



Ready for
BioData.pt
Management?



Intensive Course

Data Reuse

Daniel Faria, Jorge Oliveira, Gil Poiares-Oliveira



I – Hands On

Learning Outcomes:

- Understand why data discovery is critical in personalised medicine.
- Identify relevant genomic and phenotypic datasets using:
 - Beacon Network
 - European Genome-phenome Archive (EGA)
- Interpret discovery-level metadata to evaluate dataset relevance.
- Understand access and governance implications of sensitive human data.



Data Reuse in Personalised Medicine

In personalised medicine, researchers often need to answer:

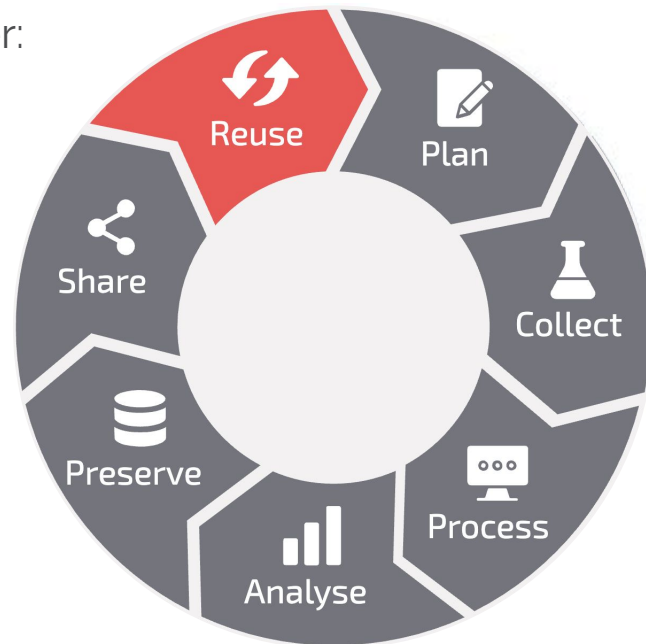
- Does any dataset exist that contains patients with a specific variant, disease, or phenotype?

However:

- Human genomic data cannot be openly shared.
- Data discovery must balance findability and privacy.

Two key tools addressing this:

- Beacon Network → Variant-centric discovery
- EGA → Dataset-centric discovery & controlled access



Beacon

- Beacon is an API (sometimes extended with a user interface) that allows for **data discovery of genomic and phenoclinic data**.
- Allows researchers to get information about the **presence/absence of a given, specific, genomic mutation** in a set of data, from patients of a given disease or from the population in general
- Currently in version 2 that also includes:
 - More informative queries, like **filtering** by gender or age;
 - An option to trigger the next step in the **data access** process, e.g. who to contact or which are the data use conditions;
 - **Annotations** about the variants found, among which the expert/clinician conclusion about the pathogenicity of a given mutation in a given individual or its role in producing a given phenotype; Information about cohorts.

Beacon

1. Create a query

Check for the presence of genomic variants, or look for detailed information through structured queries.



GRCh38 | 9:21975098_21967753del

2. Beacon API search

Beacon looks for the response, accommodating a wide range of data types, such as genomic and clinical data.



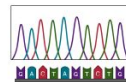
Patients data



Biosample data



Clinical Information



Genomic Information



3. Get the response

Get the results to the query:

- Boolean (yes/no).
- Count.
- Case-level data.



Obtained results

Yes / No

12 patients found

Obtain access to case data

<https://beacon.biodata.pt/>

 **Variant** Region Phenoclinic Cohorts

LOG IN



Beacon
Project

Training UI

chr : pos ref > alt



 Query Examples

Beacon Networks

Do you have information about Allele C at position 39,963,793 on Chromosome 13?

<https://beacon-network-demo.ega-archive.org>

<https://beacon-network.org>



Beacon Network

Search all beacons for allele

GRCH37 • 13 : 32906732 G > C

Response: All None
Found 0
Not Found 34
Not Applicable 21

Organisation: All None
GATC Lab, LLC Bioreason
Austriac Genomics
Beigene Medical Services
BGI
Bioinformatics Research Center
Bioinformatics Research Center
BRCA Exchange
Broad Institute
Centre for Genomics
Centre for Personalized Medicine
Chromosome Mapping Hub
Cytoscape

BRCA Exchange
Hosted by BRCA Exchange
Found

ExAC
Hosted by Broad Institute
Found

MyVariantInfo
Hosted by The European Research Institute
Found

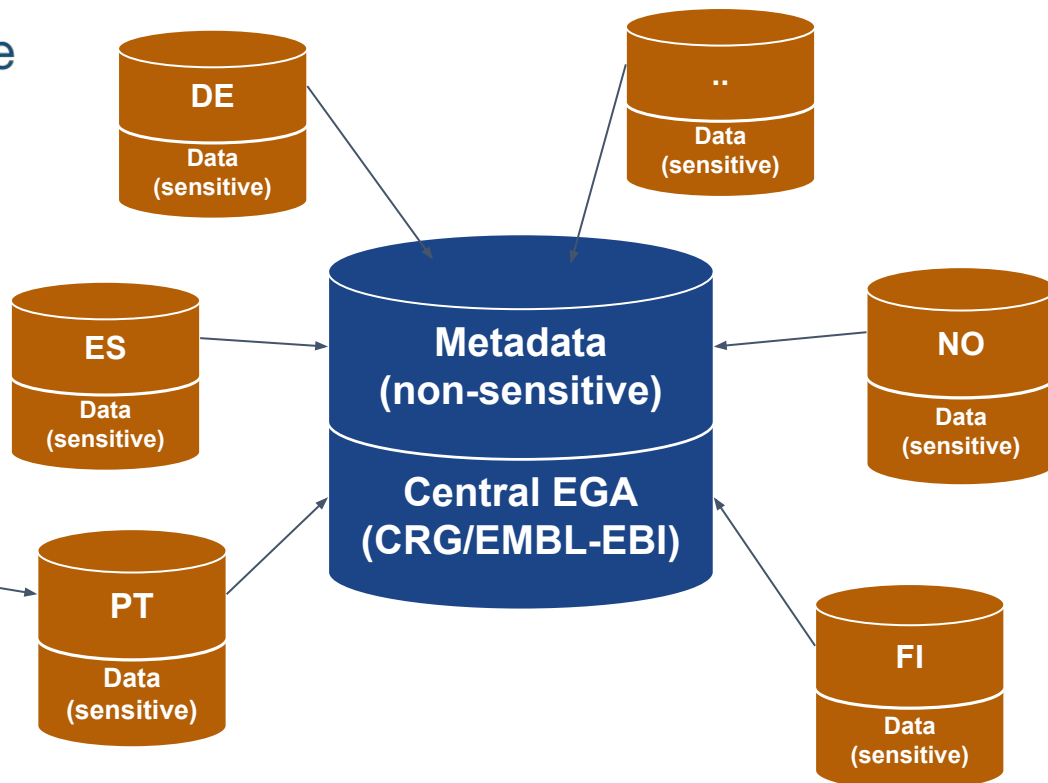
MyVariantInfo - CADD
Hosted by The European Research Institute
Found

MyVariantInfo - dbNSFP
Hosted by The European Research Institute
Found

Clinical significance	PV02:0106138
HGVScDNA	c.7879G>C
Condition ID	OMIM
Type	B/1015
Date last evaluated	
HGVSp protein	p.(Trp282Cys)
Assertion	http://evm.consortium.org/document/ENGINA_Files_2015-05-26.pdf
ClinVarAccession	CV00024475
alt	C
Condition category	Disease
Comment on clinical significance	WFO class based on posterior probability from multigene Bayesian analysis thresholds for class as per Phospho at 2000 (PMID: 18901446). Class 5 based on posterior probability < 1
Reference sequence	NAL_000559.3
Gene symbol	BRCA2
URL	http://hg19.soe.ucsf.edu/genome/hg19/chr13/32906732/chr13:32906732:G>C
chrom	13
Abbrev AA change	W282C

- Service for permanent archiving and sharing of personally identifiable genetic, phenotypic, and clinical data generated for the purposes of biomedical research projects or in the context of research-focused healthcare systems.
- Currently federated into National nodes to meets the requirements of the General Data Protection Regulation (**GDPR**).
- Data submitted to the archive is subject to **controlled access**
 - Access to the data only will be granted after a formal application procedure.

Archive for **secure** storing and sharing **sensitive** data from Portuguese **research** projects.





Intensive Course

Data Reuse

Exercise

- Pick a disease of your choice
- Explore the [Beacon networks](#) and search for genomic variants
- Identify relevant datasets in [EGA](#)
- Assess their suitability for a research question





Thank You!

Questions?