



**Prof. Dr.
Michael Baudis**

Department of Molecular Life Sciences,
University of Zurich
Swiss Institute of Bioinformatics

A project of



BioMedIT Project

SPHN Webinar Series

Beacon v2 – Towards flexible use and clinical applications for a reference genomic data protocol

2020-11-04



1992



Heidelberg

Student of medicine | doctoral thesis in molecular cytogenetics @ DKFZ (Peter Licher) | resident in clinical hematology/oncology | data, clinical studies & cancer systematics

2001



Stanford

Post-doc in hemato-pathology (Michael Cleary) | molecular mechanisms of leukemogenesis | transgenic models | expression arrays | systematic cancer genome data collection | *Progenetix* website

2003



Gainesville

Assistant professor in paediatric haematology | molecular mechanisms of leukemogenesis | focus on bioinformatics for cancer genome data analysis

2006



Aachen

Research group leader in genetics | genomic array analysis for germline alterations | descriptive analysis of copy number aberration patterns in cancer entities

2007



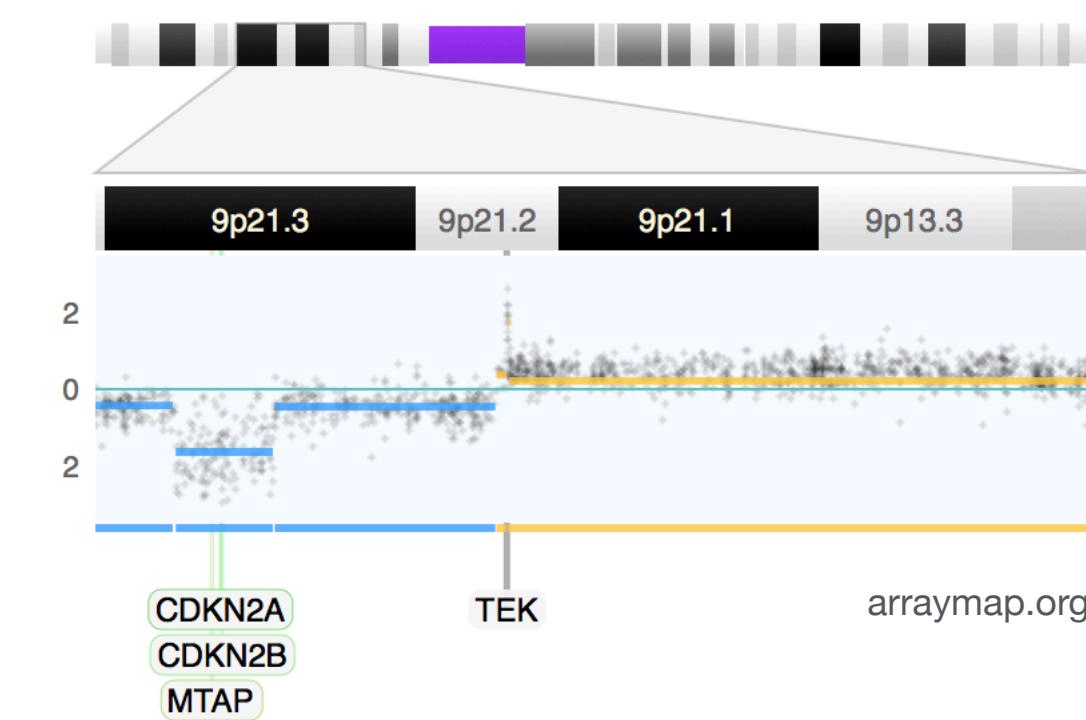
Zürich

Professor of bioinformatics @ IMLS (2015) | systematic assembly of oncogenomic data | databases and software tools | patterns in cancer genomes | *Progenetix* & *arrayMap* resources | GA4GH | SPHN

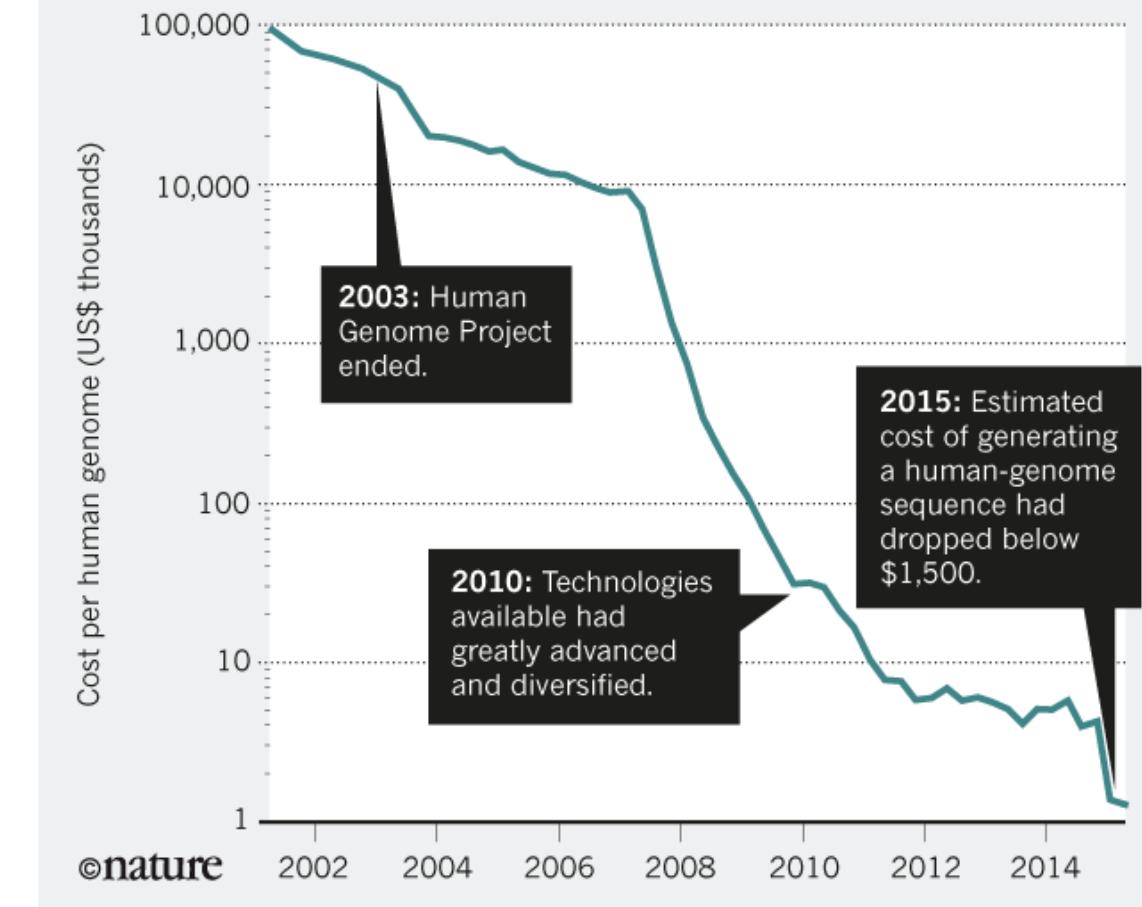


Genome screening at the core of “Personalised Health”

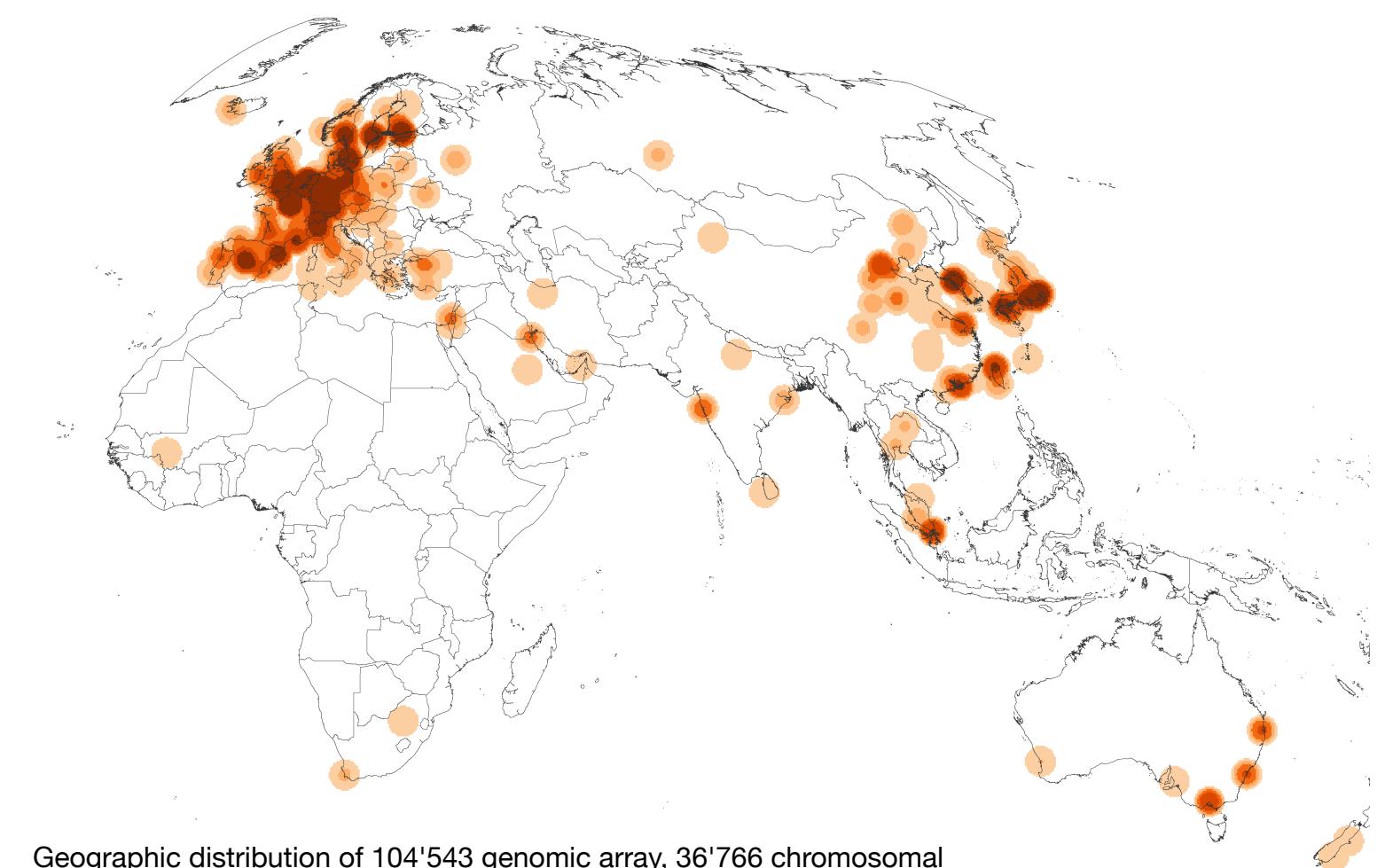
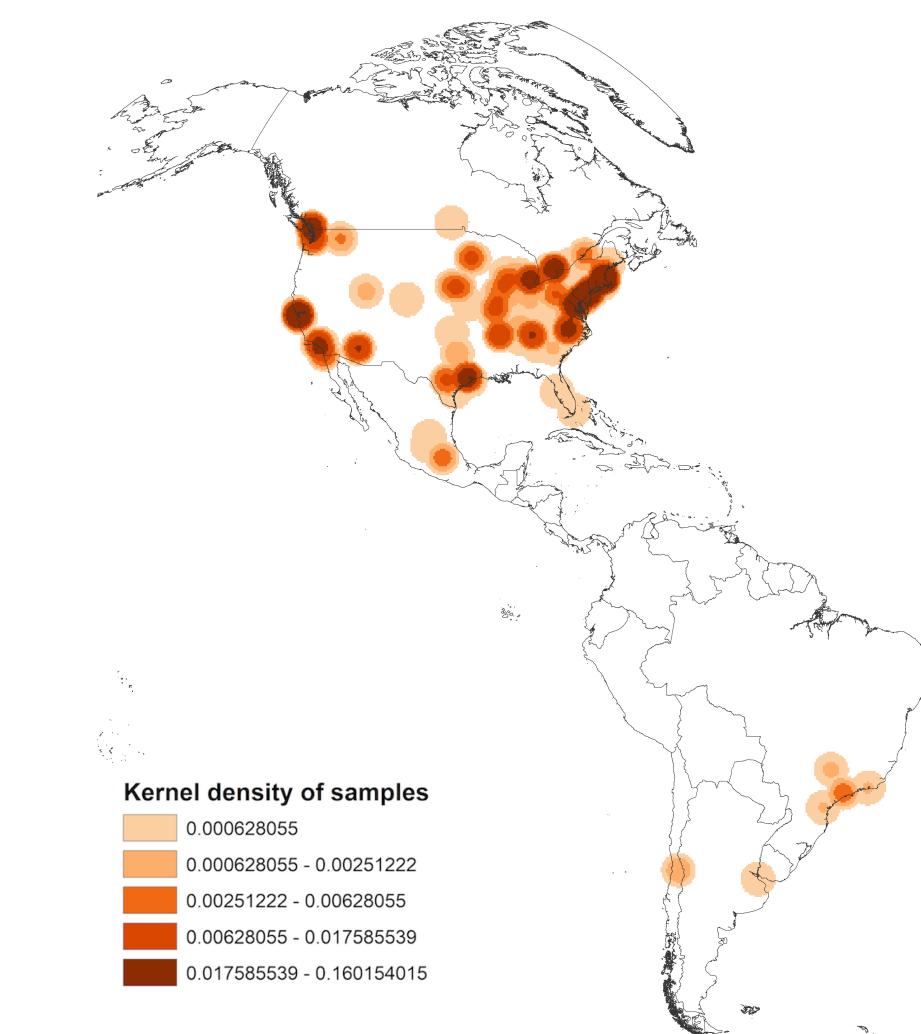
- ▶ **Genome analyses** (including transcriptome, metagenomics) are core technologies for Personalised Health™ applications
- ▶ The unexpectedly large amount of **sequence variants** in human genomes - germline and somatic/cancer - requires huge analysis efforts and creation of **reference repositories**
- ▶ **Standardized data formats** and **exchange protocols** are needed to connect these resources throughout the world, for reciprocal, international **data sharing** and **biocuration** efforts
- ▶ Our work @ UZH:
 - ▶ **cancer genome repositories**
 - ▶ **biocuration**
 - ▶ **protocols & formats**



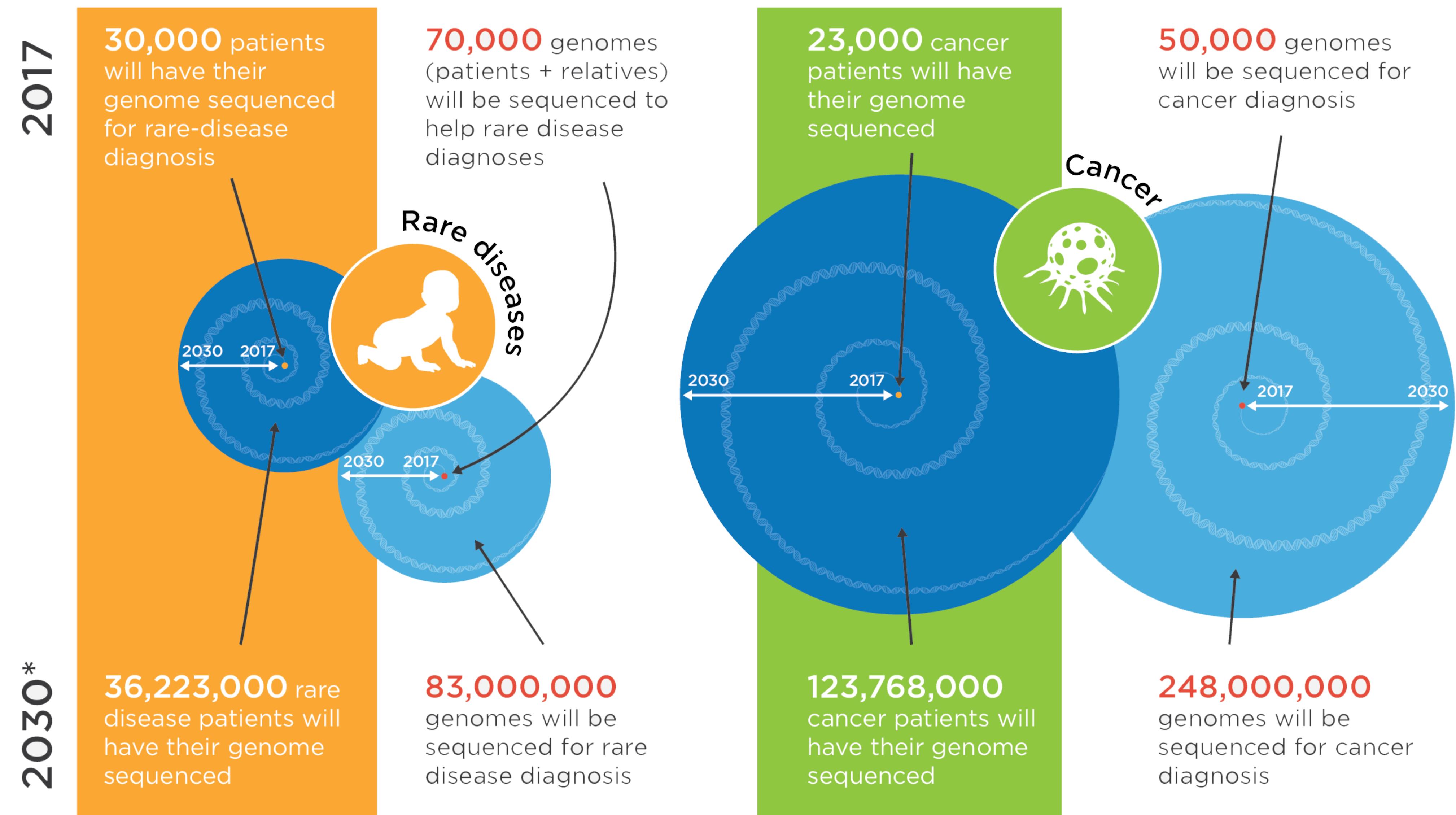
BETTER, CHEAPER, FASTER
The cost of DNA sequencing has dropped dramatically over the past decade, enabling many more applications.



The future of DNA sequencing. Eric D. Green, Edward M. Rubin & Maynard V. Olson. Nature; 11 October 2017 (News & Views)



Geographic distribution of 104'543 genomic array, 36'766 chromosomal CGH and 15'409 whole genome/exome based cancer genome datasets



* Projected figures, based on current data and known status of genomics initiatives worldwide.

The vision: Federation of data





Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

The Global Alliance for Genomics and Health

Making genomic data accessible for research and health

- January 2013 - 50 participants from eight countries
- June 2013 - White Paper, over next year signed by 70 “founding” member institutions (e.g. SIB, UZH)
- March 2014 - Working group meeting in Hinxton & 1st plenary in London
- October 2014 - Plenary meeting, San Diego; interaction with ASHG meeting
- June 2015 - 3rd Plenary meeting, Leiden
- September 2015 - GA4GH at ASHG, Baltimore
- October 2015 - DWG / New York Genome Centre
- April 2016 - Global Workshop @ ICHG 2016, Kyoto
- October 2016 - 4th Plenary Meeting, Vancouver
- May 2017 - Strategy retreat, Hinxton
- October 2017 - 5th plenary, Orlando
- May 2018 - Vancouver
- October 2018 - 6th plenary, Basel
- May 2019 - GA4GH Connect, Hinxton
- October 2019 - 7th Plenary, Boston
- October 2020 - Virtual Plenary ...

GENOMICS

A federated ecosystem for sharing genomic, clinical data

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

The Global Alliance for Genomics and Health*

SCIENCE 10 JUNE 2016 • VOL 352 ISSUE 6291



Global Alliance
for Genomics & Health

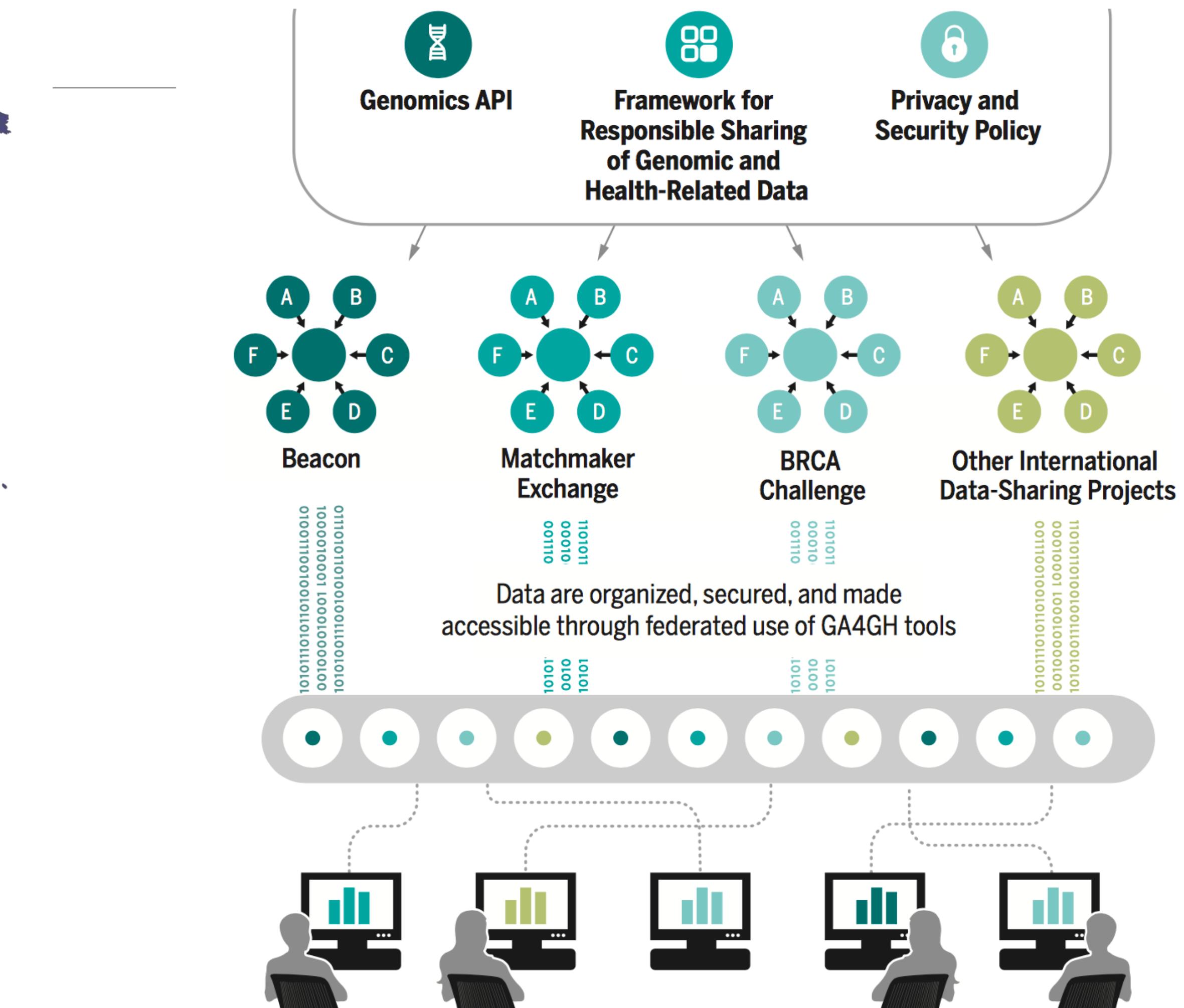


GENOMICS

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.



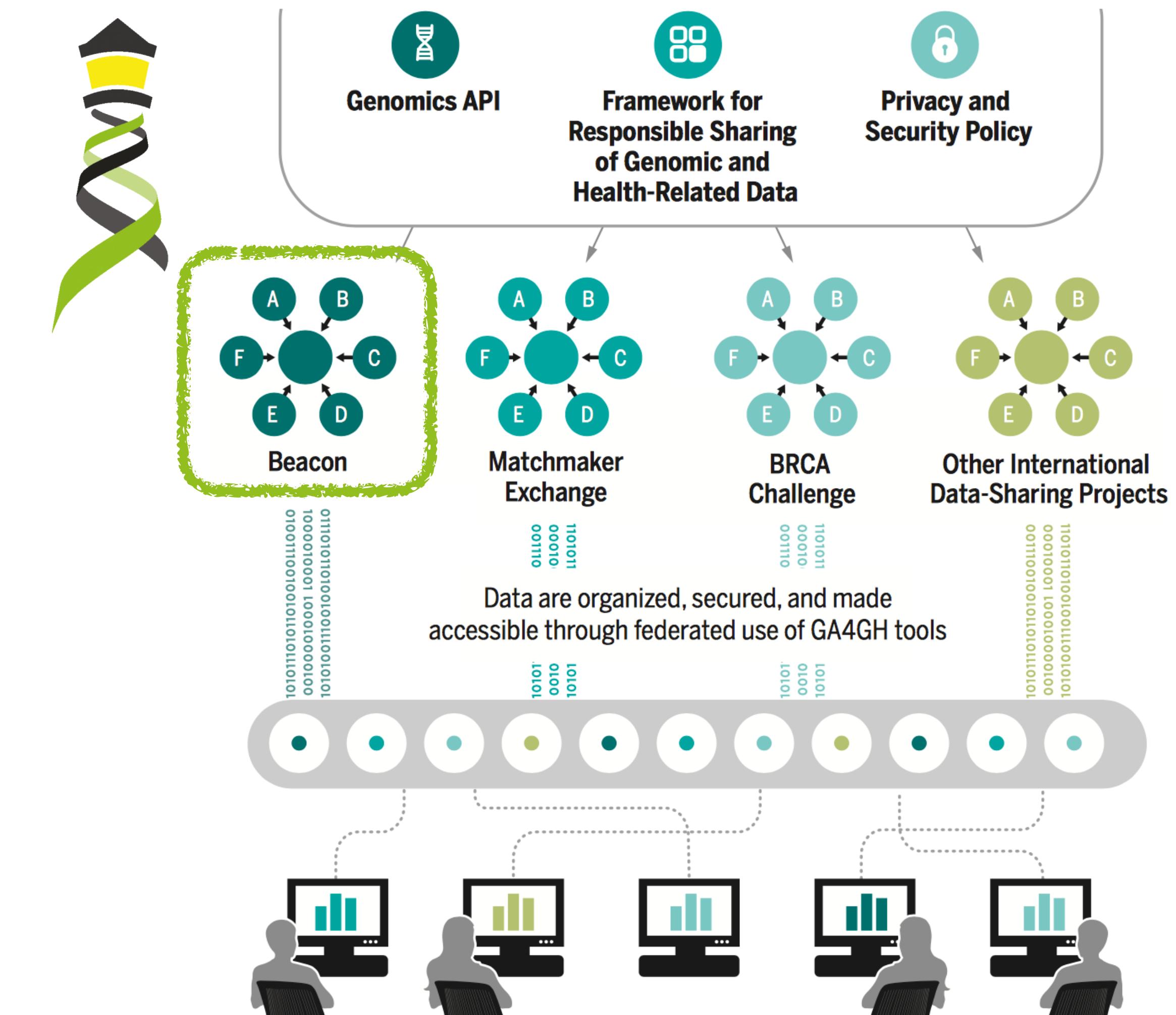


GENOMICS

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.



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.

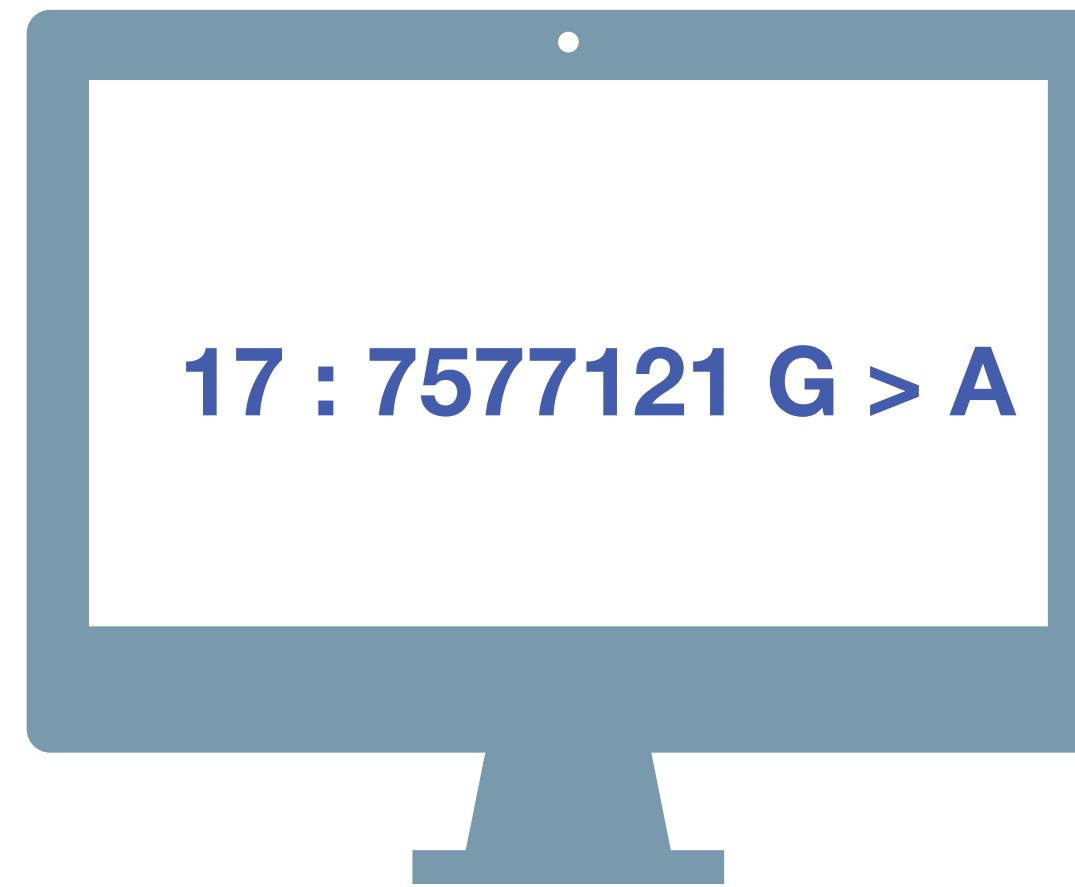
A number of more useful first versions have been suggested.

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

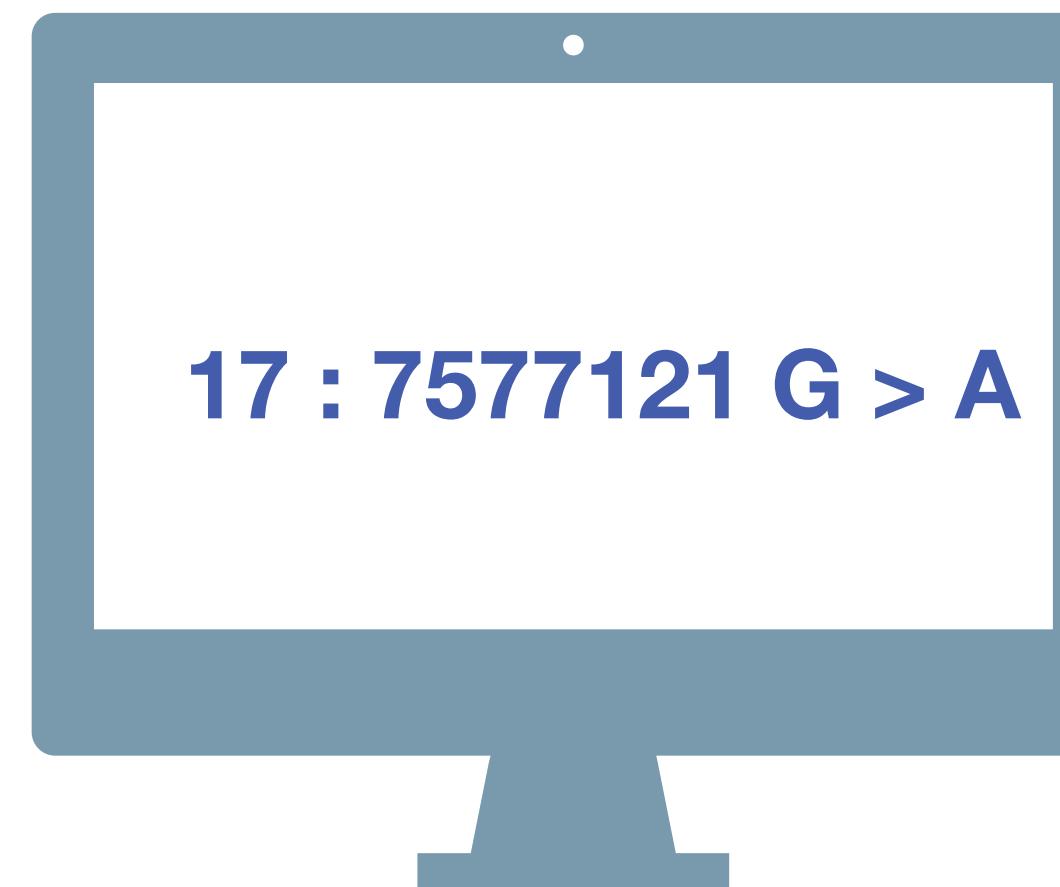
1. Specifying the chromosome ... The interface needs to specify the **accession.version** of a chromosome, or **build number**...
2. Return values ... right to **refuse** to answer without it being an error ... DOS **attack** ... or because ...especially **sensitive**...
3. Real time response ... Some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



Beacon

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

YES | NO | \0



Have you seen this variant?
It came up in my patient
and we don't know if this is
a common SNP or worth
following up.

A Beacon network federates
genome variant queries
across databases that
support the **Beacon API**

Here: The variant has
been found in **few**
resources, and those
are from **disease**
specific **collections**.

Beacon Project in 2016

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



Beacon Network

Search Beacons

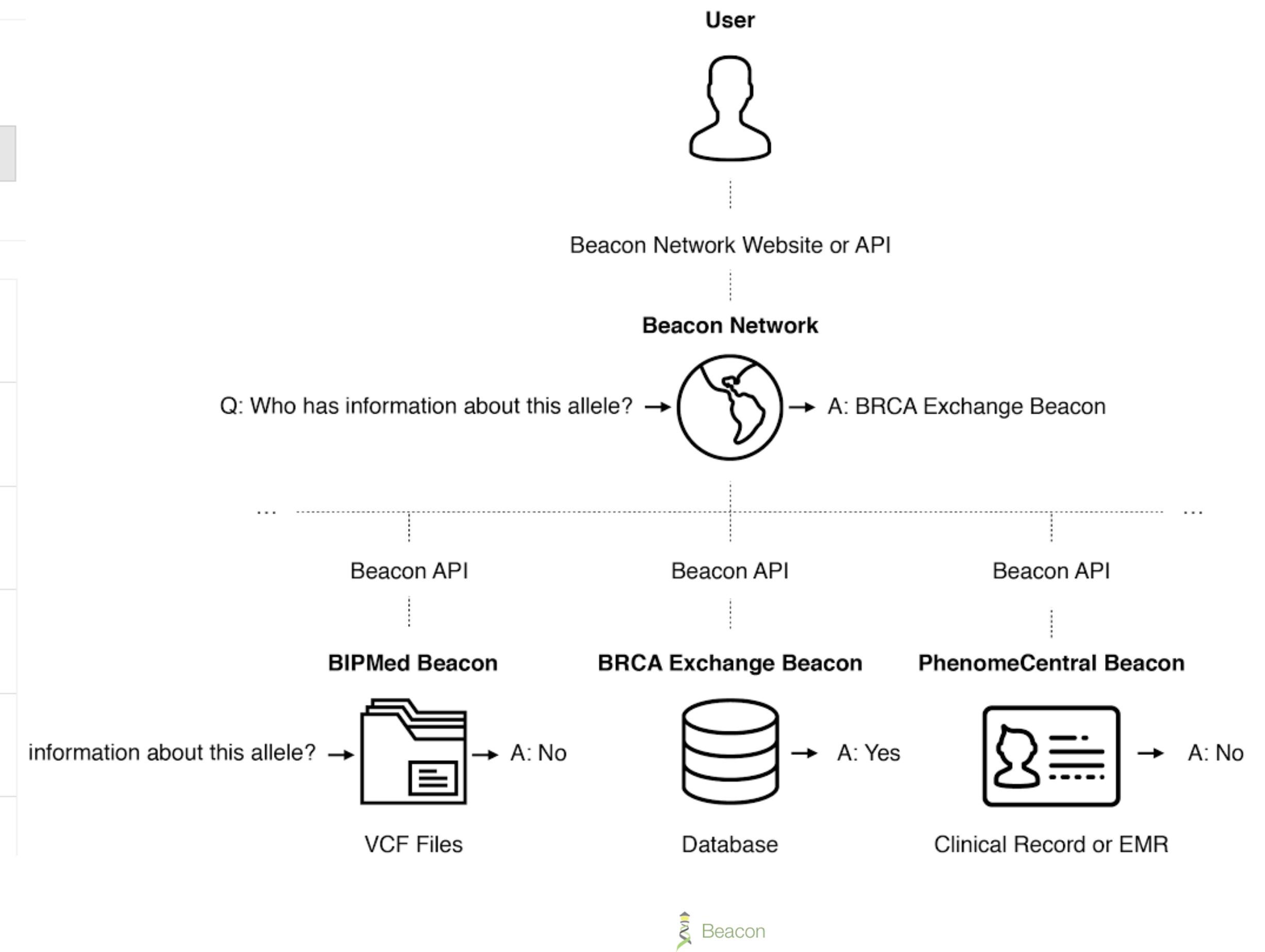
Search all beacons for allele

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 Laboratories
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 ...

Organization	Allele	Response
BioReference	10:118969015 C / CT	Found
Catalogue of Somatic Mutations in Cancer	10:118969015 C / CT	Found
Cell Lines	10:118969015 C / CT	Found
Conglomerate	10:118969015 C / CT	Found
COSMIC	10:118969015 C / CT	Found
dbGaP: Combined GRU Catalog and NHLBI Exome Seq...	10:118969015 C / CT	Found



35+

Organizations

90+

Beacons

200+

Datasets

100K+

In

Releases

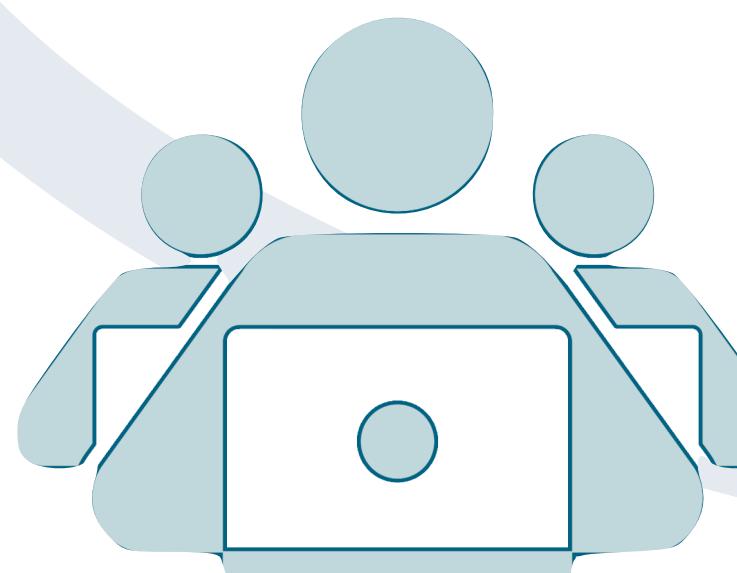
Date	Tag	Title
2018-01-24	v0.4.0	Beacon
2016-05-31	v0.3.0	Beacon

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 a link to its implementation studies, mentions Europe as the region, and lists Jordi Rambla, Juha Tornroos, and Gary Saunders as 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)



ELIXIR Beacons

EMBL-EBI



elixir
FINLAND

elixir
SWEDEN

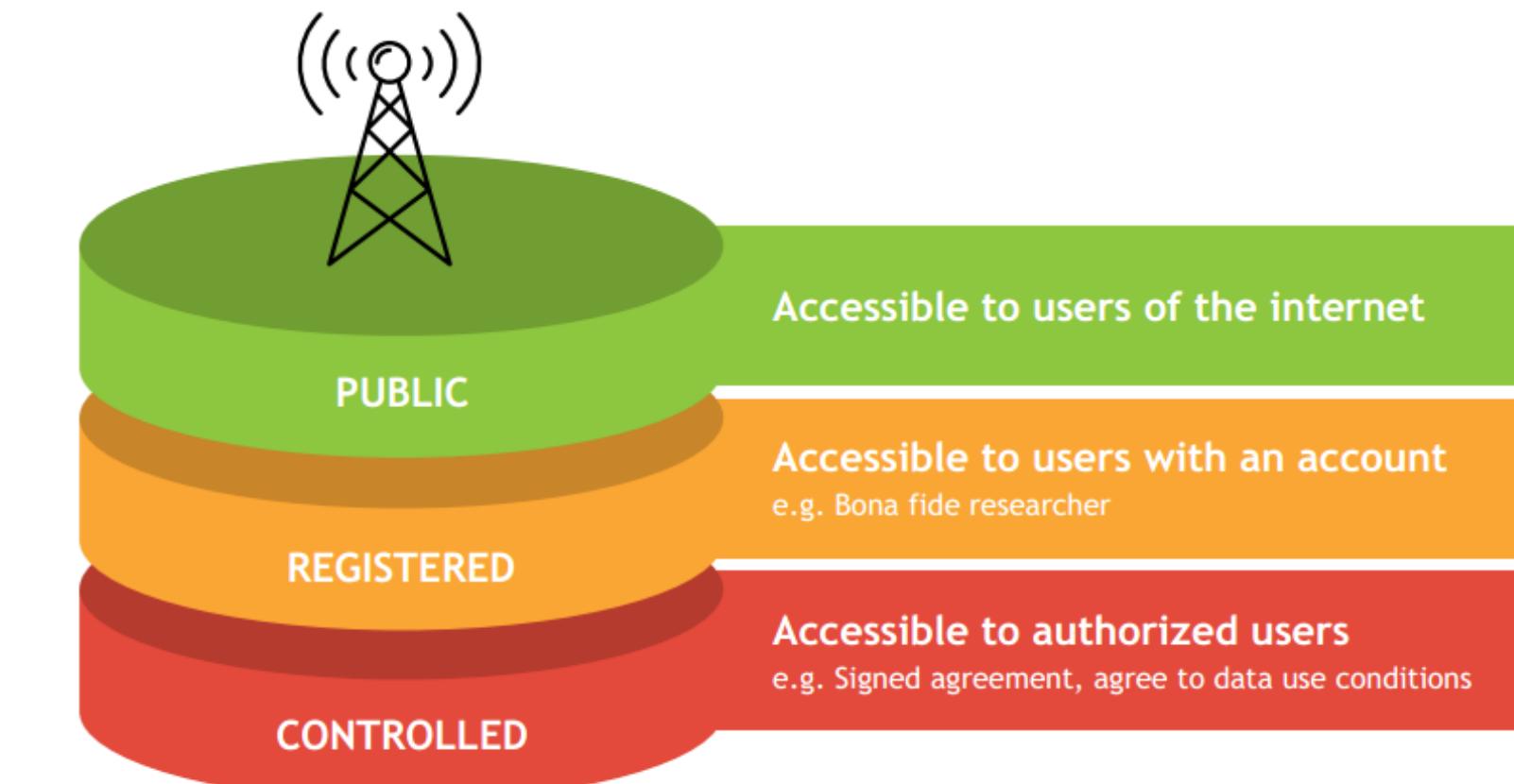
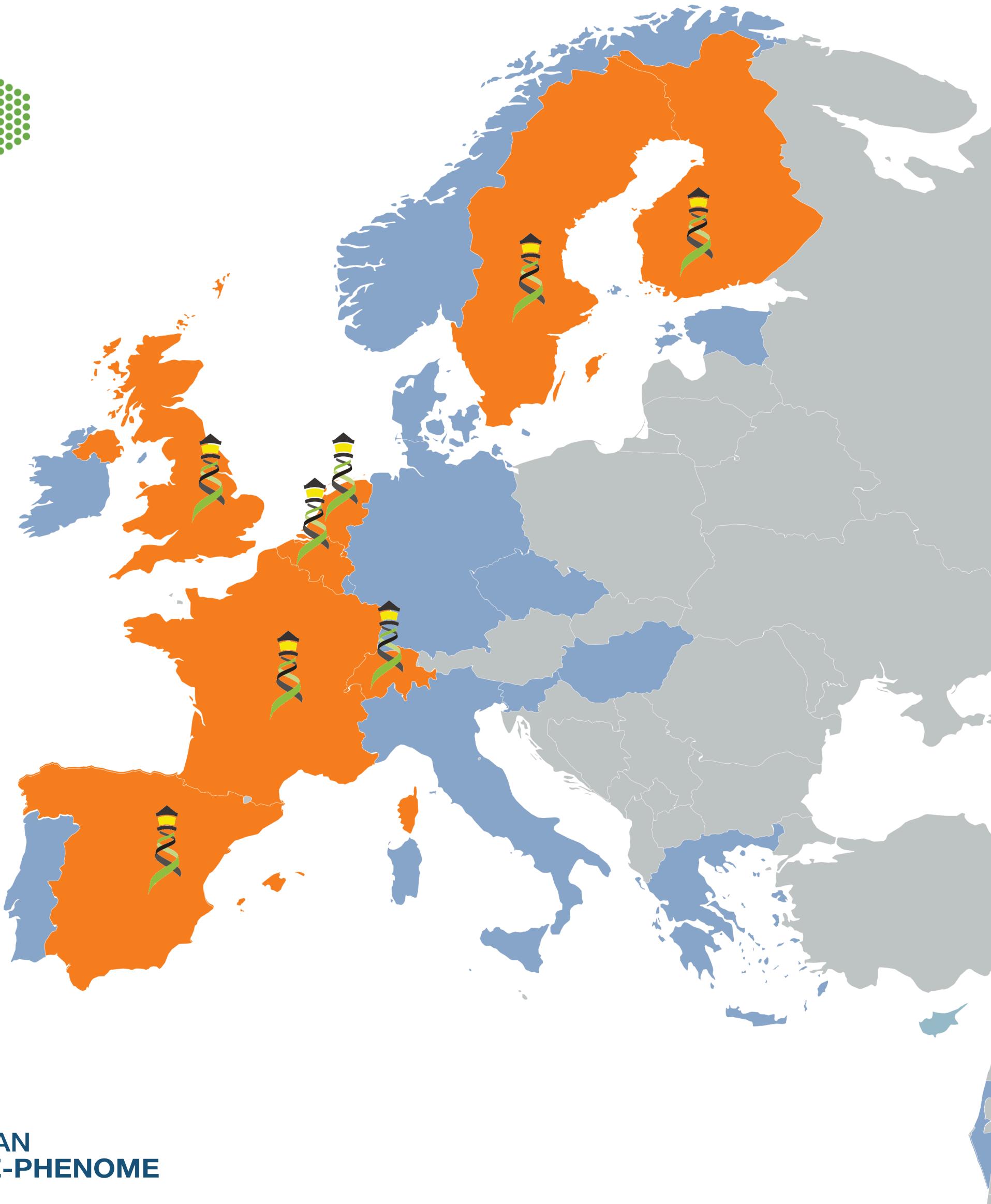
elixir
BELGIUM

elixir
NETHERLANDS

elixir
SWITZERLAND

elixir
SPAIN

EUROPEAN
GENOME-PHENOME
ARCHIVE



elixir
AAI



Driving implementation of Beacon technology in ELIXIR Nodes

→ 9 National Nodes have lit Beacons

ELIXIR Authentication and Authorization Infrastructure (AAI)

Beacon



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

elixir

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](#)

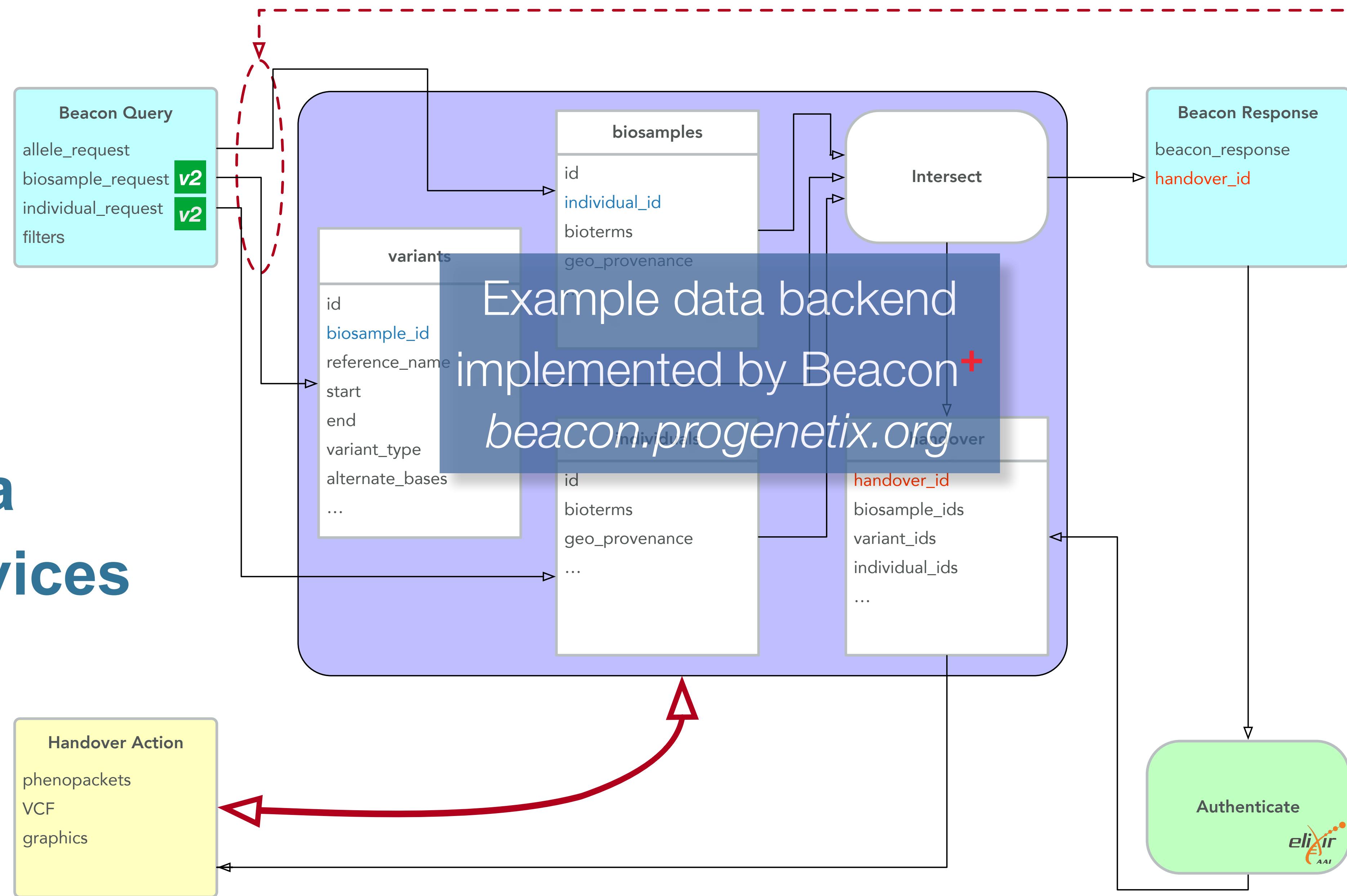
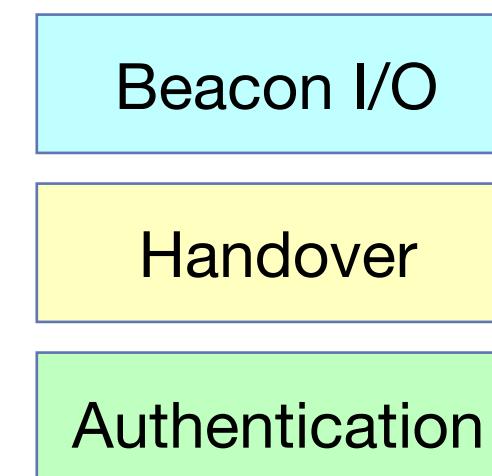
CSC - IT Center for Science Production Beacon

Beacon API Web Server based on the GA4GH Beacon API

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

Beacon & Handover

Beacons v1.1
supports data
delivery services



Handover Response

Anonymous pointer in URI and format information

- Beacon handover objects do not contain any data, but anonymous pointers to representations of the matched data in the resource
- these pointers will have to be resolved through the resource, which allows the separation of the Beacon response from the data retrieval mechanism
- there is no need for separate endpoints, greatly facilitating the implementation in different environments
- development becomes faster & more flexible

```
{  
  "callCount": 699,  
  "datasetId": "progenetix",  
  "datasetHandover": [  
    {  
      "description": "create a CNV histogram from matched callsets",  
      "handoverType": {  
        "id": "pgx:handover:cnvhistogram",  
        "label": "CNV Histogram"  
      },  
      "url": "https://progenetix.org/cgi-bin/beacondeliver.cgi?  
do=cnvhistogram&accessid=53525802-5089-43dc-8572-c0911a1df077"  
    },  
    {  
      "description": "retrieve data of the phenopackets matched by the  
query",  
      "handoverType": {  
        "id": "pgx:handover:phenopackets",  
        "label": "Phenopackets"  
      },  
      "url": "https://progenetix.org/services/phenopackets?  
do=phenopackets&accessid=70dac33c-fdf1-477d-b8e7-  
fc2fa8e7c32c&variantsaccessid=d4dda4da-a939-4c62-accc-003be237b770"  
    },  
    {  
      "description": "map variants matched by the query to the UCSC  
browser",  
      "handoverType": {  
        "id": "pgx:handover:bedfile2ucsc",  
        "label": "Variants in UCSC"  
      },  
      "url": "http://genome.ucsc.edu/cgi-bin/hgTracks?  
org=human&db=hg38&position=chr9:21531306-22492891&hgt.customText=https  
://progenetix.org/tmp/d4dda4da-a939-4c62-accc-003be237b770.bed"  
    }  
  ],  
}
```



Features and Possibilities of the current Beacon Specification

Beyond "testing the willingness for data sharing"...

- precise variant queries (chr17: 7673767 C>T)
- range queries ("any variant from here to there")
- variant frequencies
- structural genome variants, e.g. CNVs ("any deletion overlapping CDKN2A CDR coordinates")
- delivery of any kind of data matching a given query (variants, sample information, patient data ...) utilising "**handover**" objects (anonymous links to external services with their own security / privacy implementations)
- networking of v1.n Beacons with AAI integration as demonstrated by the ELIXIR Beacon Network



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**



Beacon v2 - Areas of Change

- Separate query types for different genomic variants
 - SNPs **BeaconSnpRequest**
 - Structural Variants **BeaconCnvRequest**
 - Region **BeaconRangeRequest**
 - ...
- 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

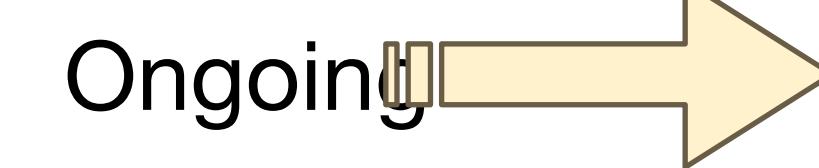
Tested and already implemented by Beacons

v2.0



GA4GH approval process ("major product update")

v2.n



incremental rollout after v2.0

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:

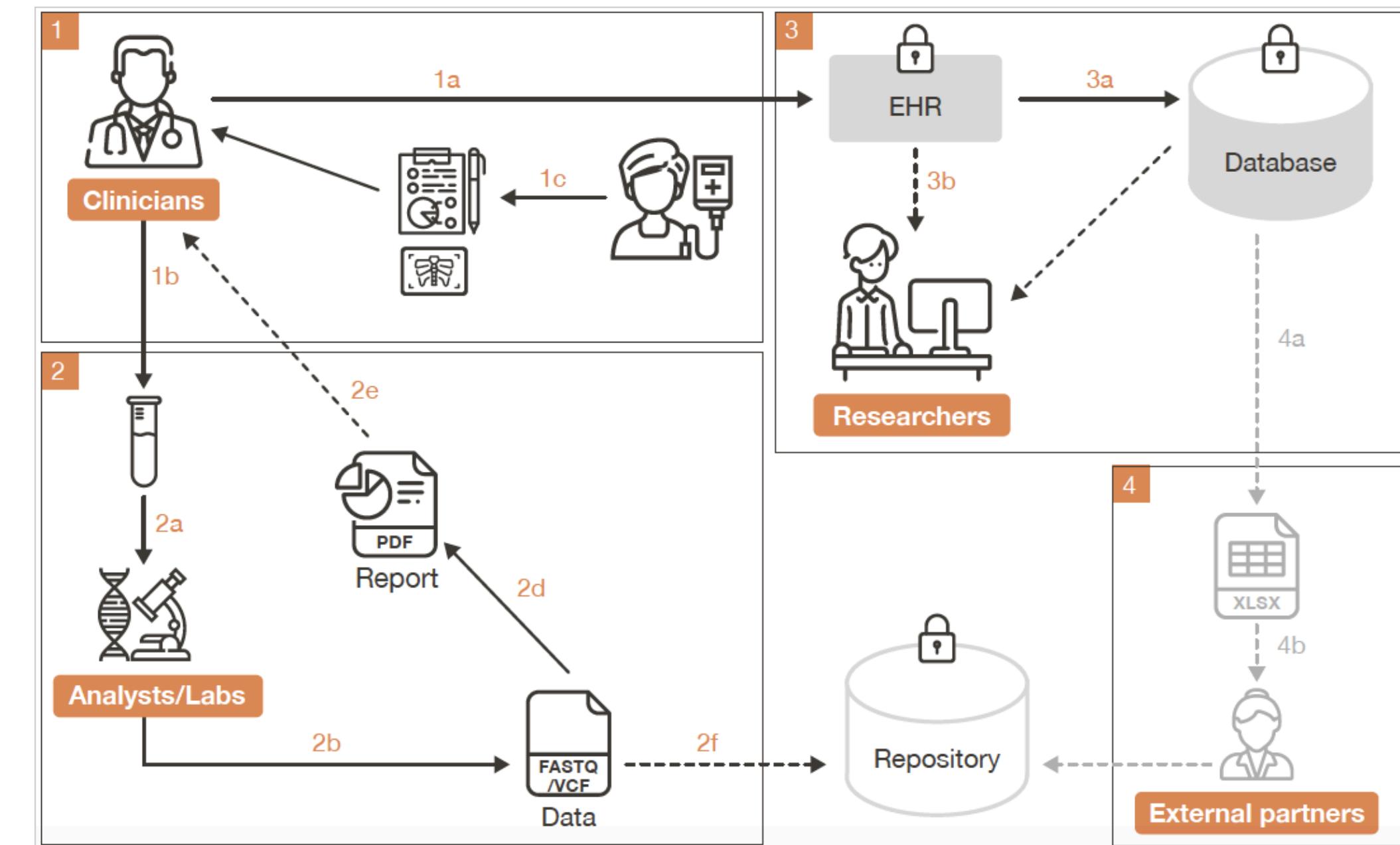
- Allowing more informative queries, like filtering by gender or age
 - Allowing to trigger the next step in the data access process, e.g. who to contact or which are the data use conditions
 - Jumping to another system where the data could be accessed, e.g. if the Beacon is internal to a hospital, to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

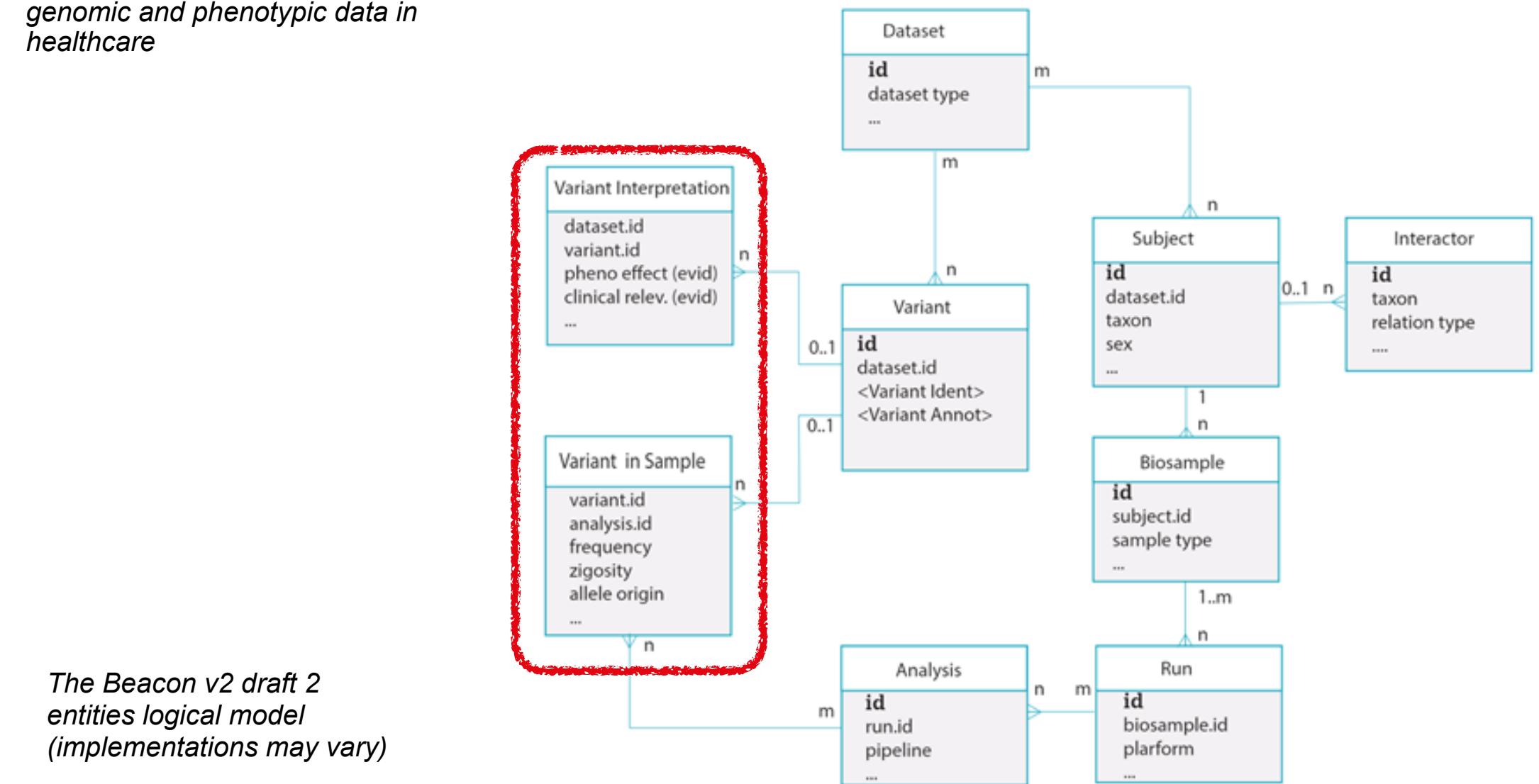
The GA4GH Beacon group started a set of meetings and interviews with GA4GH Driver Projects and with ELIXIR partners in order to determine the scope of the next generation Beacon. The goal was to be useful without breaking the simplicity that made Beacon version 1 successful.

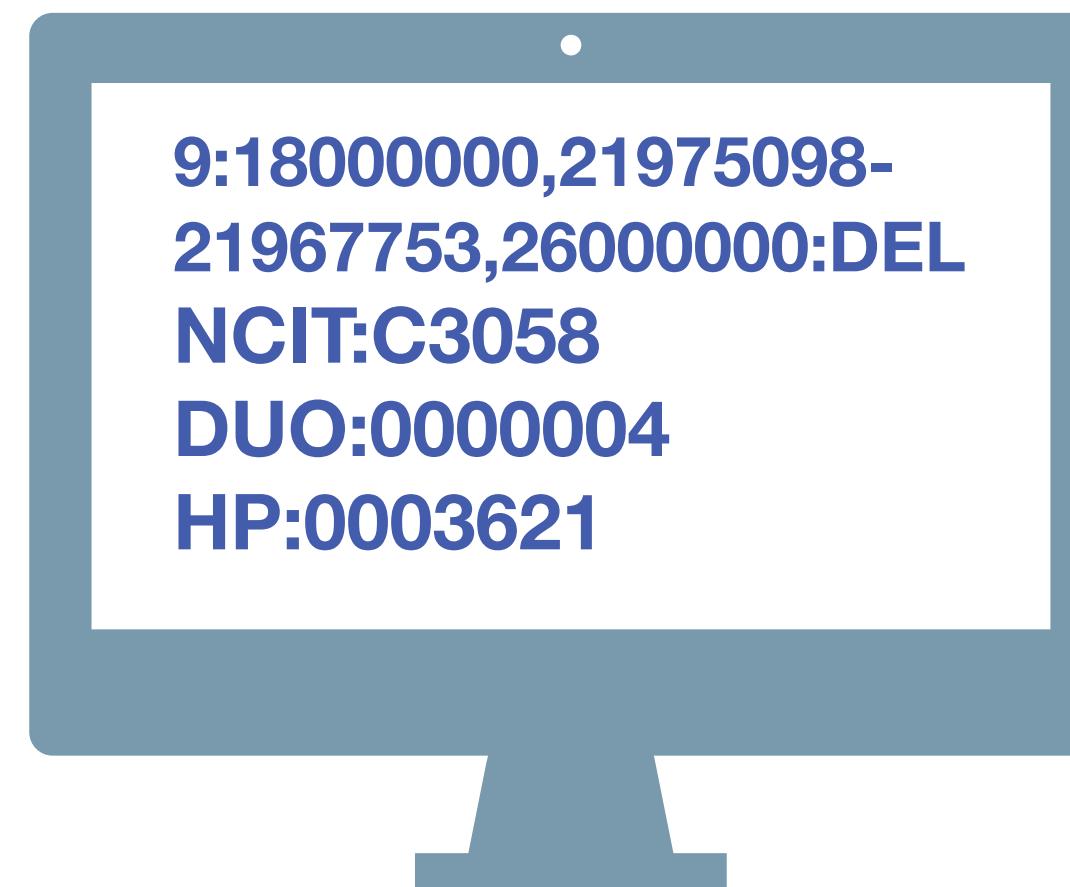
Interviews were conducted with the following GA4GH Driver Projects:

- 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





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".

DX Ontologies

Hierarchical NCIt Neoplasm Core replaces heterogeneous primary annotations

- heterogeneous and inconsistent diagnostic annotations are common in clinical reports and research studies ("text", ICD-10, ICD-O 3, OncoTree, domain-specific classifications)
- highly variable granularity of annotations is a major road block for comparative analyses and large scale data integration
 - ▶ "Colorectal Cancer" or "Rectal Mucinous Adenoca."
- initiatives and services such as Phenopackets, MONDO, OXO ... rely on and/or provide mappings to hierarchical ontologies

NCIt Neoplasm Core coded display (excerpt) for samples in the Progenetix cancer genome data resource allows sample selection on multiple hierarchy levels →



	Subsets	Samples
□	▼ NCIT:C3262: Neoplasm	88844
□	▼ NCIT:C3263: Neoplasm by Site	84747
□	▼ NCIT:C156482: Genitourinary System Neoplasm	11616
□	▼ NCIT:C156483: Benign Genitourinary System Neoplasm	219
□	▼ NCIT:C4893: Benign Urinary System Neoplasm	90
□	▼ NCIT:C4778: Benign Kidney Neoplasm	90
□	NCIT:C159209: Kidney Leiomyoma	1
□	NCIT:C4526: Kidney Oncocytoma	82
□	NCIT:C8383: Kidney Adenoma	7
□	▼ NCIT:C7617: Benign Reproductive System Neoplasm	129
□	▼ NCIT:C4934: Benign Female Reproductive System Neoplasm	129
□	▼ NCIT:C2895: Benign Ovarian Neoplasm	58
□	▼ NCIT:C4510: Benign Ovarian Epithelial Tumor	58
□	▼ NCIT:C40039: Benign Ovarian Mucinous Tumor	58
□	NCIT:C4512: Ovarian Mucinous Cystadenoma	58
□	▼ NCIT:C4060: Ovarian Cystadenoma	58
□	NCIT:C4512: Ovarian Mucinous Cystadenoma	58
□	▼ NCIT:C3609: Benign Uterine Neoplasm	71
□	▼ NCIT:C3608: Benign Uterine Corpus Neoplasm	71
□	NCIT:C3434: Uterine Corpus Leiomyoma	71
□	▼ NCIT:C156484: Malignant Genitourinary System Neoplasm	11171
□	▼ NCIT:C157774: Metastatic Malignant Genitourinary System Neoplasm	2
□	▼ NCIT:C146893: Metastatic Genitourinary System Carcinoma	2
□	NCIT:C8946: Metastatic Prostate Carcinoma	2
□	▼ NCIT:C164141: Genitourinary System Carcinoma	10561
□	▼ NCIT:C146893: Metastatic Genitourinary System Carcinoma	2
□	NCIT:C8946: Metastatic Prostate Carcinoma	2
□	▼ NCIT:C3867: Fallopian Tube Carcinoma	19

Beacon v2 Filters

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

- Beacon v2 "filters" assumes inclusion of child terms when using hierarchical classifications
 - ➡ implicit *OR* with otherwise assumed *AND*
 - implementation of hierarchical annotations overcomes some limitations of "fuzzy" disease annotations
 - data *handover* (Beacon v1.1+) enables further data exploration and export scenarios



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

Standardized Data

Data re-use depends on standardized, machine-readable metadata

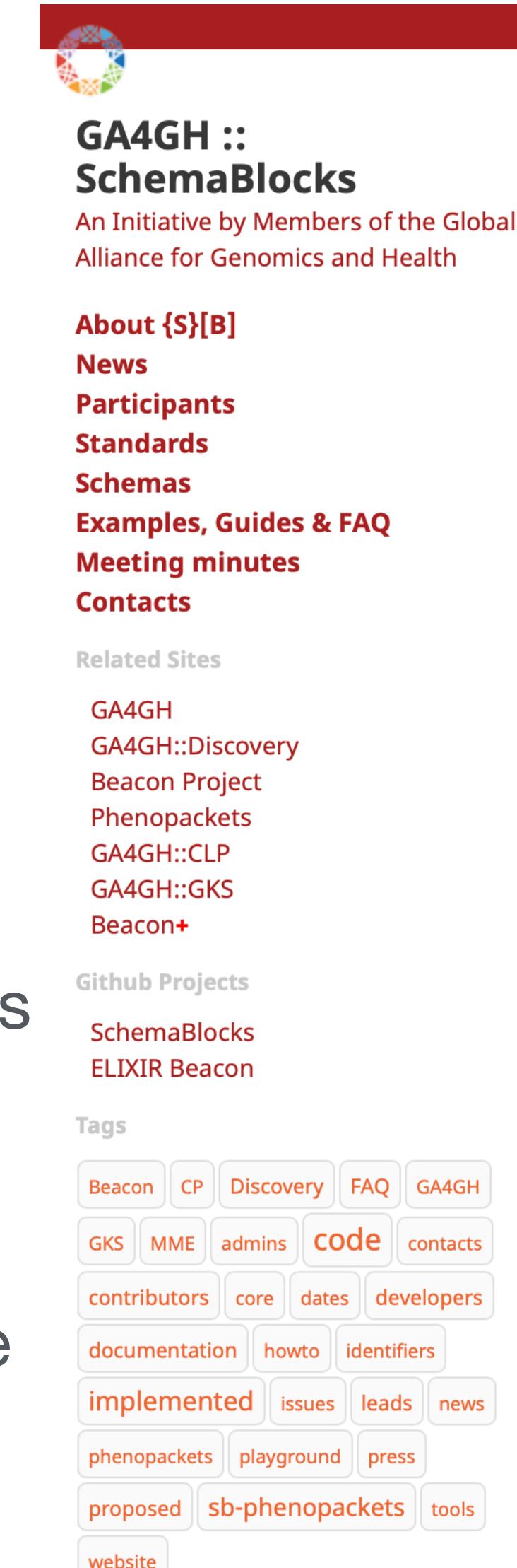
- Multiple international initiatives (ELIXIR, GA4GH, MONARCH...) and resource providers (EBI, NCBI ...) work on the generation and implementation of data annotation standards
- emerging / established principles are the use of hierarchical coding systems where individual codes are represented as CURIEs
- other formats for non-categorical annotations based on international standards, e.g.
 - ISO (ISO 8601 time & period, ISO 3166 country codes ...)
 - IETF (GeoJSON ...)
 - W3C (CURIE ...)
- these standards become pervasive throughout GA4GH's ecosystem (e.g. Phenopackets ...)

```
"data_use_conditions" : {  
    "label" : "no restriction",  
    "id" : "DUO:0000004"  
},  
  
"provenance" : {  
    "material" : {  
        "type" : {  
            "id" : "EFO:0009656",  
            "label" : "neoplastic sample"  
        }  
    },  
    "geo" : {  
        "label" : "Zurich, Switzerland",  
        "precision" : "city",  
        "city" : "Zurich",  
        "country" : "Switzerland",  
        "latitude" : 47.37,  
        "longitude" : 8.55,  
        "geojson" : {  
            "type" : "Point",  
            "coordinates" : [  
                8.55,  
                47.37  
            ]  
        },  
        "IS0-3166-alpha3" : "CHE"  
    },  
    {  
        "age": "P25Y3M2D"  
    }  
}
```

GA4GH {S}[B] SchemaBlocks

Standardized formats and data schemas for developing an "Internet of Genomics"

- “cross-workstreams, cross-drivers” initiative to document GA4GH object **standards** and **prototypes**
- launched in December 2018
- documentation and implementation examples provided by GA4GH members
- not a rigid, complete data schema
- object **vocabulary** and **semantics** for a large range of developments
- ▶ **Beacon** as contributor and user



The screenshot shows the GA4GH SchemaBlocks homepage. At the top left is the GA4GH logo. Below it, the title "GA4GH :: SchemaBlocks" is displayed, followed by the subtitle "An Initiative by Members of the Global Alliance for Genomics and Health". To the right is a navigation menu with links: "About {S}[B]", "News", "Participants", "Standards", "Schemas", "Examples, Guides & FAQ", "Meeting minutes", and "Contacts". Below the menu is a section titled "Related Sites" with links to "GA4GH", "GA4GH::Discovery", "Beacon Project", "Phenopackets", "GA4GH::CLP", "GA4GH::GKS", and "Beacon+". Further down is a "Github Projects" section with links to "SchemaBlocks" and "ELIXIR Beacon". At the bottom is a "Tags" section containing a grid of labels: Beacon, CP, Discovery, FAQ, GA4GH, GKS, MME, admins, code, contacts, contributors, core, dates, developers, documentation, howto, identifiers, implemented, issues, leads, news, phenopackets, playground, press, proposed, sb-phenopackets, tools, and website.

GA4GH SchemaBlocks Home

SchemaBlocks is a “**cross-workstreams, cross-drivers**” initiative to document GA4GH object standards and prototypes, as well as common data formats and semantics.

Launched in December 2018, this project is still to be considered a “community initiative”, with developing participation, leadership and governance structures. At its current stage, the documents can **not** be considered “**authoritative GA4GH recommendations**” but rather represent documentation and implementation examples provided by GA4GH members.

While future products and implementations may be completely based on *SchemaBlocks* components, this project does not attempt to develop a rigid, complete data schema but rather to provide the object vocabulary and semantics for a large range of developments.

The SchemaBlocks site can be accessed through the permanent link schemablocks.org. More information about the different products & formats can be found on the workstream sites. For reference, some of the original information about recommended formats and object hierarchies is kept in the [GA4GH Metadata repositories](#).

For more information on GA4GH, please visit the [GA4GH Website](#).

SchemaBlocks Repositories

The SchemaBlocks Github organisation contains several specifically scoped repositories. Please use the relevant *Github Issues* to and/or GH pull requests comment and contribute there.

@mbaudis 2019-11-19: [more ...](#)

SchemaBlocks “Status” Levels

SchemaBlocks schemas (“blocks”) provide recommended blueprints for schema parts to be re-used for the development of code based “products” throughout the GA4GH ecosystem. We propose a labeling system for those schemas, to provide transparency about the level of support those schemas have from {S}[B] participants and observers.

@mbaudis 2019-07-17: [more ...](#)

SchemaBlocks {S}[B] Mission Statement

SchemaBlocks aims to translate the work of the workstreams into data models that:

- Are usable by other internal GA4GH deliverables, such as the Search API.
- Are usable by Driver Projects as an exchange format.
- Aid in aligning the work streams across GA4GH.
- Do not create a hindrance in development work by other work streams.

@mbaudis 2019-03-27: [more ...](#)



GA4GH {S}[B] SchemaBlocks

Standardized formats and data schemas for developing an "Internet of Genomics"

- “cross-workstreams, cross-drivers” initiative to document GA4GH object **standards** and **prototypes**
- launched in December 2018
- documentation and implementation examples provided by GA4GH members
- not a rigid, complete data schema
- object **vocabulary** and **semantics** for a large range of developments
- ▶ **Beacon** as contributor and user

Biosample sb-phenopackets ↗

{S}[B] Status [i]	implemented
Provenance	◦ Phenopackets
Used by	◦ Phenopackets
Contributors	◦ GA4GH Data Working Group ◦ Jules Jacobsen ◦ Peter Robinson ◦ Michael Baudis ◦ Melanie Courtot ◦ Isuru Liyanage
Source (v1.0.0)	◦ raw source [JSON] ◦ Github

Attributes

Type: object

Description: A Biosample refers to a unit of biological material from which the genomic DNA, RNA, proteins) for molecular analyses (e.g. sequencing, array hybridization, spectrometry) are extracted.

Examples would be a tissue biopsy, a single cell from a culture or single cell ger fraction from a gradient centrifugation.

Several instances (e.g. technical replicates) or types of experiments (e.g. genome experiments) may refer to the same Biosample.

FHIR mapping: Specimen.

Properties

Property	Type
ageOfIndividualAtCollection	https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/Age.json [SRC] [HTML]
ageRangeOfIndividualAtCollection	https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/AgeRange.json [SRC] [HTML]
description	string
diagnosticMarkers	array of https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/OntologyClass.json [SRC] [HTML]
histologicalDiagnosis	https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/OntologyClass.json [SRC] [HTML]
htsFiles	array of https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/HtsFile.json [SRC] [HTML]
id	string
individualId	string
isControlSample	boolean
phenotypicFeature	array of https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/PhenotypicFeature.json [SRC] [HTML]
procedure	https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/Procedure.json [SRC] [HTML]
sampledTissue	https://schemablocks.org/schemas/sb-phenopackets/v1.0.0/Tissue.json [SRC] [HTML]

HtsFile sb-phenopackets ↗

{S}[B] Status [i]	implemented
Provenance	◦ Phenopackets
Used by	◦ Phenopackets
Contributors	◦ Jules Jacobsen ◦ Peter Robinson
Source (v1.0.0)	◦ raw source [JSON] ◦ Github

Attributes

Type: object

Description: A file in one of the HTS formats (<https://samtools.github.io/hts-specs>)

Properties

Property	Type
description	string
genomeAssembly	string
htsFormat	
individualToSampleIdentifiers	object
uri	string

Beacon Test/Driver Implementations

CRG COVID19 Beacon & Progenetix



The CRG "COVID" Beacon

Beyond humans and limited data delivery



CRG Viral Beacon - SNP Query

- SNP
- Region
- Feature
- Motifs
- AA Aminoacid
- Info
- Pipeline
- Warnings

Total number of samples: 42490
Updated: 2020-06-09 14:37:34.806227

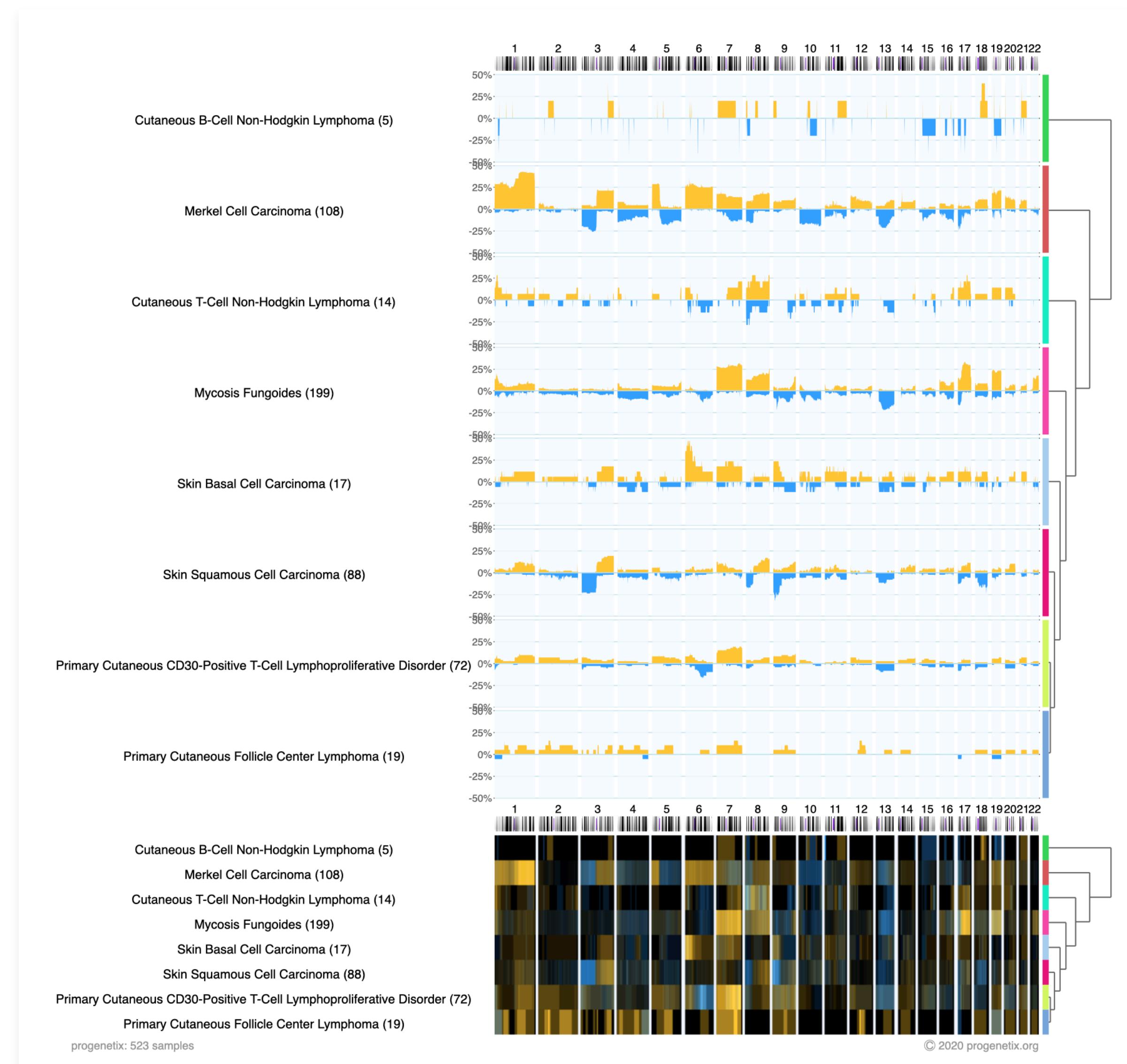
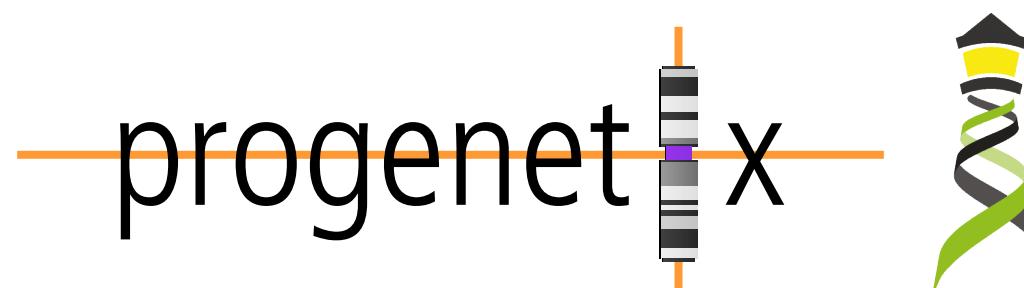


Variant Annotation							Variant Frequency per Dataset					
VARIANT TYPE	GENE REGION	EFFECT	FUNCTIONAL CLASS	LOCUS NAME	LOCUS ID	AMINOACID CHANGE	ENA-READS	ONT	GISAID	ENA-CONSENSUS		
SNP	CODING	SYNONYMOUS_CODING	SILENT	orf1ab,ORF1ab	GU280_gp01	S2839	0.6772	0.0334	0.1955	0.0844		
Metadata (# of results: 131)												
FREQUENCY PER RUN	HOST AGE	HOST SEX	GEO ORIGIN	DATASET	DISEASE	DISEASE OUTCOME	SAMPLE TYPE	COLLECTION DATE	RUN PLATFORM	RUN ID	SAMPLE ID	CALLER
-	-	-	-	ONT	-	-	-	-	MinION	ERR4007730	ERS4399630	-
-	-	-	-	ONT	-	-	-	-	MinION	ERR4018418	ERS4399631	-
-	-	-	-	ONT	-	-	-	-	MinION	ERR4018419	ERS4399632	-
-	-	-	-	ONT	-	-	-	-	MinION	ERR4018420	ERS4399635	-
-	-	-	-	ONT	-	-	-	-	MinION	ERR4018421	ERS4399633	-
-	-	-	-	ONT	-	-	-	2020-04-05	MinION	ERR4082913	ERS4535973	-
-	-	-	-	ONT	-	-	-	2020-03-30	MinION	ERR4085126	ERS4538418	-
-	-	-	-	ONT	-	-	-	2020-03-29	MinION	ERR4085129	ERS4535837	-
-	-	-	North America:USA:Washington:-	ONT	COVID-19	-	-	2020-03-19	GridION	SRR11637329	SRS6559200	-
-	-	-	North America:USA:Washington:-	ONT	COVID-19	-	-	2020-03-19	GridION	SRR11637332	SRS6559197	-
-	-	-	North America:USA:Washington:-	ONT	COVID-19	-	-	2020-03-13	MinION	SRR11637341	SRS6559188	-
-	-	-	North America:USA:Washington:-	ONT	COVID-19	-	-	2020-01-26	MinION	SRR11637346	SRS6559183	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-28	MinION	SRR11648018	SRS6569430	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-23	MinION	SRR11648029	SRS6569419	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-19	MinION	SRR11648032	SRS6569416	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-16	MinION	SRR11648034	SRS6569414	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-17	MinION	SRR11648037	SRS6569411	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-16	MinION	SRR11648038	SRS6569410	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-15	MinION	SRR11648041	SRS6569407	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-18	MinION	SRR11648043	SRS6569405	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-18	MinION	SRR11648053	SRS6569395	-
-	-	-	North America:USA:Connecticut:-	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-03-19	MinION	SRR11648055	SRS6569393	-
-	-	-	Asia:India:Gujarat:Surat	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-04-27	Ion Torrent S5	SRR11745202	SRS6619115	-
-	-	-	Asia:India:Gujarat:Surat	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-04-26	Ion Torrent S5	SRR11745204	SRS6619113	-
-	-	-	Asia:India:Gujarat:Surat	ONT	COVID-19	-	oro/naso-pharyngeal swab	2020-04-23	Ion Torrent S5	SRR11745211	SRS6619107	-

Beaconized Progenetix

From Beacon Query to Explorative Analyses of CNV Patterns

- On top of the Progenetix resource
 - 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...)



progenetix: 523 samples

© 2020 progenetix.org

[About Progenetix](#)[Cancer CNV Profiles](#)[Search Samples](#)[Publication DB](#)[Services](#)[Upload & Plot](#)[Documentation](#)[Beacon⁺](#)[Baudisgroup @ UZH](#)

Search Samples

[CDKN2A Deletion Example](#)[MYC Duplication](#)[TP53 Del. in Cell Lines](#)

This example shows the query for CNV deletion variants overlapping the CDKN2A gene's coding region with at least a single base, but limited to "highly focal" hits (here i.e. $\leq \sim 1\text{Mbp}$ in size). The query is against the Progenetix and arrayMap collections. It can be modified e.g. through changing the position parameters or diagnosis.

 [Gene Spans](#) [Cytoband\(s\)](#)**Reference name**

9

(Structural) Variant Type

DEL (Deletion)

Start or Position

21500001-21975098

End (Range or Structural Var.)

21967753-22500000

Cancer Classification(s)

NCIT:C3058: Glioblastoma (4358)

Biosample Type **Filters** **Filter Logic**

AND

City Select... [Query Beacon](#)

[About Progenetix](#)[Cancer CNV Profiles](#)[Search Samples](#)[Publication DB](#)[Services](#)[Upload & Plot](#)[Documentation](#)[Beacon⁺](#)[Baudisgroup @ UZH](#)

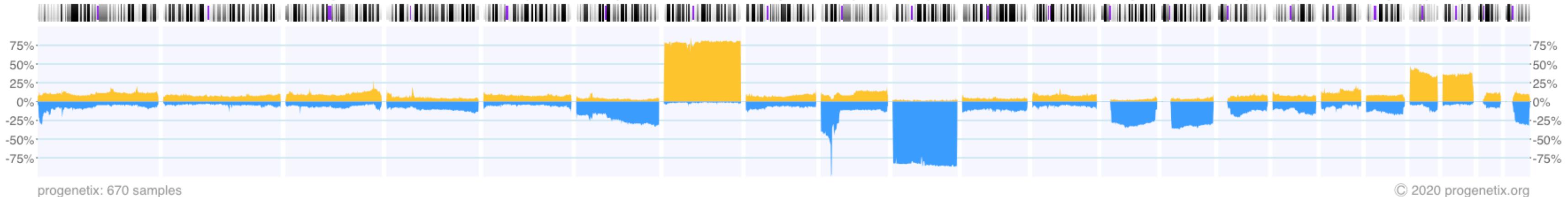
Assembly: GRCh38 Chro: 9 Start: 21500001-21975098 End: 21967753-22500000 Type: DEL Filters: NCIT:C3058

progenetix

Samples: 668

Variants: 286

Calls: 675

 $f_{alleles}$: 0.000088[Phenopackets ↗](#)[Callsets Variants ↗](#)[Variants in UCSC ↗](#)[UCSC region ↗](#)[JSON Response ↗](#)[Visualization options](#)[Results](#)[Biosamples](#)[Biosamples Map](#)[Variants](#)

Subsets	Subset Samples	Query Matches	Subset Match Frequencies
icdot-C71.4	4	1	0.250
icdom-94403	4274	664	0.155
NCIT:C3058	4358	664	0.152
icdot-C71.1	14	2	0.143
icdot-C71.9	6684	651	0.097
NCIT:C3796	84	4	0.048
icdom-94423	84	4	0.048
icdot-C71.0	1712	14	0.008



About Progenetix

Cancer CNV Profiles

Search Samples

Publication DB

Services

Upload & Plot

Documentation

Beacon⁺

Baudisgroup @ UZH

Search Samples

Assembly: GRCh38 Chro: 9 Start: 21500001-21975098 End: 21967753-22500000 T

progenetix

Samples: 668

Variants: 286

Calls: 675

f_{alleles}: 0.000088

[Phenopackets](#)

[Callsets Variants](#)

[Variants in UCSC](#)

[UCSC region](#)

[JSON Response](#)

Results

Biosamples

Biosamples Map

Variants

```
{  
  "data": [  
    {  
      "biosamples": [  
        {  
          "externalReferences": [  
            {  
              "id": "PMID:23079654"  
            }  
          ],  
          "histologicalDiagnosis": {  
            "id": "NCIT:C3058",  
            "label": "Glioblastoma"  
          },  
          "id": "pgxbs-kftvgk8h",  
          "sampledTissue": {  
            "id": "UBERON:0001869",  
            "label": "cerebral hemisphere"  
          },  
          "variants": [  
            {  
              "_id": "5bab578b727983b2e00ca99e",  
              "biosample_id": "pgxbs-kftvgk8h",  
              "callset_id": "pgxcs-kftvmlzx",  
              "digest": "9:21548871-21999595:DEL",  
              "end_max": 21999595.0,  
              "end_min": 21999595.0,  
              "info": {  
                "cnv_length": 450724,  
                "cnv_value": null  
              },  
              "mate_name": null,  
              "reference_name": "9",  
              "start_max": 21548871.0,  
              "start_min": 21548871.0,  
              "updated": "2018-09-26 09:50:58.094031",  
              "variant_type": "DEL",  
              "variantset_id": "AM_VS_GRCH36"  
            }  
          ]  
        }  
      ],  
      "id": "pxf__pgxind-kftx2am8",  
      "subject": "pgxind-kftx2am8"  
    }  
  ]  
}
```



Search Samples



About Progenetix

Cancer CNV Profiles

Search Samples

Publication DB

Services

Upload & Plot

Documentation

Beacon⁺

Baudisgroup @ UZH

Assembly: GRCh38 Chro: 9 Start: 21500001-21975098 End: 21967753-22500000 Type: DEL Filters: NCIT:C3058

progenetix

Samples: 668

Phenopackets

UCSC region

Variants: 286

Callsets Variants

JSON Response

Calls: 675

Variants in UCSC

f_{alleles}: 0.000088

Visualization options

Results

Biosamples

Biosamples Map

Variants

JSON

[Download Response](#)

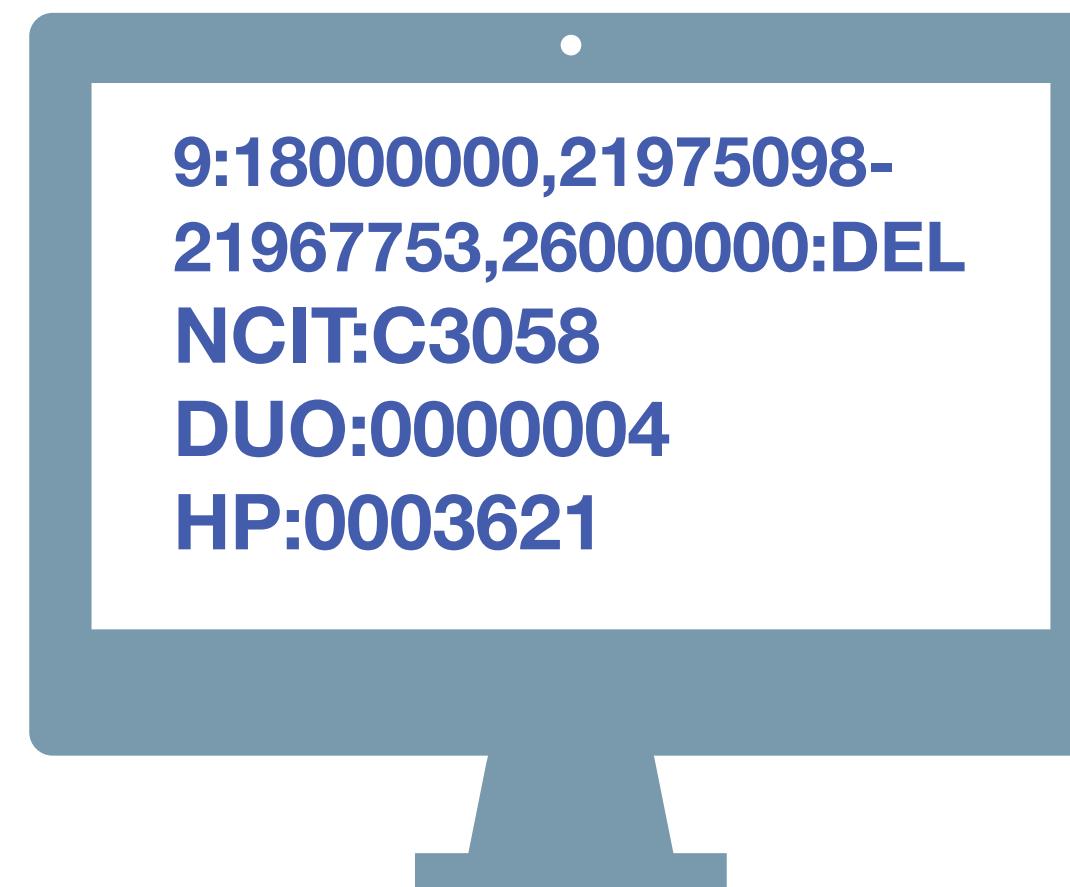
Int. ID	Digest	Callset	Biosample	Chr.	Ref. Base(s)	Alt. Base(s)	Type
5bab578b727983b2e00ca99e	9:21548871-21999595:DEL	pgxcs-kftvmlzx	pgxbs-kftvgk8h	9			DEL
5bab578d727983b2e00cb505	9:21958233-21999595:DEL	pgxcs-kftvmm5j	pgxbs-kftvgk90	9			DEL
5bab5793727983b2e00cdc18	9:21958233-21999595:DEL	pgxcs-kftvmmjj	pgxbs-kftvgka5	9			DEL
5bab5794727983b2e00ce2c6	9:21791897-21999595:DEL	pgxcs-kftvmmlu	pgxbs-kftvgkae	9			DEL
5bab5794727983b2e00ce49a	9:21958233-21999595:DEL	pgxcs-kftvmmmb	pgxbs-kftvgkaf	9			DEL



Page 1 of 135

Beacon, GA4GH & SPHN

- GA4GH has become the major "go to" international organization for the development of data exchange standards and implementation guidelines for genomics & health data
- essential for its success are national partner organizations such as SPHN which have the opportunity to shape the development of GA4GH through contributions, but importantly can benefit through the alignment with international, "cutting edge" standards developments, thereby avoiding duplicate efforts & resource waste
- The **early** adoption of protocols and standards such as Beacon and Phenopackets drives innovation and efficient data use, both for biomedical researchers and for clinicians
- While the direct benefit of e.g. local Beacon installations may be limited compared to legacy systems, it opens the door for scaled integration with outside systems
- **Beacon v2** specifically is being developed with **clinical requirements** in mind and will cover a broad range of use cases in precision medicine, rare diseases and cancer
- The active participation of SPHN in GA4GH development projects supports a leading position for Swiss biomedical research and personalized health applications

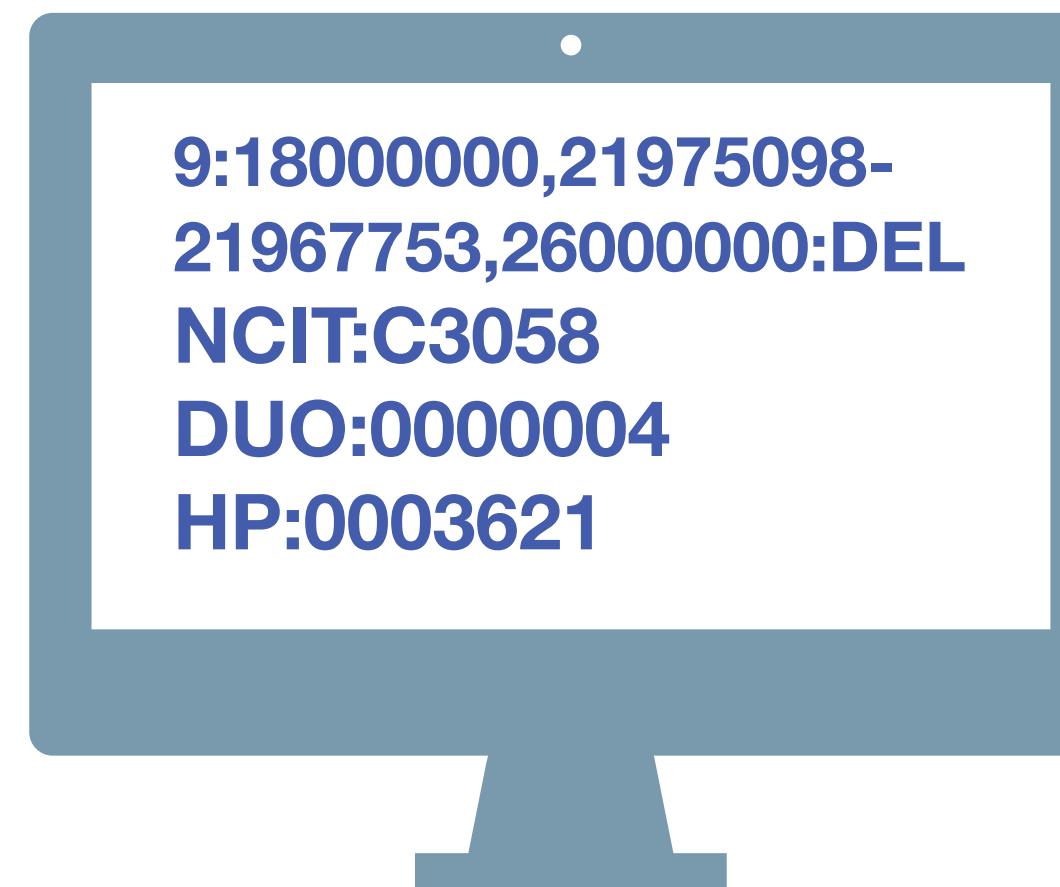


Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?



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 v2 - Towards Flexible Use and Clinical Applications



The original Beacon protocol had been designed to be:

- **Simple:** focus on robustness and easy implementation
- **Federated:** maintained by individual organizations and assembled into a network
- **General-purpose:** used to report on any variant collection
- **Aggregative:** provide a boolean (or quantitative) answer about the observation
- **Privacy protecting:** queries do not return information about single individuals

Sites offering *beacons* can scale through aggregation *Beacon Networks*, which distribute queries among a potentially large number of international *beacons* and assemble the results. Since 2015 the development of the Beacon protocol has been led by *ELIXIR* in collaboration with international participants. Recent versions of the *Beacon* protocol have expanded the framework to:

- providing a framework for other types of genome variation data (i.e. range queries)
- allowing for data delivery using *handover* protocol, e.g. to link with clinical environments and allow for data delivery and visualisation services

beacon-project.io

Beacon Protocol for Genomic Data Sharing

Beacon+ About Progenetix Help

Search Samples

CNV Request Allele Request Range Query All Fields

CNV Example

This example shows the query for CNV deletion variants overlapping the CDKN2A gene's with at least a single base, but limited to "focal" hits (here i.e. $\leq \sim 2\text{Mbp}$ in size). The query uses the arrayMap collection and can be modified e.g. through changing the position parameters or adding additional filters.

This query type is for copy number queries ("variantCNVrequest"), e.g. using fuzzy range end positions to capture a set of similar variants.

Dataset

arraymap X

Genome Assembly GRCh38 / hg38

Reference name 9 **(Structural) Variant Type** DEL (Deletion)

Start or Position 21000001-21975098 **End (Range or Structural Var.)** 21967753-23000000

Cancer Classification(s) NCIT:C3058: Glioblastoma (2119) X

City Select... 21000001 21975098 21967753 23000000

Query Beacon

beacon-project.io

ga4gh-beacon / specification-v2

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

master Go to file Add file Clone

sdelatorrep committed 52474fc ... 46 commits 1 branch 0 tags

.gitignore Create .gitignore 2 months ago

.travis.yml Do not check for empty response 6 months ago

LICENSE Add license 6 months ago

README.md fix typo beacon-project website last month

beacon.yaml Fix error (#18) last month

README.md

Beacon API specification

build passing

license Apache 2

The Beacon protocol defines an open standard for genomics data discovery, developed by members of the Global Alliance for Genomics & Health. It provides a framework for public web services responding to queries against genomic data collections, 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](#).

About
GA4GH Beacon v2 specification.
ga4gh beacon openapi

Readme
Apache-2.0 License

Releases
No releases published [Create a new release](#)

Packages
No packages published [Publish your first package](#)

Contributors 3

sdelatorrep sdelatorrep
mbaudis mbaudis
blankdots blankdots

github.com/ga4gh-beacon/



beacon.progenetix.org/ui/

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

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 services against genomic data collections, for instance from population based or disease specific repositories.

Baudisgroup @ UZH
 (Ni Ai)
 Michael Baudis
 (Haoyang Cai)
 Paula Carrio Cordo
 Bo Gao
 Qingyao Huang
 (Saumya Gupta)
 (Nitin Kumar)
 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

The original Beacon protocol has five main characteristics:

- **Simple:** focus on robustness and ease of implementation.
- **Federated:** maintained by a community of independent sites offering *beacons*.
- **General-purpose:** used for copy number queries ("variantCNVrequest"), for example to capture a set of similar variants.
- **Aggregative:** provide a summary of results across multiple sites.
- **Privacy protecting:** query results are aggregated and do not reveal individual participant data.

Beacon v2 - Towards Flexibility

(Structural) Variant Type: DEL (Deletion)
End (Range or Structural Var.): 21967753-23000000

Cancer Classification(s): NCIT:C3058: Glioblastoma (2119)

City: Select...
 21000001-21975098
 21967753-23000000

Query Beacon

{S}{B} and GA4GH
 Melanie Courtot
 Helen Parkinson
 many more ...

beacon.progenetix.org/ui/

Beacon API Leads

Jordi Rambla
 Anthony Brooks
 Juha Törnroos

Discovery WS

Michael Baudis (Beacon)
 Marc Fiume (Networks)

ELIXIR

Gary Saunders
 David Lloyd
 Serena Scollen

Sabela de la Torre Pernas

The Beacon protocol defines an open standard for genomics data discovery, developed by members of the Global Alliance for Genomics & Health. It provides a framework for public web services responding to queries against genomic data collections, 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/specification

Contributors: sdelatorrep (3)

mbaudis (1)

blankdots (1)





University of
Zurich ^{UZH}



Global Alliance
for Genomics & Health



Prof. Dr. Michael Baudis
Institute of Molecular Life Sciences
University of Zurich
SIB | Swiss Institute of Bioinformatics
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland

arraymap.org
progenetix.org
info.baudisgroup.org
sib.swiss/baudis-michael
imls.uzh.ch/en/research/baudis
beacon-project.io
schemablocks.org