



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:

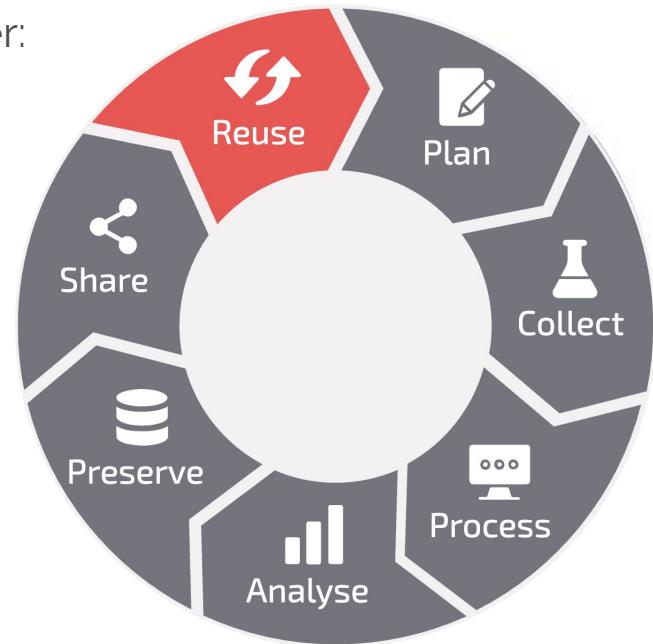
- Does any dataset exist that contains patients with a specific variant, disease, or phenotype?

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

4

- 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

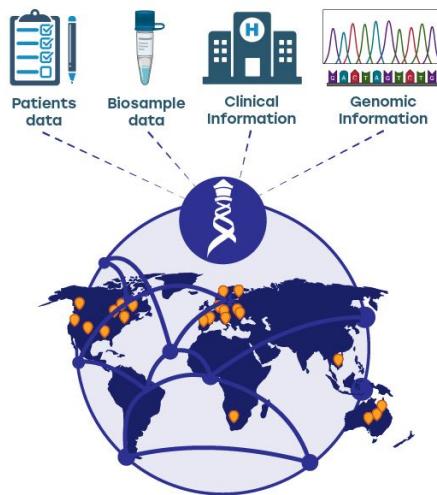
1. Create a query

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



2. Beacon API search

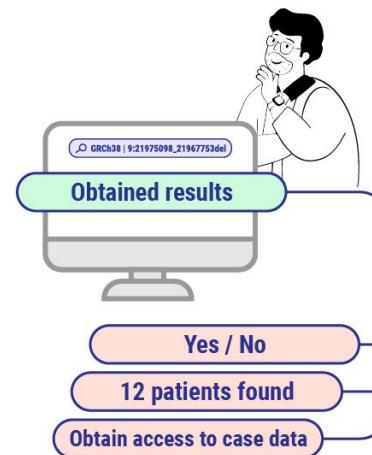
Beacon looks for the response, accommodating a wide range of data types, such as genomic and clinical data.



3. Get the response

Get the results to the query:

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



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



Variant Region Phenoclinic Cohorts

LOG IN



Beacon
Project | Training UI

chr : pos ref > alt



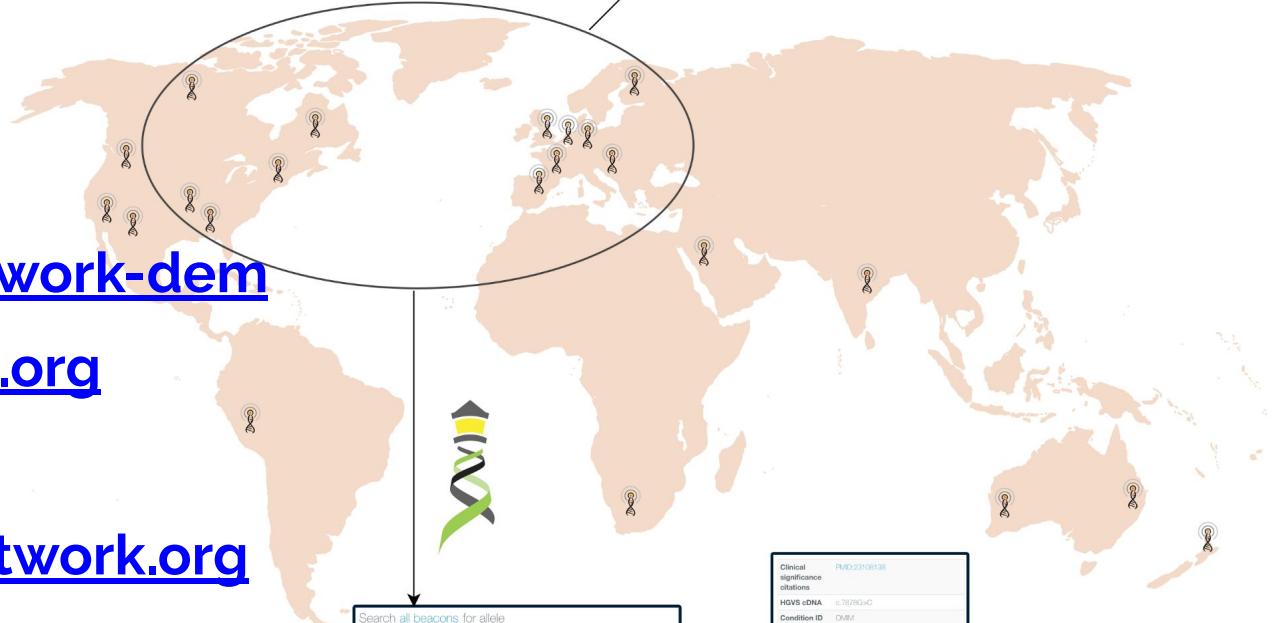
💡 Query Examples

Beacon Networks

[https://beacon-network-dem.
ega-archive.org](https://beacon-network-dem.ega-archive.org)

<https://beacon-network.org>

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



Search all beacons for allele

GRCh37 - 13 : 32996732 G > C

Response All None

Found 34 Not Found 34 Not Applicable 21

Organization All None

- MUTdb - UC Berkeley
- MyVariant.info - Wellcome Trust Sanger Institute
- MyVariant.info - Wellcome Trust Sanger Institute
- MyVariant.info - Wellcome Trust Sanger Institute
- BRCA Exchange
- ExAC
- MyVariant.info - The Genome Research Institute
- MyVariant.info - CADD
- MyVariant.info - dbNSFP

Show Metadata Found

Show Metadata Found

Show Metadata Found

Show Metadata Found

Clinical significance	PV/D/23106138
citations	HGVS cDNA: c.767Q>C
Condition ID	DMX
type	
Date last evaluated	8/19/15
HGVs protein	p.Thr239Ser
Assertion method	http://remesconforum.org/documents/ENSGVA_Rules_2015-03-28.pdf
citation	ClinVarAccession: C_000244475
alt	C
Condition category	Disease
Comment on clinical significance	WGS case based on posterior probability from multifactorial likelihood analysis. The variant has a posterior probability of a 2030 (PMID: 18951445). Class 3 based on posterior probability = 1
Reference sequence	NM_000563
Gene symbol	BRCA2
URL	http://hgvs.consortium.ncbi.nlm.nih.gov/HGVS/variant/submit/
chrom	13
Abbrev AA	W269K
change	





EUROPEAN
GENOME-PHENOME
ARCHIVE

<https://ega-archive.org/>

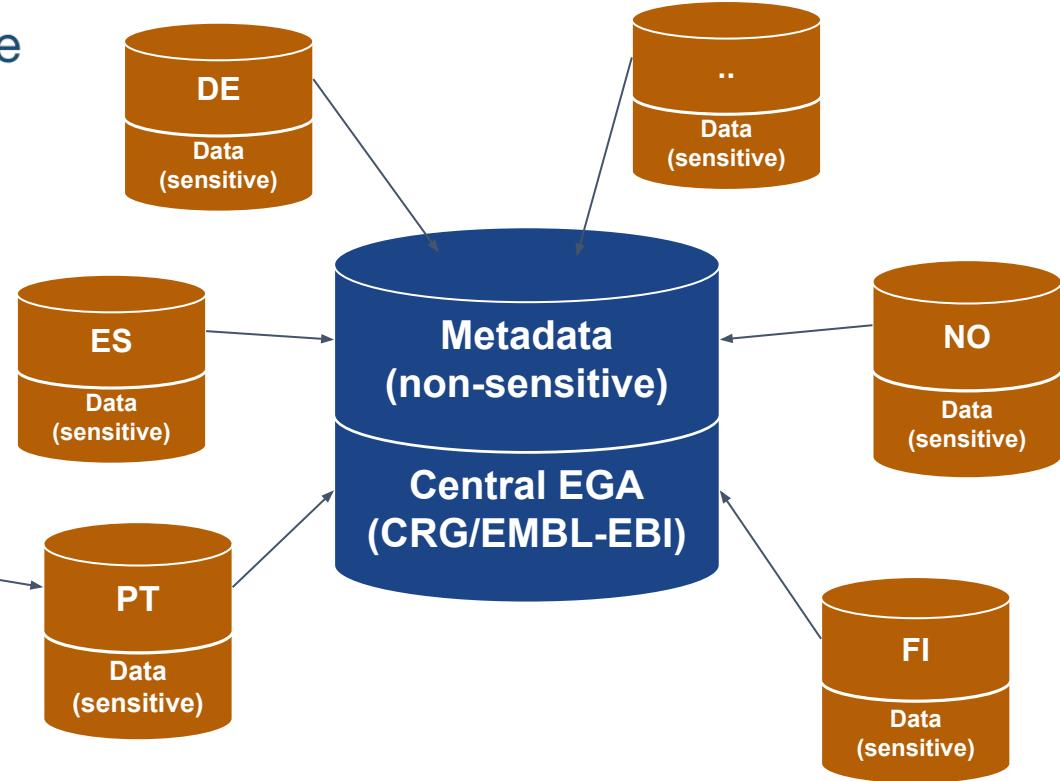
- Service for permanent archiving and sharing of personally identifiable genetic, phenotypic, and clinical data generated for the purposes of biomedical research projects or in the context of research-focused healthcare systems.
- Currently federated into National nodes to meets the requirements of the 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.





Federated European Genome-phenome Archive Portugal

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





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



Ready for
BioData.pt
Management?



Intensive Course

Data Sharing & Reuse

Thank You!

Questions?