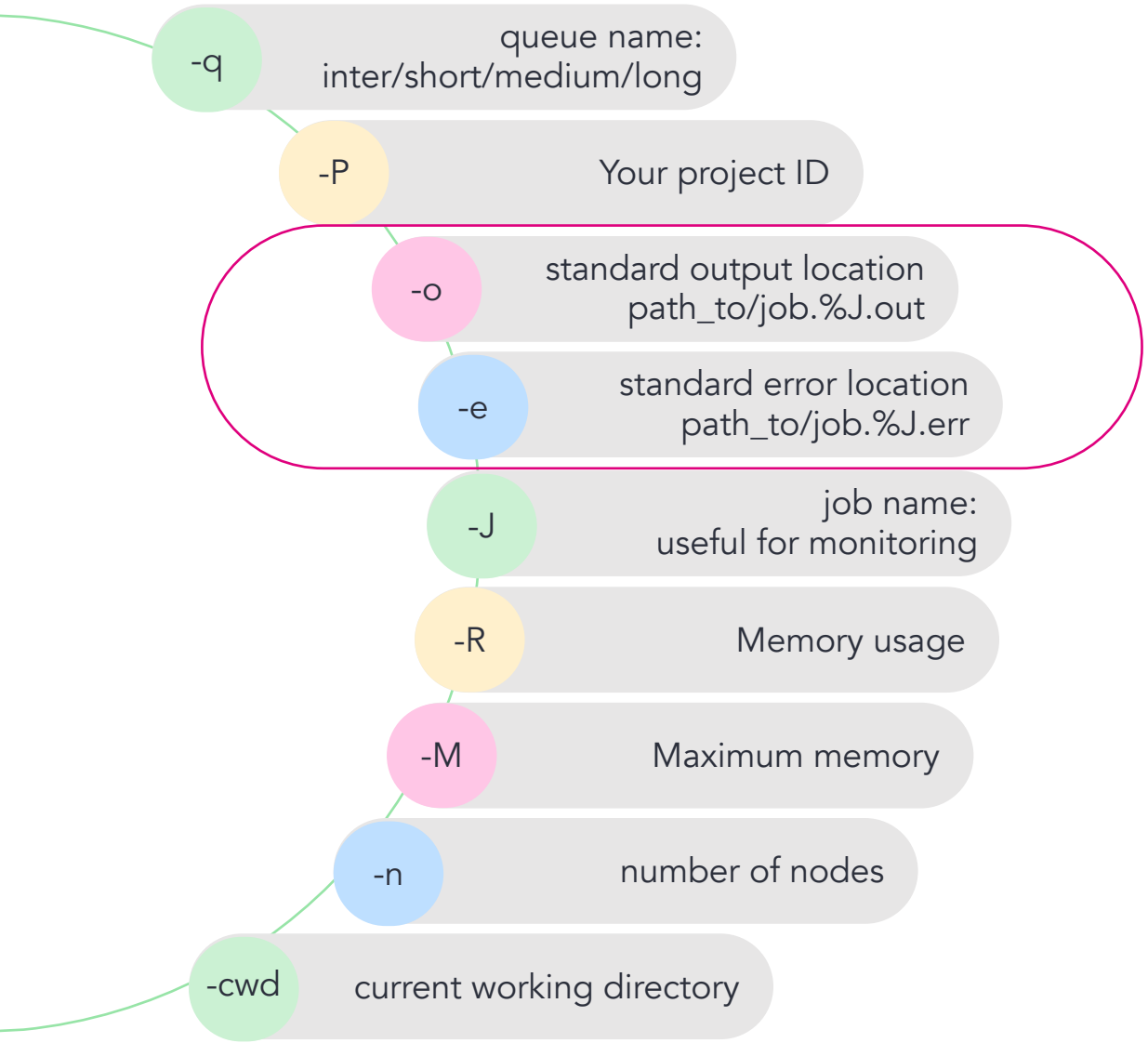
https://re-docs.genomicsengland.co.uk/hpc_memory/

Setting your memory request



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

Creating a job - parameters



bsub

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

A job on a batch queue – shell script

```
#!/bin/bash
# Include your job submission details as #BSUB headers
#BSUB -q <your_queue>
#BSUB -P <yourProject>
#BSUB -o <path_to/job.%J.out>
#BSUB -e <path_to/job.%J.err>
#BSUB -J <jobName>
#BSUB -R "rusage[mem=1000] span[hosts=1]"
#BSUB -M <max_memory_in_MB>
#BSUB -n <number_of_cores>
#BSUB -cwd <"your_dir">

# Set your temp directory as the re_scratch folder export
TMPDIR=/re_scratch/re_gecip/<your_GECIP>/<your_username> export
TMPDIR=/re_scratch/re_discovery_forum/<your_discovery_forum_folder>/<your_username>

# Load any required modules from the HPC
 module load <moduleName>

# The actual script you want to run
```

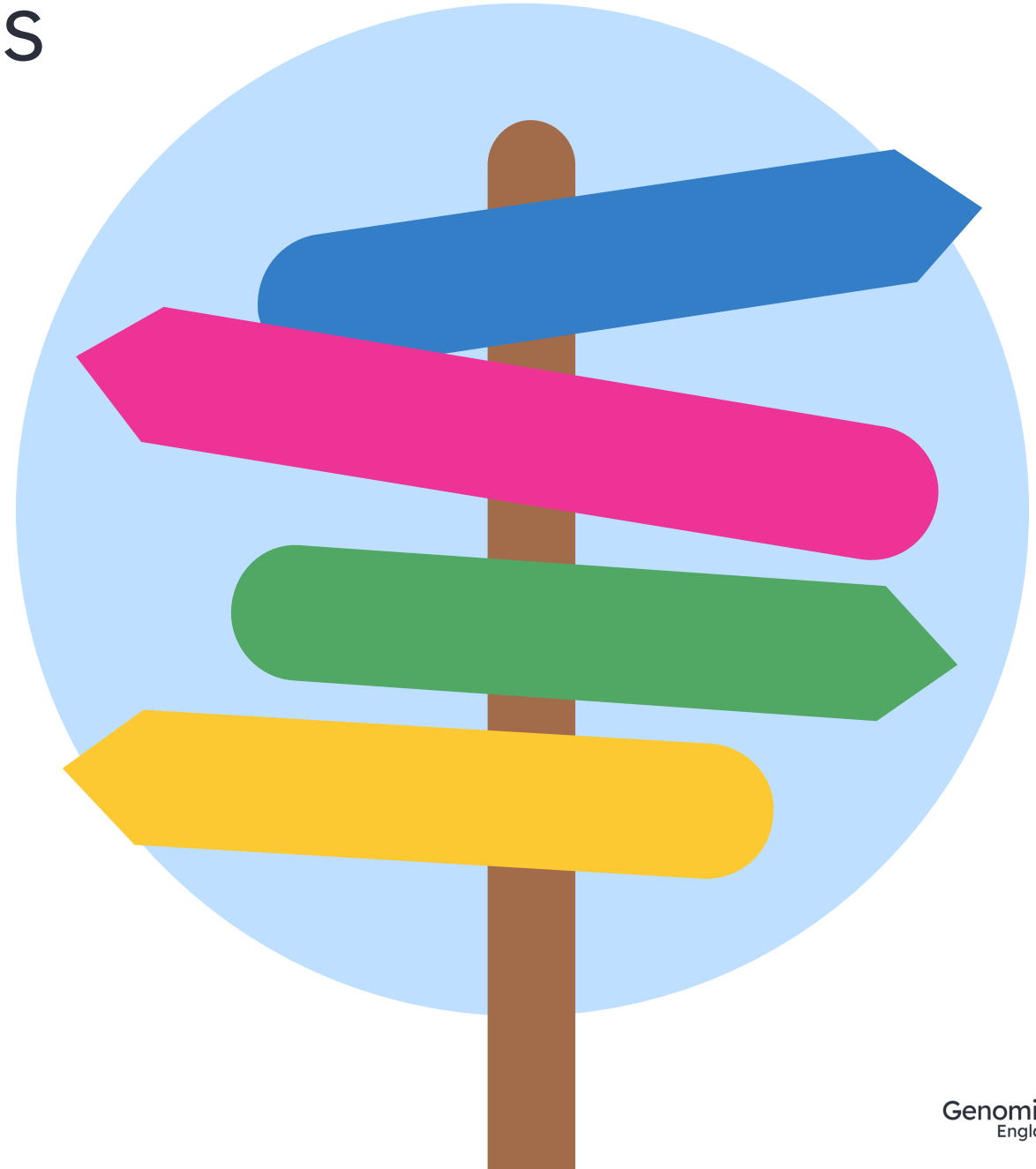
```
bsub < my_submission_script.sh
```

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

HPC batch job and checking output demo

Task – HPC batch jobs

1. Write a script to submit a batch job to the short queue using 100 MB of memory and 1 CPU
2. In your job, load bcftools and use it to query AggV2 as in the previous exercise
3. Submit the job to the HPC
4. Look at the time spent and the memory usage of the job, how much memory would you request next time?



Workflows on the HPC

Pre-built workflows/scripts to…

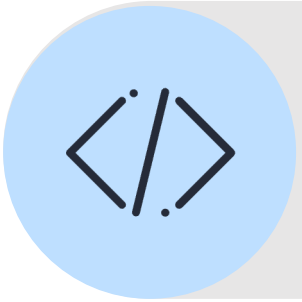
GWAS (common variants)
AVT (rare variants)

Small variant
Structural variant

survival

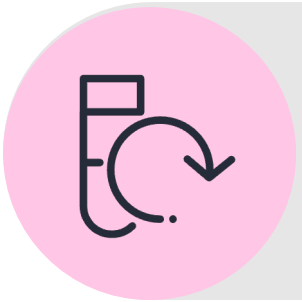
VEP

Workflows/scripts provide



Code that runs with only minor tweaks to add your input

Step-by-step instructions for use



Example input data and submission scripts

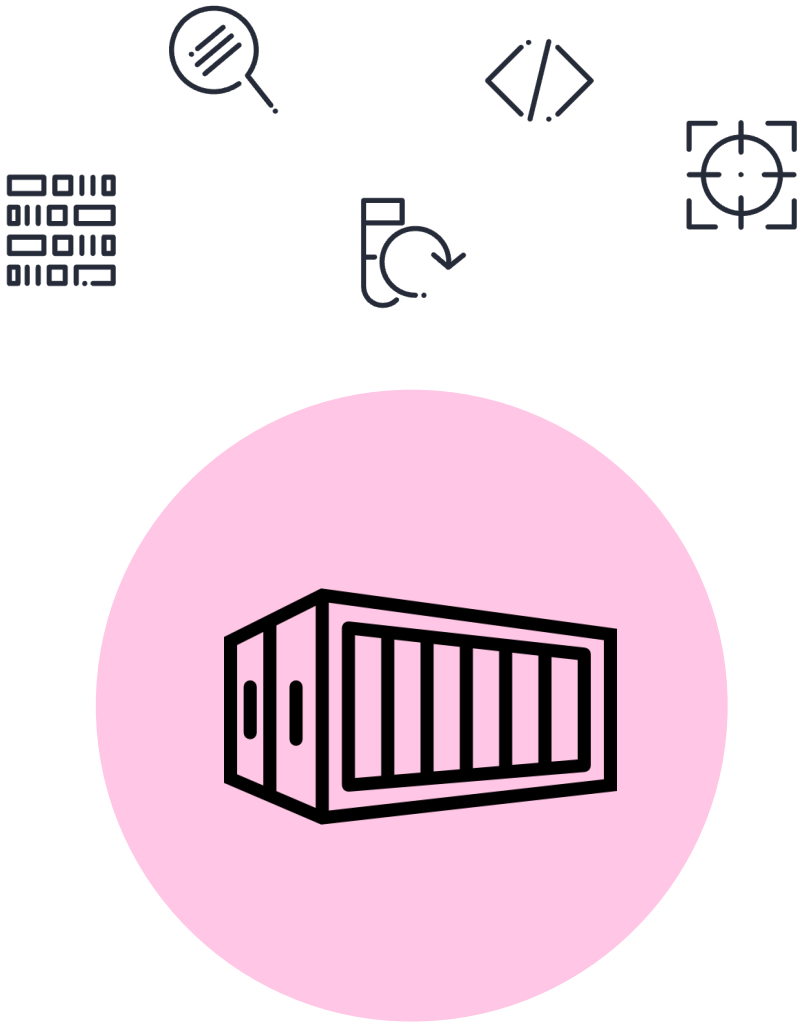
Optimised for use on our HPC and with our data

Output in standard or easy-to-interpret formats

Customisable options

Workflows demo

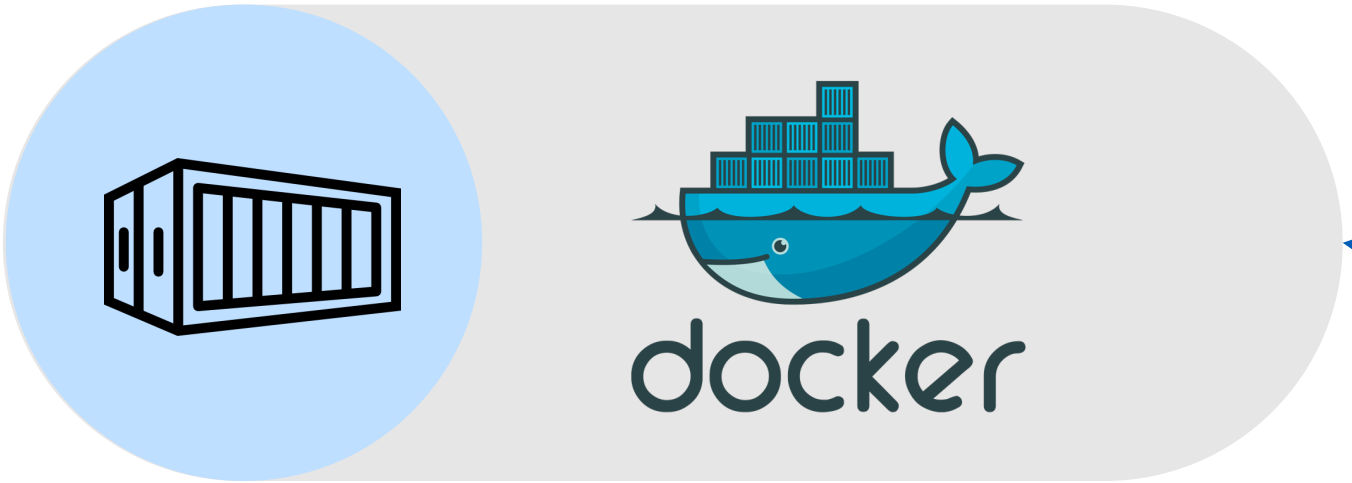
Containers

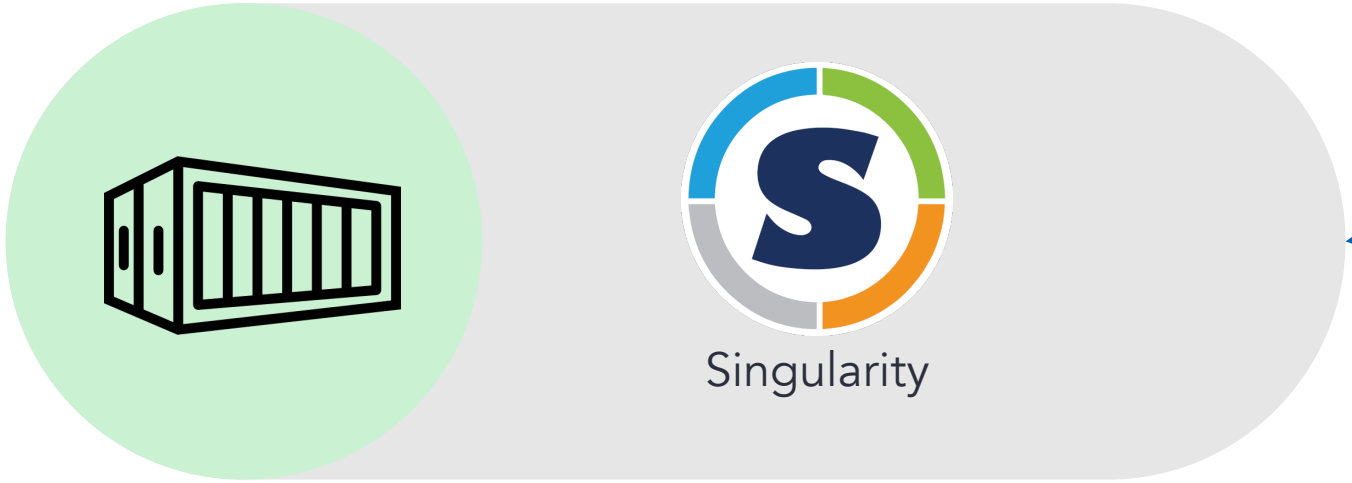


Different environment

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

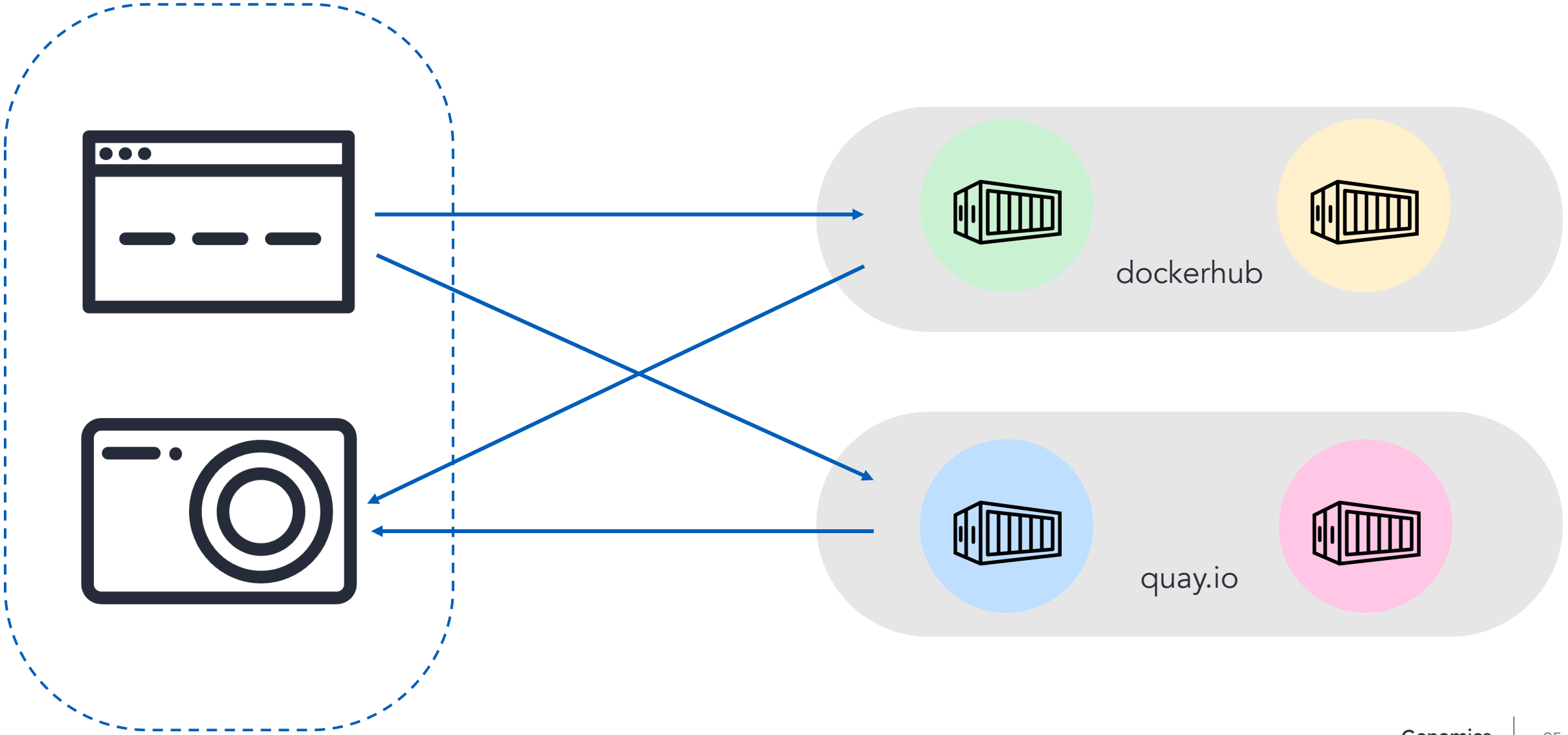
Container types





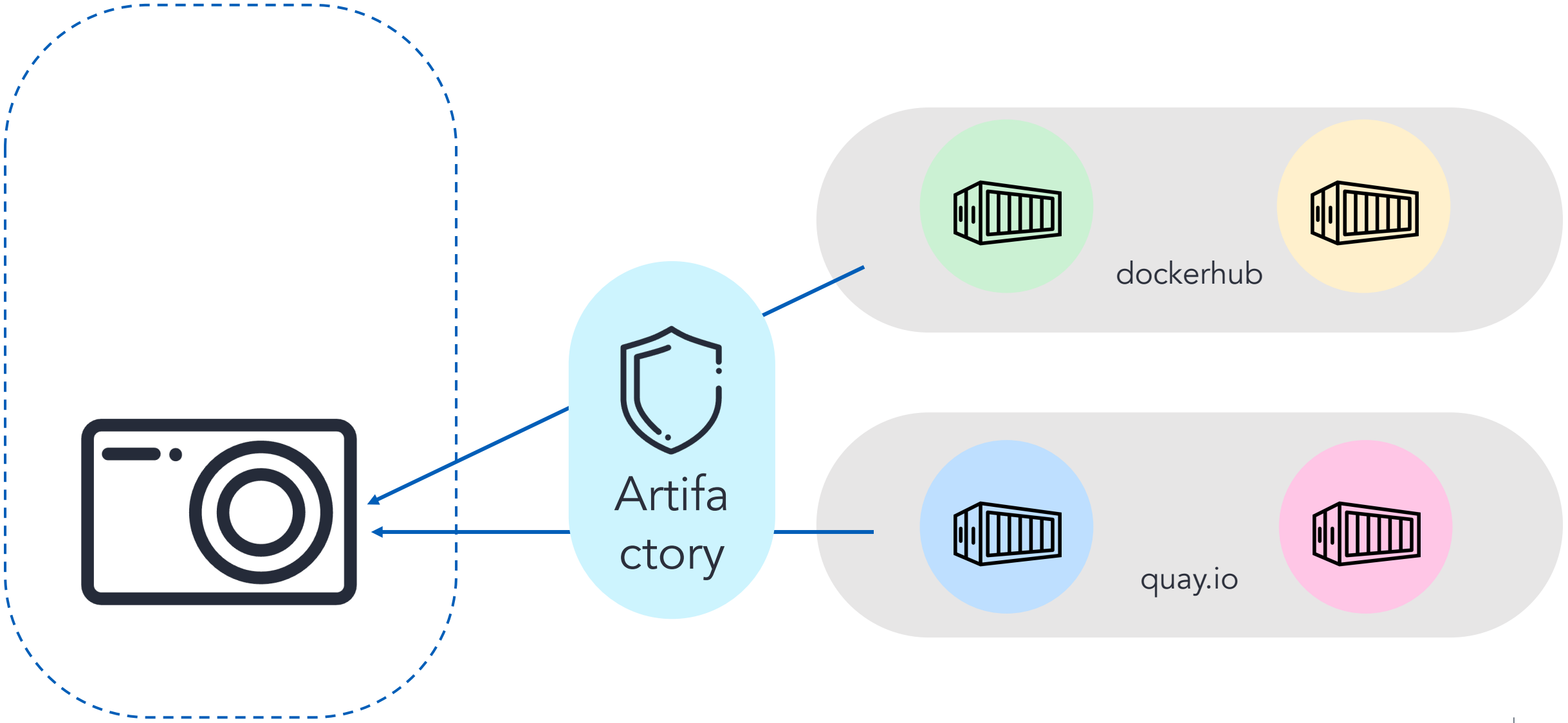
https://re-docs.genomicsengland.co.uk/hpc_containers/

Container repositories



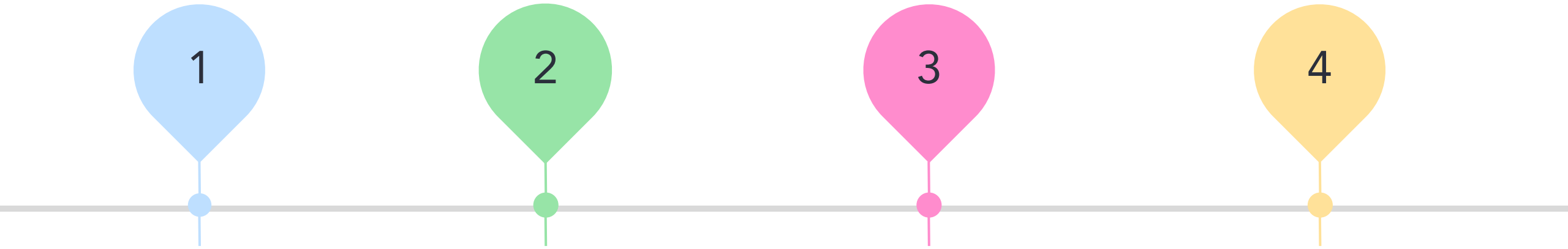
https://re-docs.genomicsengland.co.uk/hpc_containers/

Containers in the RE



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

Steps to import container

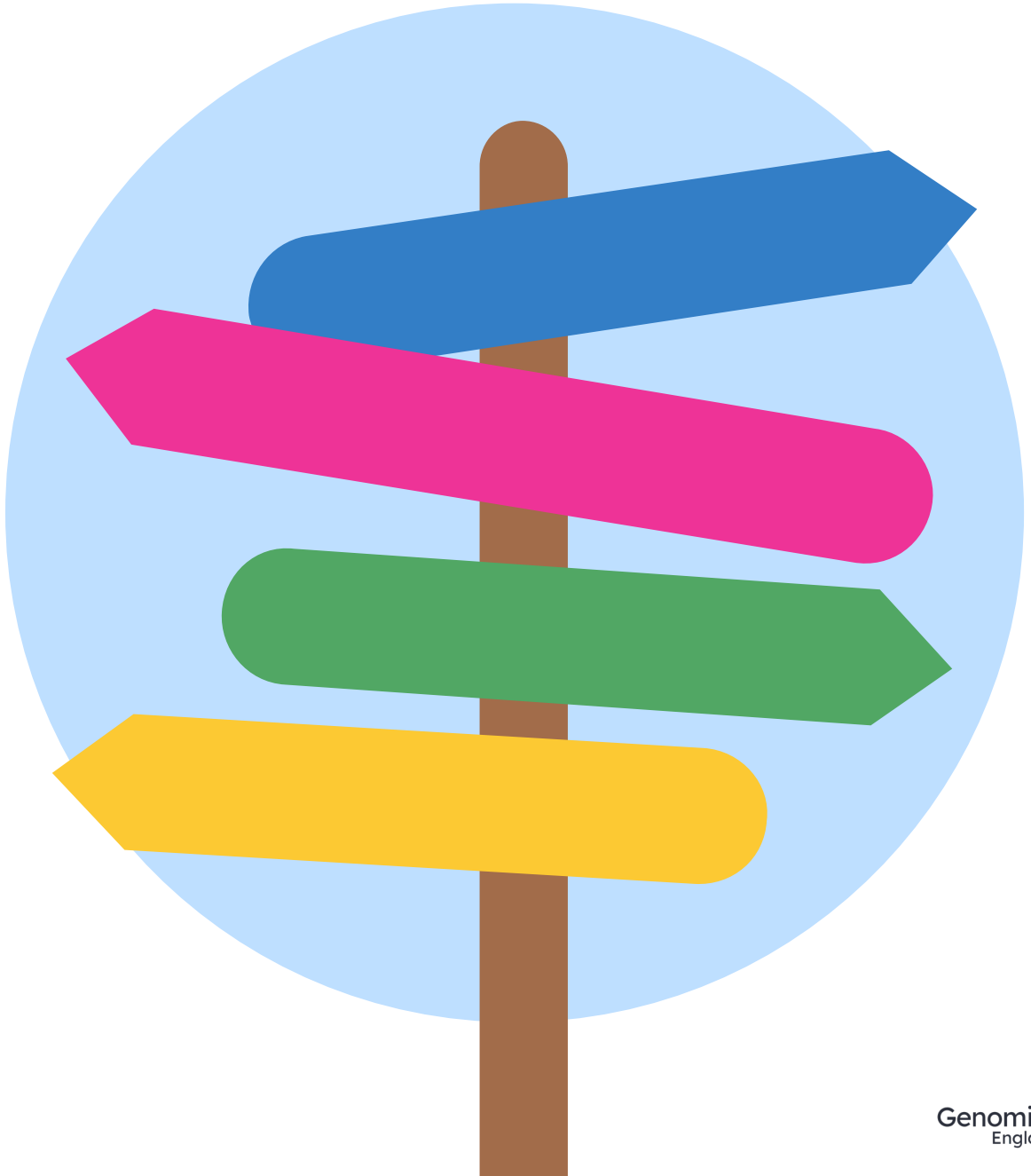


1. Find container in repository: identify docker or quay.io container location
2. modify import location to include artifactory redirect
3. pull container from an interactive job on the HPC
4. mount data and run analysis

Containers demo

Task – Containers

1. Pull the bcftools container from dockerhub: https://hub.docker.com/r/biocontainers/bcftools
2. Mount the chr21 vcf from the 1000 Genomes high coverage to the container
3. Query the locus 21:46228958 to find 1000 Genomes participants with alternative alleles at this locus, using the bcftools container (you can copy the query from the AggV2 demo)



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."

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

The Airlock

Data in the RE

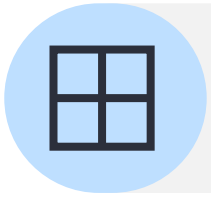
Outside world

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

What should go through the Airlock?

Figures

Statistics or numbers for your text



Tables

Notes on the data

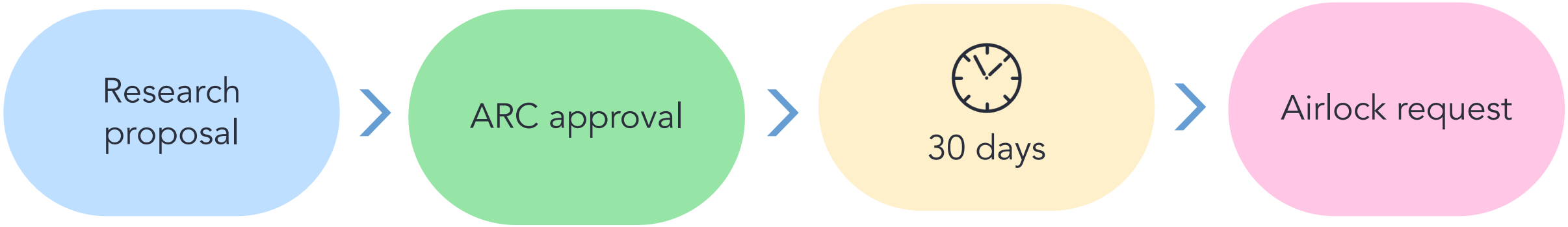
Airlock rules

Approved research project

Participants cannot be identified

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

Approved research project



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

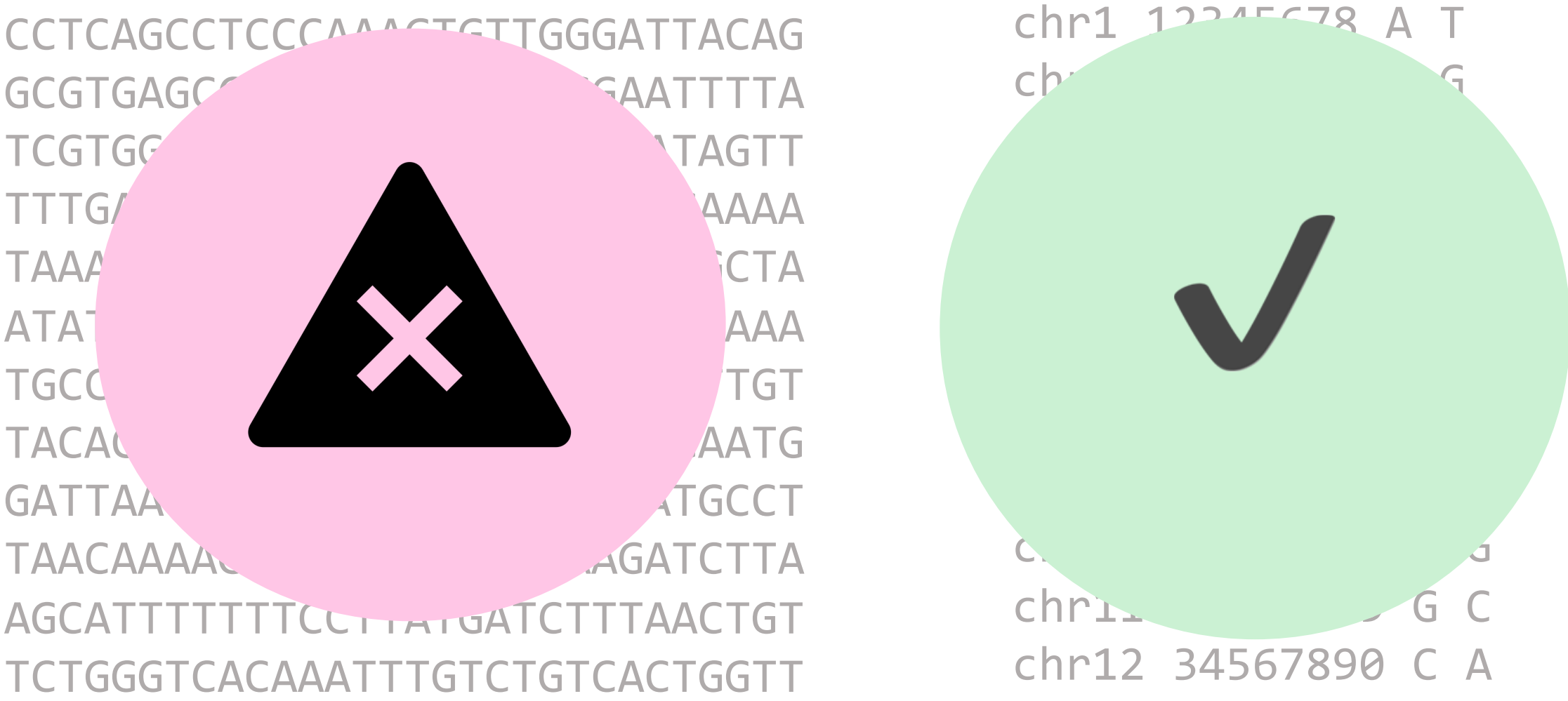
Participant and sample IDs

?

Public sample IDs can be used with non-identifiable data (eg tumour mutational signatures). These allow meta-analyses between studies. These always require discussion with the Airlock committee.

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

Nucleotides



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

Rule of five

Fewer than five participants

Five or more participants

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

Summary and test statistics

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

Forms in Airlock

- Export findings
- Export analysis scripts and software
- Contact clinical team and/or report potential diagnosis

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

Export findings

User group	Research Network or Discovery Forum
File(s) to export	Up to 20 files, up to 5GB
List of participant IDs	To check for consent for research
Description of file content	Describe everything, particularly unlabelled and non-standard
Research registry ID	To check it matches your registered project
Cancer/rare disease	
Aim of export	Publication, abstract, scientific talk, external analysis
Transfer request justification	Where you plan to publish and share

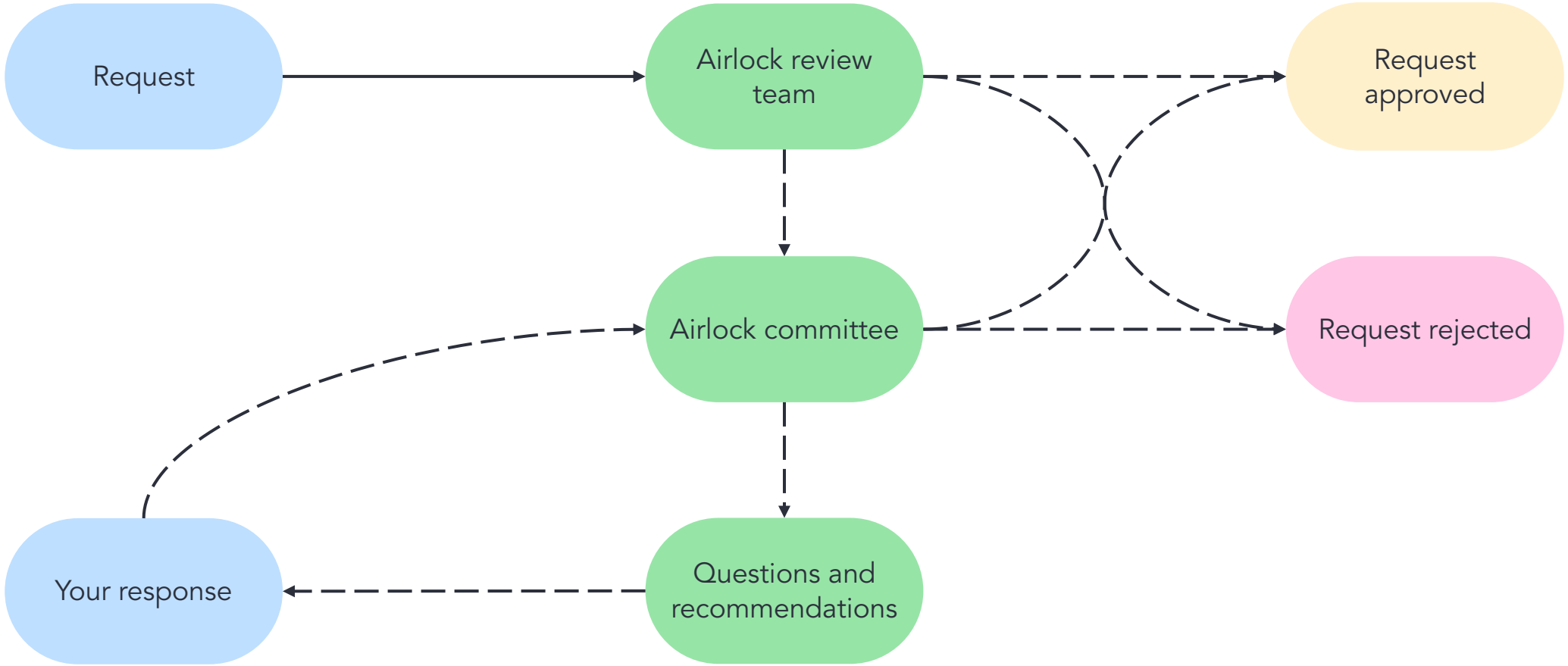
https://re-docs.genomicsengland.co.uk/airlock_tool/

Export analysis scripts and software

User group	Research Network or Discovery Forum
File(s) to export	Up to 20 files, up to 5GB, check to remove any hard-coded data
Description of file content	What the software does and any data included in the scripts
Research registry ID	To check it matches your registered project
Transfer request justification	If you plan to publish and share
Programming language	If multiple, select the most common
Software license	Who owns the software, if not you

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

What happens to my request?



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

Clinical collaboration

Examples of collaboration include:

- Request for patient consent to publish paper
- Inviting a clinician as a co-author on a paper
- Request for further health information or clinical tests
- To discuss with the clinician a potential diagnostic variant
- To offer laboratory tests to investigate in more detail whether a particular variant is likely to be diagnostic or not

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

Clinical collaboration

User group	Research Network or Discovery Forum
Research registry ID	To check it matches your registered project
What is your expertise in this area?	What research has led to this finding? Who is your PI?
Do you wish to be put in touch with the clinical team?	Yes or no, you can also use this form for submitting diagnostic discoveries
Why?	What is your reason for contact? What do you want from the clinician, and what are you offering them?

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

Airlock demo

Open Q&A/clinic

Questions

No such thing as a stupid question

Helpers

Roel Bevers
Senior Bioinformatician - Research Services

Christian Bouwens
Bioinformatician - Research Services

Alex Ho
Bioinformatician BRSC

Matthieu Vizuete-Forster
Bioinformatician - Research Services

Wrap-up

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

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

Training sessions 2024

10/12 Introduction to the RE

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/

Materials from past training all online

Feedback

Thank you

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