



ELIXIR Beacon Project

Concepts | Status | History | Outlook

Michael Baudis :: ELIXIR-CH:: ELIXIR All Hands :: 2020-06-10



Global Alliance
for Genomics & Health

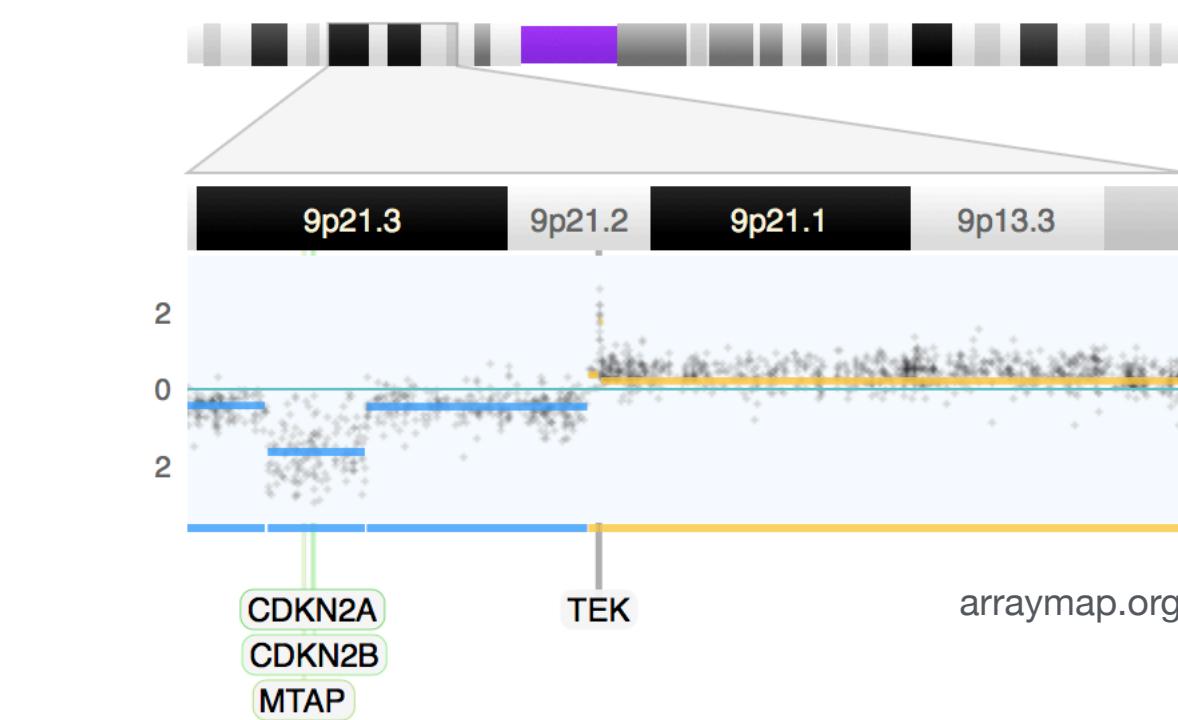


University of
Zurich UZH

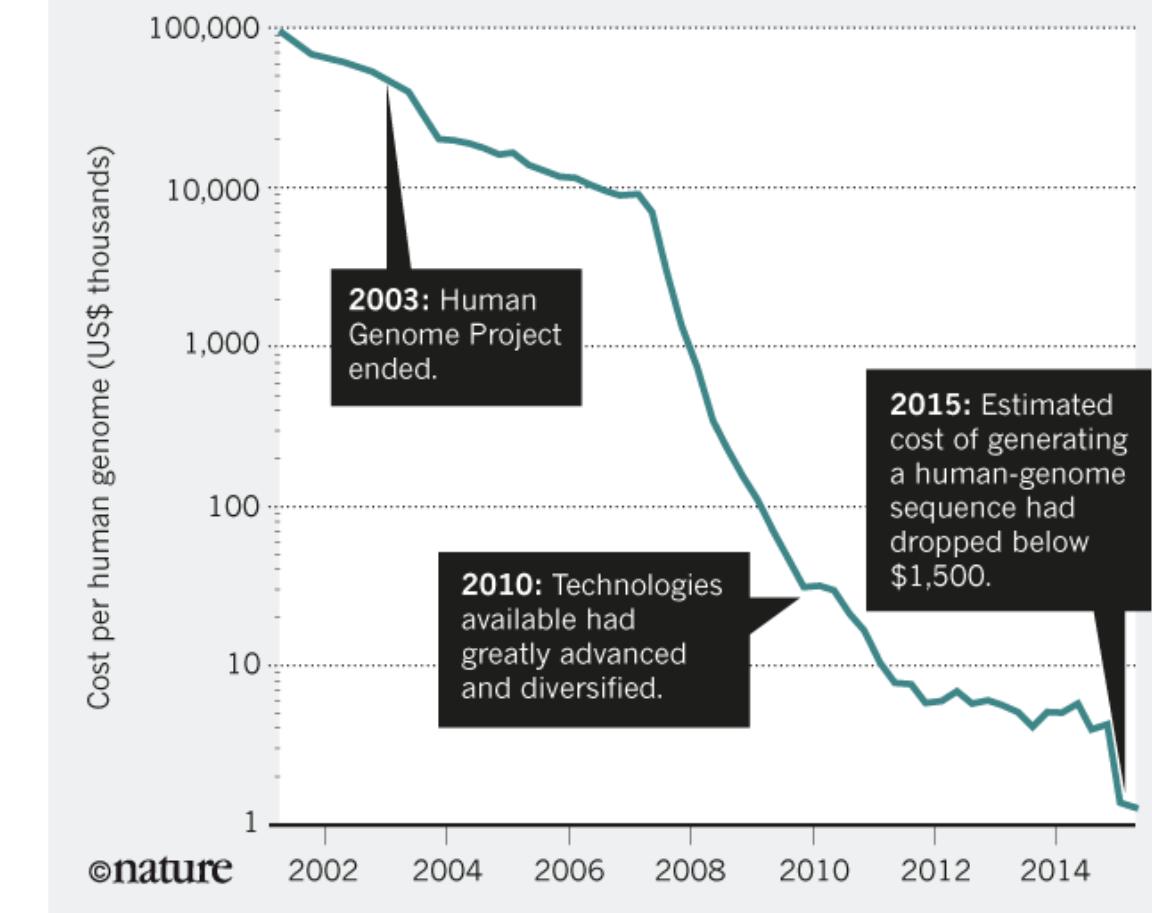


Genome screening at the core of “Personalised Health”

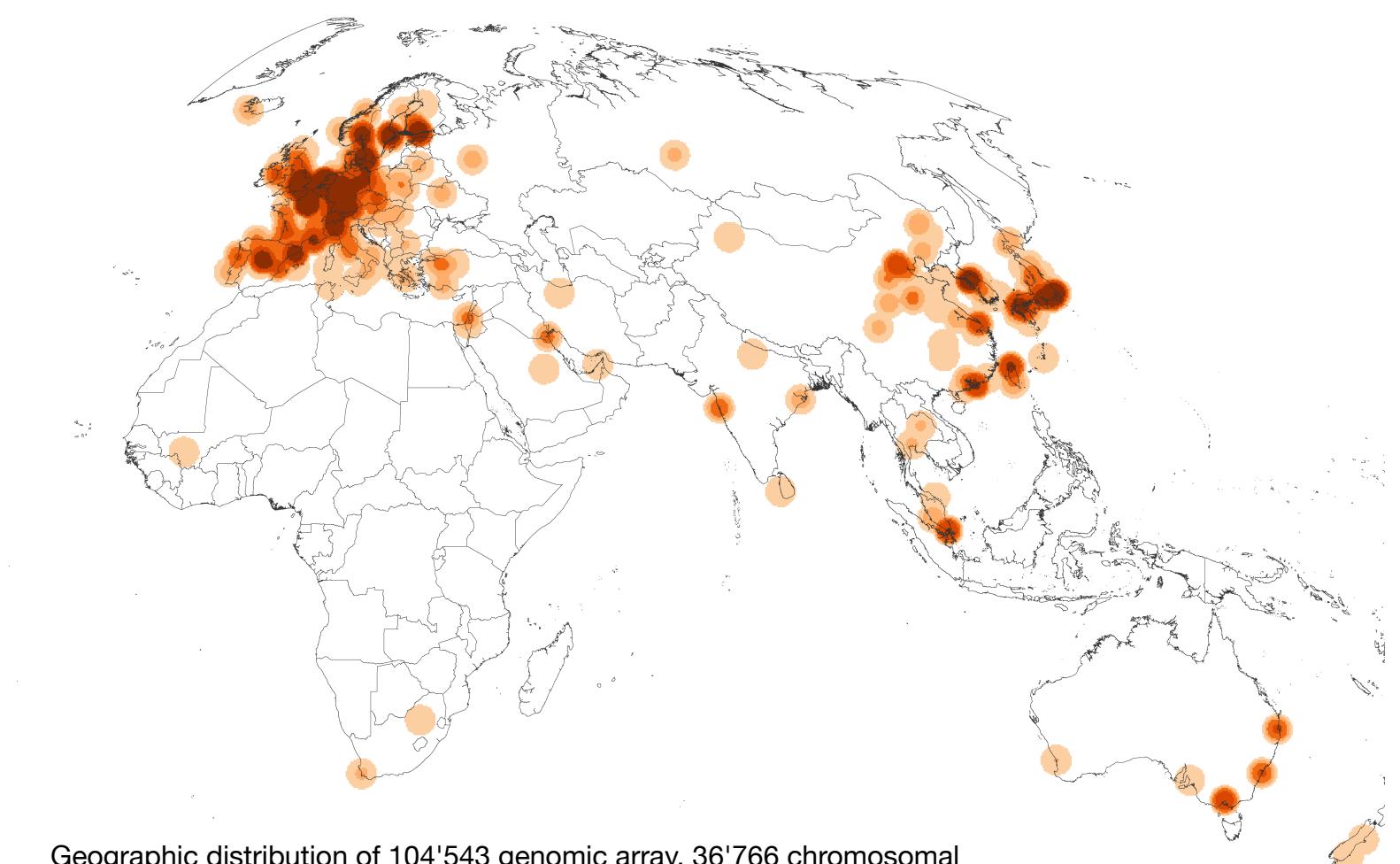
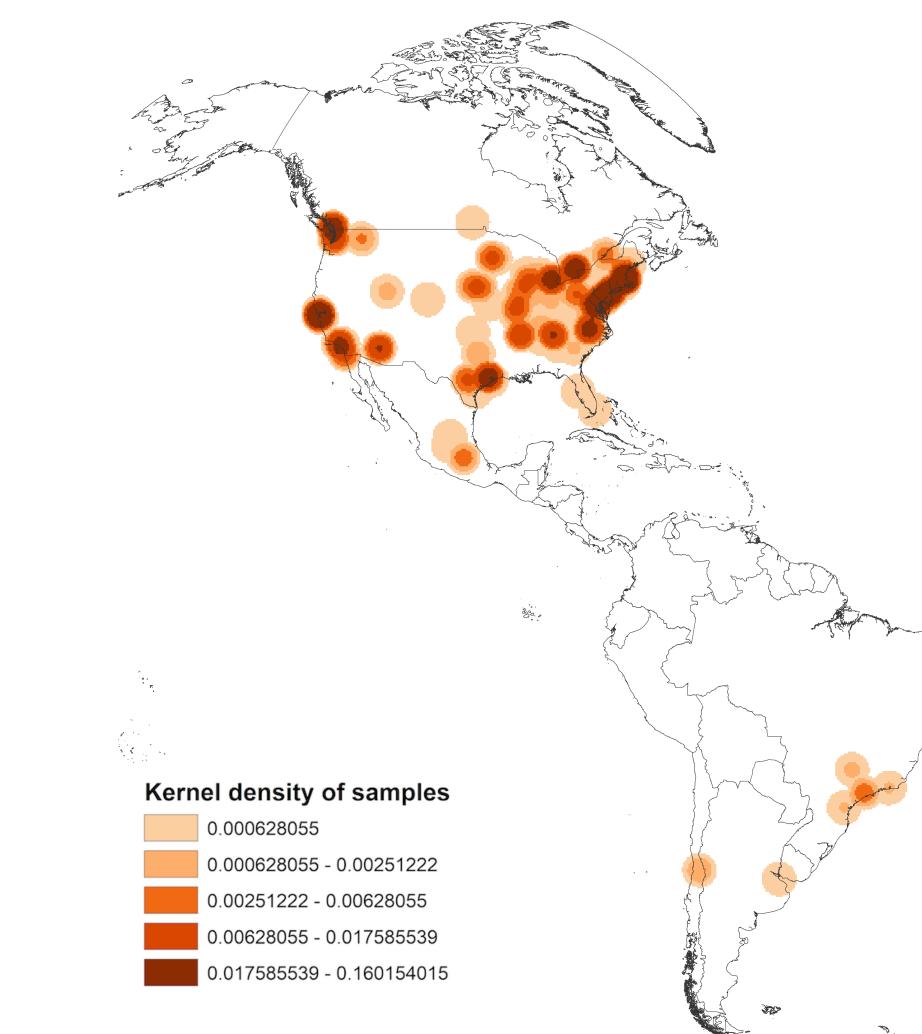
- ▶ **Genome analyses** (including transcriptome, metagenomics) are core technologies for Personalised Health™ applications
- ▶ The unexpectedly large amount of **sequence variants** in human genomes - germline and somatic/cancer - requires huge analysis efforts and creation of **reference repositories**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

The vision: Federation of data



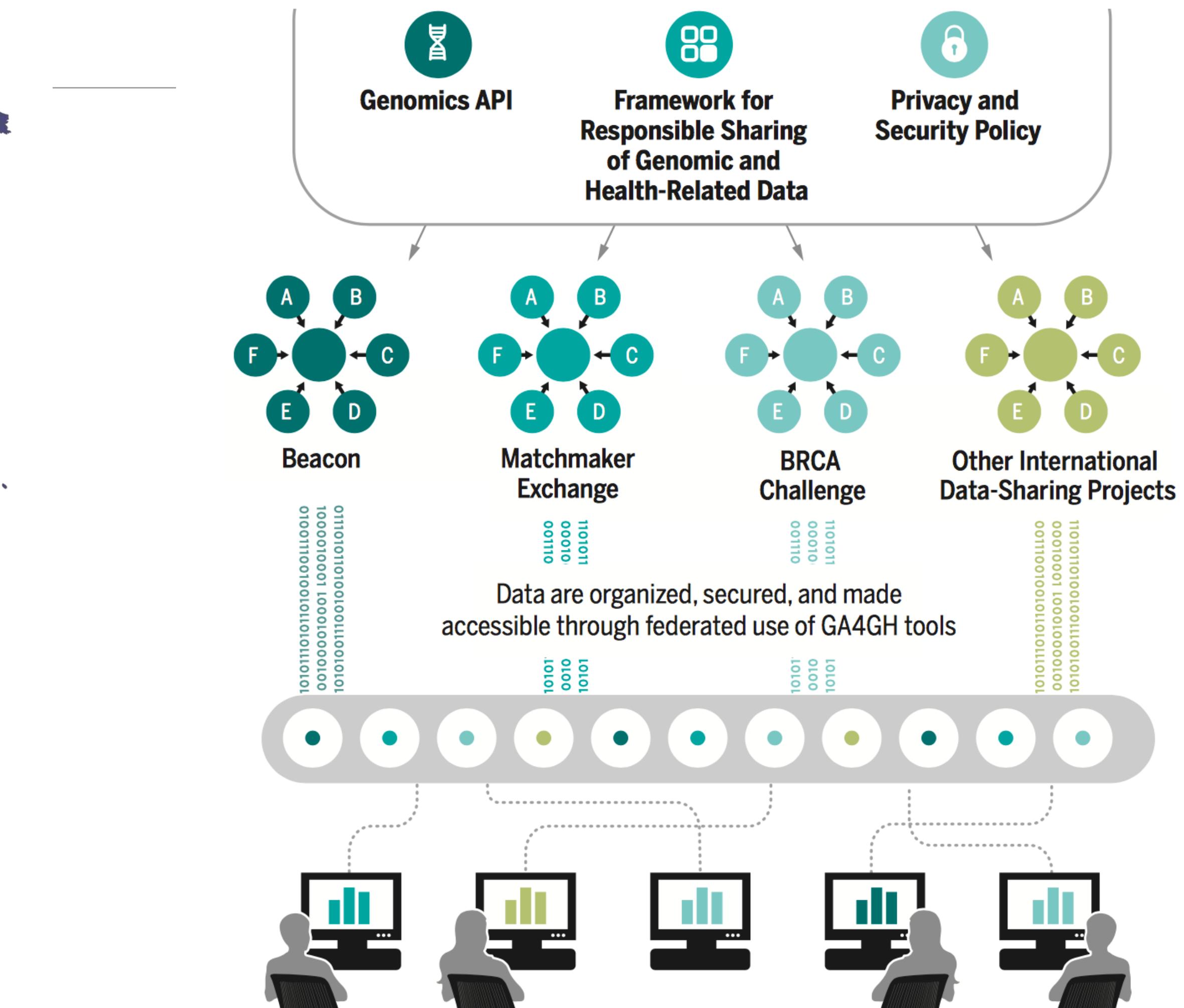


GENOMICS

A federated ecosystem for sharing genomic, clinical data

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

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



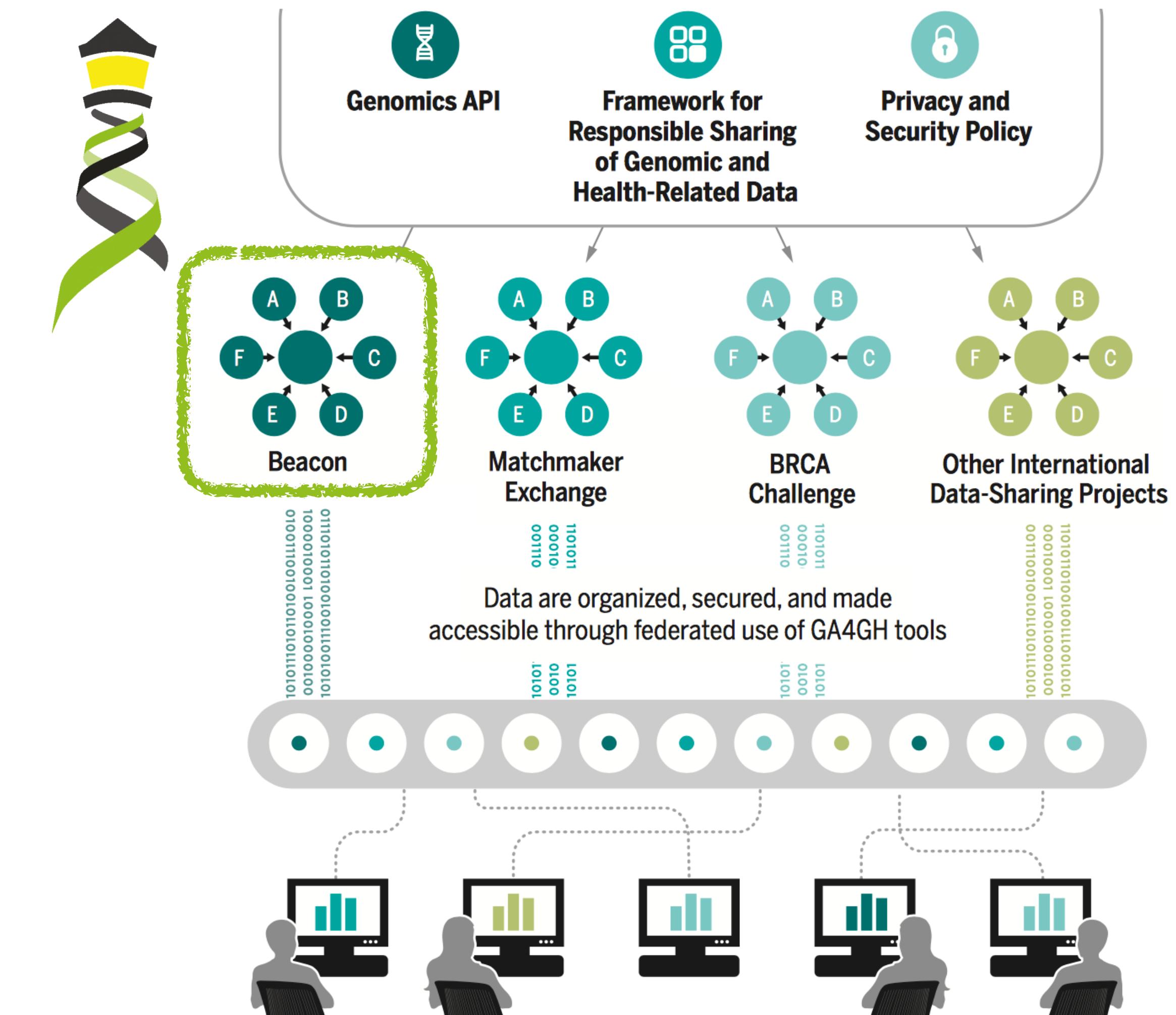


GENOMICS

A federated ecosystem for sharing genomic, clinical data

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

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



Introduction

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

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

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

Utility

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

A number of more useful first versions have been suggested.

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

“I would personally recommend all those be held for version 2, when the beacon becomes a service.”
Jim Ostell, 2014

Implementation

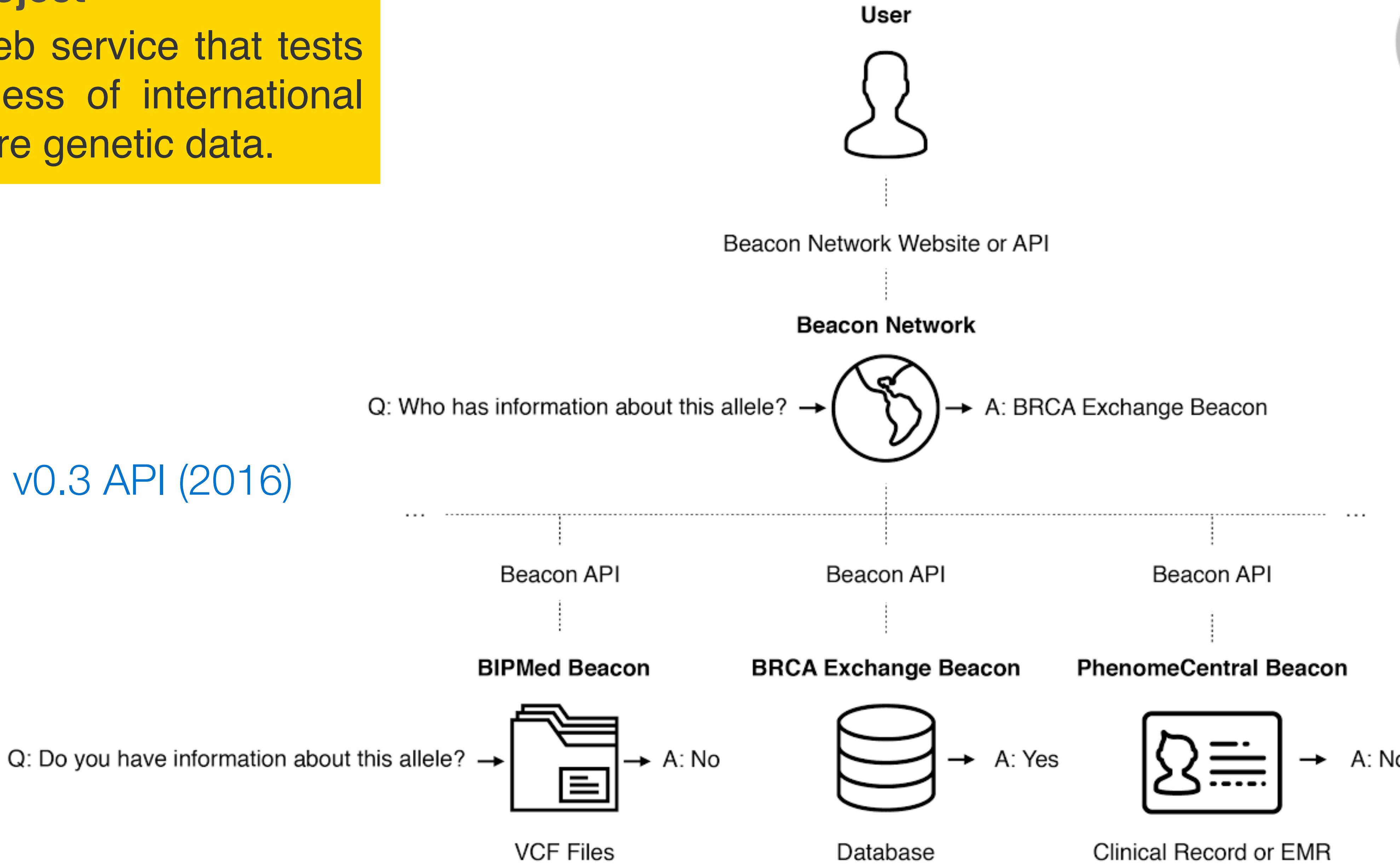
1. Specifying the chromosome ... The interface needs to specify the **accession.version** of a chromosome, or **build number**...
2. Return values ... right to **refuse** to answer without it being an error ... DOS **attack** ... or because ...especially **sensitive**...
3. Real time response ... Some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

Beacon Project

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

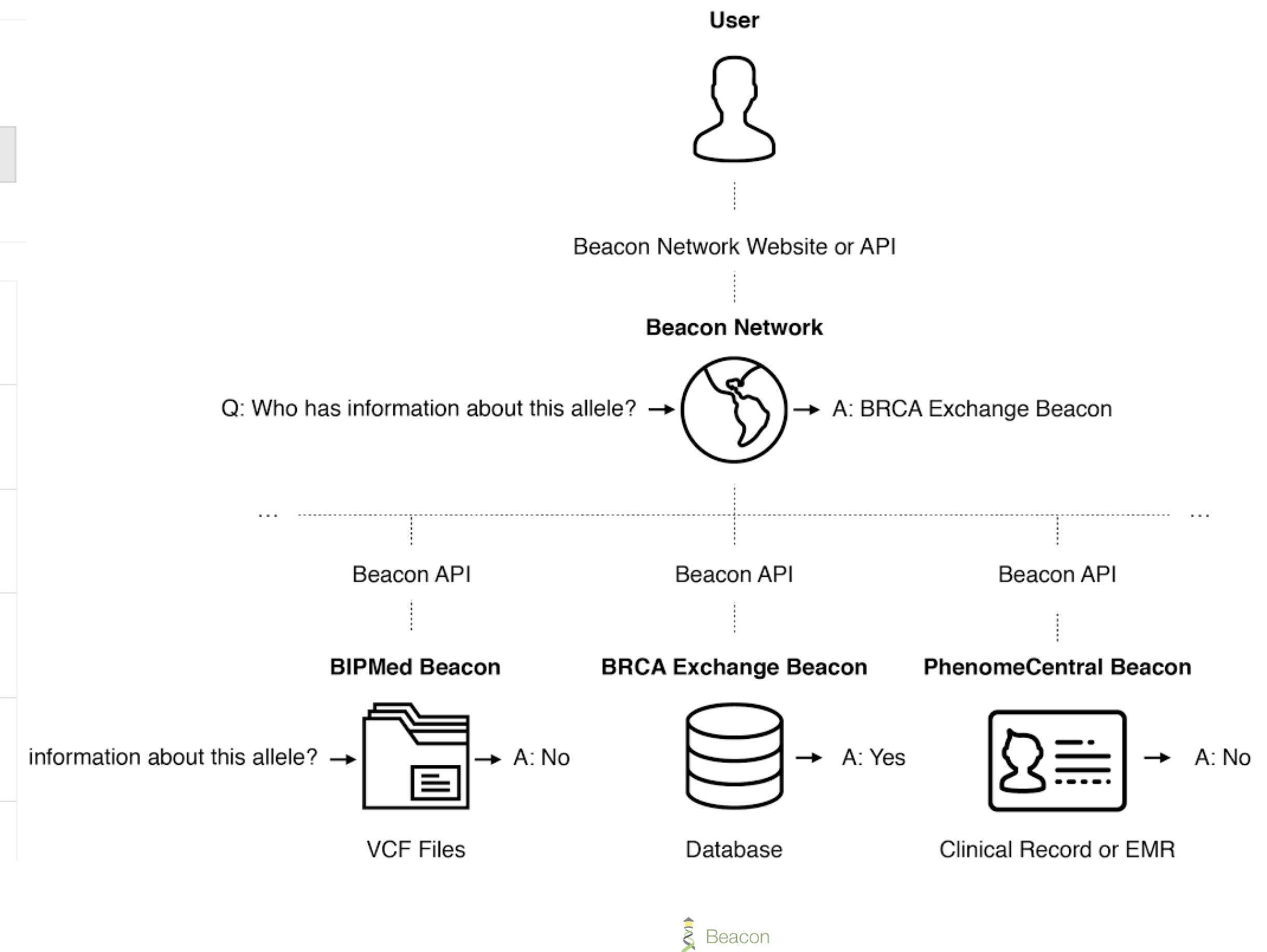
GRCh37 ▾ 10:118969015 C / CT

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



2016++: 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 the 'Driver Projects' section of the GA4GH website. It features a red circular icon with a stylized DNA helix. Below it, the heading 'Driver Projects' is displayed. A text block explains that GA4GH Driver Projects are real-world genomic data initiatives that help development efforts and pilot tools. Stakeholders around the world advocate, mandate, implement, and use our frameworks and standards in local contexts. To the right, there is a card for 'ELIXIR Beacon'. The card includes the ELIXIR logo, the text 'ELIXIR Beacon', the URL 'www.elixir-europe.org', the location 'Europe', and the 'Champions: Serena Scollen, Ilkka Lappalainen, Michael Baudis'.

Beacon v0.4 forward

- **structural variations** (DUP, DEL) in addition to SNV
- **metadata** queries
- layered **authentication** system using **ELIXIR** AAI
- **quantitative** responses
- Beacon queries as entry for **data delivery** (outside Beacon protocol)
- Ubiquitous **deployment** (e.g. throughout **ELIXIR** network)

Beacon+ Concept

- Implementation of cancer beacon prototype, backed by arrayMap data
- structural variations
- quantitative queries
- metadata
- current version uses GA4GH schema compatible, non-SQL database backend (MongoDB)

Beacon arrayMap

Beacon v0.4 implementation for arrayMap.



Reference name	<input type="text" value="9"/>
Start	<input type="text" value="42049214"/>
Length	<input type="text" value="1000"/>
Assembly ID	<input type="text" value="GRCh36"/>
Dataset Ids	<input type="text" value="(9440/3) 9440/3: Glioblastoma, NOS (2047)"/>
Alternate bases	<input type="text" value="DEL (Deletion)"/>
Confidence Interval (Start position)	<input type="text" value="500"/>
Confidence Interval (End position)	<input type="text" value="500"/>
Match type	<input type="text" value="Complete"/>

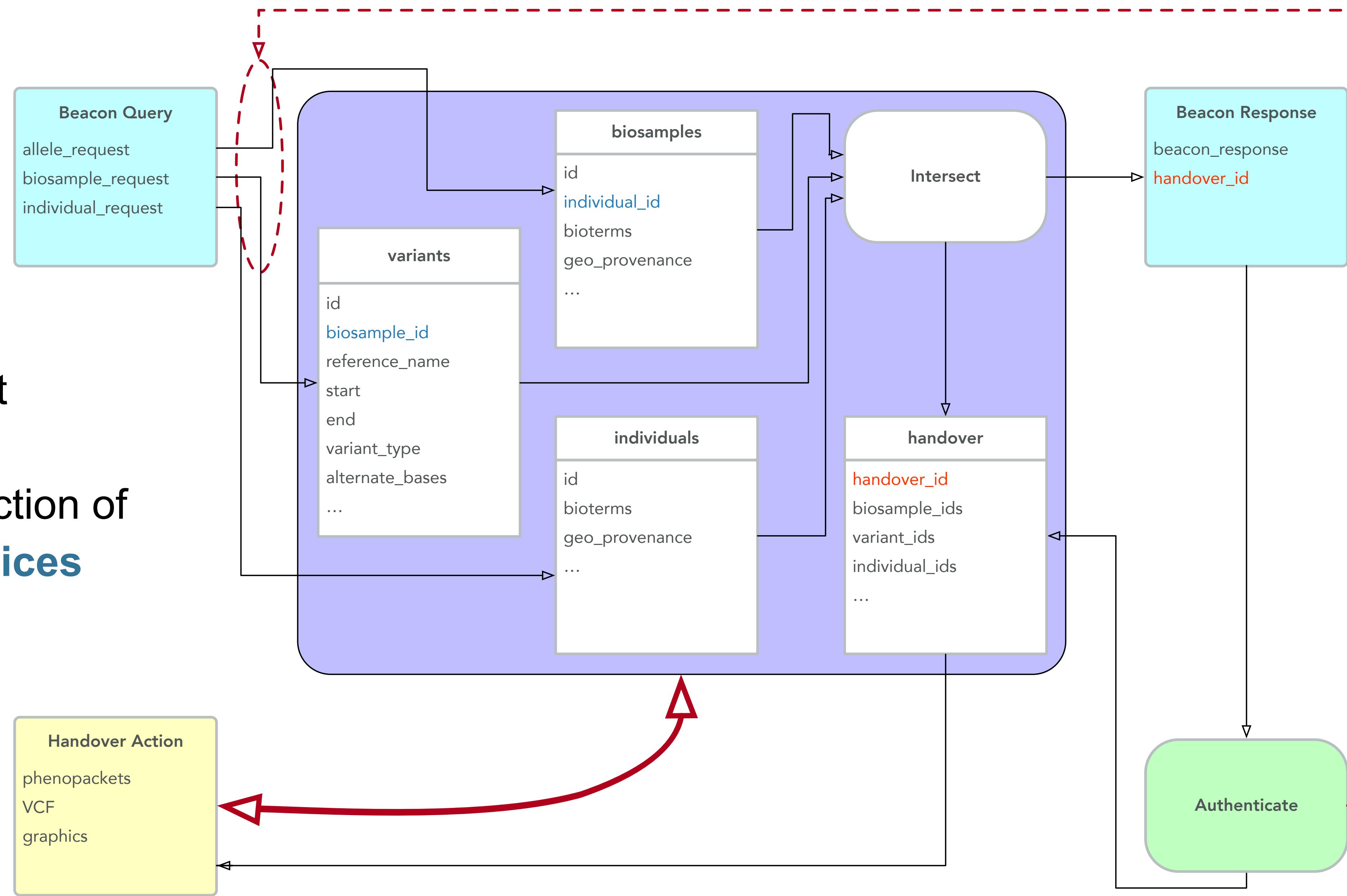
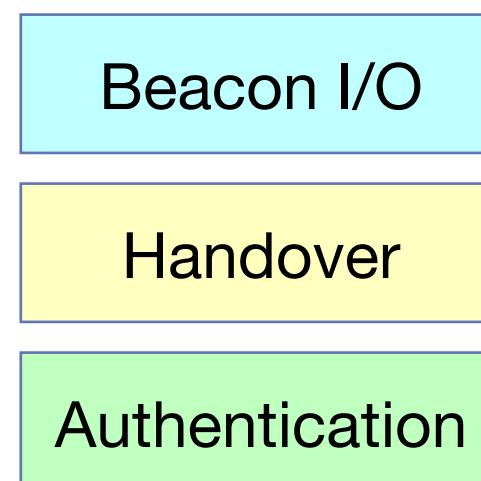
SIB

Beacon Query **Beacon Info**

arrayMap

Beacon & Handover

Beacons to support advanced types of queries and connection of **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**





2018 Progress

- Release of v.1.0 specification
 - basis for **GA4GH Beacon v.1.0**
 - GA4GH certification (first product)
- Building flexible **Beacon Network(s)**
- Beacon paper under review
- Post-2018 planning, documented in **5-year plan**
- Widening Node participation & additional use cases:
 - ELIXIR Norway joined Implementation Study, to light Beacon(s) against **Marine data**
 - Widen this out to more use cases? Plants, other...



ELIXIR Beacon Project Plan 2019-2021

- Extend the Beacon **protocol** to become the **reference ELIXIR Data Discovery product**, through expanding query options and providing richer responses, with view on biomedical applications, and in alignment with developing GA4GH standards
- Deliver ELIXIR Beacon Network as an established **ELIXIR service**
- Leverage ELIXIR Nodes to increase **data flow** through Beacon services
- Actively support the **integration** of the Beacon API with **human data resources** throughout ELIXIR, with particular view on National Genomics Initiatives, biobanks, and **Human Data Communities** (such as Rare Diseases and Human CNV)

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](#)



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



IMPLEMENTATION CHALLENGES

- ▶ interpretation and counting issues, especially w/ multi-allelic variants from VCF and "borderline" structural variants (e.g. INDEL vs. DEL, DUP:TANDEM vs. DUP...) - much of this related to using VCF instead of a variant model
- ▶ limited recognition & uptake of "recent" API versions
- ▶ only recent emergence of modern Beacon network aggregator
- ▶ protocol as "moving target" - while v1.n potential for widespread use, the rapid move towards v2 limits dissemination
- ▶ lack of well documented, flexible Beacon distributions
 - ▶ the "reference implementation" is limited & more for drop-in testing ("get your VCF online") than a scalable solution
 - ▶ other implementations (Ensembl, Beacon+ ...) more part of own ecosystem than for easy distribution

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.
- Tested and already implemented by Beacons **v2.0**
- Ongoing **v2.n**



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.



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

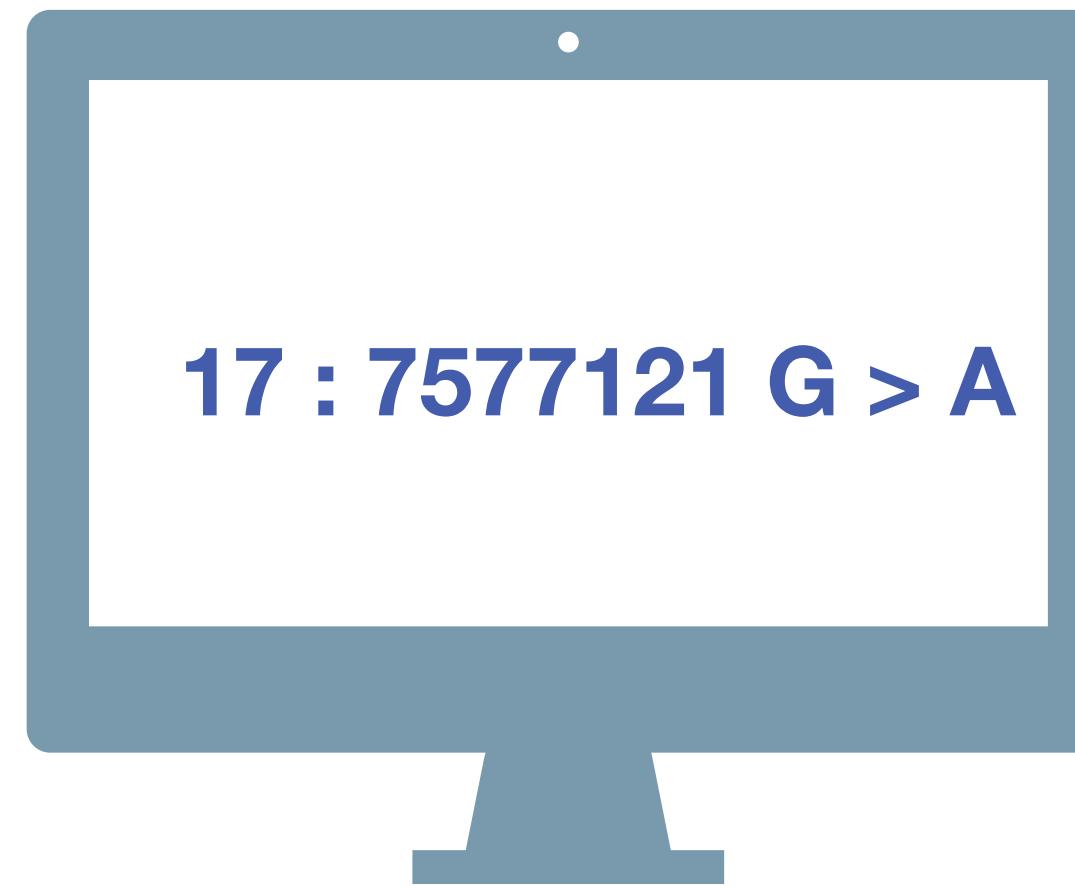
Our Beacon platform is being used for the rapid testing of queries and responses - both v1.n and v2.0.a - against a number of partially large-scale genome datasets.

- 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.



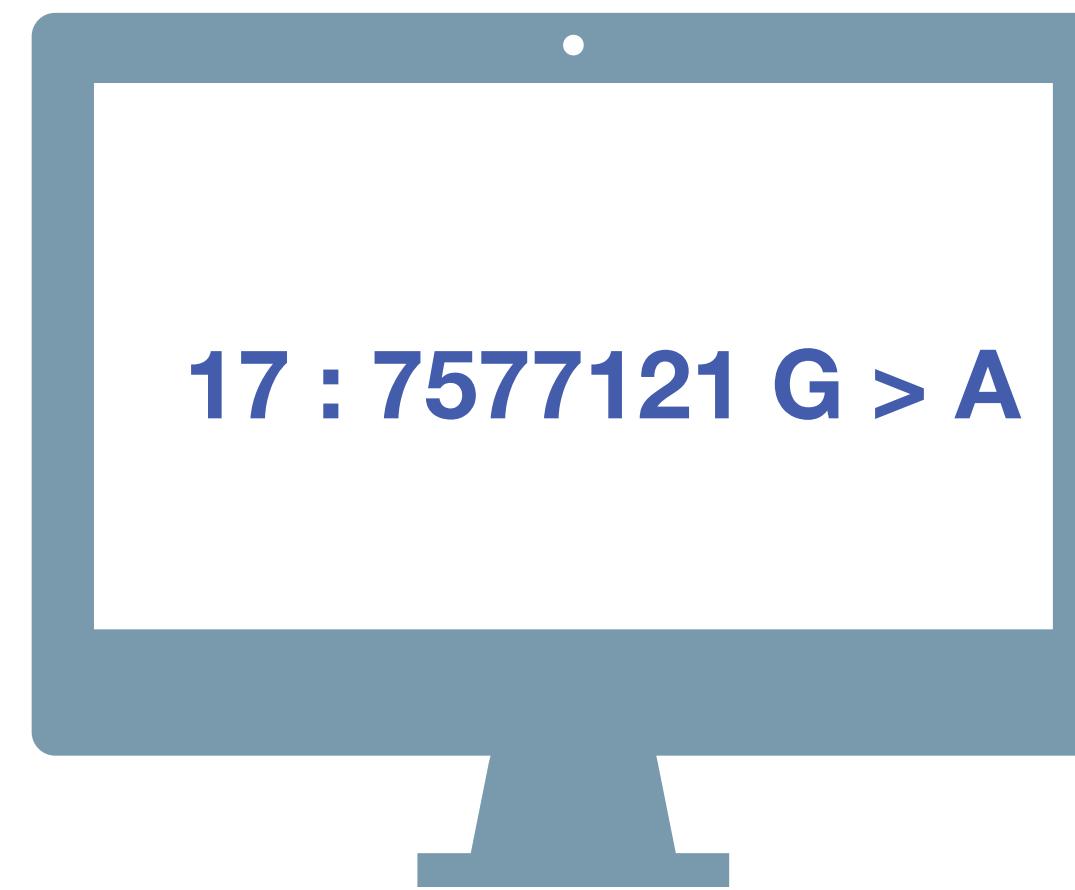
```
{
  "callset_id": "cs-cellosaurus:CVCL_EI02",
  "info": {
    "cellosaurus": {
      "cell_line": "BT474-LAPRa",
      "id": "CVCL_EI02",
      "cellosaurus_variant_name": "TP53 p.Glu285Lys (c.853G>A)"
    },
    "clinvar": {
      "gene_id": "7157",
      "allele_id": "410258",
      "assembly": "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"
    }
  }
}
```



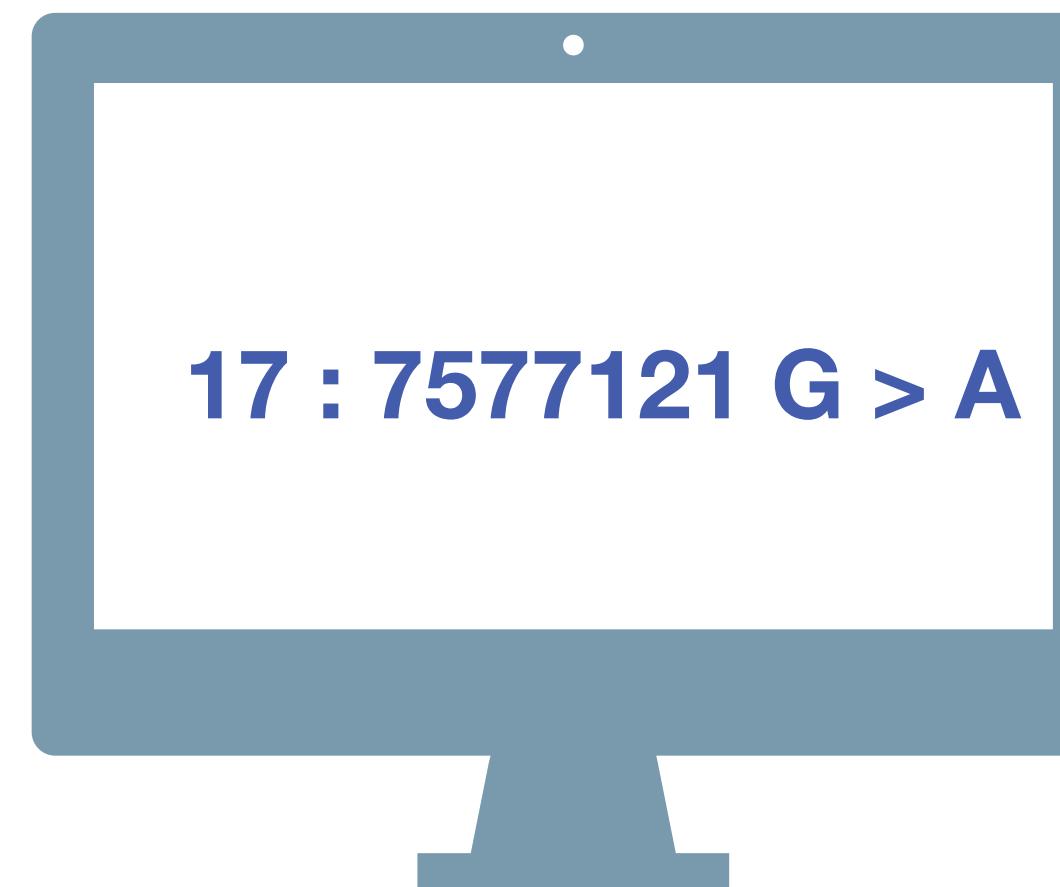
Beacon

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

YES | NO | \0



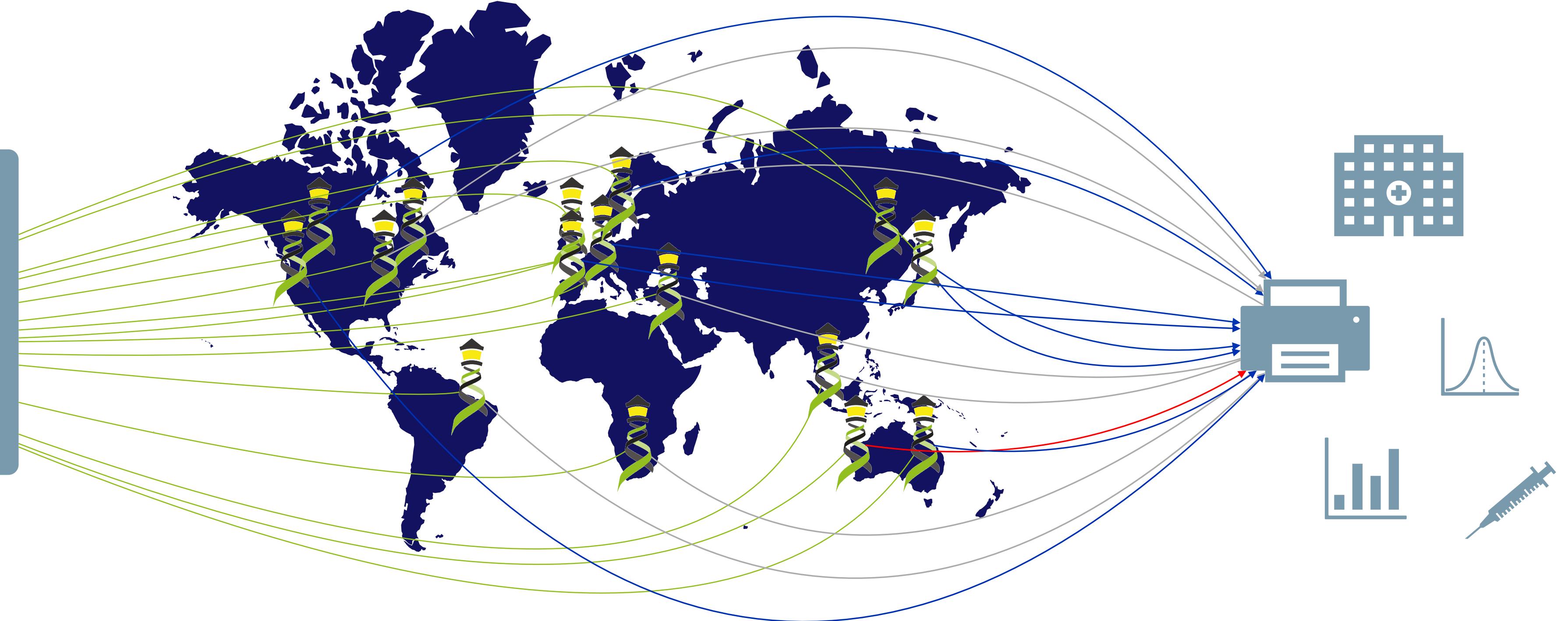
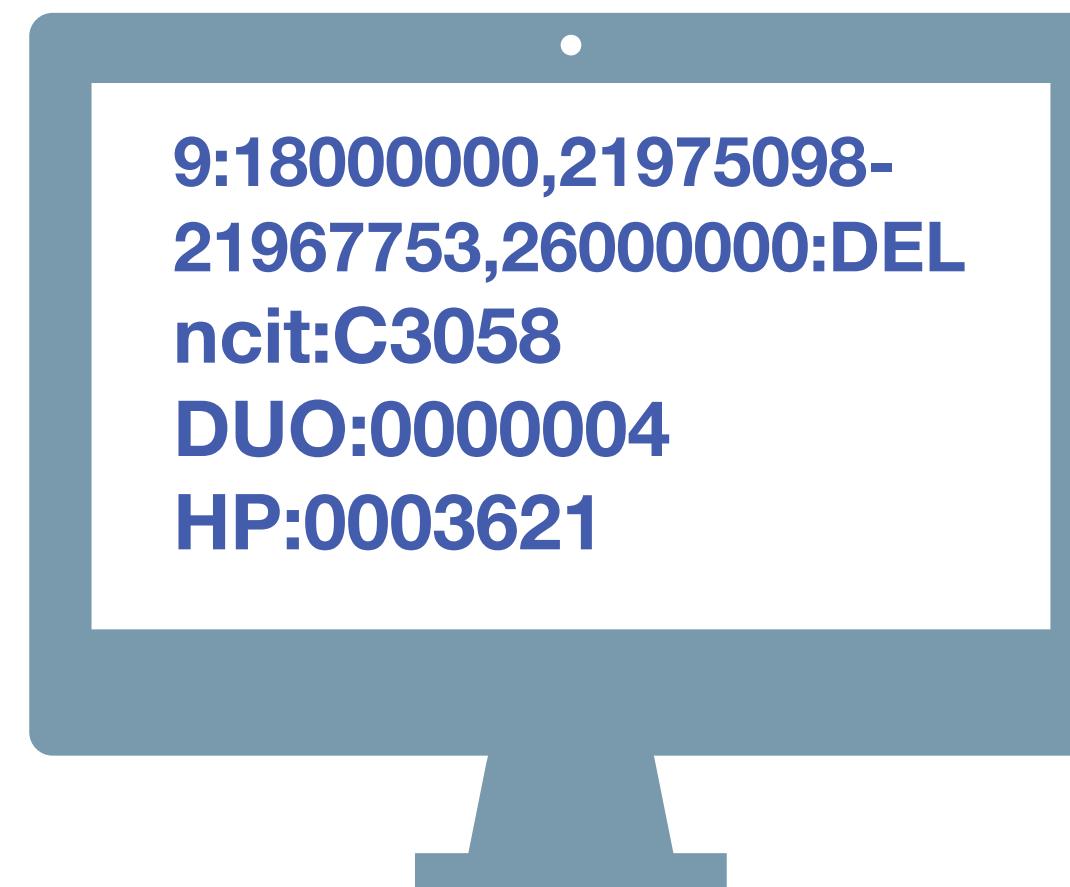
A Beacon network federates
genome variant queries
across databases that
support the ***Beacon API***



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**.

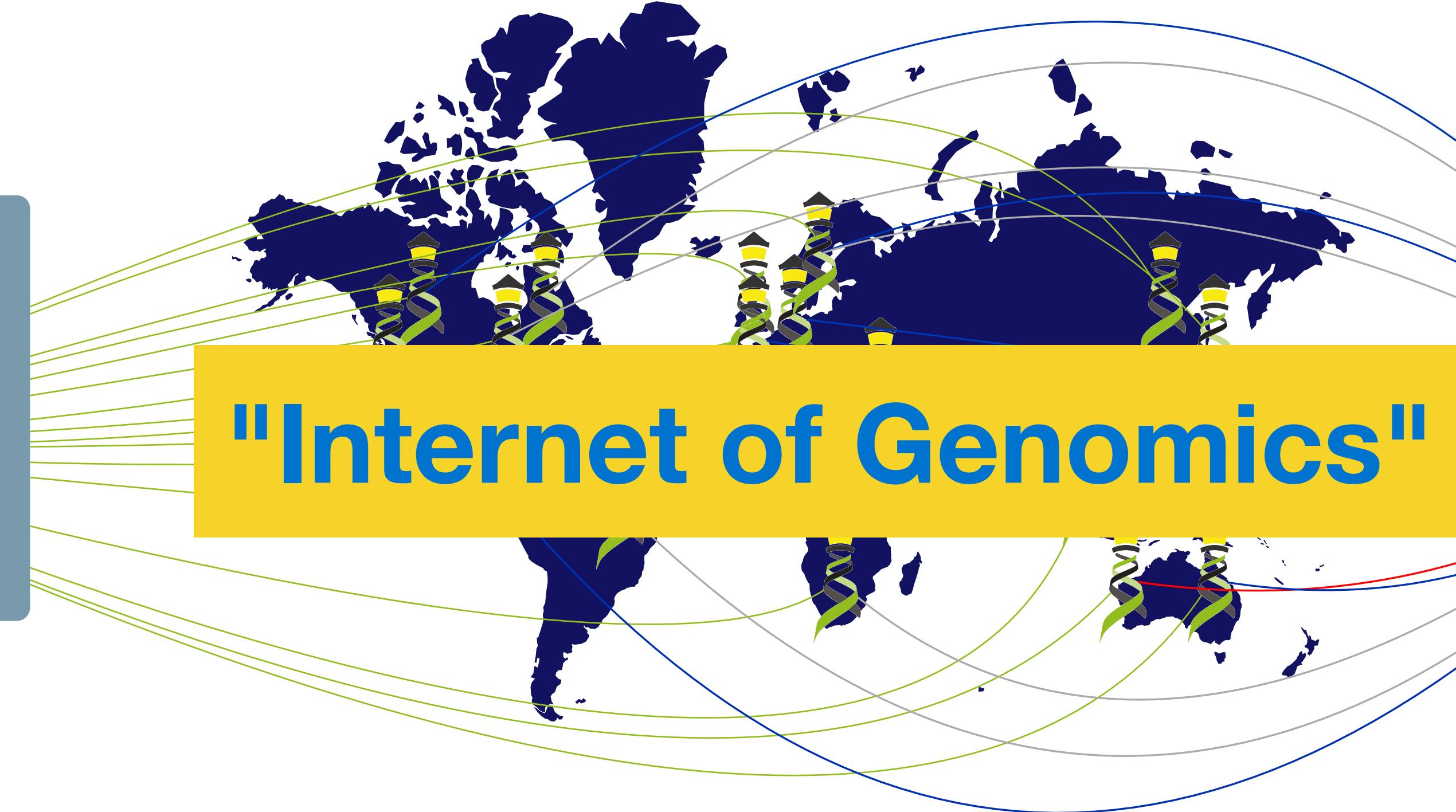
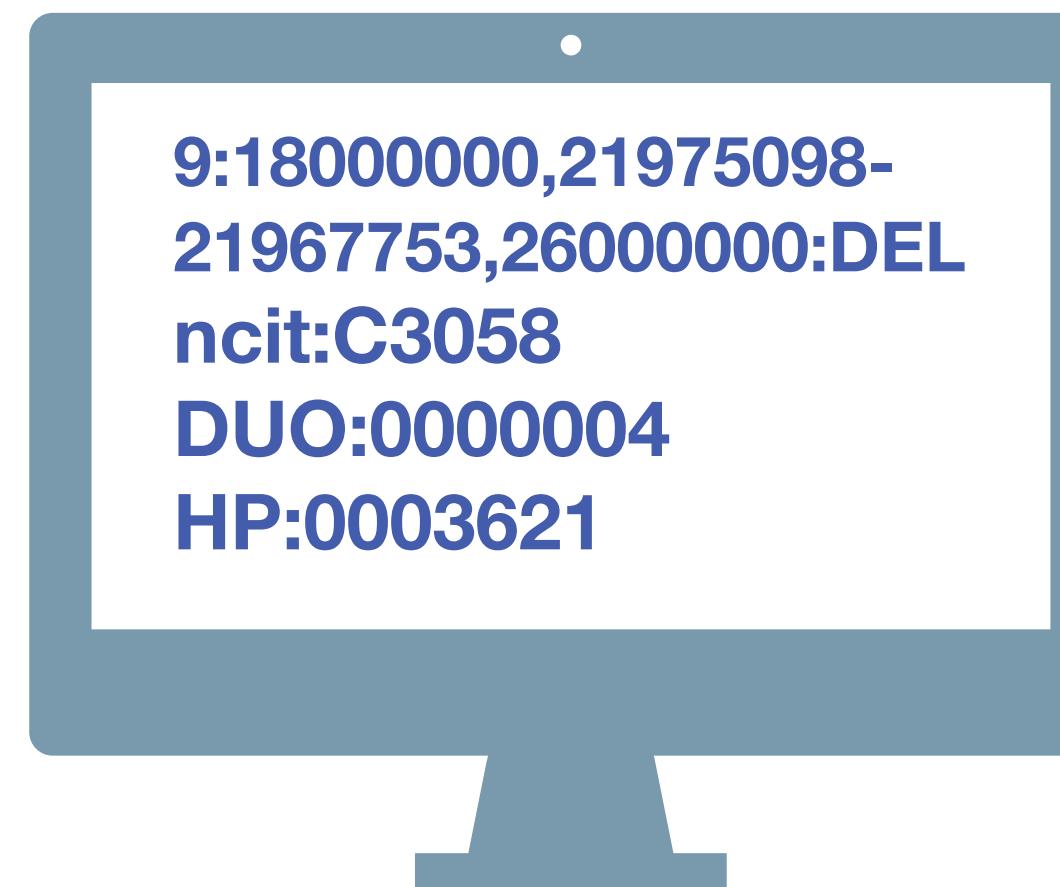


Have you seen deletions in this region on chromosome 9 in Glioblastomas from a juvenile patient, in a dataset with unrestricted access?

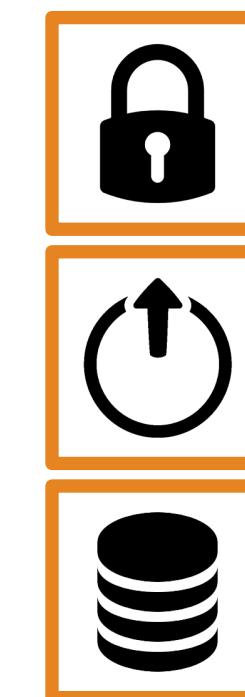


Beacon v2 API

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

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.



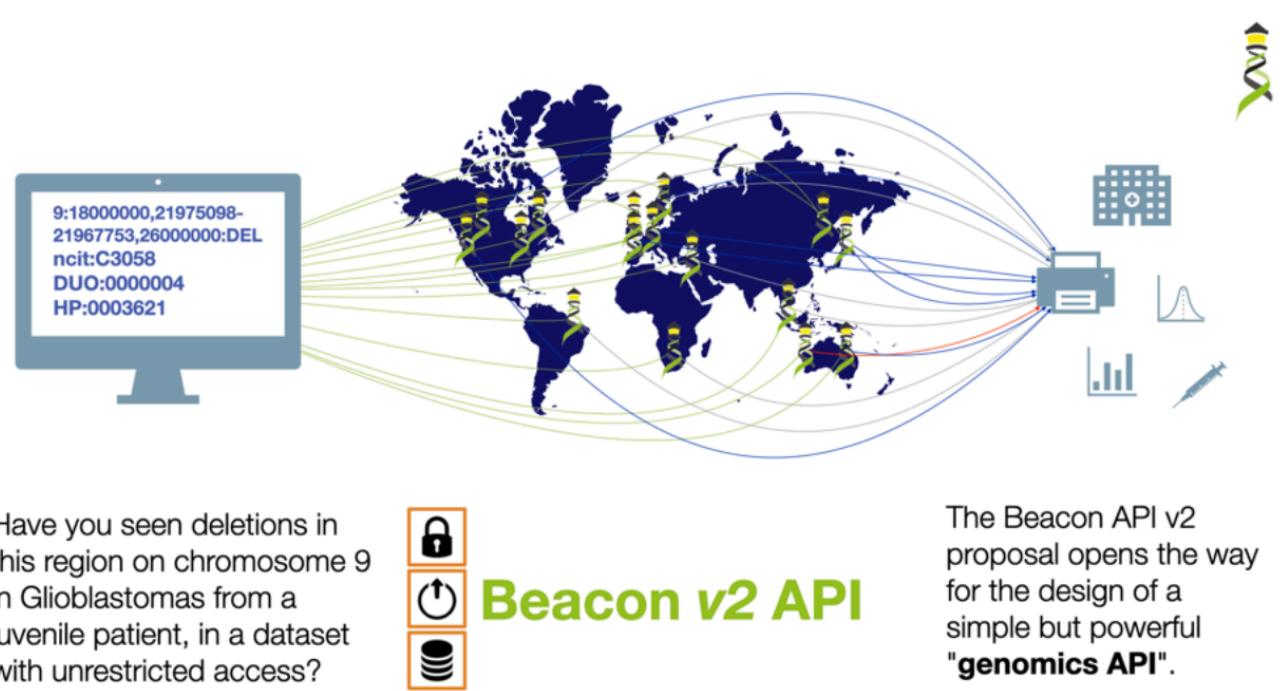
The original Beacon protocol had been designed to be:

- **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?

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

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

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)

Session Schedule

15 min	Beacon - where are we now, and how did we get here?	<i>Michael Baudis</i>
15 min	ELIXIR Beacon Network: what does it offer to users?	<i>Juha Tornroos</i>
15 min	Beacon v2 - status, Use Cases, and implementations: examples from ES	<i>Jordi Rambla</i>
10 min	Beacon is secure: what does this mean, how does this interact with GA4GH, and how will this work with v2	<i>Dylan Spalding</i>
10 min	The Beacon API Compliance Tester: do the Beacons really speak the same language?	<i>Malin Klang</i>
5 min	Refreshment break	
25 min	Beacon and Covid-19	<i>Jordi Rambla</i>
10 min	Beacon as a discovery tool for human data outside of genomics: Proteomics	<i>Juan Antonio Vizcaino</i>
15 min	Beacon in the ELIXIR Programme - Node perspective from FI	<i>Juha Törnroos</i>