



ELIXIR Beacon Project

Concepts | Status | History | Outlook

Michael Baudis :: RDA 2021 :: 2021-04-20

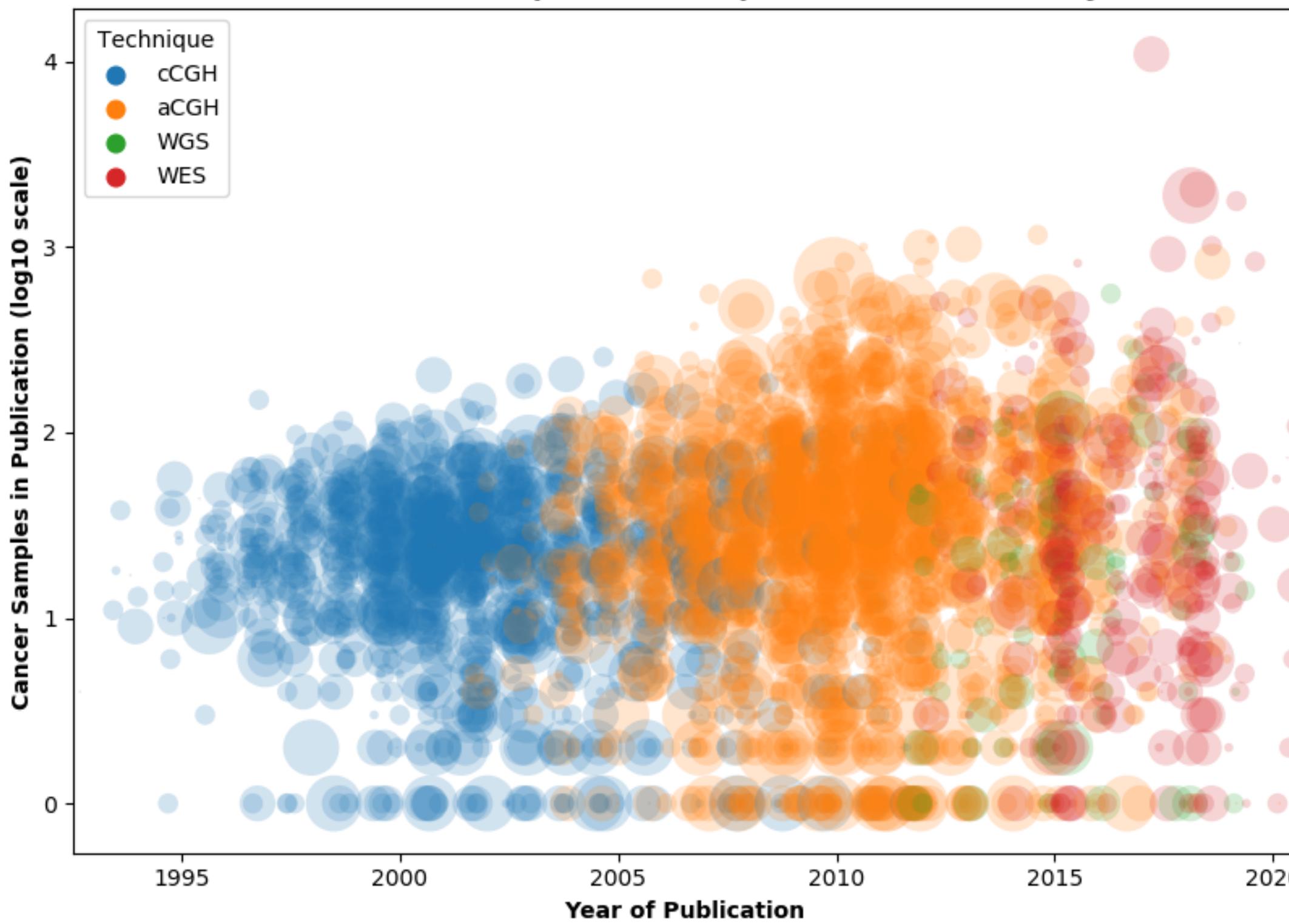


Global Alliance
for Genomics & Health



University of
Zurich UZH

Number of tumor samples for each publication across the years



Map of the geographic distribution (by first author affiliation) of the 104'543 genomic array, 36'766 chromosomal CGH and 15'409 whole genome/exome based cancer genome datasets.

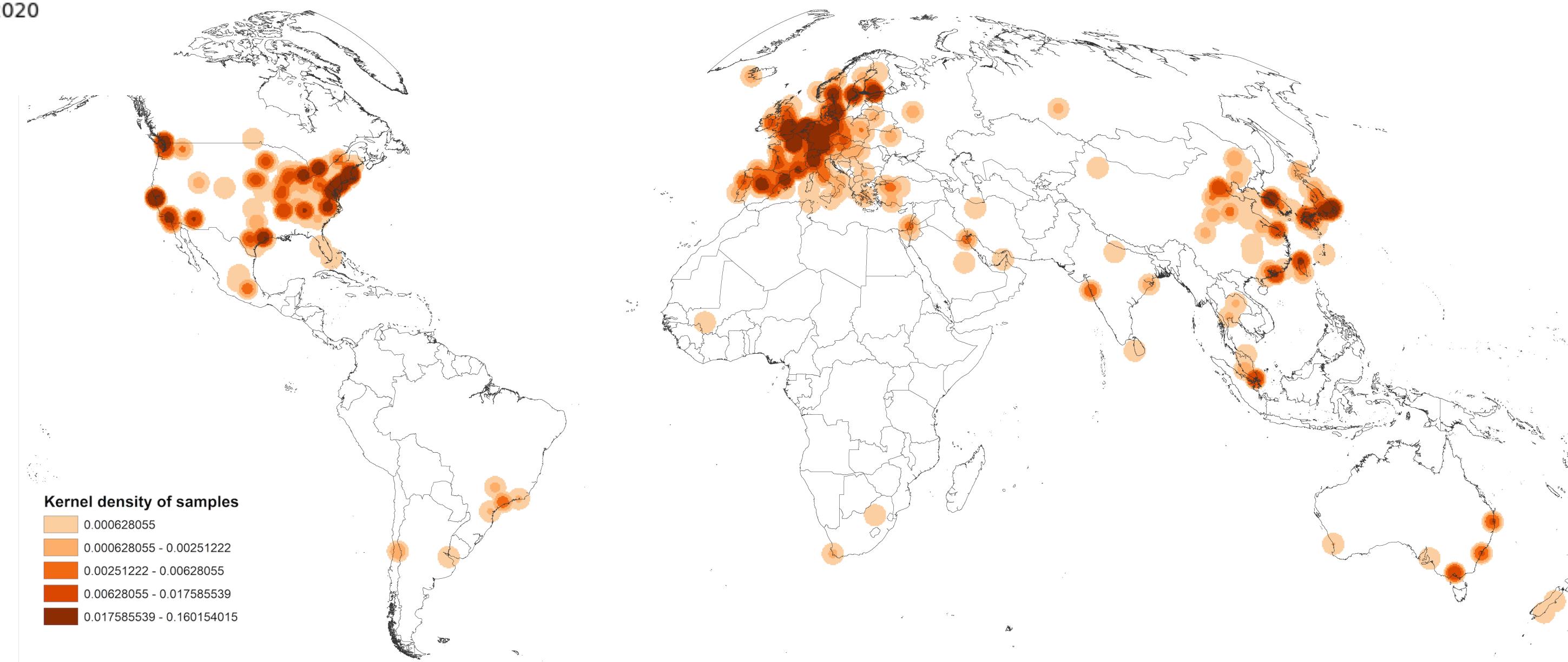
The numbers are derived from the 3'240 publications registered in the Progenetix database.



Publication Landscape of Cancer CNV Profiling

Publication statistics for cancer genome screening studies. The graphic shows our assessment of publications reporting whole-genome screening of cancer samples, using molecular detection methods (chromosomal CGH, genomic array technologies, whole exome and genome sequencing).

For the years 1993-2018, we found 3'229 publications reporting 174'530 individual samples in single series from 1 to more than 1000 samples. Y-axis and size of the dots correspond to the sample number; the color codes indicate the technology used.



The vision: Federation of data



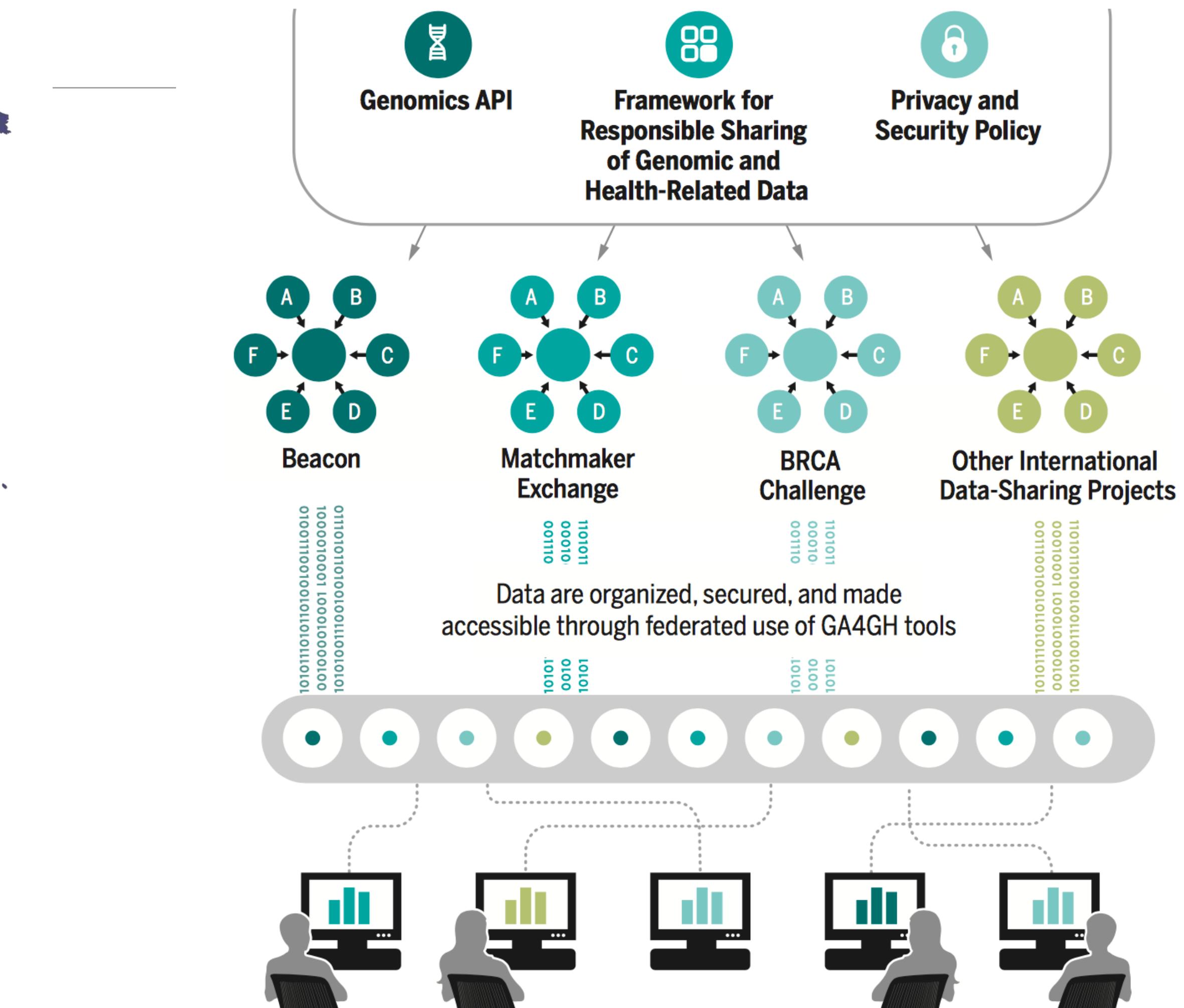


GENOMICS

A federated ecosystem for sharing genomic, clinical data

Silos of genome data collection are being transformed into seamlessly connected, independent systems

A federated data ecosystem. To share genomic data globally, this approach furthers medical research without requiring compatible data sets or compromising patient identity.



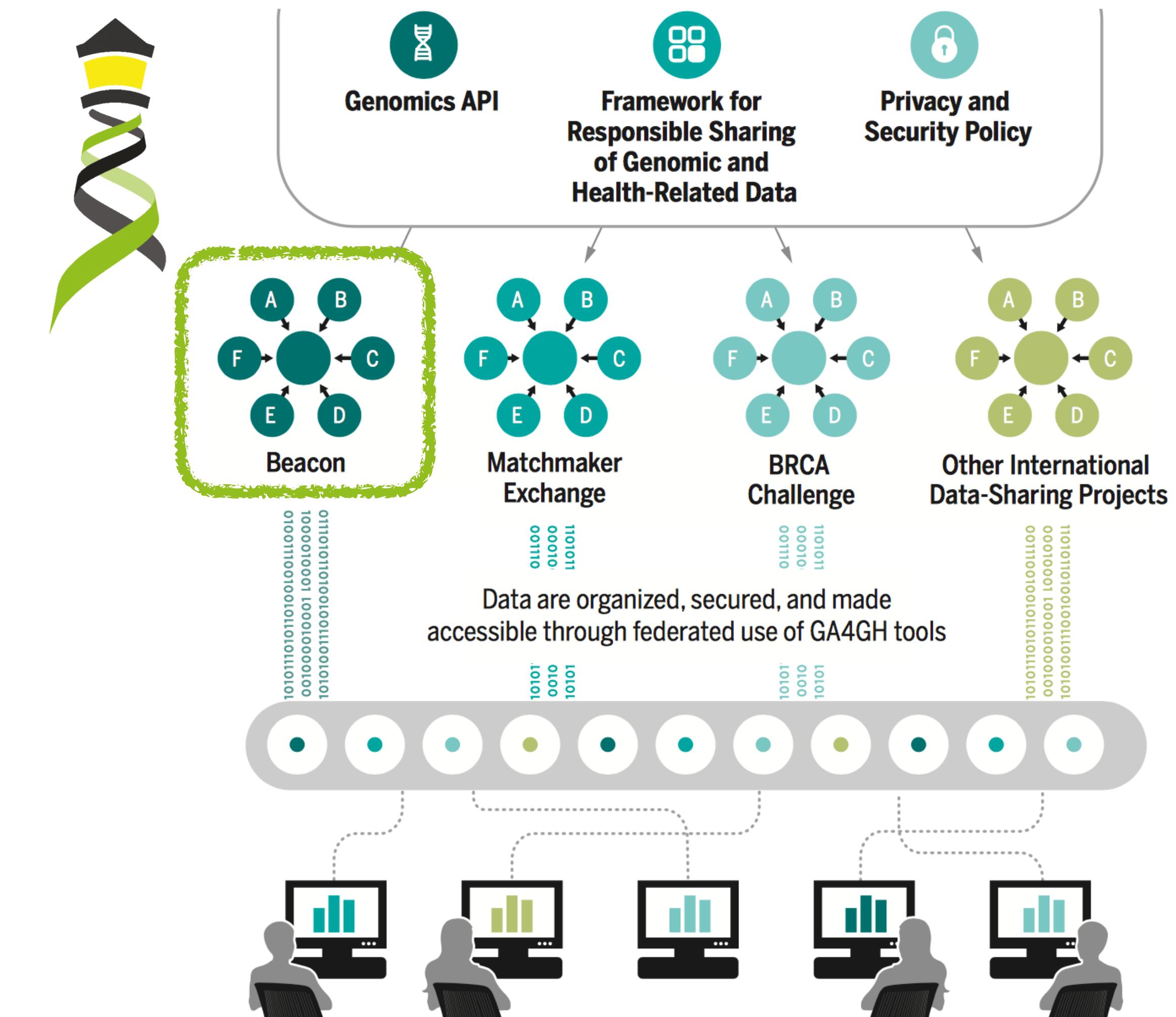


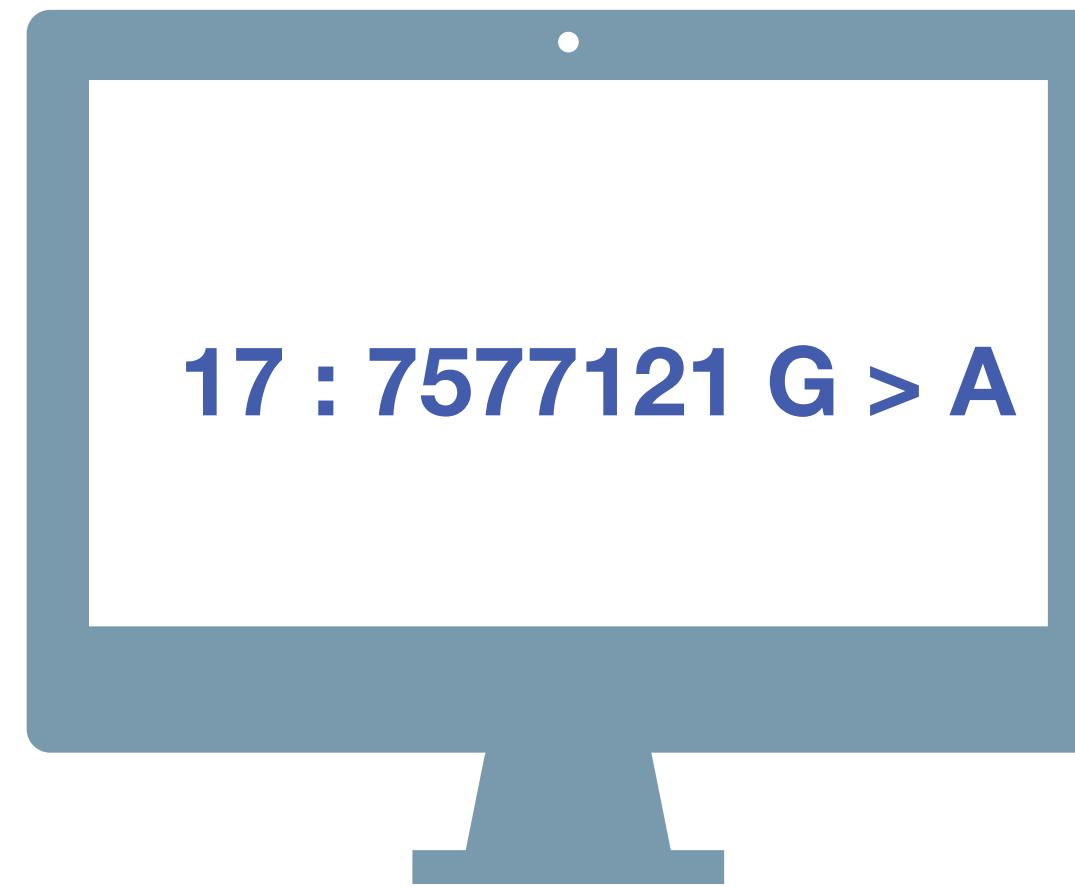
GENOMICS

A federated ecosystem for sharing genomic, clinical data

Silos of genome data collection are being transformed into seamlessly connected, independent systems

A federated data ecosystem. To share genomic data globally, this approach furthers medical research without requiring compatible data sets or compromising patient identity.





Beacon

A **Beacon** answers a query for a specific genome variant against individual or aggregate genome collections

YES | NO | \0

Introduction

... I proposed a challenge application for all those wishing to seriously engage in *international* data sharing for human genomics. ...

1. Provide a **public web service**
2. Which accepts a query of the form “Do you have **any** genomes with an “**A**” at position **100,735** on chromosome **3**?”
3. And responds with one of “**Yes**” or “**No**” ...

“Beacon” because ... people have been scanning the universe of human research for *signs of willing participants in far reaching data sharing*, but ... it has remained a **dark** and **quiet** place. The hope of this challenge is to 1) *trigger the issues* blocking groups ... in way that isn’t masked by the ... complexities of the science, fully functional interfaces, and real issues of privacy, and to 2) in *short order* ... see *real beacons of measurable signal* ... from *at least some sites* ... Once your “GABeacon” is shining, you can start to take the *next steps to add functionality* to it, and *finding the other groups* ... following their GABeacons.

Utility

Some have argued that this simple example is not “useful” so nobody would build it. Of course it is not the first priority for this application to be scientifically useful. ...intended to provide a *low bar for the first step of real ... engagement*. ... there is some utility in ...locating a rare allele in your data, ... not zero.

A number of more useful first versions have been suggested.

1. Provide *frequencies of all alleles* at that point
2. Ask for all alleles seen in a gene *region* (and more elaborate versions of this)
3. Other more complicated queries



"I would personally recommend all those be held for version 2, when the beacon becomes a service."

Jim Ostell, 2014

Implementation

1. Specifying the chromosome ... The interface needs to specify the *accession.version* of a chromosome, or *build number*...
2. Return values ... right to *refuse* to answer without it being an error ... DOS attack ... or because ...especially *sensitive*...
3. Real time response ... Some sites suggest that it would be necessary to have a “*phone home*” response ...

Beacon Project in 2016

An open web service that tests the willingness of international sites to share genetic data.



Beacon Network

Search Beacons

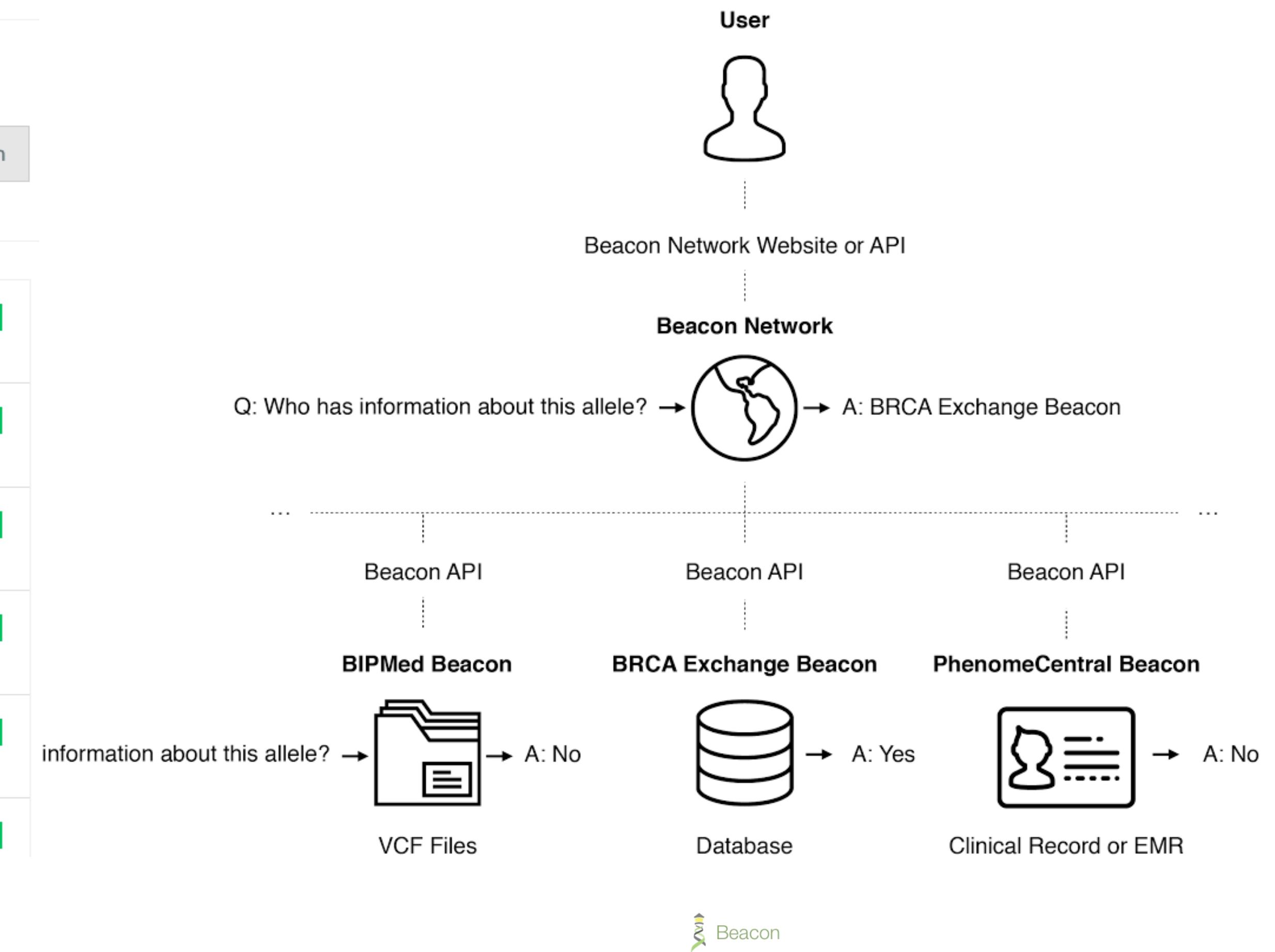
Search all beacons for allele

GRCh37 ▾ 10:118969015 C / CT Search

Response All None
 Found 16
 Not Found 27
 Not Applicable 22

Organization All None
 AMPLab, UC Berkeley
 BGI
 BioReference Laborato...
 Brazilian Initiative on ...
 BRCA Exchange
 Broad Institute
 Centre for Genomic R...
 Centro Nacional de A...
 Curoverse
 EMBL European Bio...
 Global Alliance for G...
 Google
 Institute for Systems ...
 Instituto Nacional de ...

BioReference	Hosted by BioReference Laboratories	Found
Catalogue of Somatic Mutations in Cancer	Hosted by Wellcome Trust Sanger Institute	Found
Cell Lines	Hosted by Wellcome Trust Sanger Institute	Found
Conglomerate	Hosted by Global Alliance for Genomics and Health	Found
COSMIC	Hosted by Wellcome Trust Sanger Institute	Found
dbGaP: Combined GRU Catalog and NHLBI Exome Seq...		Found



35+

Organizations

90+

Beacons

200+

Datasets

100K+

In

Releases

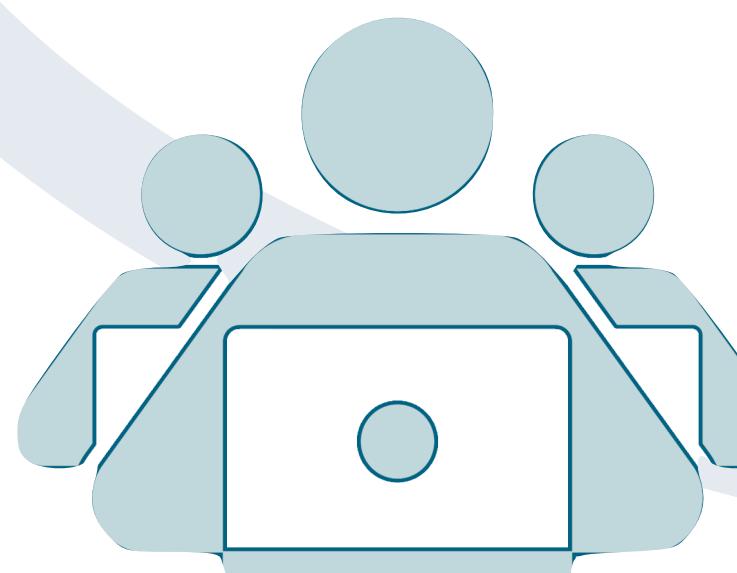
Date	Tag	Title
2018-01-24	v0.4.0	Beacon
2016-05-31	v0.3.0	Beacon

GA4GH & ELIXIR - Partnering for Standards and Tools in Genomics and Health



Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

Simplify the way people search for and request access to potentially identifiable data in international and national genomic data resources



ELIXIR - Making Beacons Biomedical



- Authentication to enable non-aggregate, patient derived datasets
 - ELIXIR AAI with compatibility to other providers (OAuth...)
 - Scoping queries through "biodata" parameters
 - Extending the queries towards clinically ubiquitous variant formats
 - cytogenetic annotations, named variants, variant effects
- Beacons as part of local, secure environments
 - local EGA ...
- Beacon queries as entry for **data delivery**
 - handover to stream and download using htsget, VCF, EHRs
- Interacting with EHR standards
 - FHIR translations for queries and handover ...

ELIXIR Beacon Project

- Driver project on GA4GH roadmap
- aligns with Discovery Work Stream
- strong impact on GA4GH developments as a concrete, funded project

The screenshot shows two cards. The left card is titled 'Driver Projects' and contains text about real-world genomic data initiatives. The right card is titled 'ELIXIR Beacon' and provides links to its implementation studies and champions.

Driver Projects
GA4GH Driver Projects are real-world genomic data initiatives that help guide our development efforts and pilot our tools. Stakeholders around the globe advocate, mandate, implement, and use our frameworks and standards in their local contexts.

ELIXIR Beacon
<https://www.elixir-europe.org/about/implementation-studies/beacons>

Europe
Champions: Jordi Rambla, Juha Tornroos, Gary Saunders

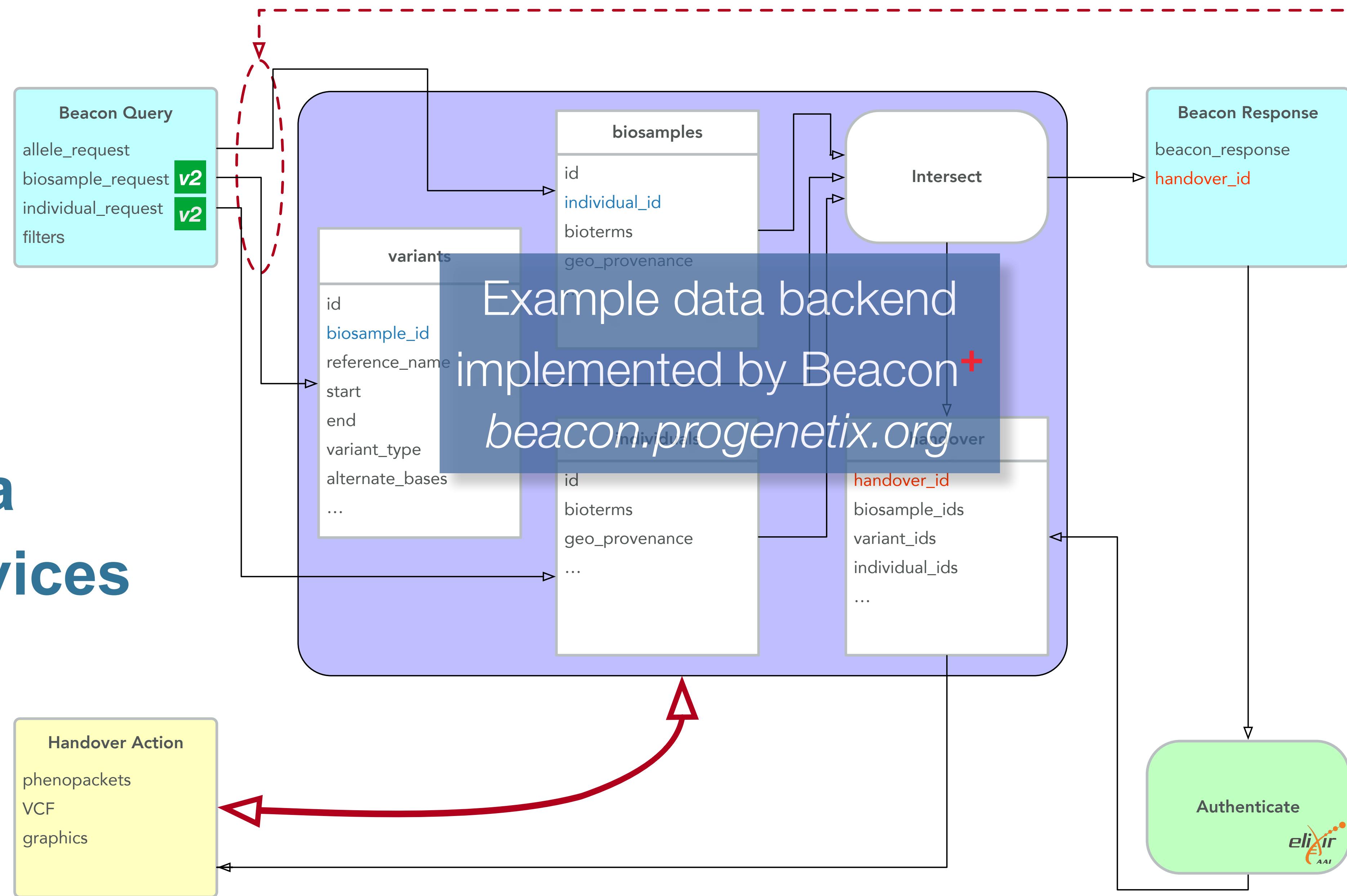
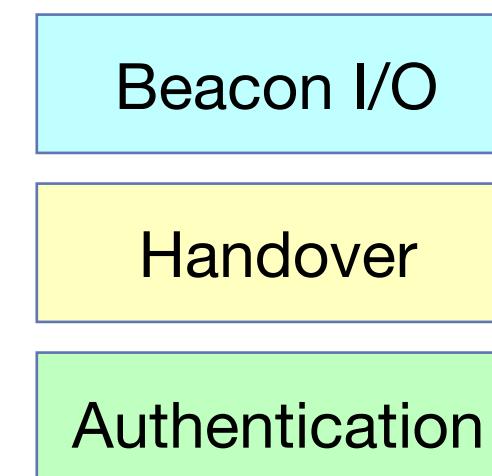
v1.1 and roadmap

- structural variations** (DUP, DEL) in addition to SNV
- ... more structural queries (translocations/fusions...)
- Beacon queries as entry for **data handover** (outside Beacon protocol)
- layered authentication system using **ELIXIR AAI**
- v2** **filters** for phenotypic & technical metadata
- v2** Extended quantitative responses
 - Ubiquitous **deployment** (e.g. throughout ELIXIR network)



Beacon & Handover

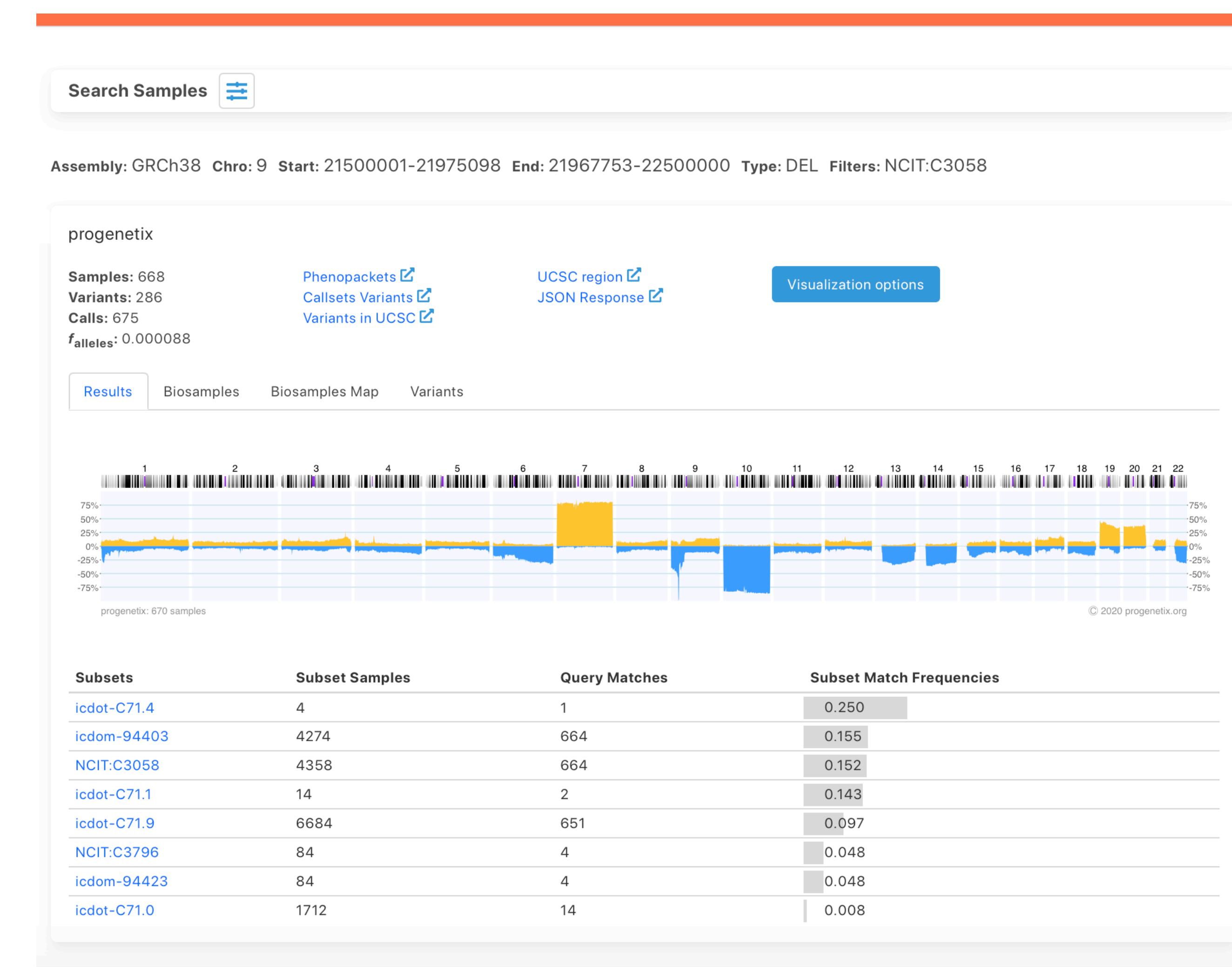
Beacons v1.1
supports data
delivery services



Beacon+ by Progenetix

From Beacon Query to Explorative Analyses of CNV Patterns

- Since 2016 the Progenetix resource has been used to model options for Beacon development
 - ▶ 138334 individual samples from 698 cancer types
- The consistent use of hierarchical diagnostic codes allows the use of Beacon "filters" for histopathological/clinically scoped queries
- Beacon's **handover** protocols can be utilized for data retrieval and, well, handing over to additional services, e.g.
 - ▶ **downloads**
 - ▶ **visualization**
 - ▶ **use of external services (UCSC browser display...)**



ELIXIR Beacon Network



- developed under lead from ELIXIR Finland
- **authenticated access** w/ ELIXIR AAI
- **incremental extension**, starting with ELIXIR Beacon resources adhering to the **latest specification** (contrast to legacy networks)
- service details provided by individual Beacons, using **GA4GH service-info**
- **registration service**
 - integrator** throughout ELIXIR Human Data
 - starting point for "**beyond ELIXIR**" **feature rich** federated Beacon services

GRCh38 ▾ 17 : 7577121 G > A

[Example variant query](#) [Advanced Search](#)

baudisgroup at UZH and SIB
Progenetix Cancer Genomics Beacon+

Beacon+ provides a forward looking implementation of the Beacon API, with focus on structural variants and metadata based on the cancer and reference genome profiling data represented in the Progenetix oncogenomic data resource (<https://progenetix.org>).

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

National Bioinformatics Infrastructure Sweden
SweFreq Beacon

Beacon API Web Server based on the GA4GH Beacon API

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

LCSB at University of Luxembourg
ELIXIR.LU Beacon

ELIXIR.LU Beacon

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

Research Programme on Biomedical Informatics
DisGeNET Beacon

Variant-Disease associations collected from curated resources and the literature

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

European Genome-Phenome Archive (EGA)
EGA Beacon

This [Beacon](https://beacon-project.io/) is based on the GA4GH Beacon [v1.1.0](https://github.com/ga4gh/beacon/specification/blob/develop/beacon.yaml)

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

University of Tartu Institute of Genomics, Estonia
Beacon at the University of Tartu, Estonia

Beacon API Web Server based on the GA4GH Beacon API

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

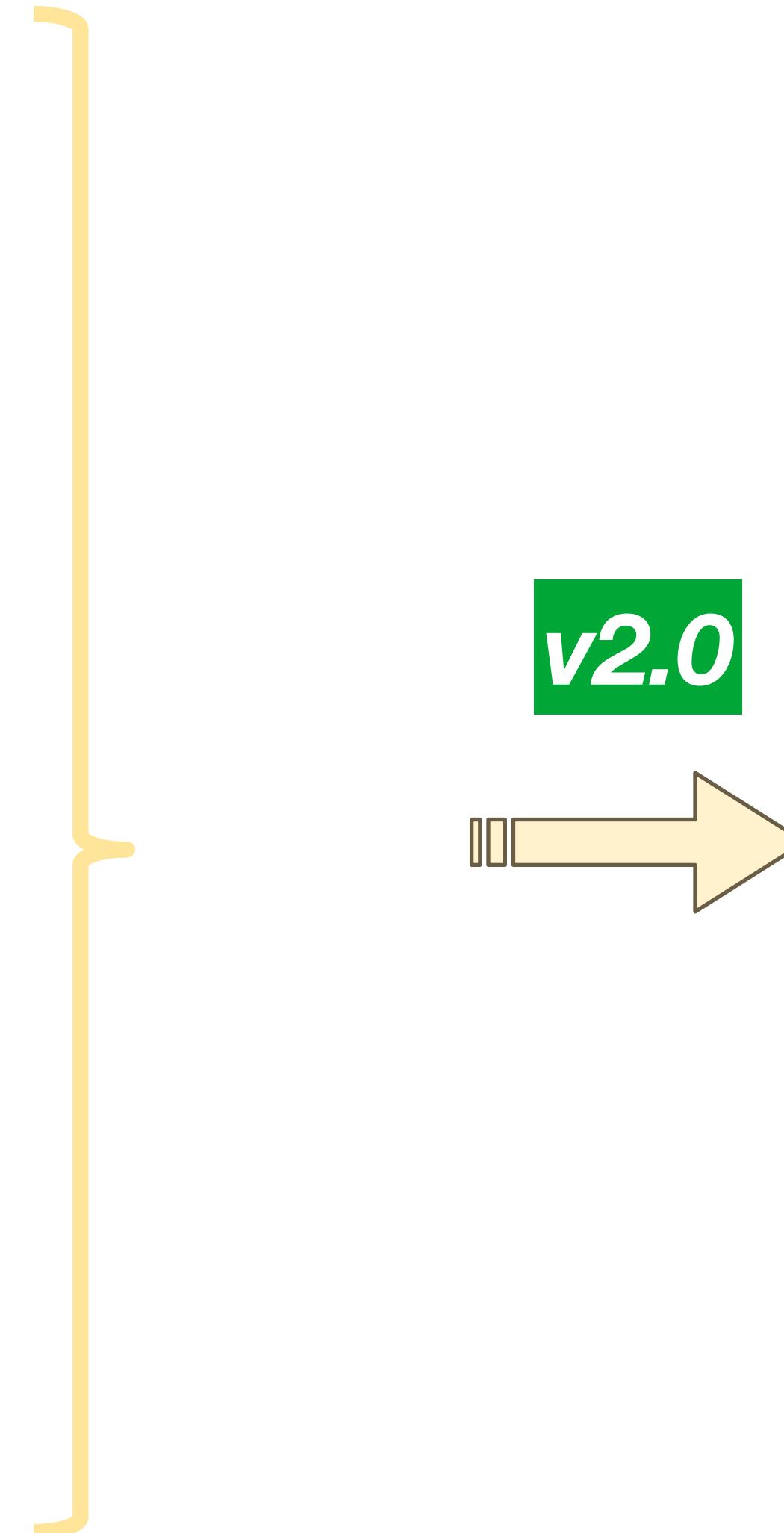
CSC - IT Center for Science Production Beacon

Beacon API Web Server based on the GA4GH Beacon API

[Visit Us](#) · [Beacon API](#) · [Contact Us](#)

Beacon v2 - Areas of Change in 2021

- Specifying query types for multiple types of genomic variants
- Access levels
- Filters
 - Simple general filter schema w/ **scoping through prefixes** (CURIEs, private implementations)
- New types of queries:
 - By sample, patient, variant effect/evidence
 - Complex queries? (stakeholder driven; e.g. EJP-RD, GEL...)
- Schema versions & Service Info
 - Negotiated queries based on individual Beacon capabilities
- Cohorts
- ...



GA4GH approval
process
("major product
update")

Beacon v2 - Clinical Beacon requirements

Authors: Jordi Rambla, Michael Baudis, Anthony J Brookes, Lauren Fromont, Claudia Vasallo, Aina Jené

The original GA4GH Beacon implementation (up to v0.3) was conceived as a protocol for sharing 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). Although with some potential benefit, e.g. in the area of rare disease diagnostics, it was *not* designed for clinical use but chiefly to foster data sharing by triggering the inquisitiveness of researchers once some data of interest is discovered in another institution. While later extensions of the protocol (v1.0 - v1.n) expanded the query and response options, this did not deviate from the general "existence of variants in resource X" paradigm.

The simplicity and success of the concept has generated the request of making it more powerful, more useful in healthcare environments. The requests include:

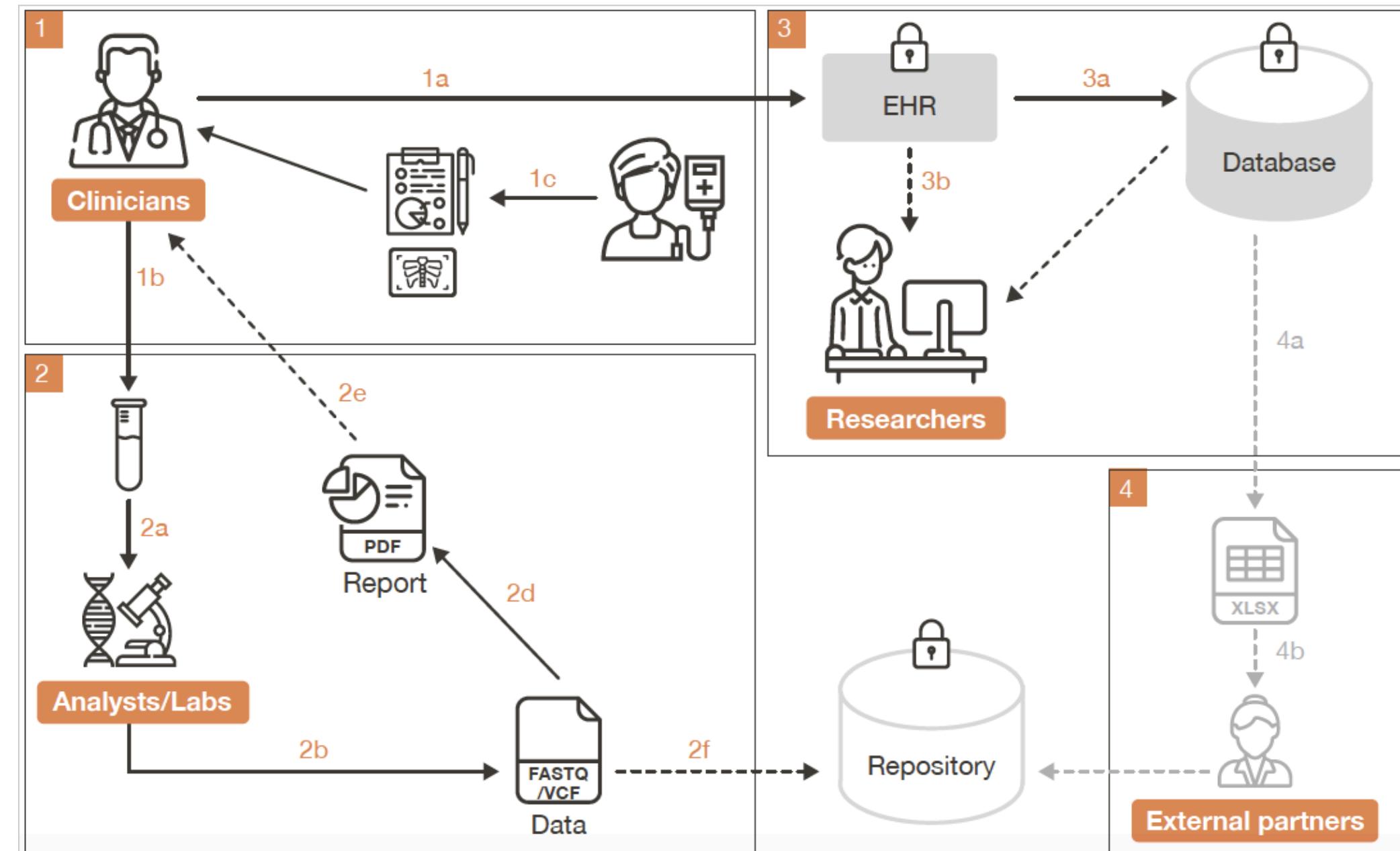
- Allowing more informative queries, like filtering by gender or age
- Allowing to trigger the next step in the data access process, e.g. who to contact or which are the data use conditions
- Jumping to another system where the data could be accessed, e.g. if the Beacon is internal to a hospital, to provide the Id of the EHR of the patients having the mutation of interest.
- Including 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.

The process

The GA4GH Beacon group started a set of meetings and interviews with GA4GH Driver Projects and with ELIXIR partners in order to determine the scope of the next generation Beacon. The goal was to be useful without breaking the simplicity that made Beacon version 1 successful.

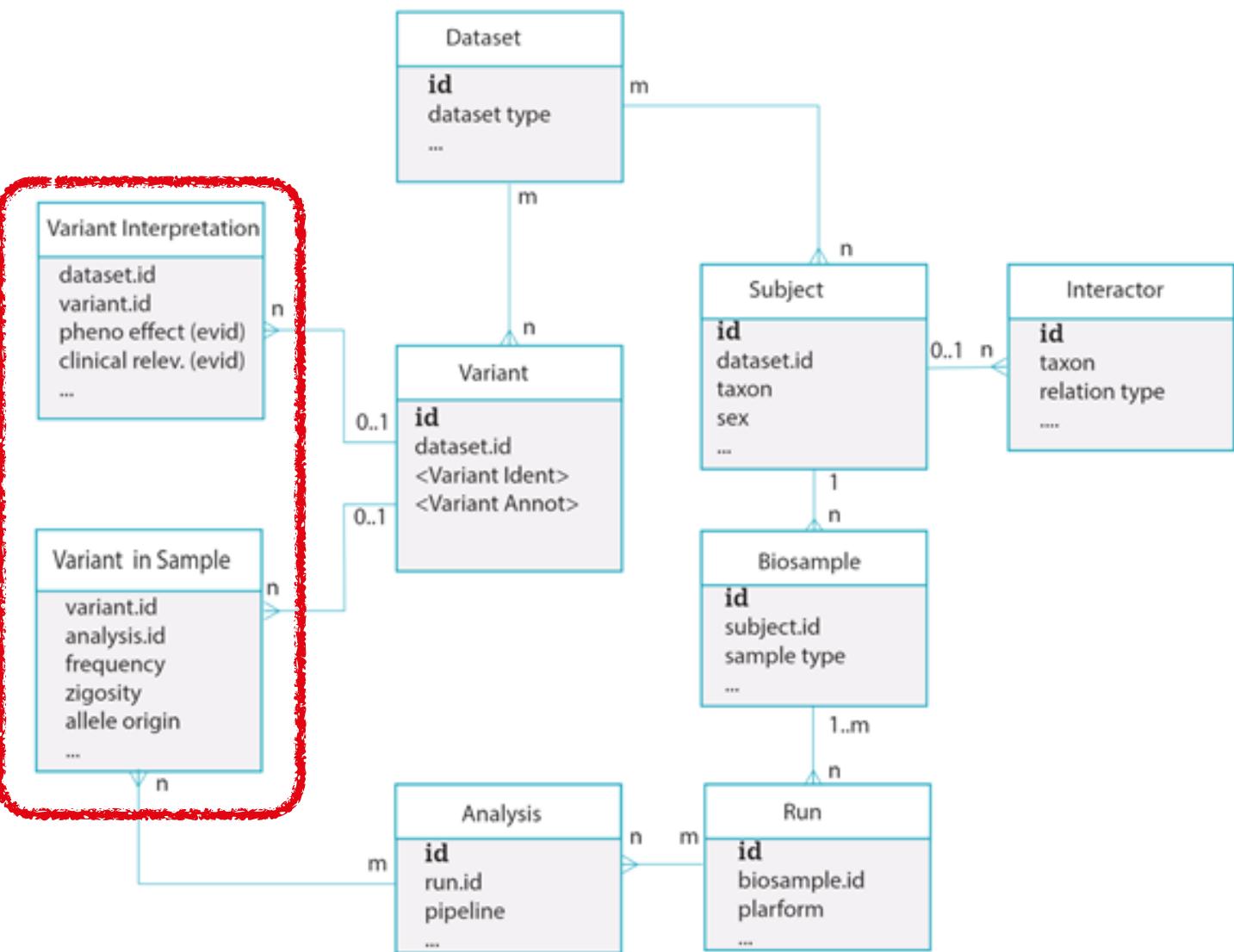
Interviews were conducted with the following GA4GH Driver Projects:

- Autism Speaks
- BRCA Exchange
- CanDIG
- EGA, ENA, EVA
- EuCanCancer
- European Joint Programme - Rare Diseases
- H3Africa
- GEM Japan
- Genomics England
- Matchmaker Exchange
- SVIP /SPHN
- VICC



Data flow and data sharing of genomic and phenotypic data in healthcare

The Beacon v2 draft 2 entities logical model (implementations may vary)



Beacon v2 Filters

Example: Use of hierarchical classification systems (here NCI ICD neoplasm core)

- Beacon v2 "filters" assumes inclusion of child terms when using hierarchical classifications
 - implicit *OR* with otherwise assumed *AND*
 - implementation of hierarchical annotation overcomes some limitations of "fuzzy" disease annotations



Beacon+ specific: Multiple term selection with OR logic

<input checked="" type="checkbox"/>	> NCIT:C4914: Skin Carcinoma	213
<input type="checkbox"/>	> NCIT:C4475: Dermal Neoplasm	109
<input checked="" type="checkbox"/>	▼ NCIT:C45240: Cutaneous Hematopoietic and Lymphoid Cell Neoplasm	310

Filters: NCIT:C4914, NCIT:C4819, NCIT:C9231, NCIT:C2921, NCIT:C45240, NCIT:C6858, NCIT:C3467, NCIT:C45340, NCIT:C7195, NCIT:C3246, NCIT:C7217



progenetix

Variants: 0 *f*alleles: 0 Callsets Variants ↗ UCSC region ↗ Calls: 0 Legacy Interface ↗ Samples: 523 [Show JSON Response](#)

Results **Biosamples**

Id	Description	Classifications	Identifiers	DEL	DUP	CNV
PGX_AM_BS_MCC01	Merkel cell carcinoma	icdot-C44.9 Skin, NOS icdom-82473 Merkel cell carcinoma NCIT:C9231 Merkel Cell Carcinoma	PMID:9537255	0.116	0.104	0.22
PGX_AM_BS_MCC02	Merkel cell carcinoma	icdot-C44.9 Skin, NOS icdom-82473 Merkel cell carcinoma NCIT:C9231 Merkel Cell Carcinoma	PMID:9537255	0.154	0.056	0.21
PGX_AM_BS_MCC03	Merkel cell carcinoma	icdot-C44.9 Skin, NOS icdom-82473 Merkel cell carcinoma NCIT:C9231 Merkel Cell Carcinoma	PMID:9537255	0.137	0.21	0.347
PGX_AM_BS_MCC04	Merkel cell carcinoma	icdot-C44.9 Skin, NOS icdom-82473 Merkel cell carcinoma NCIT:C9231 Merkel Cell Carcinoma	PMID:9537255	0.158	0.056	0.214
PGX_AM_BS_MCC05	Merkel cell carcinoma	icdot-C44.9 Skin, NOS icdom-82473 Merkel cell carcinoma NCIT:C9231 Merkel Cell Carcinoma	PMID:9537255	0.107	0.327	0.434



Beacon Scouts: Structural Variants

Re-defining & scoping variant queries

- contributors from different "stakeholder" areas
 - clinical genomics / rare diseases
 - variant repository (Ensembl)
 - cancer research resource
 - cancer variant annotation repositories
- close integration with ELIXIR h-CNV group
- process involved discussions about semantics of variant types, e.g.
 - DUP as CNV or in place
 - DEL as CNV from which size
- general attempt to use Sequence Ontology classes as guidance, but no still ambiguities / lack of terms

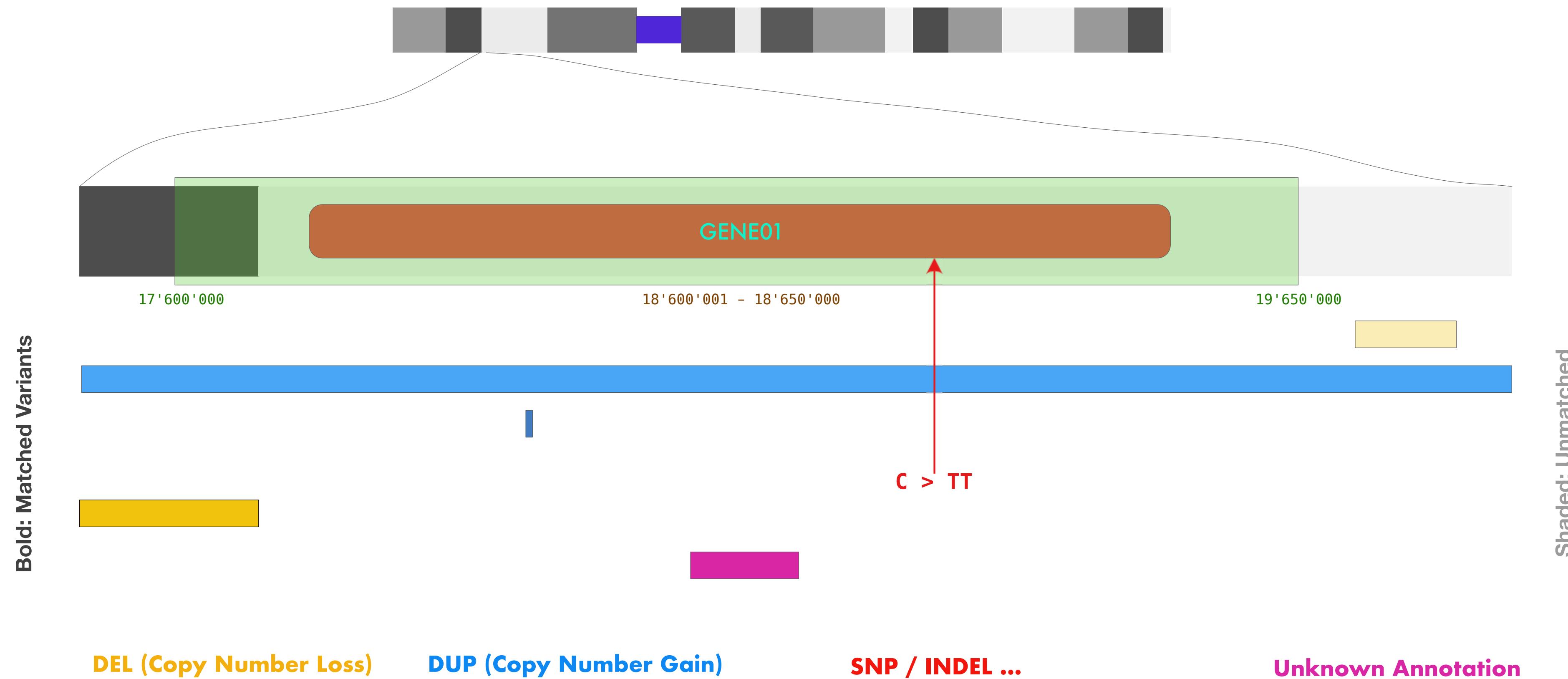
Beacon Scouts: Structural Variants Use Cases & Examples

This document develops a set of structural variant types and associated query formats which will be supported by the Beacon protocol. The focus of the initial development is on the possibly limited, but unambiguous definition of query formats, driven and documented through real-world use cases.

References	1
Conventions Followed in the Document	1
Use of Positional Parameters	2
Variant Types, Documentation and Example Queries	2
INS (Insertion)	2
DEL (Deletion)	3
DUP (Duplication)	5
Amp (DUP more than 2) CN type of approach	8
LOH (loss of representation of second allele, with or without copy number change)	8
INV (inversion)	8
TL (Translocation)	9
Proposal: BRK (Breakpoint)	9
ME (Mobile elements insertion /deletion)	10
CNV - (non directional CNVs) - do we allow cnv queries? / complex CNVs	10
Tandem Duplication	11
Name Based Queries	11
Topics for discussion	11
Technical considerations	12
Format of GET queries	12

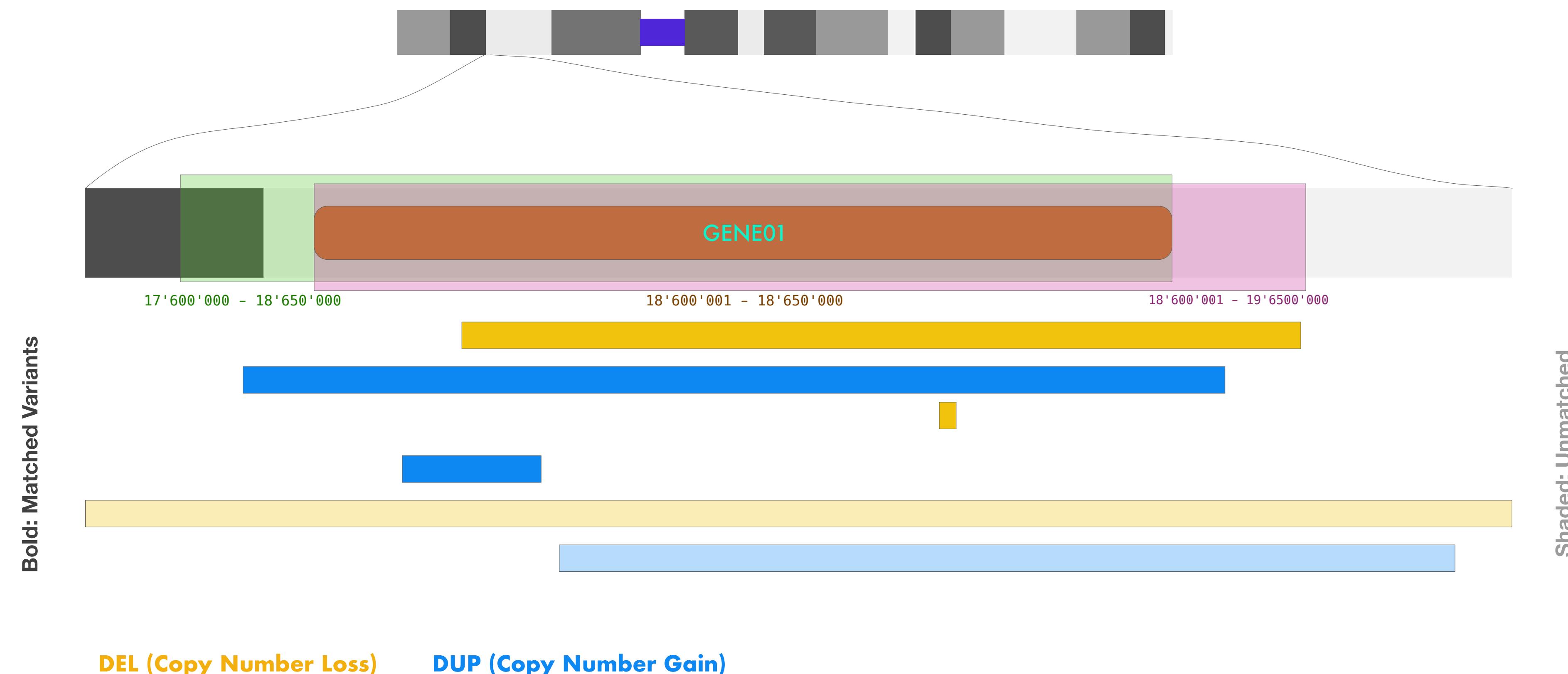
Beacon Range Query: Fishing for Variants

Example: Matching any variant, of any type, touching a gene + neighborhood



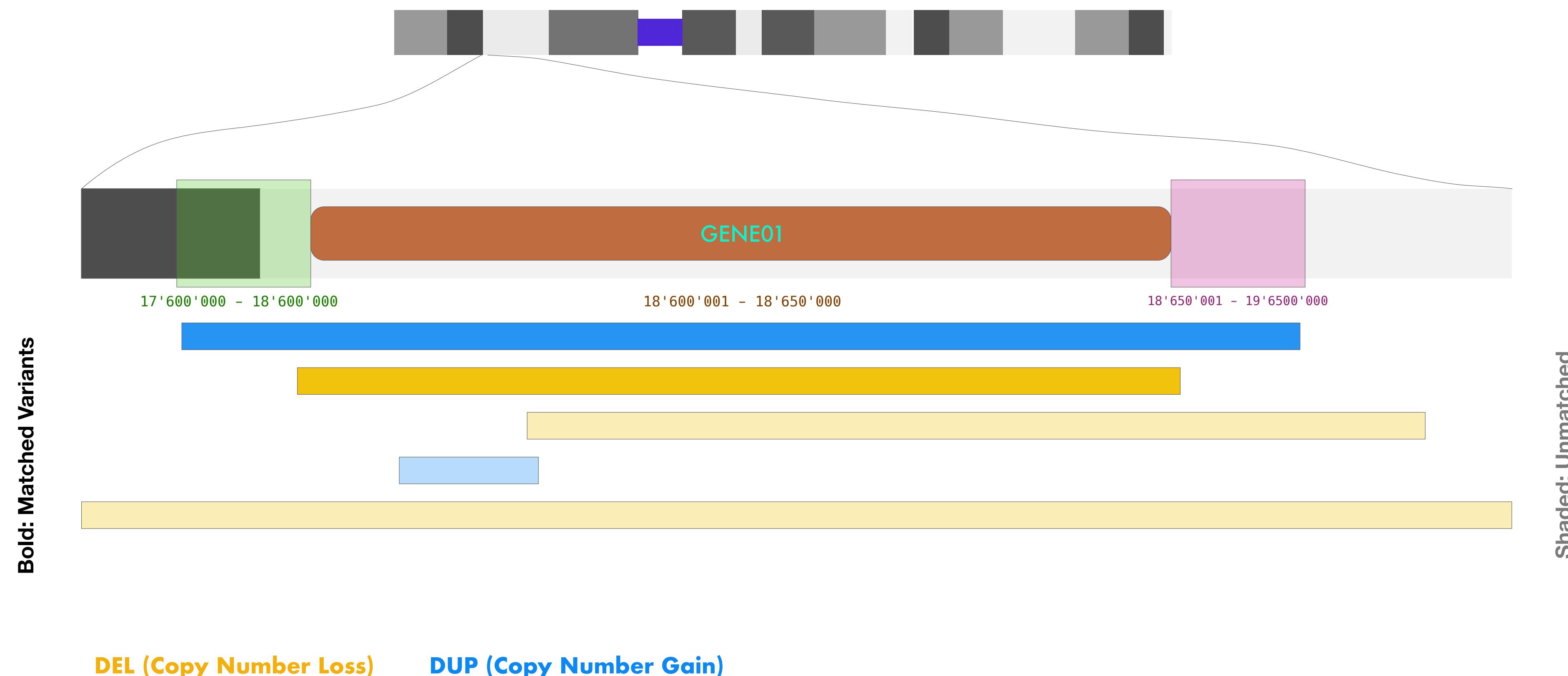
Beacon Bracket Query: Fuzzy Start and End Interval Matching

Example: Requiring *any overlap* with a region, with *implied size limit*



Beacon Bracket Query: Fuzzy Start and End Interval Matching

Example: Requiring *complete coverage* of a region, with *implied size limit*



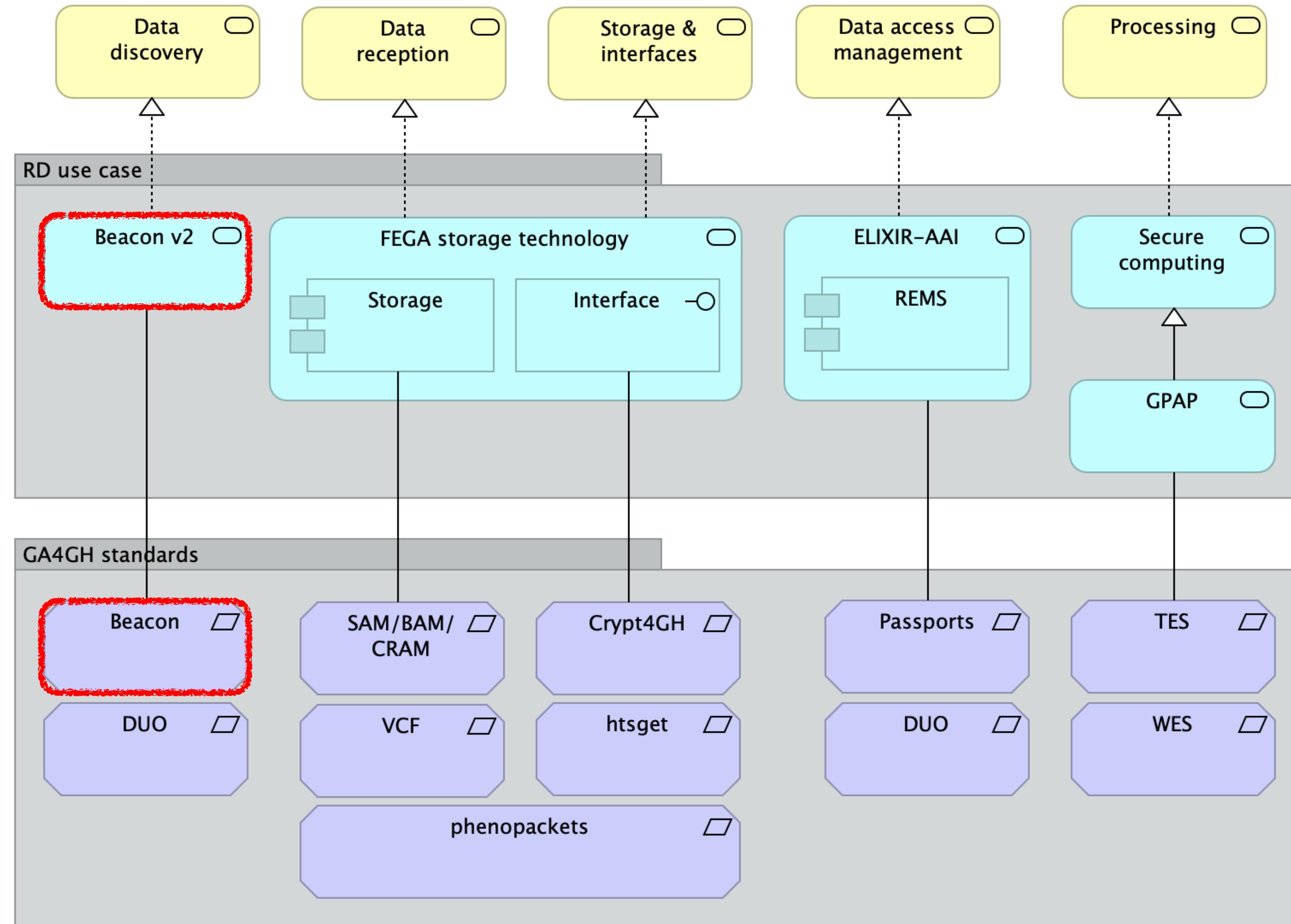


Beacon Project - Partner Engagement & Next Steps

- Working with partner communities & projects on *deploying Beacons*
 - ELIXIR hCNV Community
 - European Joint Program on Rare Diseases
 - clinical groups & data initiatives (e.g. Andalucia, Cancer Core Europe, SPHN)
 - variant annotation resources, with optional clinical components (e.g. SVIP-O)
- Improving reference implementation and standards / compliance testing
- Beacon v2 "fast forward" development
- aligning w/ GA4GH standards, through "request & adopt" => SchemaBlocks {S}[B]
- networks throughout & beyond ELIXIR



Functionalities, Standards, Components



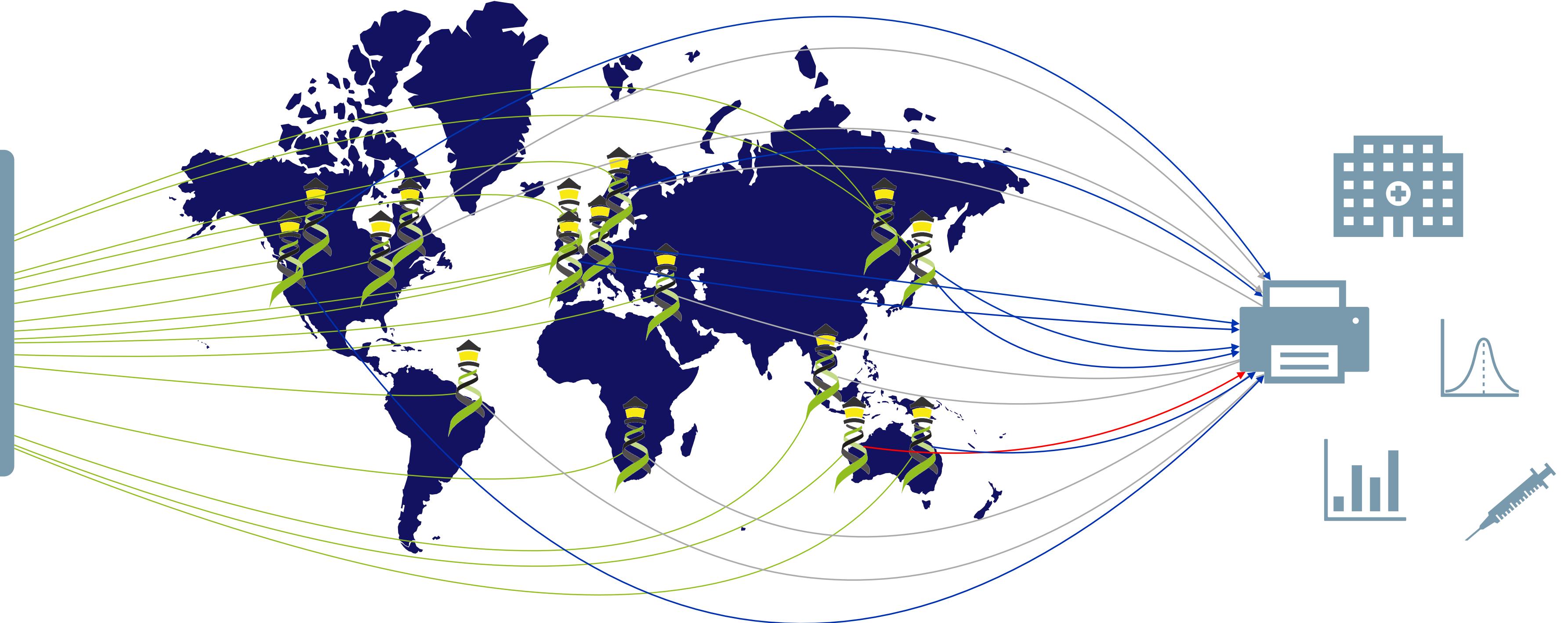
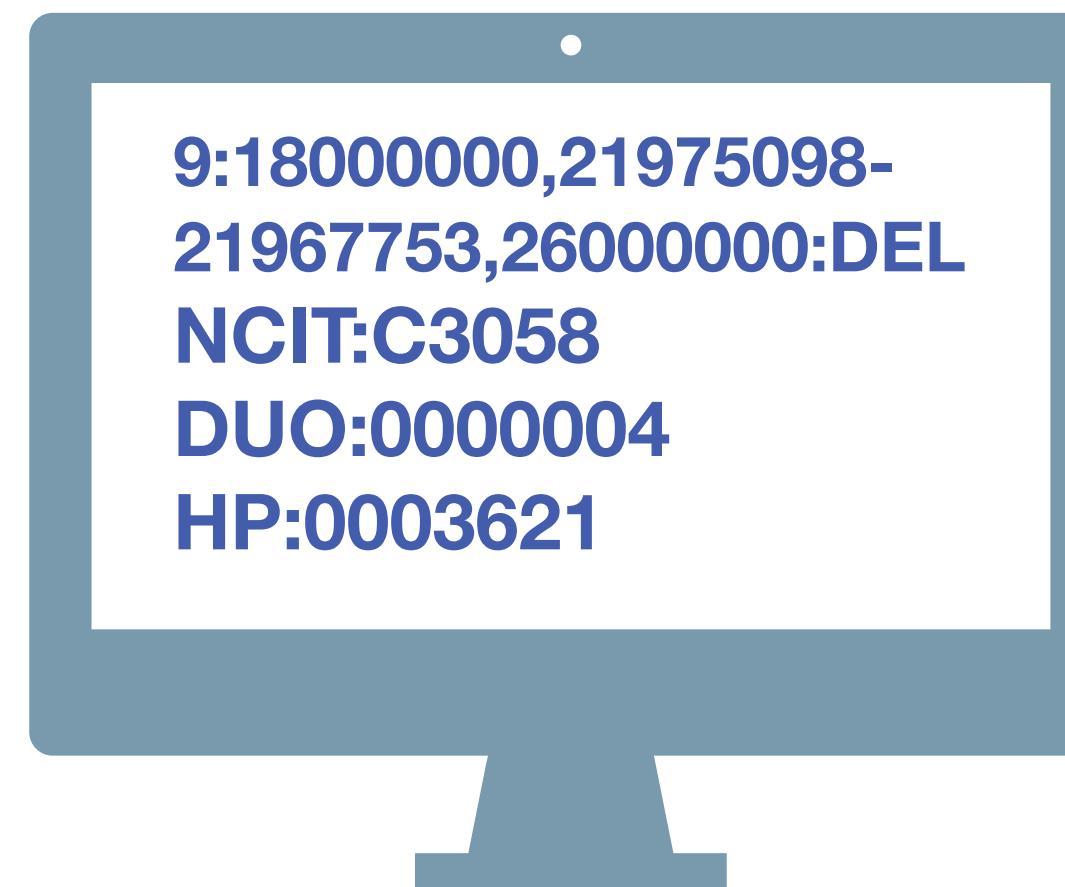
Beacon v2 Status & ToDo

Race to the Standard Submission Finish Line...



- while Beacon v1 had been accepted as GA4GH standard in 2016, the large extensions in v2 require a new approval process
- currently (April 2021) working on last issues
 - ▶ finalizing schema, query model
 - ▶ documentation
 - ▶ updated reference implementation
 - ▶ test-driving multiple independent implementations
- submission by end of June to GA4GH review process
- approval of Beacon v2 (?!?) at GA4GH plenary September 2021

Planned "Beacon Service Support" through ELIXIR
* implementations
* network

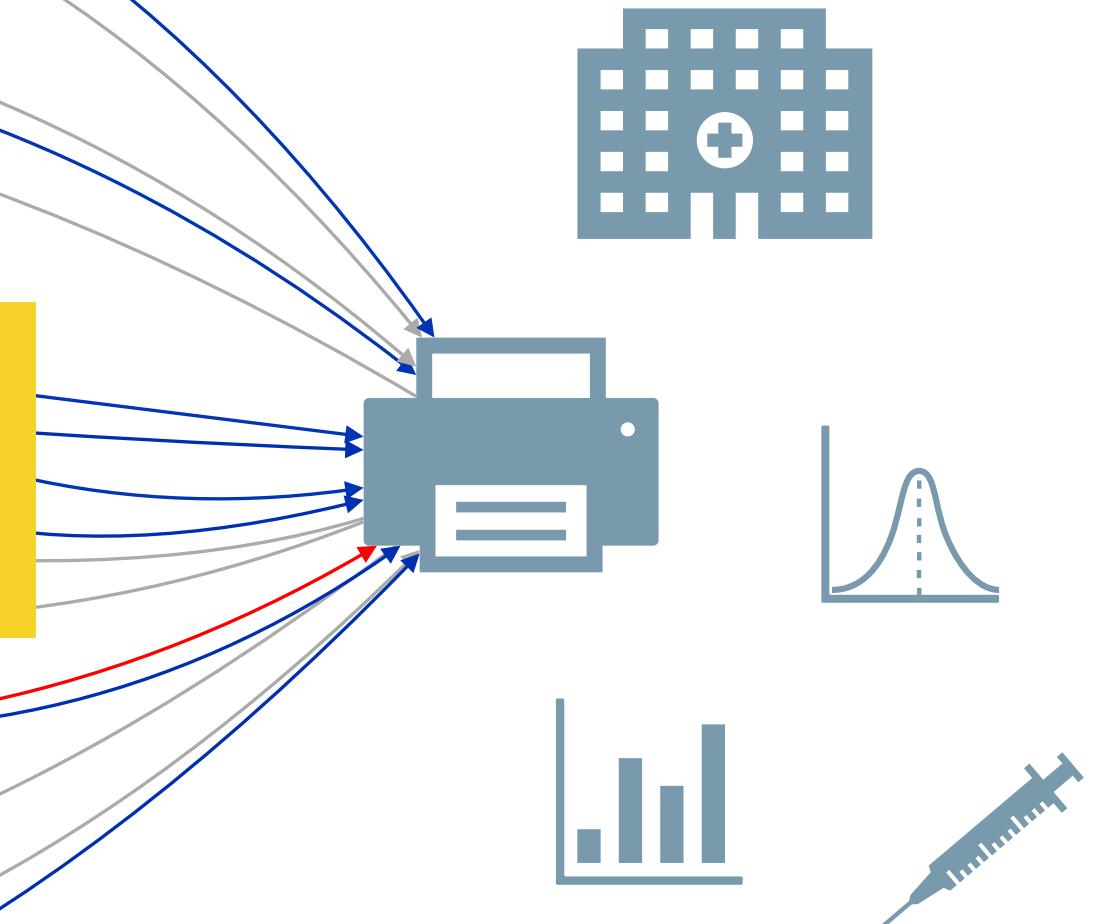
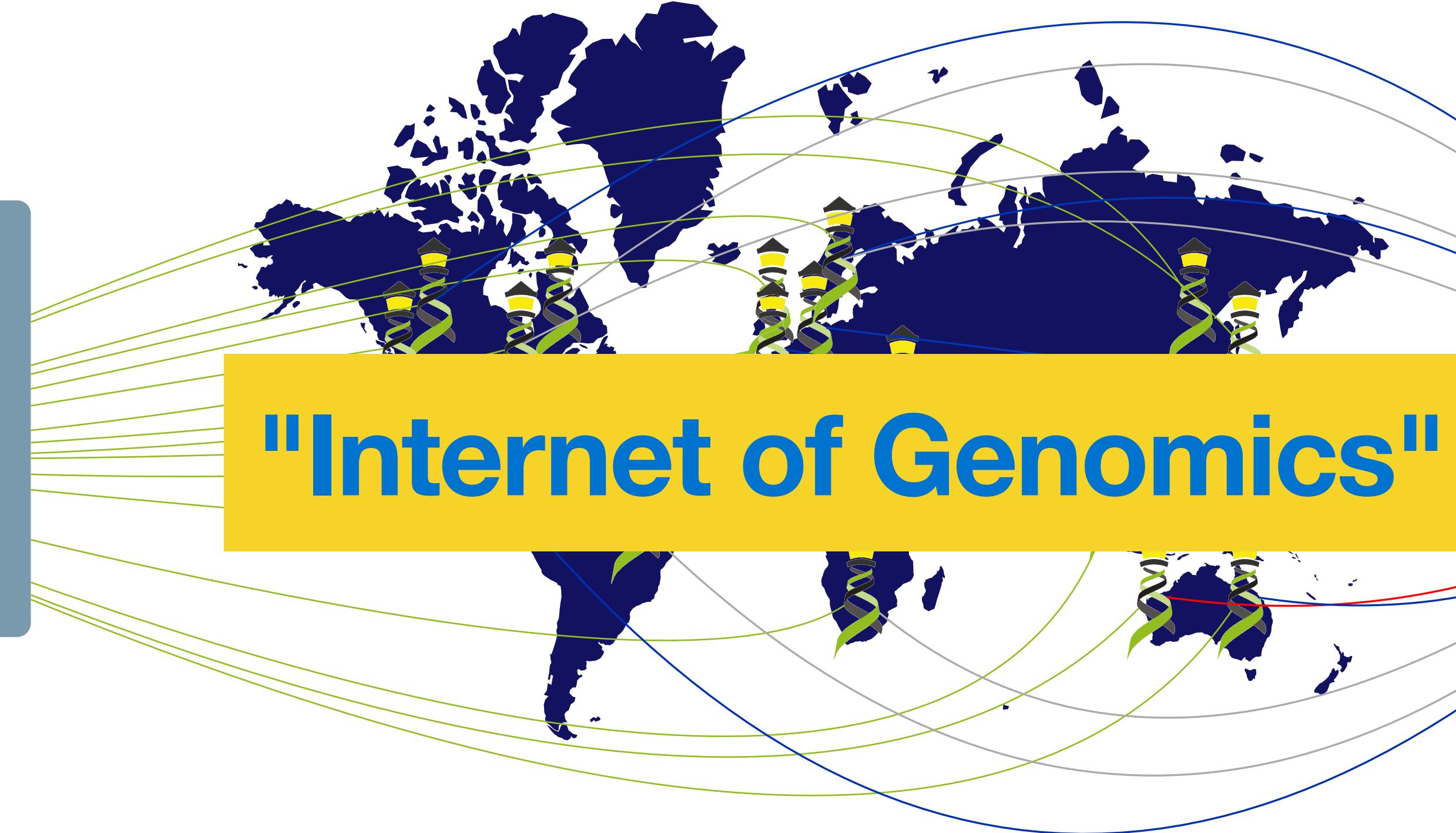
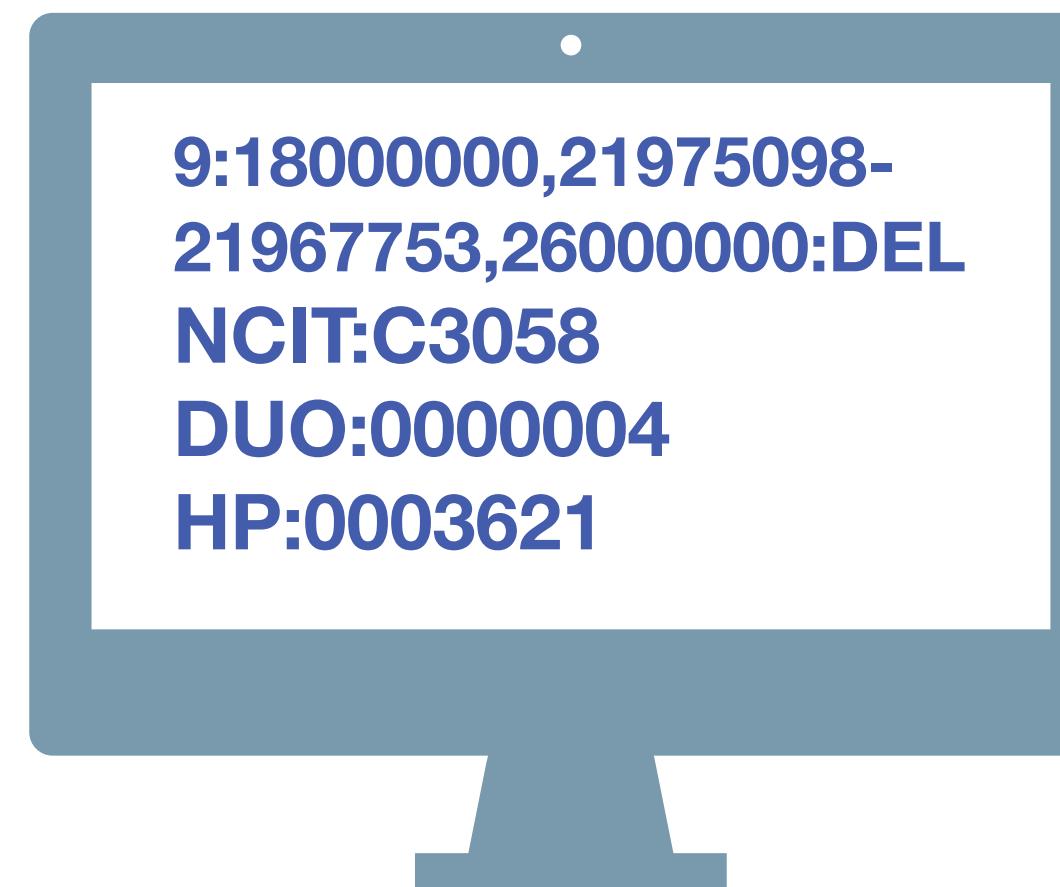


Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?

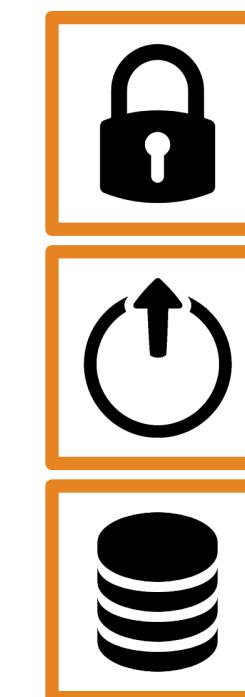


Beacon v2 API

The Beacon API v2 proposal opens the way for the design of a simple but powerful "genomics API".



Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?



Beacon v2 API

The Beacon API v2 proposal opens the way for the design of a simple but powerful "genomics API".



GA4GH Genome Beacons

A Driver Project of the Global Alliance for Genomics and Health GA4GH and supported through ELIXIR

News
Specification & Roadmap
Beacon Networks
Events

Examples, Guides & FAQ
Contributors & Teams
Contacts
Meeting Minutes

Related Sites

ELIXIR BeaconNetwork
Beacon @ ELIXIR
GA4GH
beacon-network.org
Beacon+
GA4GH::SchemaBlocks
GA4GH::Discovery

Github Projects

Beacon API and Tools
SchemaBlocks
Tags
CNV EB FAQ SV VCF beacon clinical
code compliance contacts definitions
developers development events filters
minutes network press proposal
queries releases roadmap
specification teams v2 versions
website

Beacon Protocol for Genomic Data Sharing

Beacons provide discovery services for genomic data using the Beacon API developed by the Global Alliance for Genomics and Health (GA4GH). The Beacon protocol itself standard for genomics data discovery. It provides a framework for public web service against genomic data collections, for instance from population based or disease specific repositories.



Baudisgroup @ UZH

Michael Baudis
Paula Carrio Cordo
Bo Gao
Qingyao Huang
Sofia Pfund
Rahel Paloots

Pierre-Henri Toussaint

Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?



Beacon v2 API

beacon-project.io



Beacon+

About Progenetix Help

Search Samples

CNV Request Allele Request Range Query All Fields

CNV Example

This example shows the query for CNV requests with at least a single base, but limited to arrayMap collection and can be modified

This query type is for copy number queries end positions to capture a set of similar

Gene Spans Cytoband(s)

dataset

arraymap

Genome Assembly

GRCh38 / hg38

Reference name

9

Start or Position

21000001-21975098

Cancer Classification(s)

NCIT:C3058: Glioblastoma (2119)

City

Select...

21000001_21975098



Query Beacon

progenetix.org

ga4gh-beacon / specification-v2

Code Issues 14 Pull requests Actions Wiki Security Insights ...

Go to file Add file Clone

Commits 52474fc ... 46 commits 1 branch 0 tags

Create .gitignore 2 months ago

Do not check for empty response 6 months ago

Add license 6 months ago

fix typo beacon-project website last month

Fix error (#18) last month

Readme

Apache-2.0 License

Releases

No releases published Create a new release

PI specification build pass

Sabela de la Torre Pernas

Push your first package

Contributors 3

sdelatorrep sdelatorrep mbaudis mbaudis blankdots blankdots

This repository defines an open standard for genomics data discovery, members of the Global Alliance for Genomics & Health. It work for public web services responding to queries against actions, for instance from population based or disease specific genome repositories.

This repository contains the specification for the v2 major version upgrade of the Beacon API. It is now (2020) under active development and has not seen a stable code release.

For further information, please follow the work here and consult the [Beacon Project website](#).

github.com/ga4gh-beacon/