

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

Concepts | Status | History | Outlook

Michael Baudis :: Personalized Health Technologies 2020 :: 2020-06-30



**Global Alliance
for Genomics & Health**

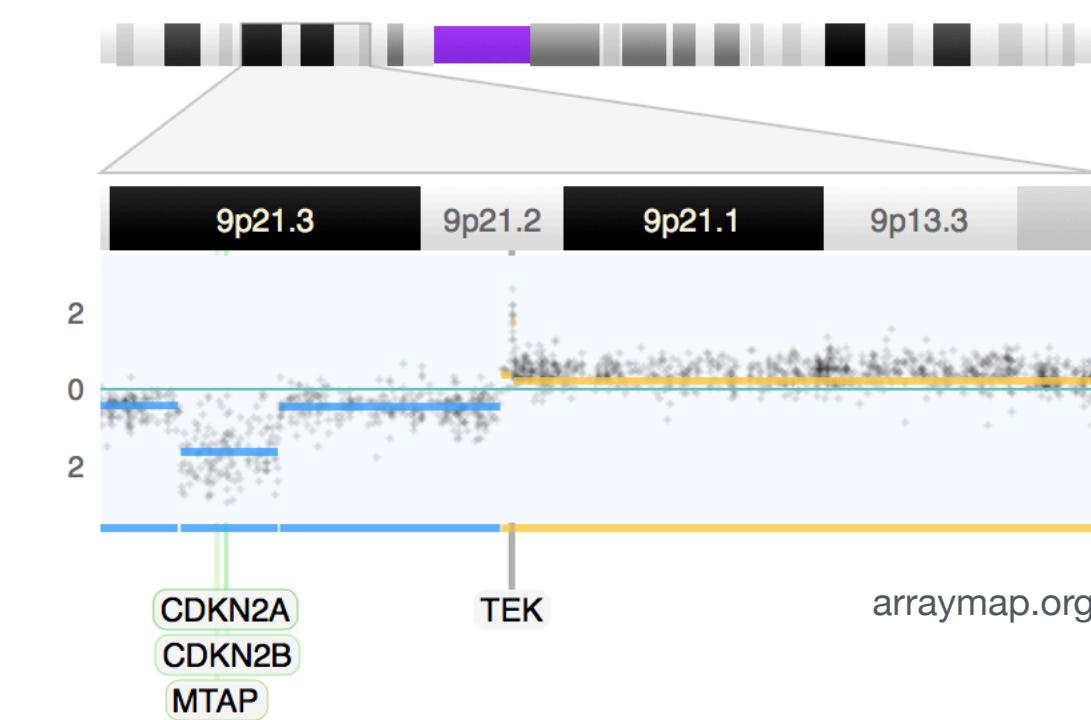


**University of
Zurich UZH**

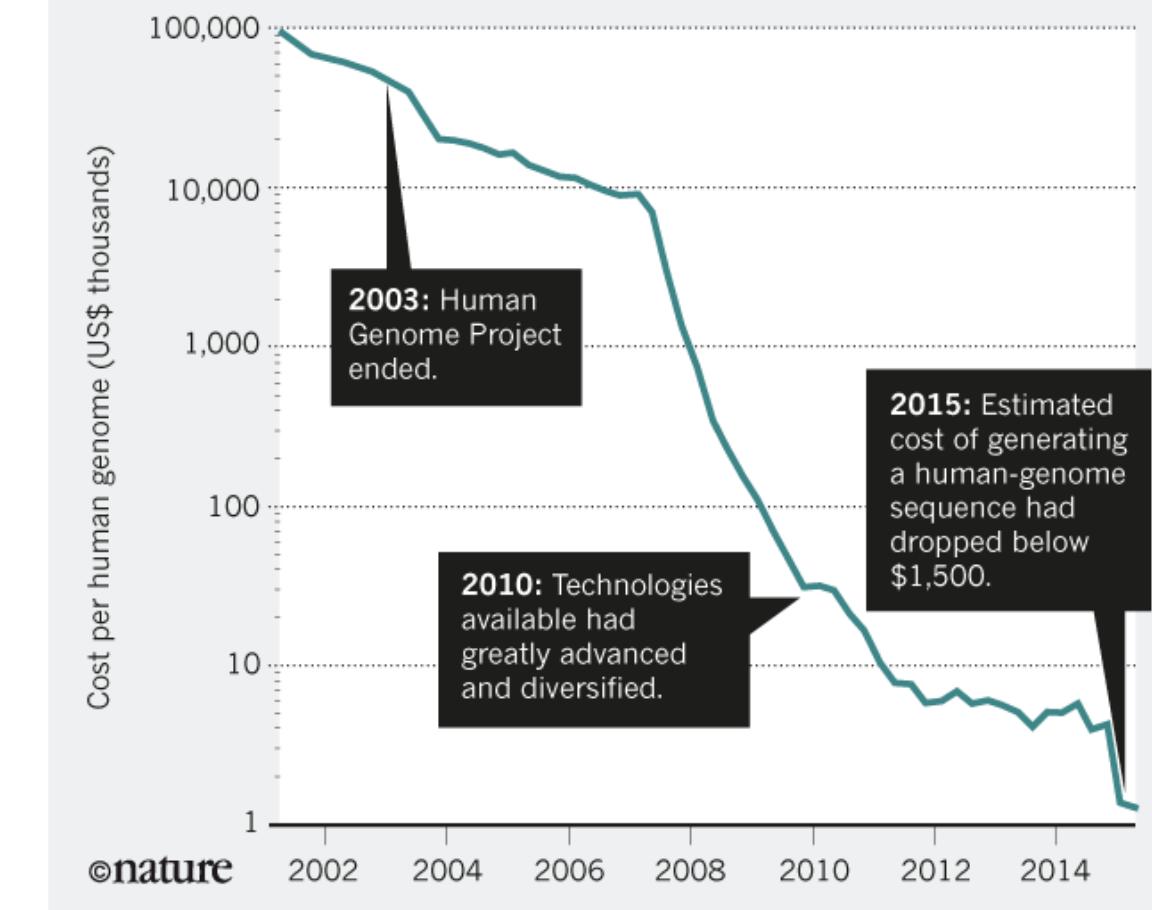


Genome screening at the core of “Personalised Health”

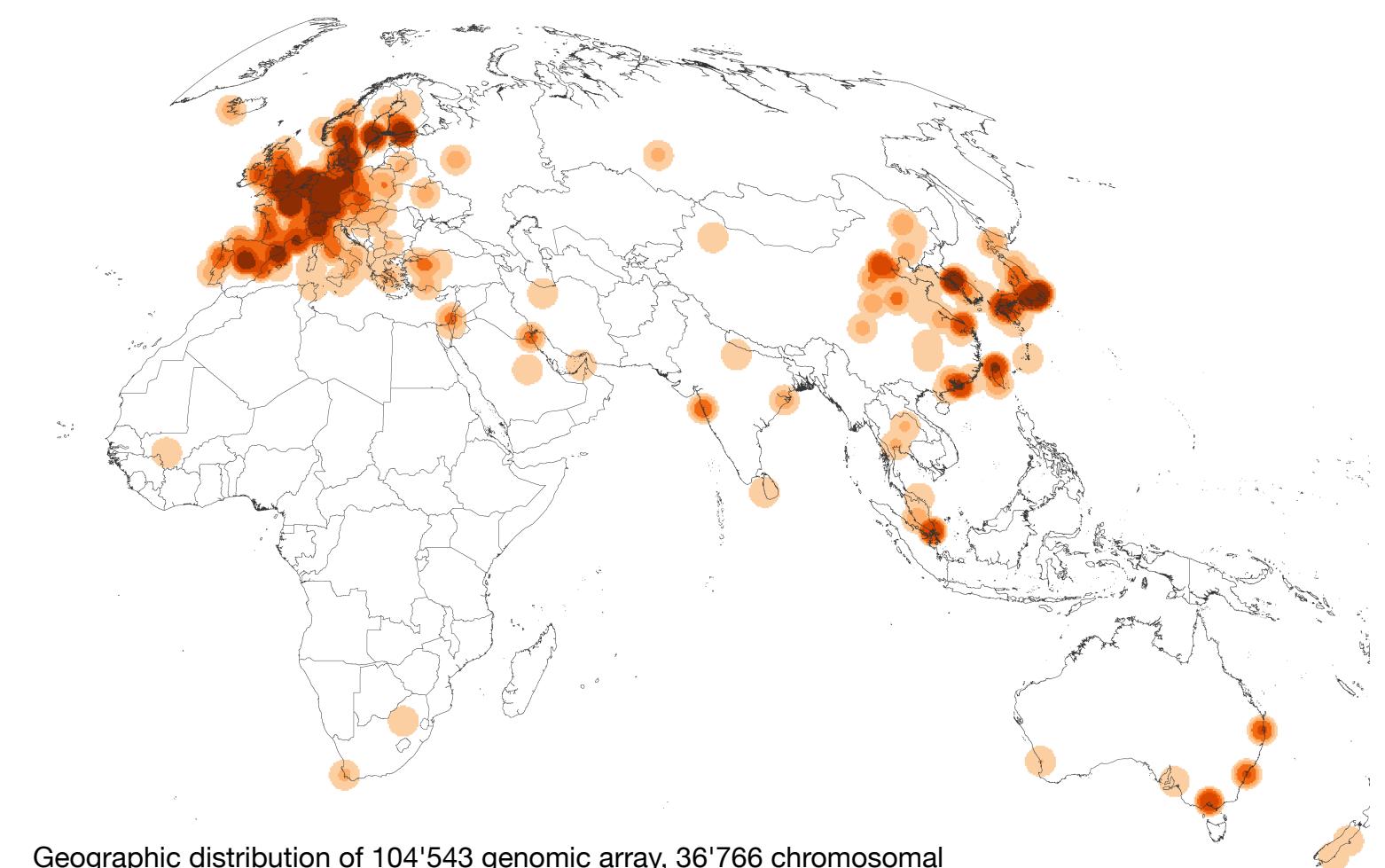
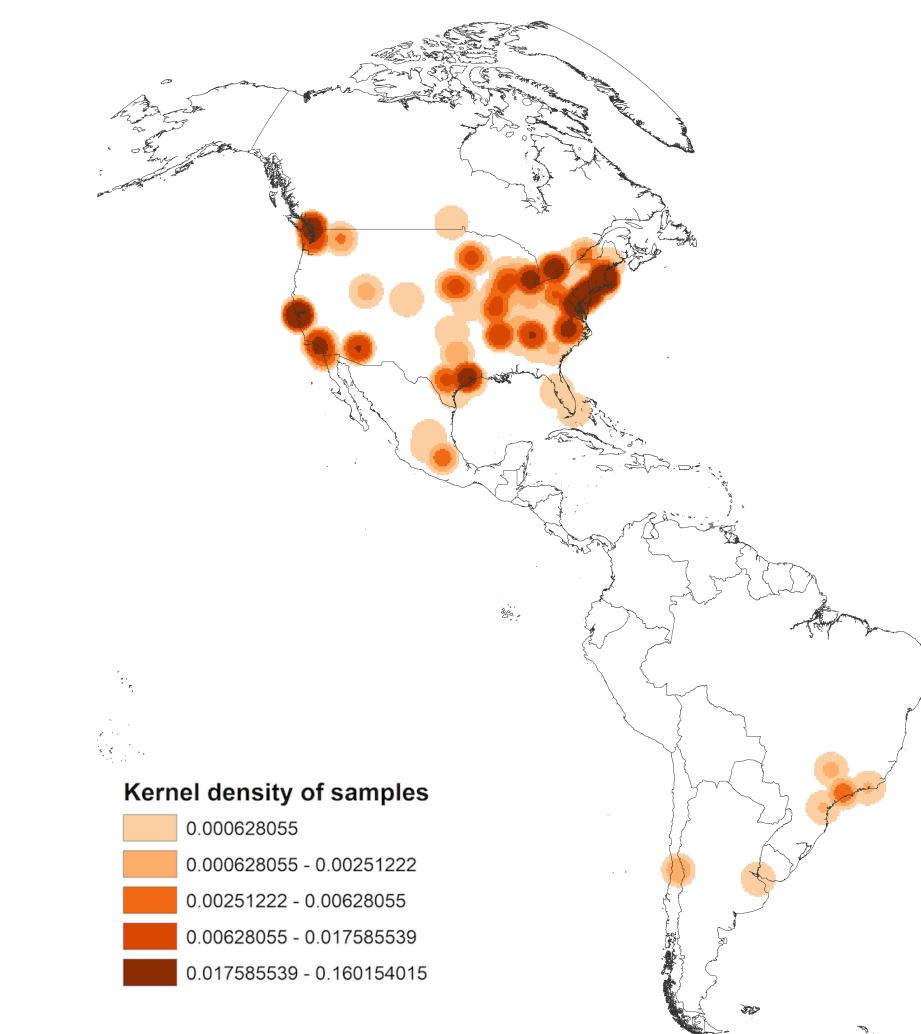
- ▶ **Genome analyses** (including transcriptome, metagenomics) are core technologies for Personalised Health™ applications
- ▶ The unexpectedly large amount of **sequence variants** in human genomes - germline and somatic/cancer - requires huge analysis efforts and creation of **reference repositories**
- ▶ **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**



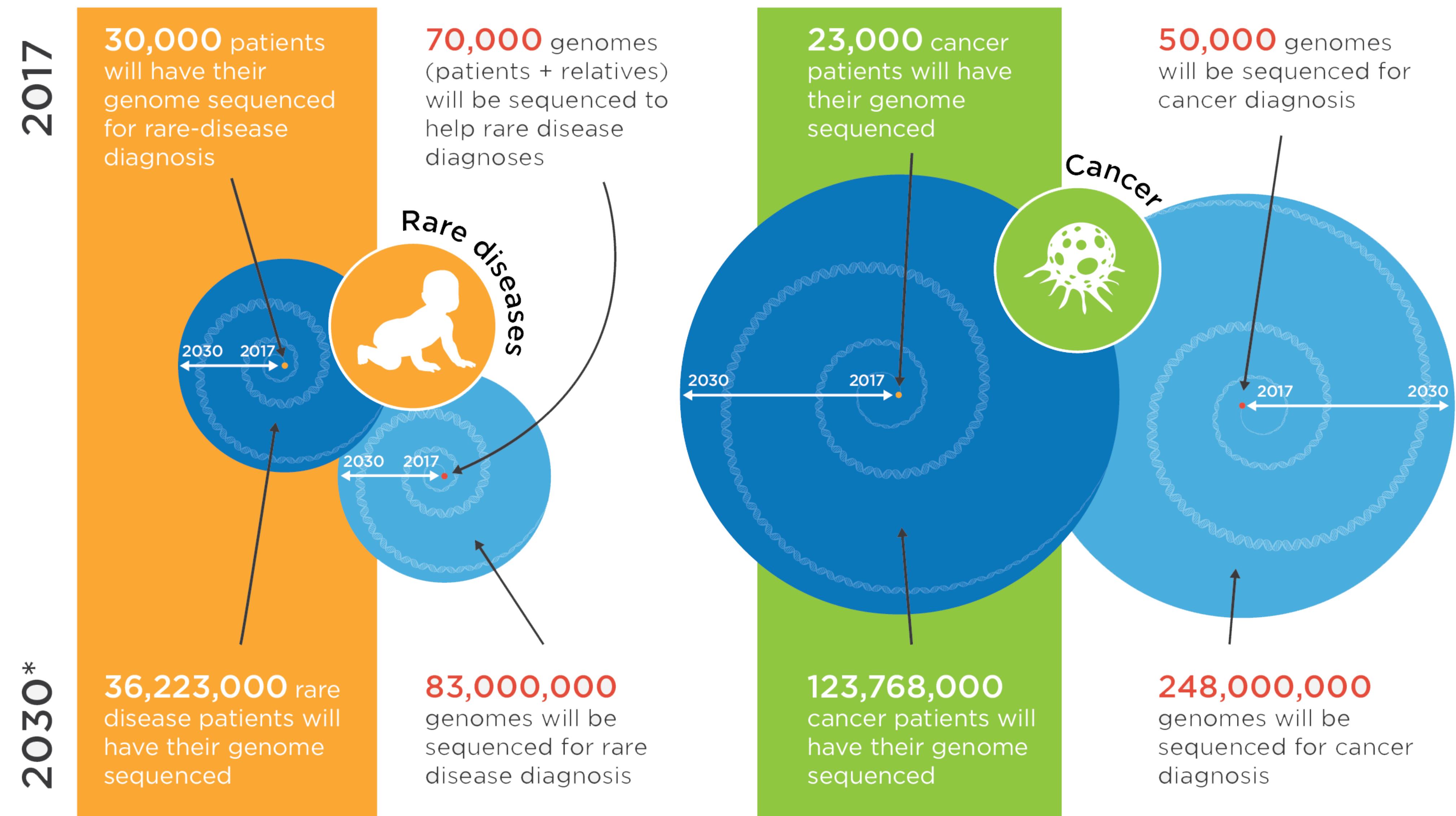
BETTER, CHEAPER, FASTER
The cost of DNA sequencing has dropped dramatically over the past decade, enabling many more applications.



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





Enabling genomic data sharing for the benefit of human health

The Global Alliance for Genomics and Health (GA4GH) is a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a **human rights framework**



**Genomic Data
Toolkit**



**Regulatory & Ethics
Toolkit**



**Data Security
Toolkit**



[VIEW OUR LEADERSHIP](#)

[MORE ABOUT US](#)

[BECOME A MEMBER](#)

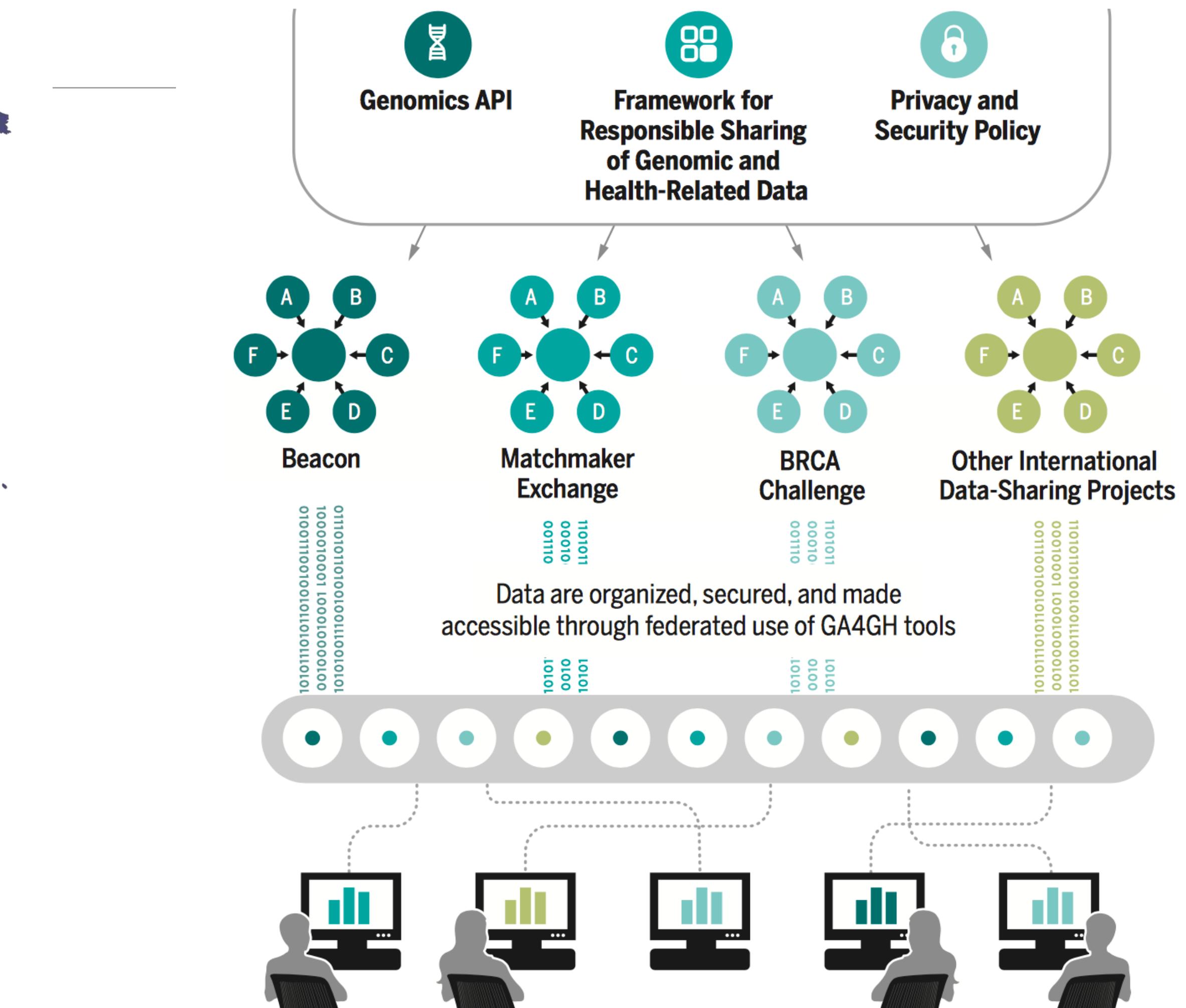


GENOMICS

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

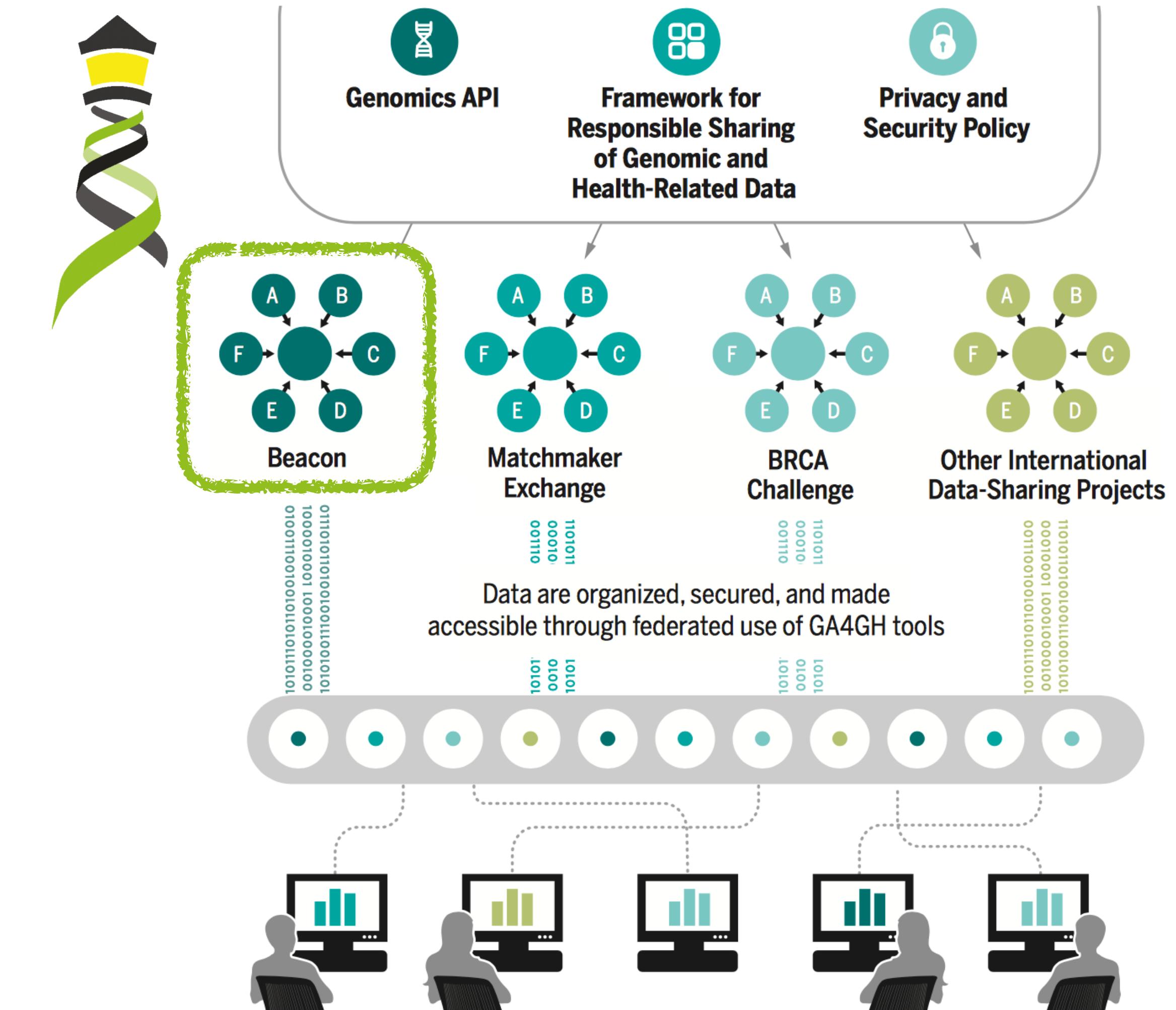
A federated ecosystem for sharing genomic, clinical data

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

The Global Alliance for Genomics and Health*

SCIENCE 10 JUNE 2016 • VOL 352 ISSUE 6264

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



DNASTACK



Global Alliance
for Genomics & Health

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.

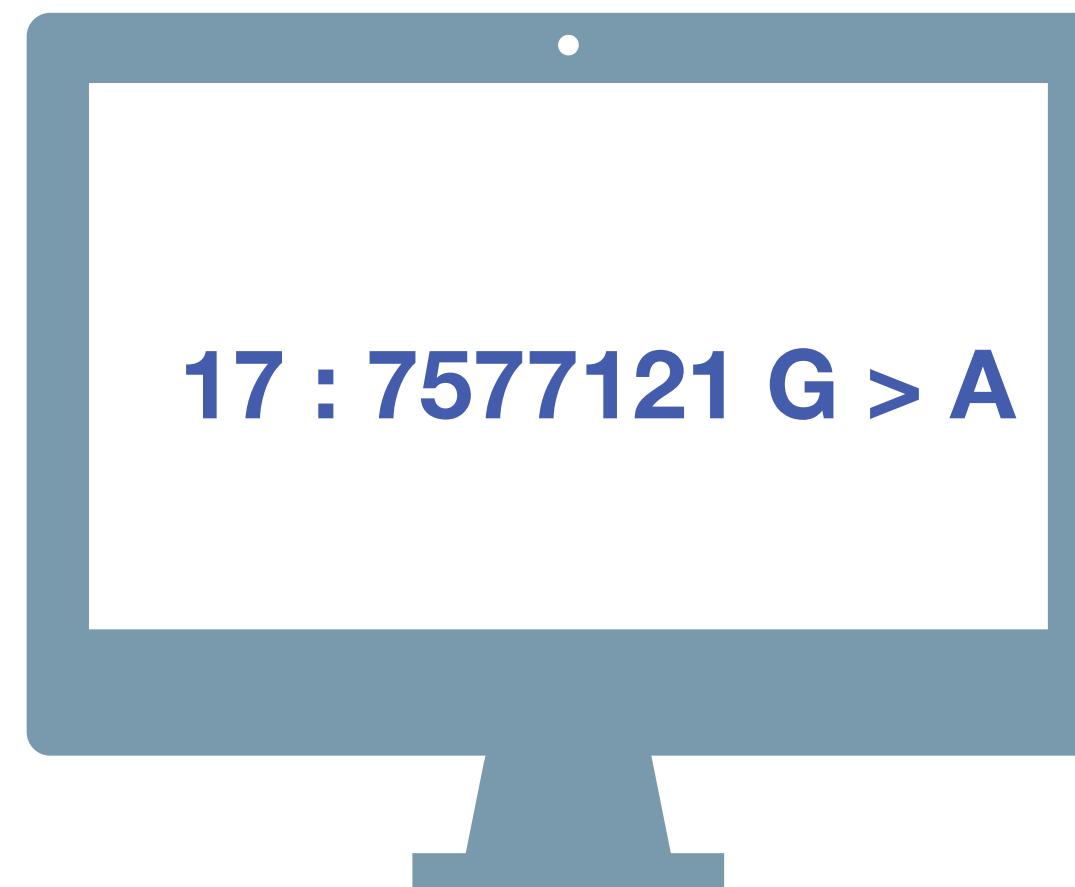
A number of more useful first versions have been suggested.

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*** ...



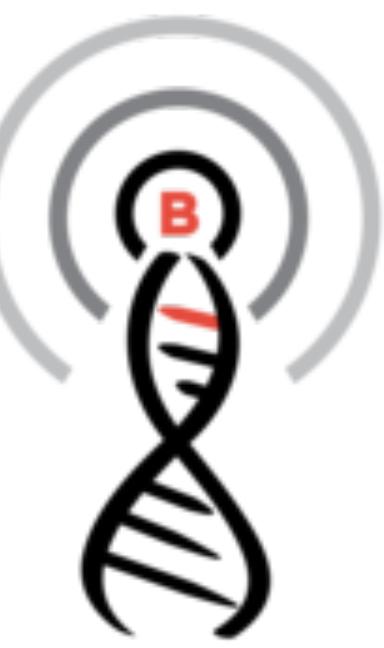
Beacon

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

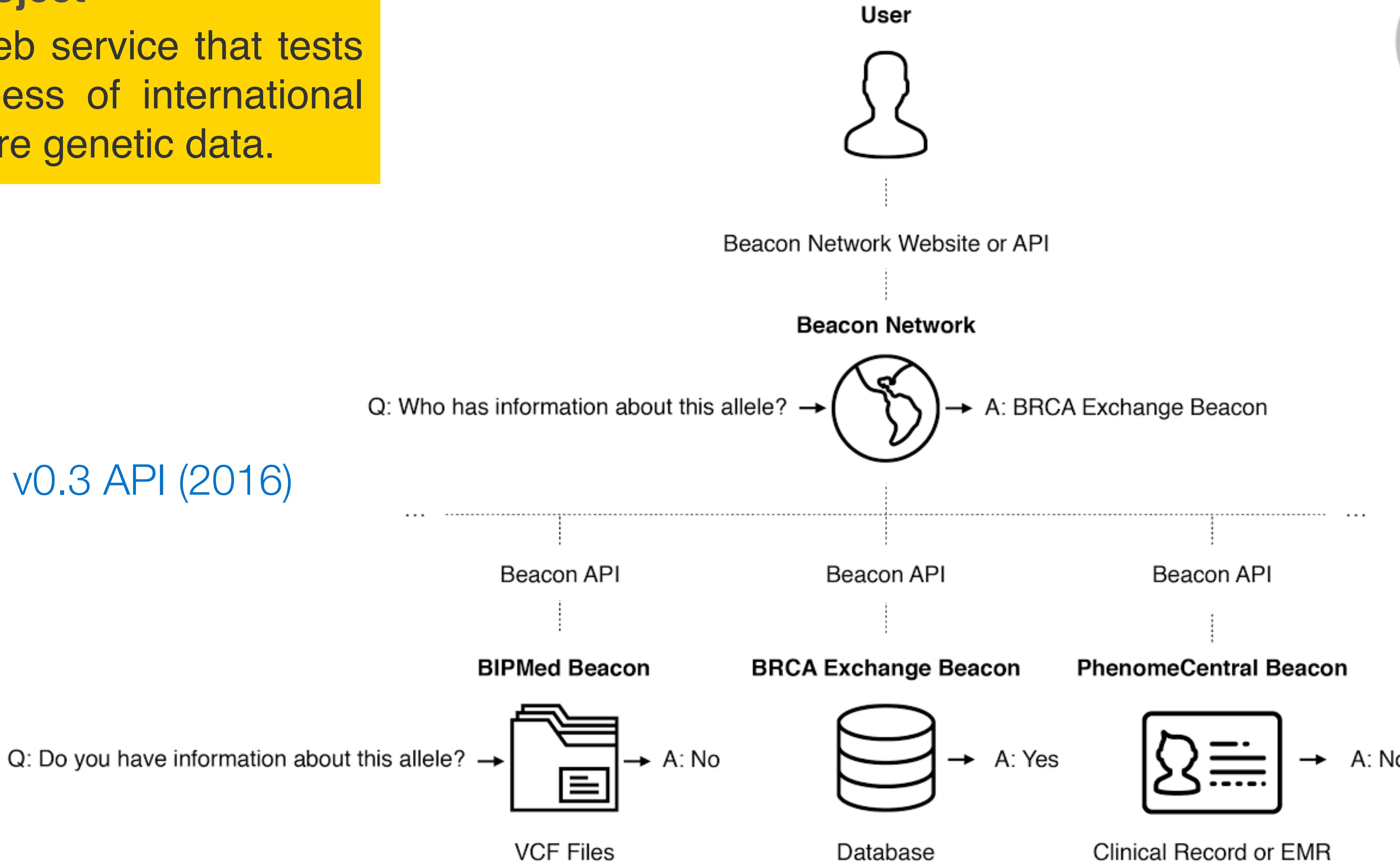
YES | NO | \0

Beacon Project

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



Beacon v0.3 API (2016)



Beacon Project

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



Beacon Network

Search Beacons

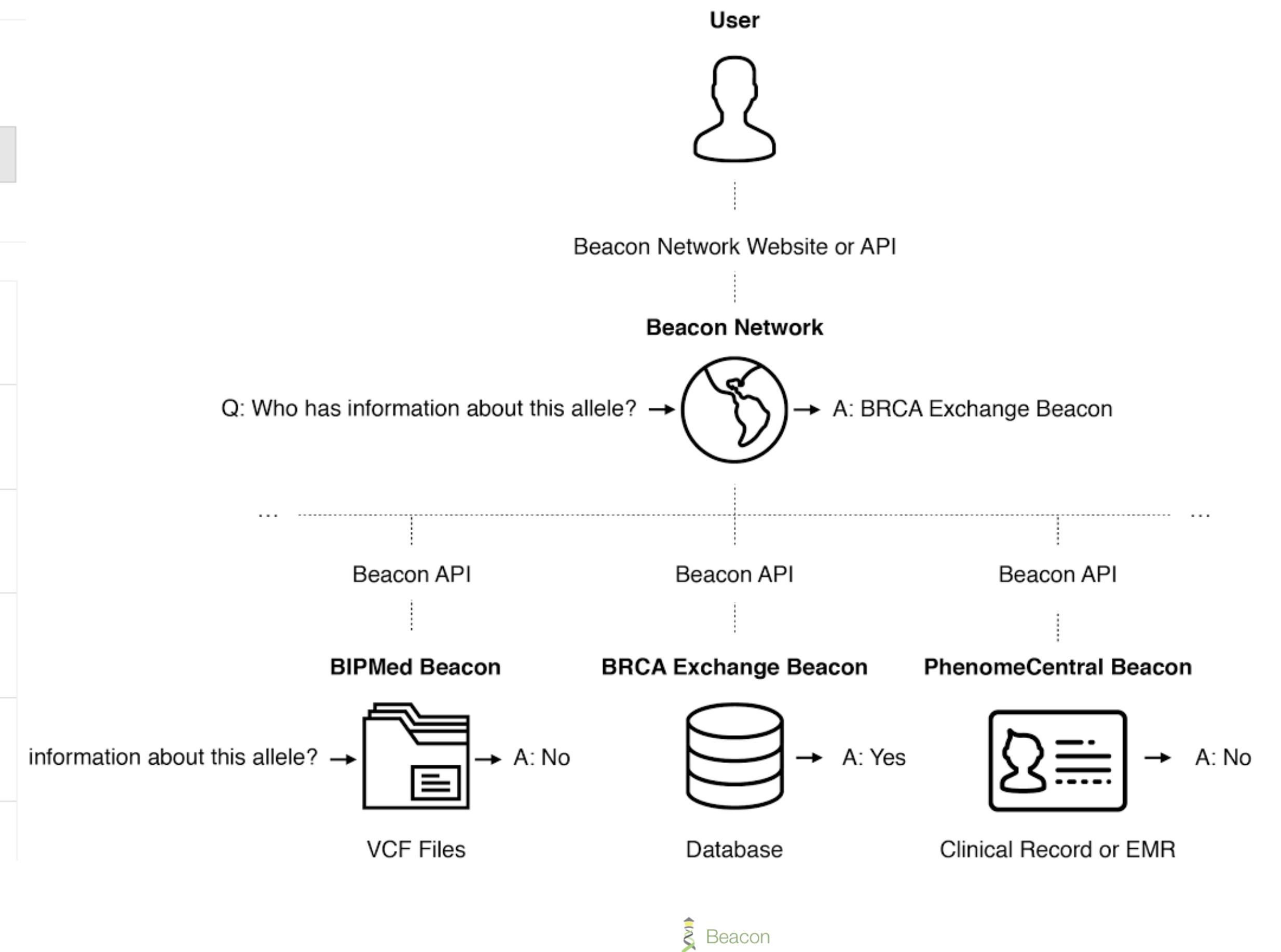
Search all beacons for allele

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 ...

		Found
	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



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

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)



ELIXIR Beacons

EMBL-EBI



elixir
FINLAND

elixir
SWEDEN

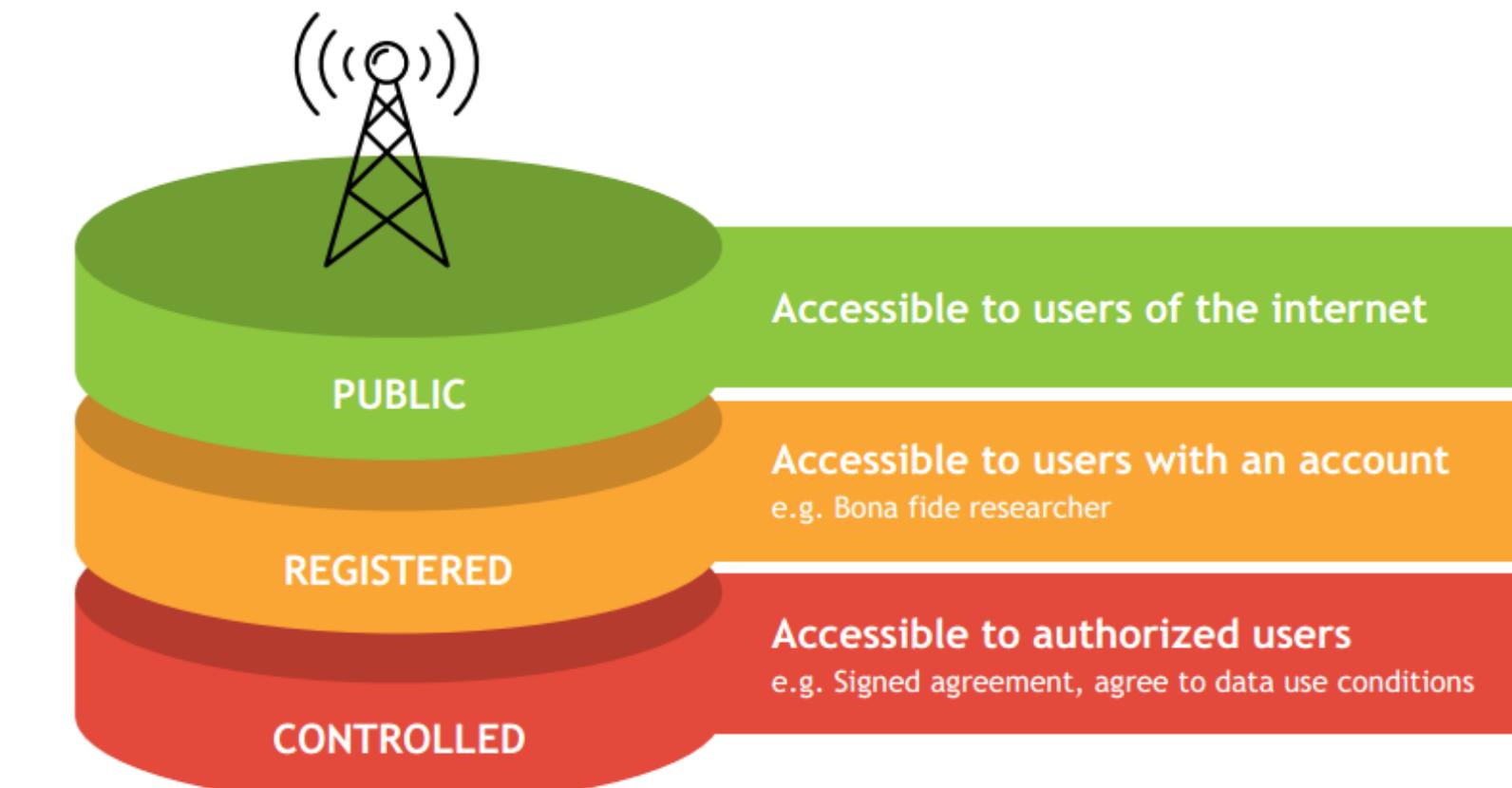
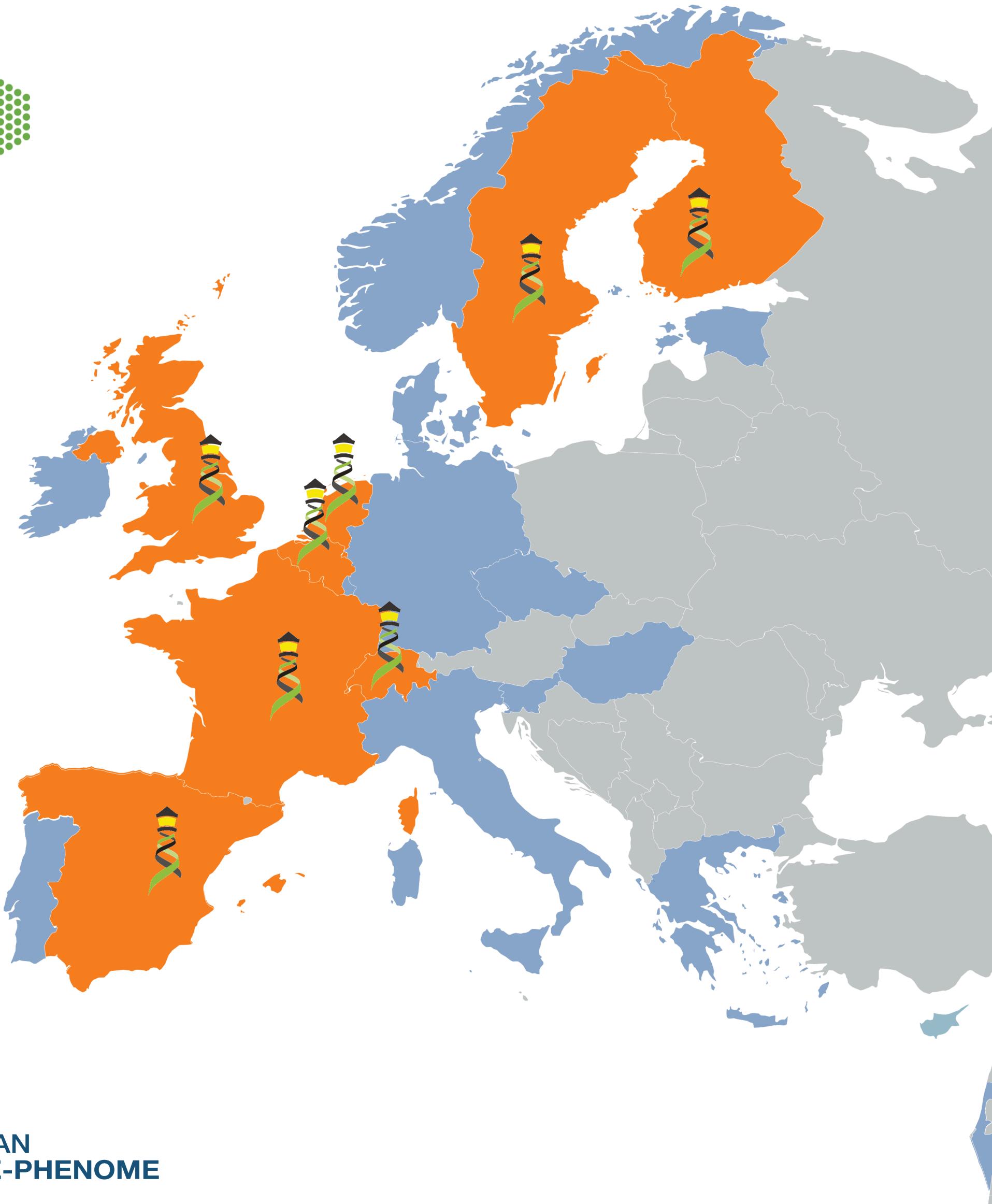
elixir
BELGIUM

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 Search Advanced Search

Example variant query

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](#)

CSC - IT Center for Science
Development Beacon

Beacon API Web Server based on the GA4GH Beacon API

[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 is based on the GA4GH Beacon API v1.1.0

[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](#)

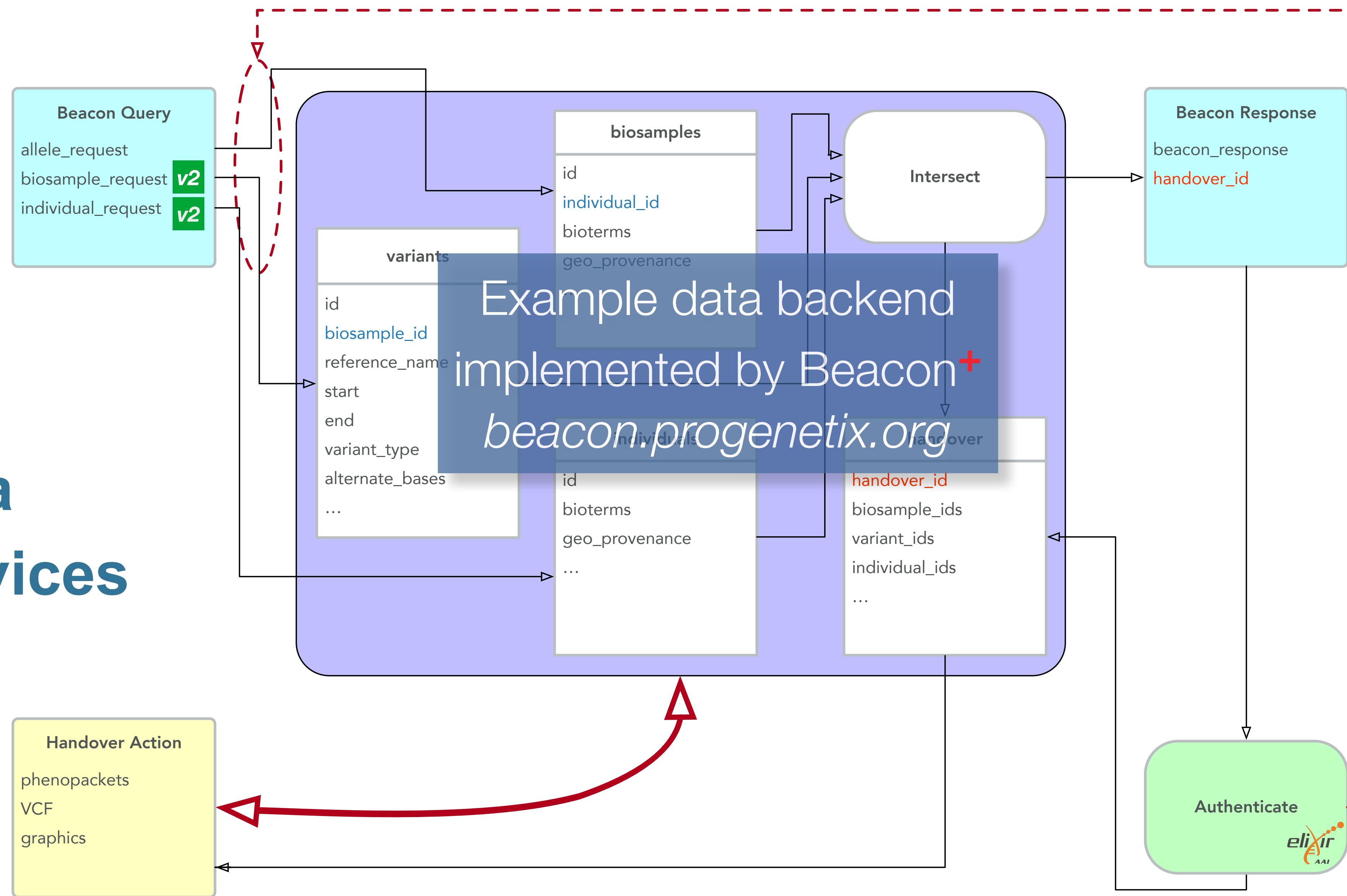
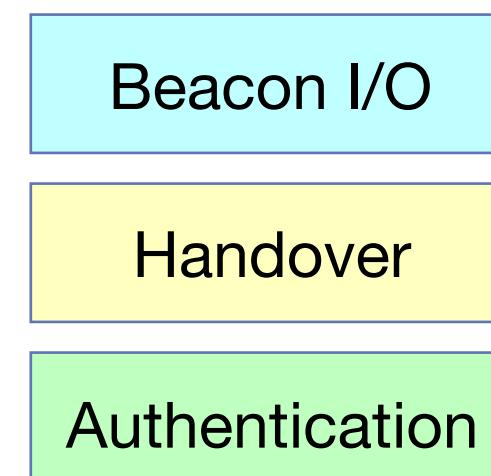
CSC - IT Center for Science
Production Beacon

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



```
{
  "alleleRequest": {
    "endMax": "26000000",
    "referenceName": "9",
    "startMax": "21975098",
    "endMin": "21967753",
    "startMin": "18000000",
    "alternateBases": "N",
    "variantType": "DEL",
    "referenceBases": "*"
  },
  "url": "https://beacon.progenetix.org/beacon/info/",
  "beaconId": "progenetix-beacon",
  "datasetAlleleResponses": [
    {
      "externalUrl": "https://beacon.progenetix.org/beacon/info/",
      "datasetId": "arraymap",
      "variantCount": 588,
      "info": {
        "distinctVarCount": 551,
        "description": "The query was against database \"arraymap\", variant collection \"variants\". 588 matched callsets for 588 distinct variants.",
        "error": null,
        "exists": true,
        "datasetHandover": [
          {
            "url": "https://beacon.progenetix.test/beaconplus-server/beacondeliver.cgi?do=biosamplesdata&accessid=5d76f88d-4012-11e9-a0b4-d9893b611ec4",
            "handoverType": { "label": "Biosamples", "id": "pgx:handover:biosamplesdata" },
            "description": "retrieve data of the biosamples matched by the query"
          },
          {
            "url": "https://beacon.progenetix.test/beaconplus-server/beacondeliver.cgi?do=callsetsvariants&accessid=5d77fb88-4012-11e9-a0b4-bb5a9c8cf98a",
            "description": "export all variants of matched callsets - potentially huge dataset...",
            "handoverType": { "label": "Callsets Variants", "id": "pgx:handover:callsetsvariants" }
          },
          {
            "handoverType": { "id": "pgx:handover:cnvhistogram", "label": "CNV Histogram" },
            "description": "create a CNV histogram from matched callsets",
            "url": "https://beacon.progenetix.test/beaconplus-server/beacondeliver.cgi?do=cnvhistogram&accessid=5d77fb88-4012-11e9-a0b4-bb5a9c8cf98a"
          },
          {
            "handoverType": { "label": "Variants", "id": "pgx:handover:variantsdata" },
            "description": "retrieve data of the variants matched by the query",
            "url": "https://beacon.progenetix.test/beaconplus-server/beacondeliver.cgi?do=variantsdata&accessid=5d6e982b-4012-11e9-a0b4-c5ce5cc21906"
          }
        ],
        "callCount": 588,
        "varResponses": [
          "9:21773941-21968713:DEL",
          "9:21732467-23813102:DEL",
          "9:21785019-21968713:DEL",
          "9:21968713-22031006:DEL"
        ]
      }
    }
  ]
}
```

Beacon Handover

- only exposure of access handle to data stored in secure system
- one-step authentication and selection of *handover* action; other scenarios possible / likely
- *handover response* **outside of Beacon protocol / system**





FEATURES AND POSSIBILITIES OF THE CURRENT BEACON SPECIFICATION

- ▶ 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



This example shows a core Beacon query, against a specific mutation in the TP53 gene, in cellosaurus, with ClinVar data.

CNV Example SNV Range Example SNV Example ClinVar Example Beacon Help

Dataset*

arraymap
progenetix
cellosaurus
dipg
BeaconSpecTest2
BeaconSpecTest

Genome Assembly*

GRCh38 / hg38

Dataset Responses

All Selected Datasets

Reference name*

17

Gene Coordinates

TP53

Cytoband(s)

17p13.1

Start

7673767

Ref. Base(s)

C

Alt. Base(s)

T

Bio-ontology

no selection
NCIT:C102872: Pharyngeal squamous cell carcinoma (2)
NCIT:C103968: Pyruvate dehydrogenase deficiency (1)
NCIT:C105555: High grade ovarian serous adenocarcinoma (75)
NCIT:C105556: Low grade ovarian serous adenocarcinoma (10)
NCIT:C111802: Dyskeratosis congenita (3)

Other Filters

additional comma-separated, prefixed filters

Beacon Query

Beacon+

Flexible Modeling of New Features

Our Beacon platform is being used for the rapid testing of queries and responses - both v1.n and v2.0.a - against a number of partially large-scale genome datasets.

- Progenetix (>100000 cancer CNV profiles)
- DIPG (childhood brain tumor study)
- NEW: Cellosaurus ClinVar annotations for evidence representation
- Brewing: COVID-19

Currently running on a Perl+MongoDB stack, a Python-based OS solution is in early development.



```
{
  "callset_id": "cs-cellosaurus:CVCL_EI02",
  "info": {
    "cellosaurus": {
      "cell_line": "BT474-LAPRa",
      "id": "CVCL_EI02",
      "cellosaurus_variant_name": "TP53 p.Glu285Lys (c.853G>A)"
    },
    "clinvar": {
      "gene_id": "7157",
      "allele_id": "410258",
      "assembly": "GRCh38",
      "cytoband": "17p13.1",
      "variant_type": "single nucleotide variant",
      "origin": "germline;somatic",
      "phenotype": "Hereditary cancer-predisposing syndrome;Li-Fraumeni syndrome;PARP Inhibitor response;not provided",
      "clinical_significance": "Pathogenic/Likely pathogenic",
      "clinvar_full_name": "NM_001126112.2(TP53):c.853G>A (p.Glu285Lys)"
    }
  },
  "start_min": 7673766,
  "reference_name": "17",
  "end_min": 7673767,
  "biosample_id": "bios-cellosaurus:CVCL_EI02",
  "alternate_bases": [
    "T"
  ],
  "digest": "17_7673767_C_T",
  "reference_bases": "C",
  "variantset_id": "cellosaurus_clinvar_GRCH38",
  "end_max": 7673767,
  "start_max": 7673766
},
{
  "digest": "17_7673767_C_T",
  "reference_bases": "C",
  "alternate_bases": [
    "T"
  ],
  "variantset_id": "cellosaurus_clinvar_GRCH38",
  "end_max": 7673767,
  "start_max": 7673766,
  "callset_id": "cs-cellosaurus:CVCL_AQ07",
  "start_min": 7673766,
  "info": {
    "cellosaurus": {
      "cellosaurus_variant_name": "TP53 p.Glu285Lys (c.853G>A)",
      "cell_line": "BT-474 Clone 5",
      "id": "CVCL_AQ07"
    },
    "clinvar": {
      "assembly": "GRCh38",
      "allele_id": "410258",
      "gene_id": "7157",
      "cytoband": "17p13.1",
      "variant_type": "single nucleotide variant",
      "phenotype": "Hereditary cancer-predisposing syndrome;Li-Fraumeni syndrome;PARP Inhibitor response;not provided",
      "origin": "germline;somatic",
      "clinvar_full_name": "NM_001126112.2(TP53):c.853G>A (p.Glu285Lys)",
      "clinical_significance": "Pathogenic/Likely pathogenic"
    }
  },
  "end_min": 7673767,
  "biosample_id": "bios-cellosaurus:CVCL_AQ07",
  "reference_name": "17"
},
{
  "alternate_bases": [
    "T"
  ],
  "reference_bases": "C",
  "digest": "17_7673767_C_T",
  "end_max": 7673767,
  "variantset_id": "cellosaurus_clinvar_GRCH38",
  "start_max": 7673766,
  "callset_id": "cs-cellosaurus:CVCL_0179",
  "info": {
    "cellosaurus": {
      "cellosaurus_variant_name": "TP53 p.Glu285Lys (c.853G>A)",
      "cell_line": "BT-474 Clone 5",
      "id": "CVCL_0179"
    },
    "clinvar": {
      "assembly": "GRCh38",
      "allele_id": "410258",
      "gene_id": "7157",
      "cytoband": "17p13.1",
      "variant_type": "single nucleotide variant",
      "phenotype": "Hereditary cancer-predisposing syndrome;Li-Fraumeni syndrome;PARP Inhibitor response;not provided",
      "origin": "germline;somatic",
      "clinvar_full_name": "NM_001126112.2(TP53):c.853G>A (p.Glu285Lys)",
      "clinical_significance": "Pathogenic/Likely pathogenic"
    }
  }
}
```

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
-
- The diagram uses yellow curly braces to group items. The first group, containing the first four bullet points, is associated with 'v2.0' and a green arrow pointing right, indicating GA4GH approval process ("major product update"). The second group, containing the last three bullet points, is associated with 'v2.n' and a yellow arrow pointing right, indicating incremental rollout after v2.0.
- v2.0
GA4GH approval process ("major product update")
- v2.n
incremental rollout after v2.0
- Tested and already implemented by Beacons
- Ongoing



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



Beacon v2 Filters

An v2 extension of the Beacon protocol will allow the query for additional data beyond genome variants, using a proposed [filters](#) extension. Such filters are thought to be prefixed attributes, where the (public or private) prefix becomes the basis of scoping the value to the correct database value.

Overview of Beacon filters

The Beacon v2 API supports the discovery of genomics and clinical datasets, and includes a powerful feature to enable the “filtering” of beacon responses by biomedical properties (e.g. phenotypes) and procedural metadata.

Filters belong to one of currently three super-classes:

- [Filters](#) correspond to classes from bio-ontologies for biomedical data or procedural metadata that are contained in public repositories such as the [Ontology Lookup Service](#) or NCBO’s [BioPortal](#). Filters are identified using the full bio-ontology term/class identifier as CURIE, e.g. “HP:0100526”. Examples of bio-ontologies are, among others:
 - Human Phenotype Ontology (HPO)
 - [HP:0100526](#): Neoplasm of the lung
 - Phenotypic Qualities Ontology (PATO)
 - [PATO:0020000](#): female genetic sex
 - National Cancer Institute Thesaurus (NCIT)
 - [NCIT:C8430](#): Ovarian Papillary Tumor
 - [NCIT:C48724](#): T2 Stage Finding
 - Experimental Factor Ontology (EFO)
 - [EFO:0009656](#): neoplastic sample
- [CustomFilters](#) are biomedical or metadata terms that are locally defined by a Beacon (e.g. not corresponding to known bio-ontology terms). As with standard [Filters](#), the Beacon v2 API is agnostic to how [CustomFilters](#) are implemented by the Beacon, so this permits maximum flexibility with regards to identifying, labelling and grouping CustomFilters. For example, related phenotype terms or experimental sets could be grouped into local “dictionaries”, which could be addressed through a local identifier. The only requirement is that each custom filter term contains a unique identifier that can be used in Beacon requests.
- [FuzzyFilters](#) are implementations of classifiers which allow for some alternatives in matching and mostly can be drop-ins where ontologies are incomplete. Logically, through the potential matching of multiple values they provide a limited alternative mechanism to allow [OR](#)-style queries.

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



GA4GH :: SchemaBlocks

An Initiative by Members of the Global Alliance for Genomics and Health

About {S}[B]

News

Participants

Standards

Schemas

Examples, Guides & FAQ

Meeting minutes

Contacts

Related Sites

GA4GH

GA4GH::Discovery

Beacon Project

Phenopackets

GA4GH::CLP

GA4GH::GKS

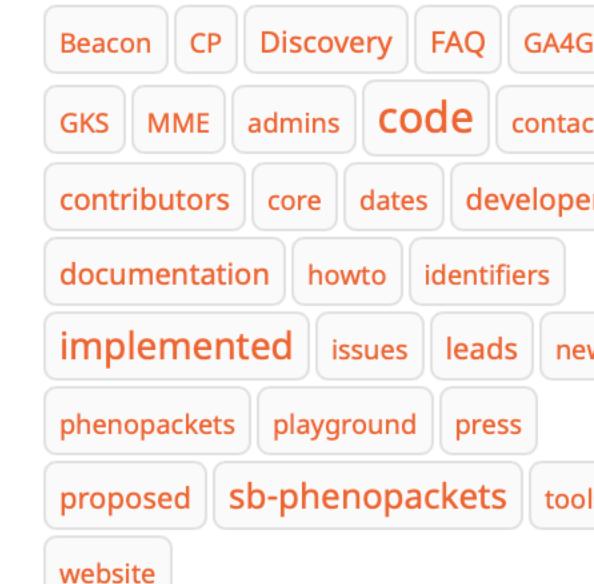
Beacon+

GitHub Projects

SchemaBlocks

ELIXIR Beacon

Tags



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

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 ...](#)

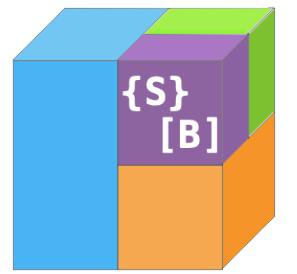
SchemaBlocks {S}[B] Mission Statement

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 ...](#)





BeaconAlleleRequest beacon ↗

{S}[B] Status [i]	implemented
Provenance	◦ Beacon API
Used by	◦ Beacon ◦ Progenetix database schema (Beacon+ backend)
Contributors	◦ Marc Fiume ◦ Michael Baudis ◦ Sabela de la Torre Pernas ◦ Jordi Rambla ◦ Beacon developers...
Source (v1.1.0)	◦ raw source [JSON] ◦ Github

Attributes

Type: object

Description: Allele request as interpreted by the beacon.

Properties

Property	Type
alternateBases	string
assemblyId	string
datasetIds	array of string
end	integer
endMax	integer
endMin	integer
mateName	https://schemablocks.org/schemas/beacon/v1.1.0/Chromosome [HTML]
referenceBases	string
referenceName	https://schemablocks.org/schemas/beacon/v1.1.0/Chromosome [HTML]
start	integer (int64)
startMax	integer
startMin	integer
variantType	string

alternateBases

- type: string

The bases that appear instead of the reference bases. Accepted values: [ACGTN]*. N is a wildcard, that denotes the position of any base, and can be used as a standalone base of any type or within a partially known sequence. For example a sequence where the first and last bases are known, but the middle portion can exhibit countless variations of [ACGT], or the bases are unknown: ANNT the Ns can take any form of [ACGT], which makes both ACCT and ATGT (or any other combination) viable sequences.

Symbolic ALT alleles (DEL, INS, DUP, INV, CNV, DUP:TANDEM, DEL:ME, INS:ME) will be represented in variantType.

Optional: either alternateBases or variantType is required.

alternateBases Value Example

assemblyId

- type: string

Assembly identifier (GRC notation, e.g. GRCh37).

assemblyId Value Example

Curie sb-vr-spec ↗

{S}[B] Status [i]	implemented
Provenance	◦ vr-spec
Used by	◦ vr-spec
Contributors	◦ Reece Hart ◦ Michael Baudis

Attributes

Type: object

Description: A CURIE is a Uniform Resource Identifier (URI) that identifies a single entity. It consists of a prefix followed by a namespace and a local identifier. The prefix is typically a well-known identifier for a specific domain, such as 'http://www.w3.org/2002/07/owl#' for the Web Ontology Language (OWL). The namespace is a URI that identifies the vocabulary or ontology being used. The local identifier is a unique identifier within that vocabulary.

VR does not impose any constraints on strings used as identifiers, the VR Specification RECOMMENDS that implementers use standard CURIEs. String CURIEs are represented as [prefix:reference](#) (W3C REC-2014-03-19) or [namespace:accession](#) or [namespace:local_id](#) colloquially.

The VR specification also RECOMMENDS that [prefix](#) be the [reference](#) component is an unconstrained string. Several instances (e.g. technical replicates) or types of experiments (e.g. genomic array as well as sequencing) 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]

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

Attributes

Type: object

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

Examples would be a tissue biopsy, a single cell from a culture for single cell genome sequencing, a fraction from a gradient centrifugation.

Several instances (e.g. technical replicates) or types of experiments (e.g. genomic array as well as sequencing) may refer to the same Biosample.

FHIR mapping: [Specimen](#).

Properties

Property	Type
checksum	<ul style="list-style-type: none"> • type: string <p>The hexadecimal encoded (Base16) checksum for the data.</p>
checksum Value Example	"77af4d6b9913e693e8d0b4b294fa62ade6054e6b2f1ffb617ac955dd63fb0182"
type	<ul style="list-style-type: none"> • type: string <p>The digest method used to create the checksum. The value (e.g. sha-256) SHOULD be listed as Hash Name String in the GA4GH Hash Algorithm Registry. Other values MAY be used, as long as implementors are aware of the issues discussed in RFC6920.</p> <p>GA4GH may provide more explicit guidance for use of non-IANA-registered algorithms in the future.</p>
type Value Example	"sha-256"

Checksum sb-checksum ↗

{S}[B] Status [i]	proposed
Provenance	◦ GA4GH DRS ('develop' branch)
Used by	◦ GA4GH DRS ◦ GA4GH TRS
Contributors	◦ Susheel Varma

Attributes

Type: object

Description: Checksum

Properties

Property	Type
checksum	string
type	string

The CRG "COVID" Beacon

Beyond humans and limited data delivery



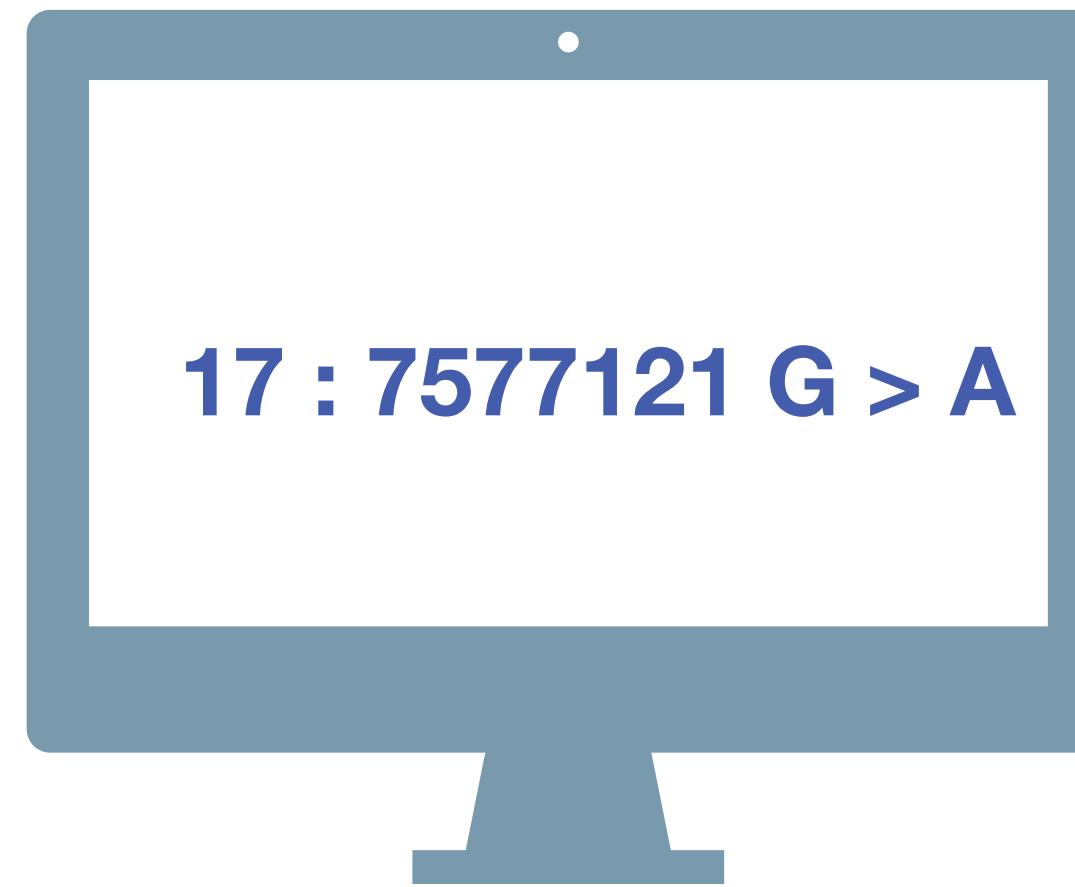
CRG Viral Beacon - SNP Query

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

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



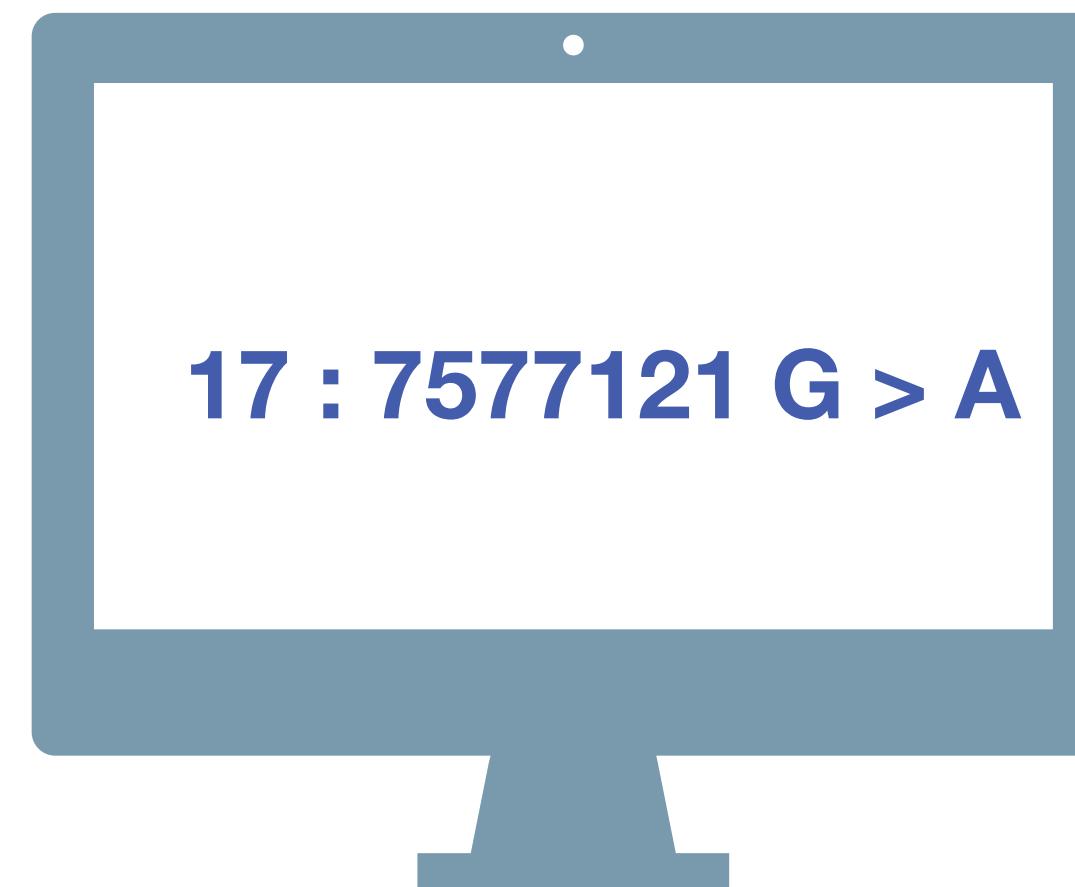
Variant Annotation							Variant Frequency per Dataset					
VARIANT TYPE	GENE REGION	EFFECT	FUNCTIONAL CLASS	LOCUS NAME	LOCUS ID	AMINOACID CHANGE	ENA-READS	ONT	GISAID	ENA-CONSENSUS		
SNP	CODING	SYNONYMOUS_CODING	SILENT	orf1ab,ORF1ab	GU280_gp01	S2839	0.6772	0.0334	0.1955	0.0844		
Metadata (# of results: 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	-



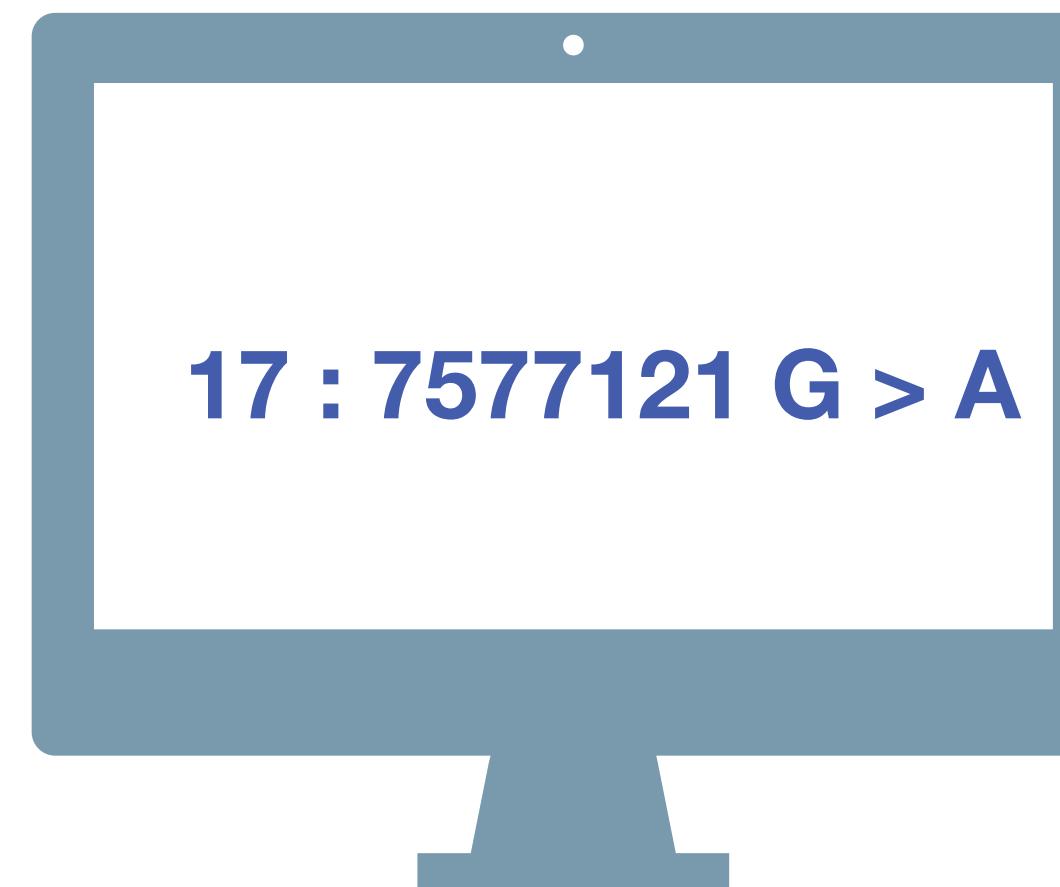
Beacon

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

YES | NO | \0



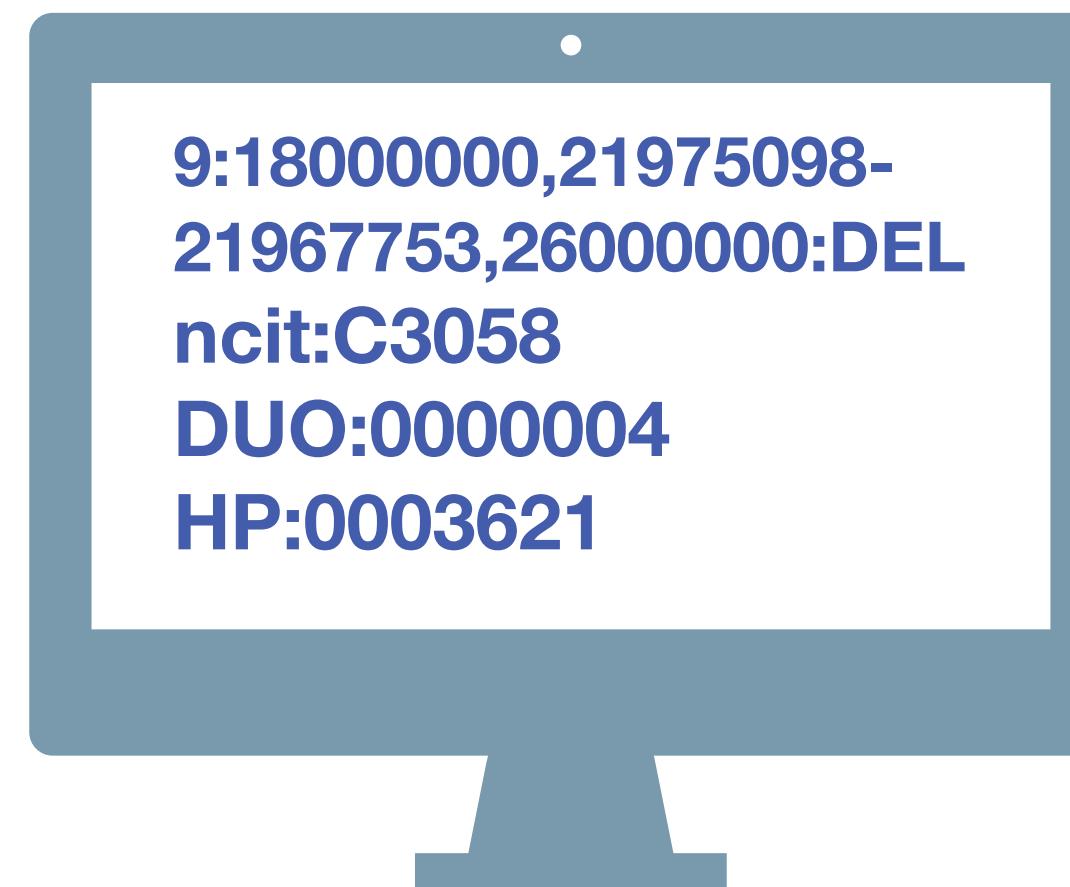
A Beacon network federates
genome variant queries
across databases that
support the ***Beacon API***



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**.

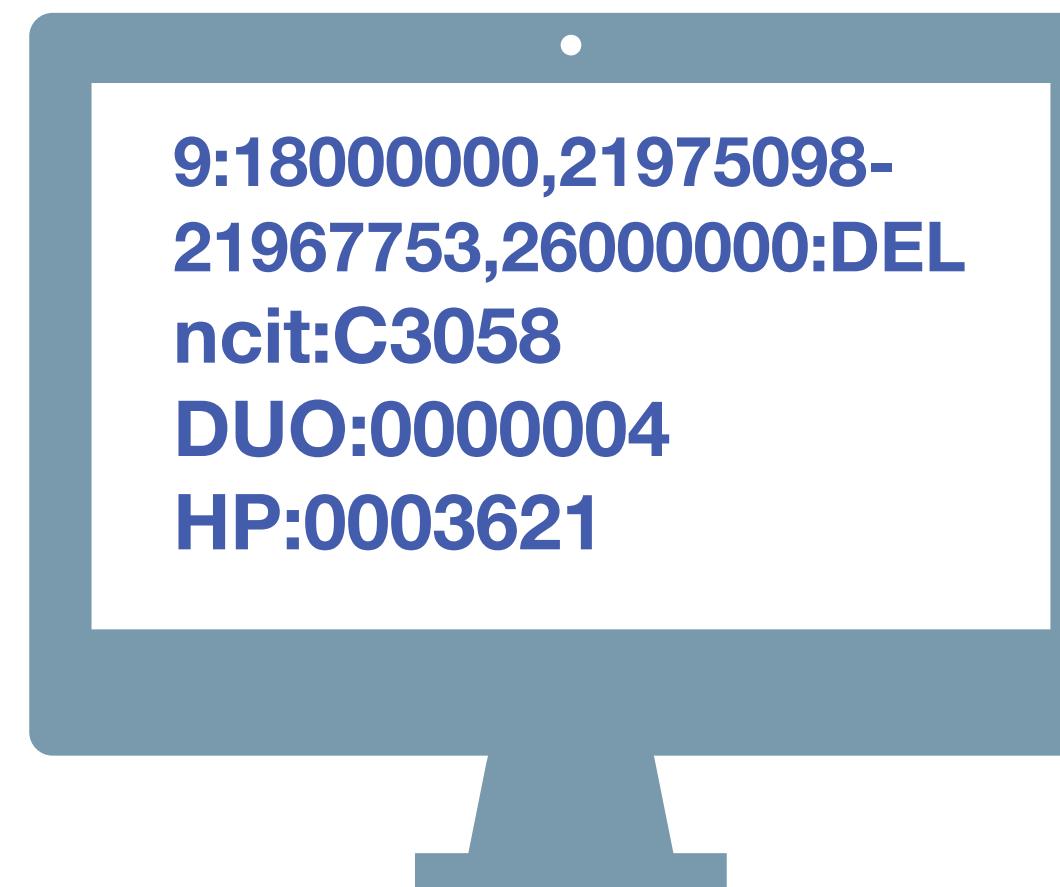


Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?



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".



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**





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 as a key driver project of the **Global Alliance for Genomics and Health (GA4GH)**. The **Beacon** protocol itself defines an open standard for genomics data discovery. 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.



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 of a variant
- **Privacy protecting:** queries do not return information about single individuals

Sites offering *beacons* can scale through aggregation **Beacon Networks**, which distribute single genome queries among a potentially large number of international *beacons* and assemble their responses.

Since 2015 the development of the Beacon protocol has been led by **ELIXIR** in collaboration with GA4GH and international participants. Recent versions of the *Beacon* protocol have expanded the original concept by e.g.:

- providing a framework for other types of genome variation data (i.e. [range queries](#) and [structural variants](#))
- allowing for data delivery using [handover](#) protocol, e.g. to link with clinical information in protected environments and allow for data delivery and visualisation services

Beacon v2 - Towards Flexible Use and Clinical 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".

beacon-project.io

Beacon+

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 "focal" hits (here i.e. $\leq \sim 4\text{Mbp}$ in size). The query is against the arrayMap collection and can be modified e.g. through changing the position parameters or data source.

Query		BND Example	SNV Example	SNV Range Example	CNV Example
Dataset*	toga				
Reference name*	5				
Genome Assembly*	GRCh38 / hg38				
(structural) variantType	DEL (Deletion)				
Gene Coordinates	[PDE4D] 5:58974663-60185597				
Start min Position*	56974663				
Start max Position	60185596				
End min Position	58974664				
End max Position	62185597				
Bio-ontology	no selection icdom:94403: Glioblastoma, NOS (2048) icdom:94423: Gliosarcoma (9) ncit:C102872: Pharyngeal squamous cell... (9) ncit:C105555: High grade ovarian serous... (78) ncit:C114656: Endometrial adenosquamous... (4)				
Beacon Query					

Response

Dataset	Assembly	Chro	Position Start Range End Range	Ref Alt Type	Bio Query	Variants Calls Samples	f_alleles	Response Context
arraymap	GRCh38	5	56974663 - 60185596 58974664 - 62185597	*	icdot:C61.9	60 47 47	0.0008	JSON UCSC [H->O] Biosamples [H->O] Callsets Variants [H->O] CNV Histogram [H->O] Individuals [H->O] Variants
toga	GRCh38	5	56974663 - 60185596 58974664 - 62185597	*	icdot:C61.9	35 27 27	0.0012	JSON UCSC [H->O] Biosamples [H->O] Callsets Variants [H->O] CNV Histogram [H->O] Individuals [H->O] Variants

Michael Baudis
Bo Gao
Qingyao Huang
Rahel Paloots
Paula Carrio Cordero
Séverine Duvaud (SIB)

beacon.progenetix.org



ELIXIR Genome

Beacons

A Driver Project of the Global
Alliance for Genomics and
Health

About...

News & Press

Contributors

Events

Examples, Guides & FAQ

Specification

Roadmap

Beacon Networks

Meeting Minutes

Contacts

Related Sites

Beacon @ ELIXIR
GA4GH

Beacon+
beacon-network.org
GA4GH::SchemaBlocks
GA4GH::Discovery
GA4GH::CLP
GA4GH::GKS

Github Projects

ELIXIR Beacon
SchemaBlocks

Tags

EB FAQ contacts definitions
developers development
minutes network press
proposal queries releases
specification versions website



Search or jump to...

Pull requests Issues Marketplace Explore



Roadmap

The ELIXIR Beacon Roadmap delineates short-, mid- and long-term objectives, to expand functional scope and reach of Beacon as a protocol and genomic data ecosystem.

Beacon Flavours

Beacons may be able to increase their functionality through the development of specific flavours. These can extend the core Beacon concept for specific use cases.

@mbaudis 2018-10-24: more ...

Bio-metadata Query Support

Future Beacon API versions will support querying for additional, non-sequence based metadata.

@mbaudis 2018-10-18: more ...

EvidenceBeacon Notes - GA4GHconnect 2018

The topic of "EvidenceBeacon" was discussed with many different attendees during the GA4GHconnect 2018 meeting and beyond, leading to some clearer picture about the (widely) different approaches.

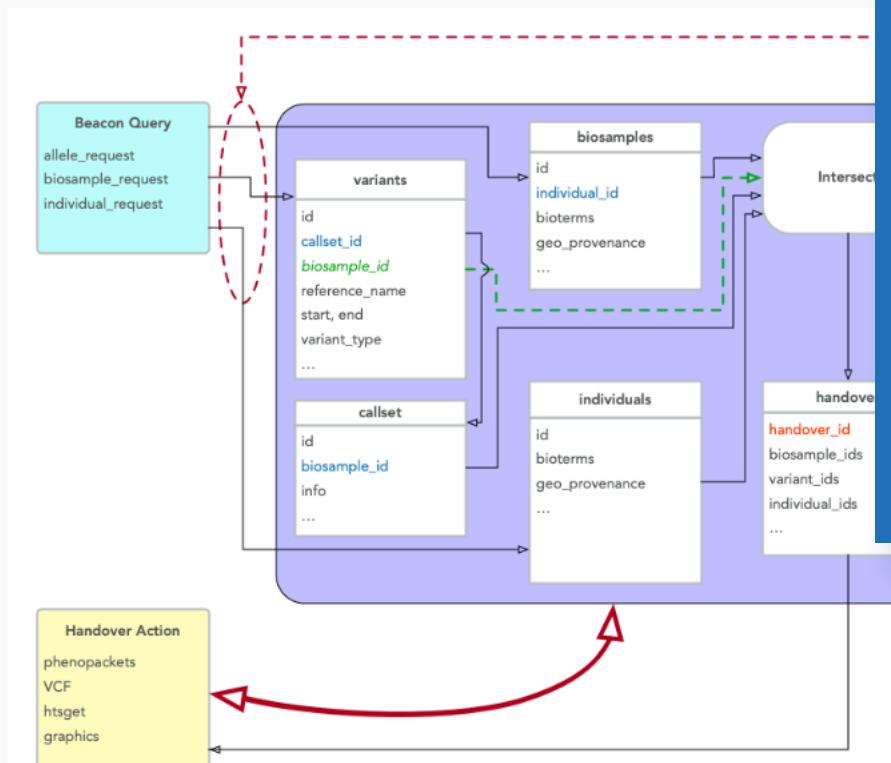
@mbaudis 2019-04-30: more ...

[H→O] Beacon Handover for Data Delivery

While the Beacon response should be restricted to aggregate data (yes/no), the usage of the protocol could be greatly expanded by providing an access mechanism that can be matched by a Beacon query.

As part of the mid-term product strategy, the ELIXIR Beacon team is evaluating the possibility of extending the protocol, in which rich data content (e.g. variant data, phenotypic information, linked services, etc.) can be provided from linked services, initiated through a Beacon query. This would involve several steps like protocol selection, authentication...). A discussion of the topic is ongoing in the developer area on Github (issue #114).

As of 2018-11-13, the **handover** concept has become part of the ongoing work.



beacon-project.io



Beacon

Beacon Project, Global Alliance for Genomics & Health.

API Leads

Jordi Rambla
Anthony Brooks
Juha Törnroos

Discovery WS

Michael Baudis (Beacon)
Marc Fiume (Networks)

ELIXIR

Gary Saunders
David Lloyd
Serena Scollen

openapi beacon ga4gh

Apache-2.0 23 stars 28 issues 41 pull requests 7 comments Updated on May 9

Projects 1 Settings

Customize pinned repositories

specification

GA4GH Beacon specification.

28 stars 23 issues

Type: All ▾ Language: All ▾ New

Top languages

- JavaScript
- Java
- HTML
- PLpgsql

Most used topics

- beacon
- ga4gh

Sabela de la Torre Pernas

Updated 9 days ago

Authenticate

elair logo

github.com/ga4gh-beacon/

github.com/ga4gh-beacon/



University of
Zurich^{UZH}



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



Global Alliance
for Genomics & Health

