progenetix

Data *Discovery* and Data *Sharing* in Biomedical Genomics

Reference Genomics Resources Powered by the GA4GH Beacon Standard

Michael Baudis
Professor of Bioinformatics
University of Zürich
Swiss Institute of Bioinformatics **SIB**
Member GA4GH Strategic Leadership Committee
Co-lead ELIXIR Beacon API Development
Co-lead ELIXIR hCNV Community

Universität Zürich[UZH]

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

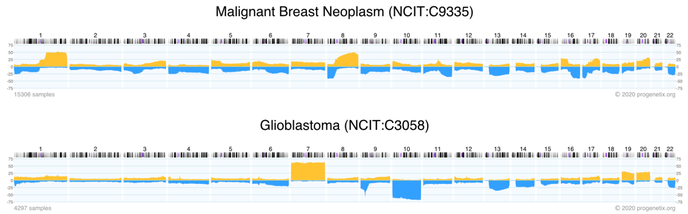
Swiss Institute of Bioinformatics

elixir

Theoretical Cytogenetics and Oncogenomics

... but what does this entail @baudisgroup?

- **patterns & markers** in cancer genomics, especially somatic structural genome variants
- bioinformatics support in **collaborative** studies
- **reference resources** for curated cancer genome variations
- bioinformatics **tools & methods**
- **standards** and **reference implementations** for data sharing in genomics and personalized health
- **open research data** "ambassadoring"

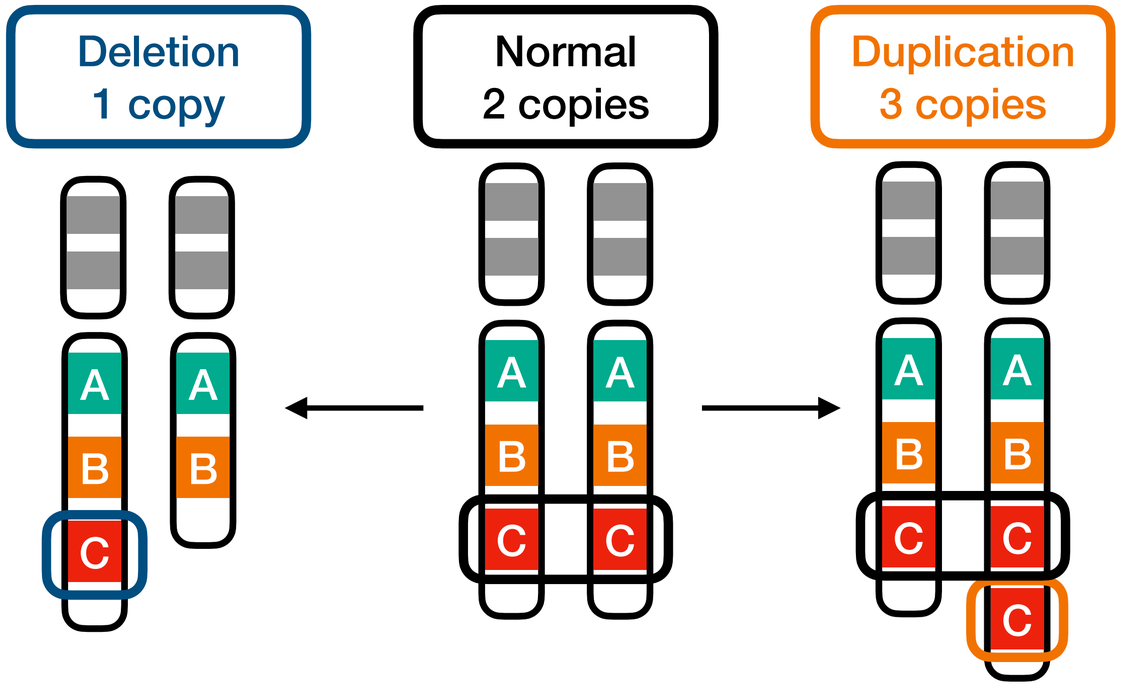


progenetix

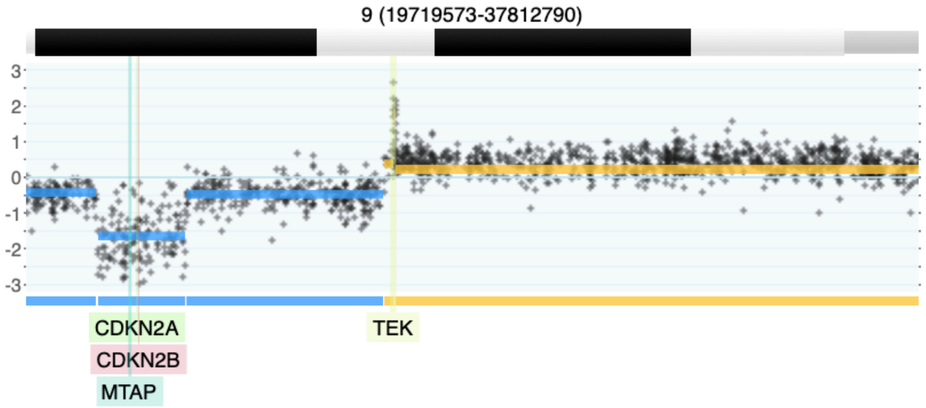
Curators

Data ~~Parasites~~

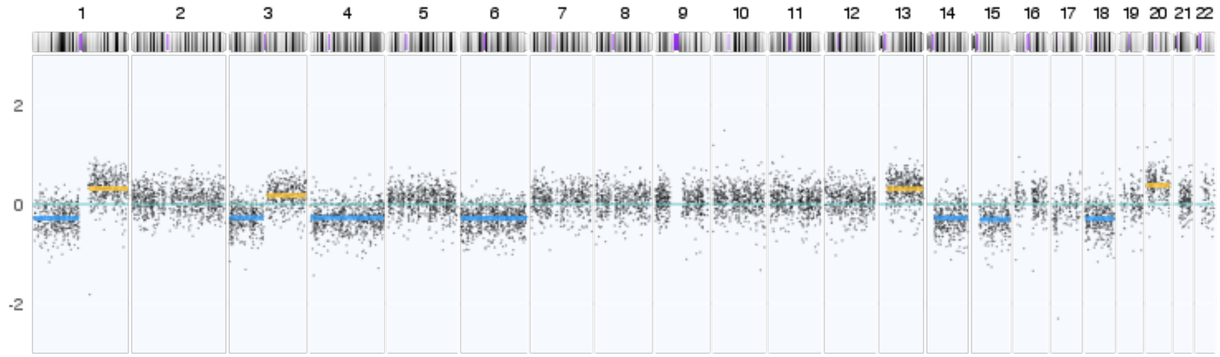
Copy Number Variant (CNV)



- Intermediate-scale genetic change
- Size: 1kb to multiple megabase
- Additional copies of sequence (duplications) and losses of genetic material (deletions)



2-event, homozygous deletion in a Glioblastoma

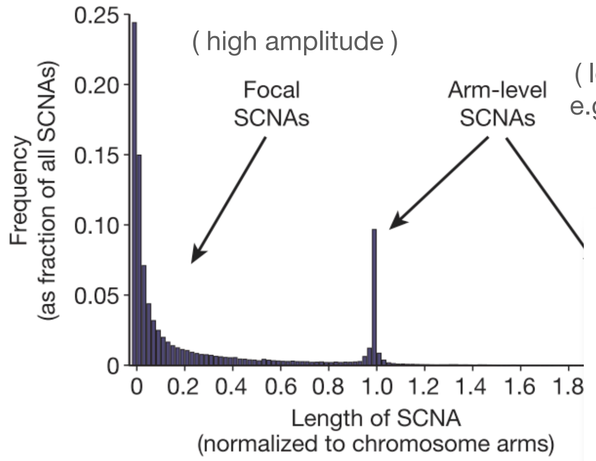


Gain of chromosome arm 13q in colorectal carcinoma

CNV Categorization Method Supporting GA4GH Standards

GA4GH Variation Representation Specification

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



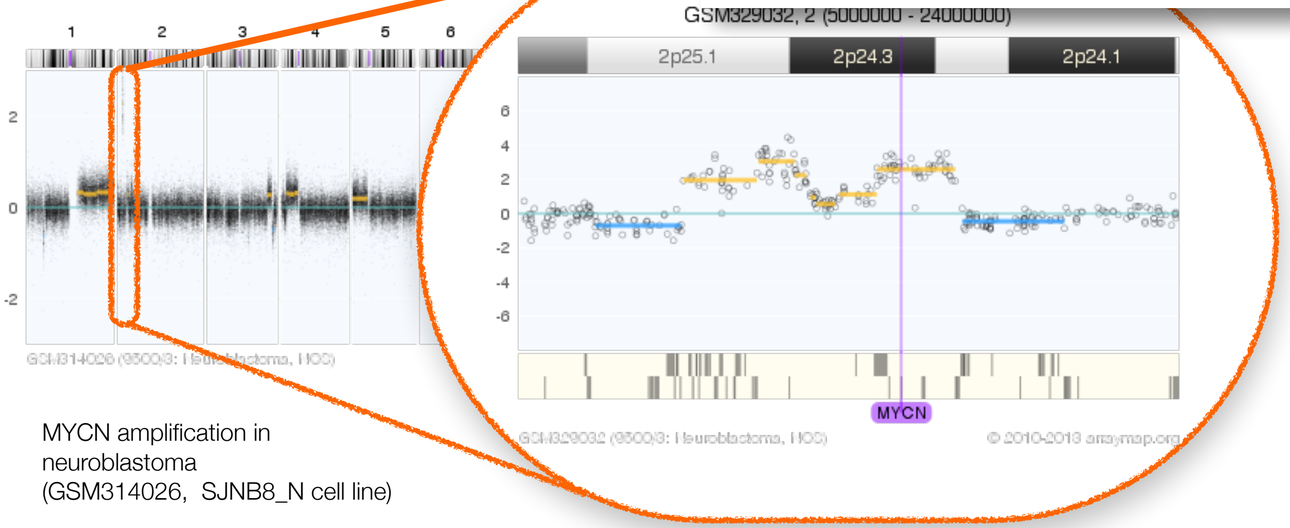
Rameen et al 20

OXFORD

Briefings in Bioinformatics, 2024, **25(2)**, 1–12
https://doi.org/10.1093/bib/bbad541
Problem Solving Protocol

labelSeg: segment annotation for tumor copy number alteration profiles

Hangjia Zhao and Michael Baudis
Corresponding author: Michael Baudis, Department of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.
Tel.: (+41) 44 635 34 86; E-mail: michael.baudis@mls.uzh.ch



MYCN amplification in neuroblastoma (GSM314026, SJNB8_N cell line)

Lead: Hangjia Zhao

CopyNumberChange

Copy Number Change captures a categorization of copies of a molecule within a system, relative to a baseline. These types of Variation are common outputs from CNV callers, particularly in the somatic domain where integral CopyNumberCount are difficult to estimate and less useful in practice than relative statements. Somatic CNV callers typically express changes as relative statements, and many xpress copy number variation are interpreted to be relative copy

r of a Location or a Feature within a system (e.g. genome, cell,

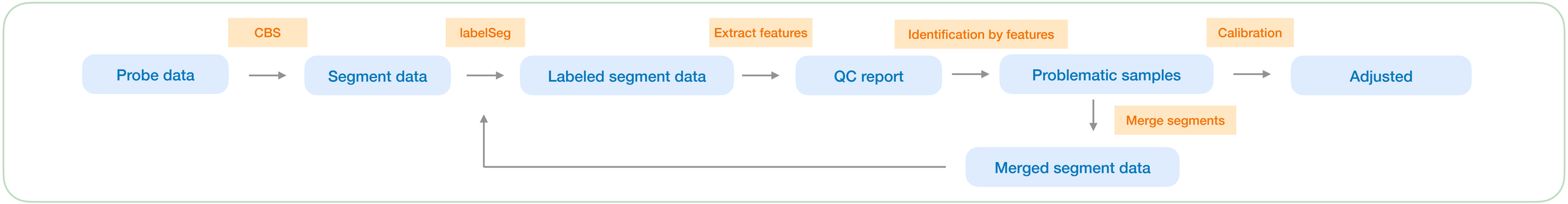
tes are inherited from Variation.

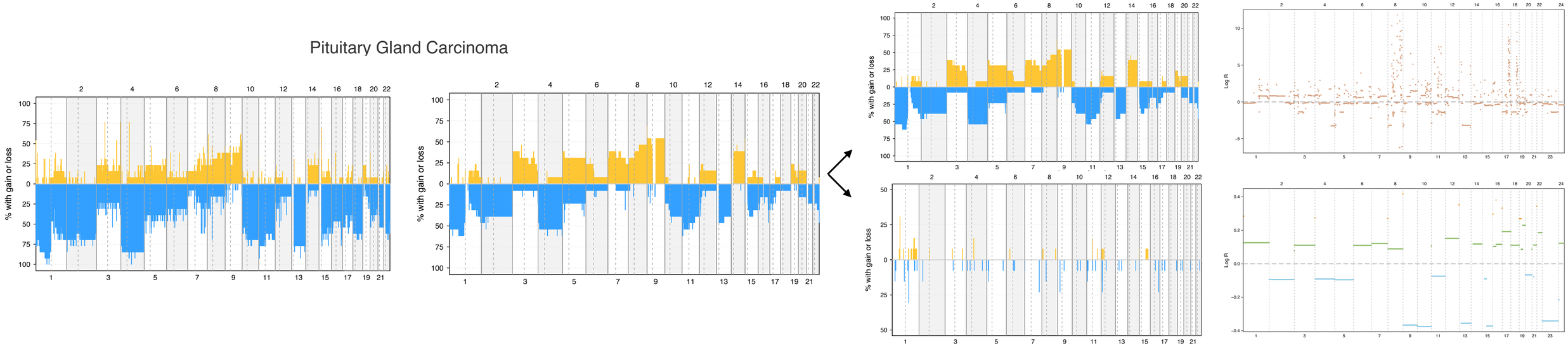
Field	Type	Limits	Description
_id	CURIE	0..1	Variation Id. MUST be unique within document.
type	string	1..1	MUST be “CopyNumberChange”
subject	Location | CURIE | Feature	1..1	A location for which the number of systemic copies is described.
copy_change	string	1..1	MUST be one of “efo:0030069” (complete genomic loss), “efo:0020073” (high-level loss), “efo:0030068” (low-level loss), “efo:0030067” (loss), “efo:0030064” (regional base ploidy), “efo:0030070” (gain), “efo:0030071” (low-level gain), “efo:0030072” (high-level gain).

Pipeline Development

improve CNV calling in large numbers of heterogeneous cancer samples

nextflow



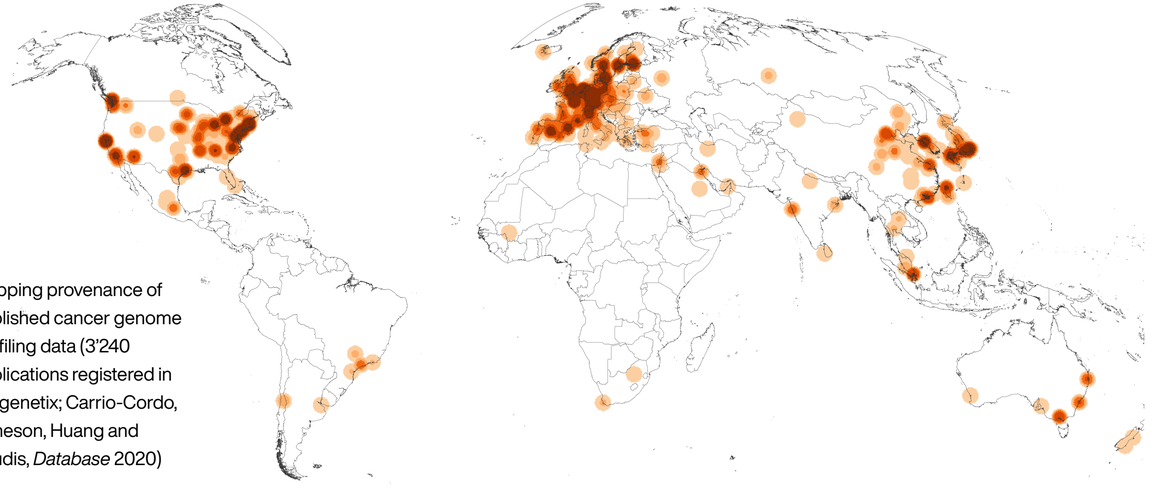


Lead: Hangjia Zhao

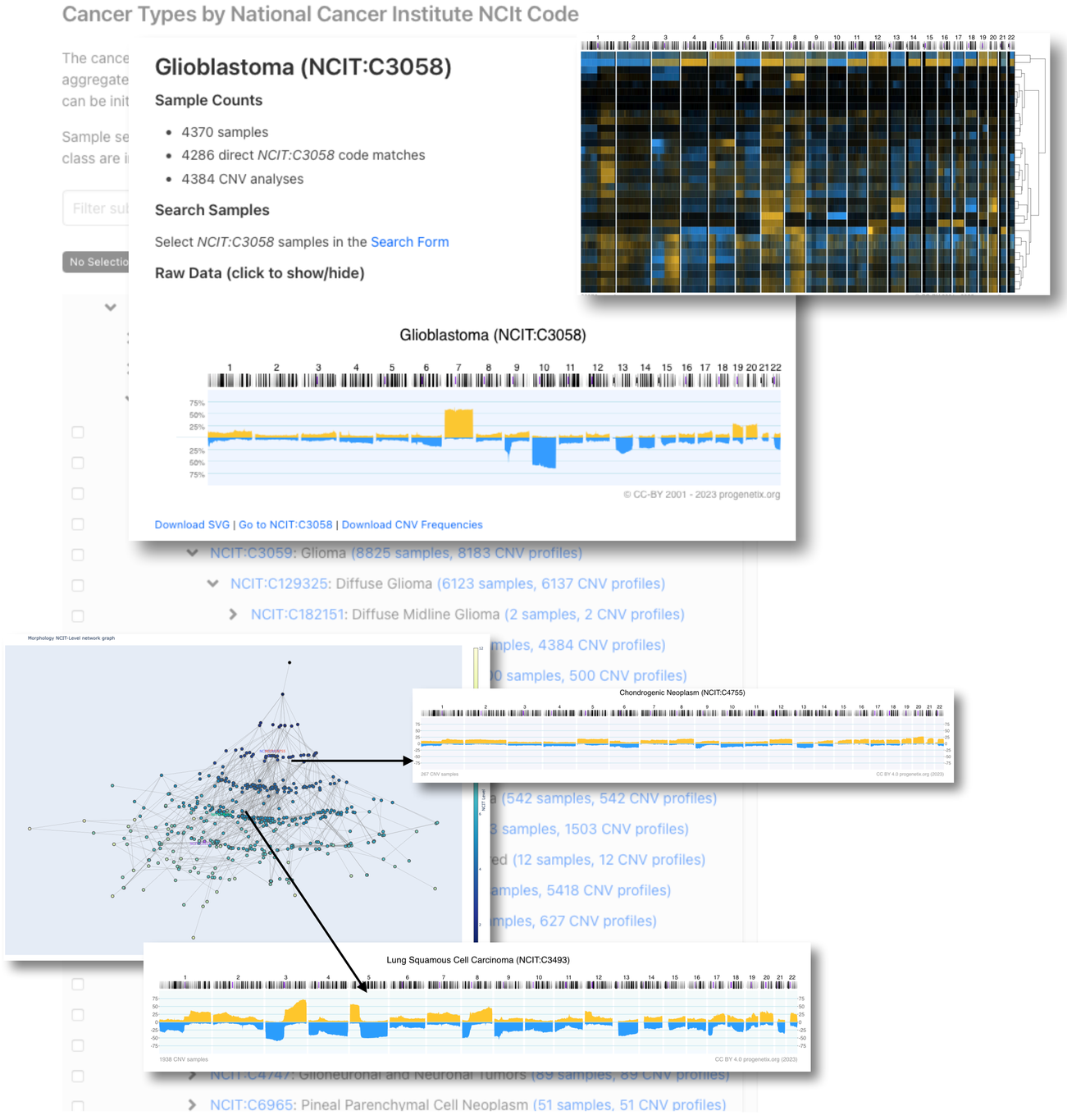
progenetix.org @ UZH

Cancer Genomics Reference Resource

- over **240'000 cancer CNV profiles**
- more than **1100 diagnostic types**
- runs on a **GA4GH Beacon API**
- **Open Research Data** practices



Mapping provenance of published cancer genome Profiling data (3'240 publications registered in Progenetix; Carrio-Cordo, Acheson, Huang and Baudis, *Database* 2020)



Universität Zürich[UZH]

progenetix

SIB Swiss Institute of Bioinformatics

progenetix.org

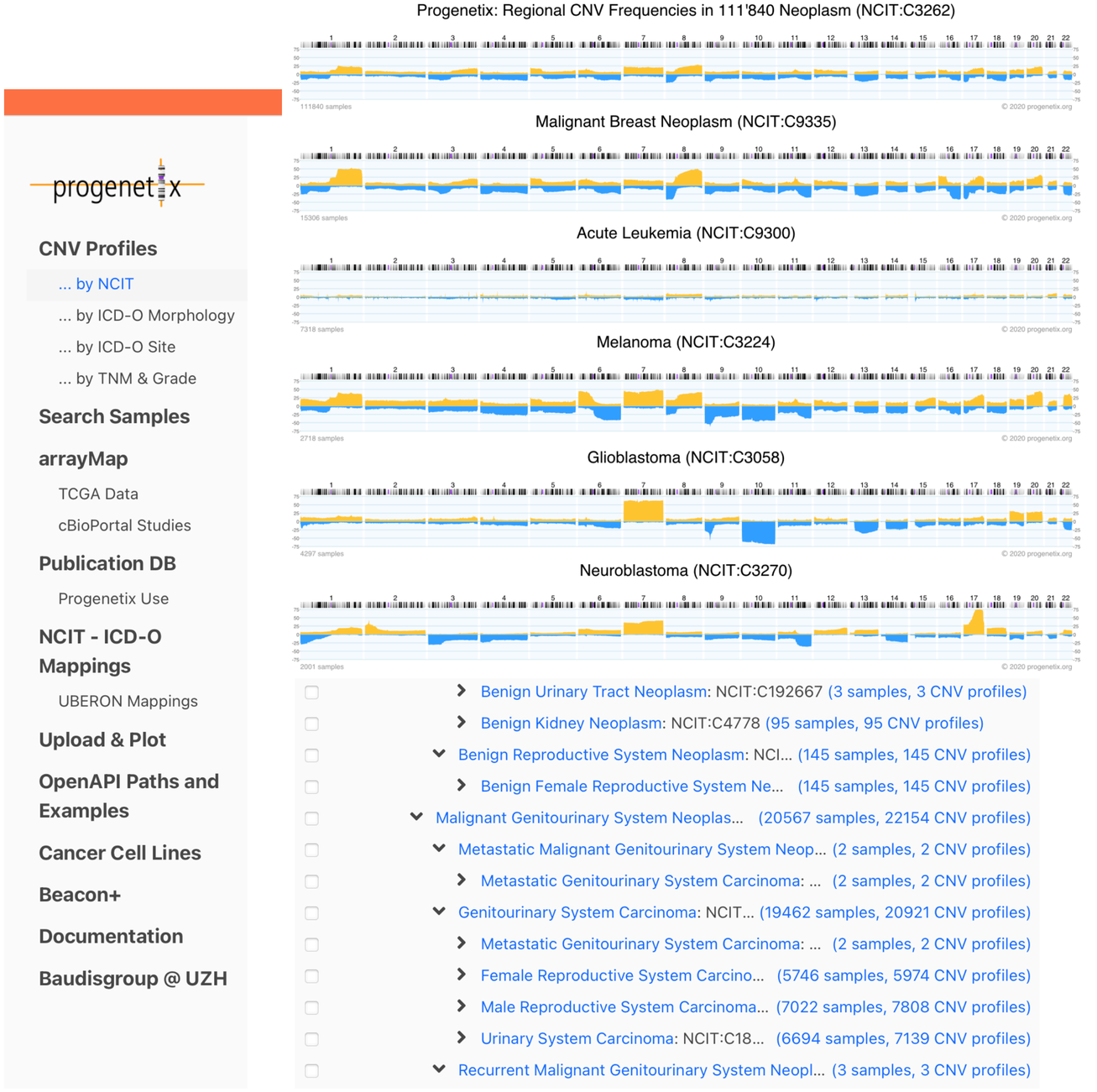
Cancer Genomics Reference Resource

- ***open*** resource for oncogenomic profiles
- over **240'000 cancer CNV profiles**
- SNV data for some series (e.g. TCGA)
- more than **1100 diagnostic types**
- inclusion of reference datasets (e.g. TCGA, GENIE, cBioPortal)
- standardized encodings (e.g. NCIt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core clinical data (TNM, sex, survival ...)
- data mapping services

Universität Zürich[UZH]

progenetix

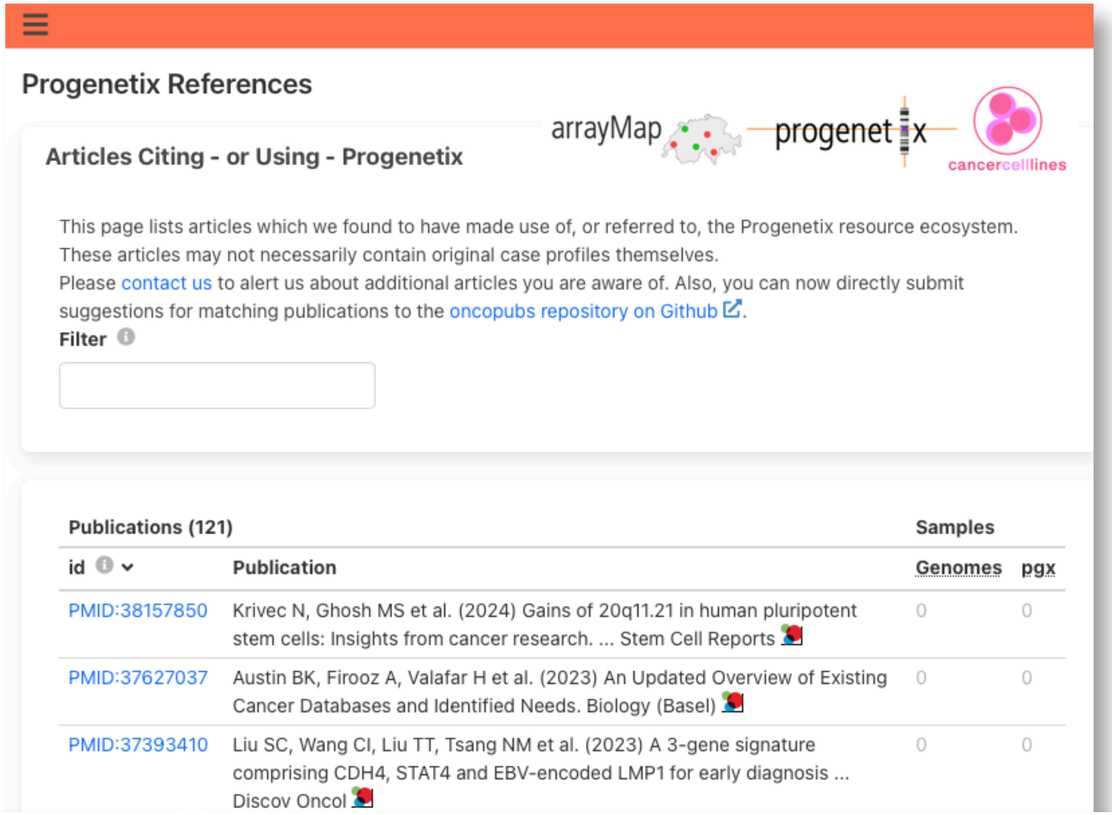
SIB Swiss Institute of Bioinformatics

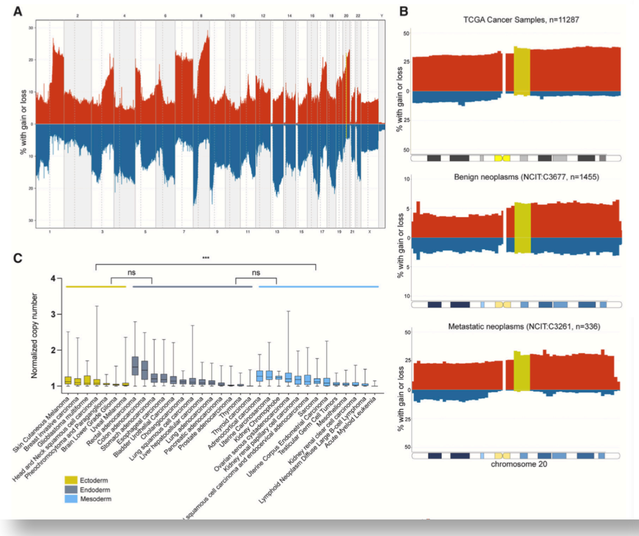


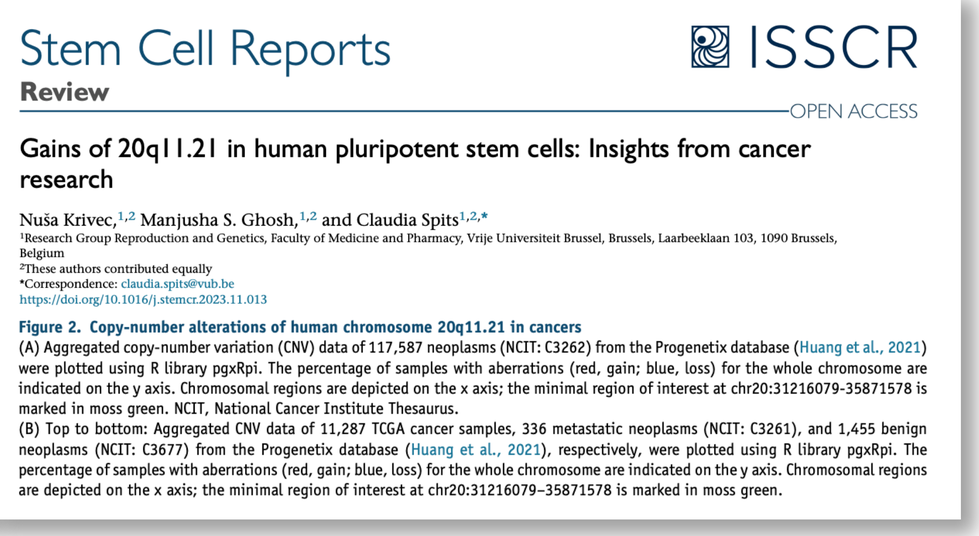
Progenetix Use

- CNV data is used e.g. as reference data in cancer genomics studies
- diagnosis specific CNV profiles serve as "fast look-up" in clinical genomics laboratories
- we loosely track publications in in our literature database but there is no systematic check-back mechanism...

Example: 2025 article using Progenetix' ***pgxRpi*** Beacon/R interface to retrieve & visualize 117'587 cancer CNV profiles for a study into pluripotent stem cells' genomics






Stem Cell Reports
ISSCR
Review
OPEN ACCESS

Gains of 20q11.21 in human pluripotent stem cells: Insights from cancer research

Nuša Krivec,[1,2] Manjusha S. Ghosh,[1,2] and Claudia Spits[1,2,*]
[1]Research Group Reproduction and Genetics, Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Brussels, Laarbeeklaan 103, 1090 Brussels, Belgium
[2]These authors contributed equally
*Correspondence: claudia.spits@vub.be

https://doi.org/10.1016/j.stemcr.2023.11.013

Figure 2. Copy-number alterations of human chromosome 20q11.21 in cancers

(A) Aggregated copy-number variation (CNV) data of 117,587 neoplasms (NCIT: C3262) from the Progenetix database (Huang et al., 2021) were plotted using R library pgxRpi. The percentage of samples with aberrations (red, gain; blue, loss) for the whole chromosome are indicated on the y axis. Chromosomal regions are depicted on the x axis; the minimal region of interest at chr20:31216079-35871578 is marked in moss green. NCIT, National Cancer Institute Thesaurus.

(B) Top to bottom: Aggregated CNV data of 11,287 TCGA cancer samples, 336 metastatic neoplasms (NCIT: C3261), and 1,455 benign neoplasms (NCIT: C3677) from the Progenetix database (Huang et al., 2021), respectively, were plotted using R library pgxRpi. The percentage of samples with aberrations (red, gain; blue, loss) for the whole chromosome are indicated on the y axis. Chromosomal regions are depicted on the x axis; the minimal region of interest at chr20:31216079–35871578 is marked in moss green.

Different Approaches to Data Sharing

progenetix

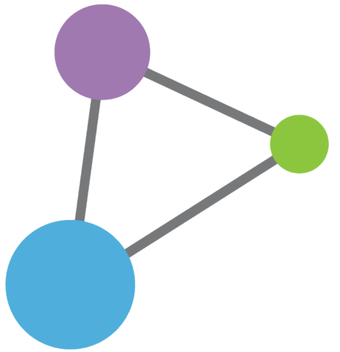
Centralized Genomic Knowledge Bases

European Genome-Phenome Archive

Data Commons
Trusted, controlled repository of multiple datasets

SPHN Swiss Personalized Health Network

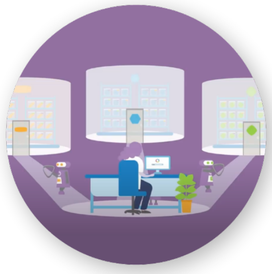
Hub and Spoke
Common data elements, access, and usage rules



Linkage of distributed and disparate datasets

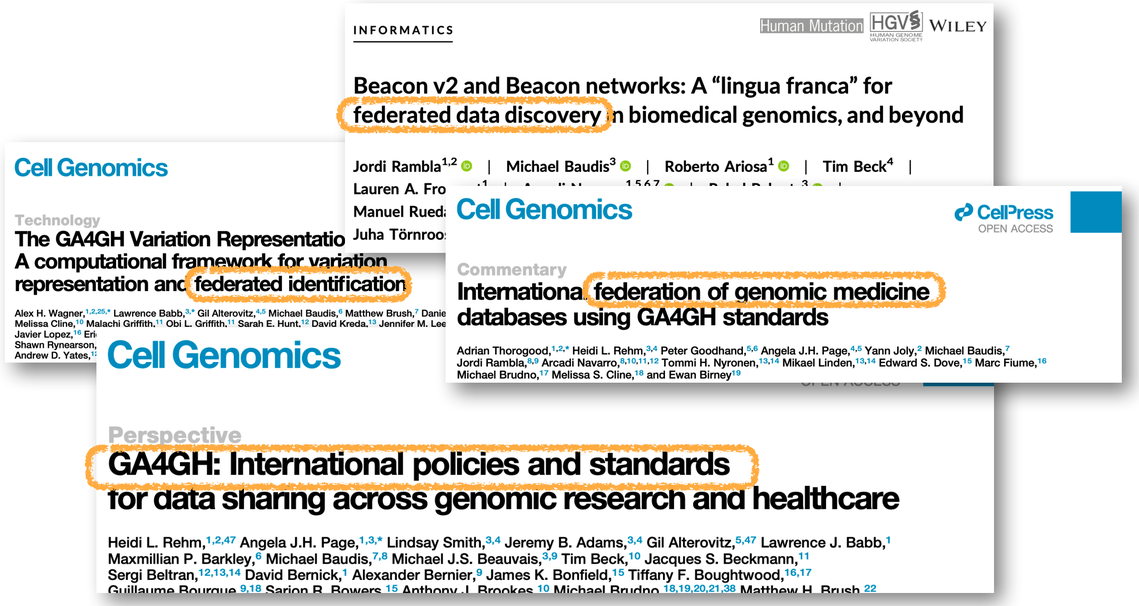
GA4GH STANDARDS

From Data Copying to
Data Discovery & Visiting

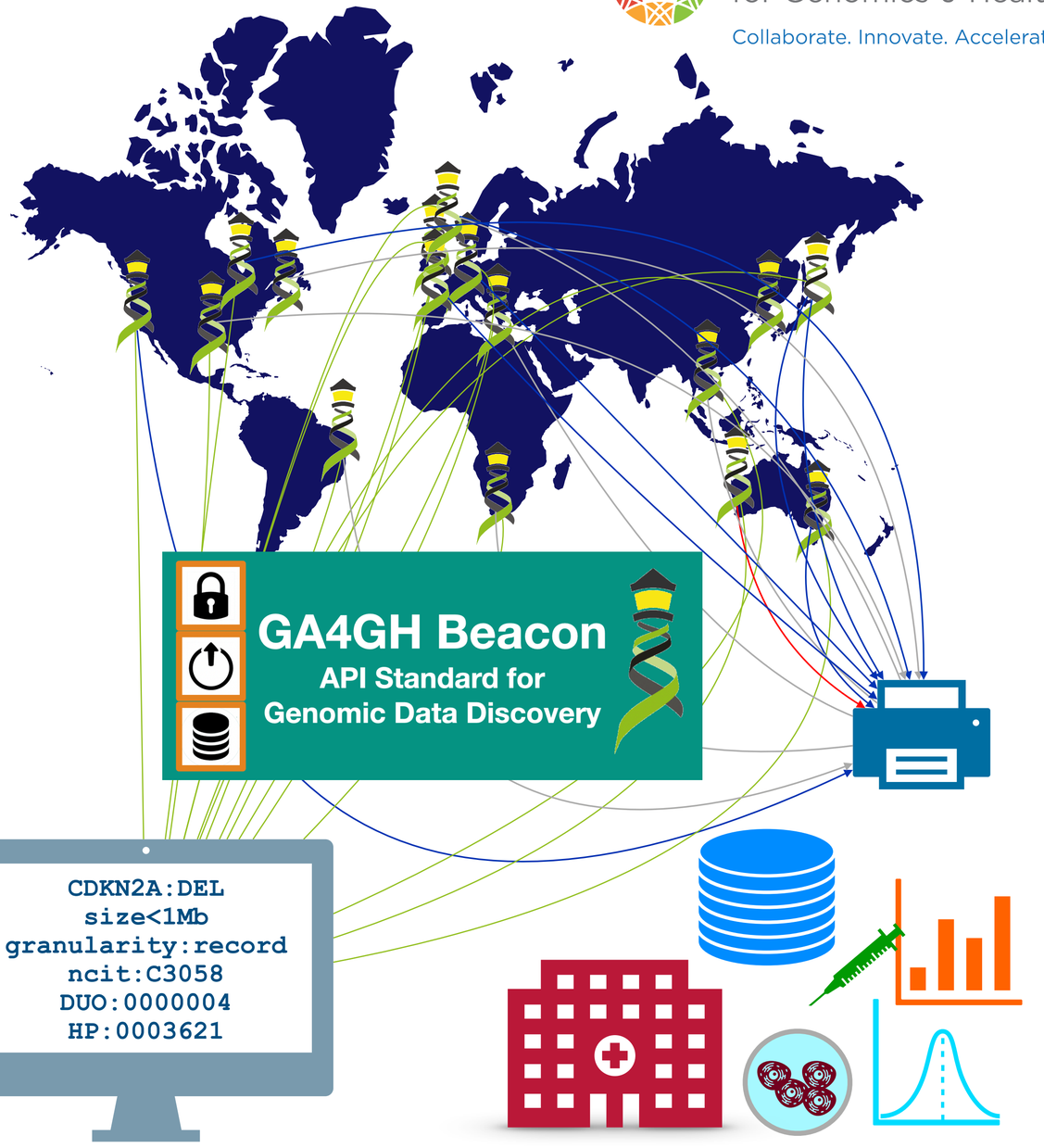


- **Growing genomics data**: research, clinics and national projects
- **Data federation** allows utilization of rare and population related data **for research** and **precision medicine**
- The Global Alliance for Genomics and Health **GA4GH** started operation in 2014 to promote the “responsible sharing of genomic and health-related data” through **standards and policies**

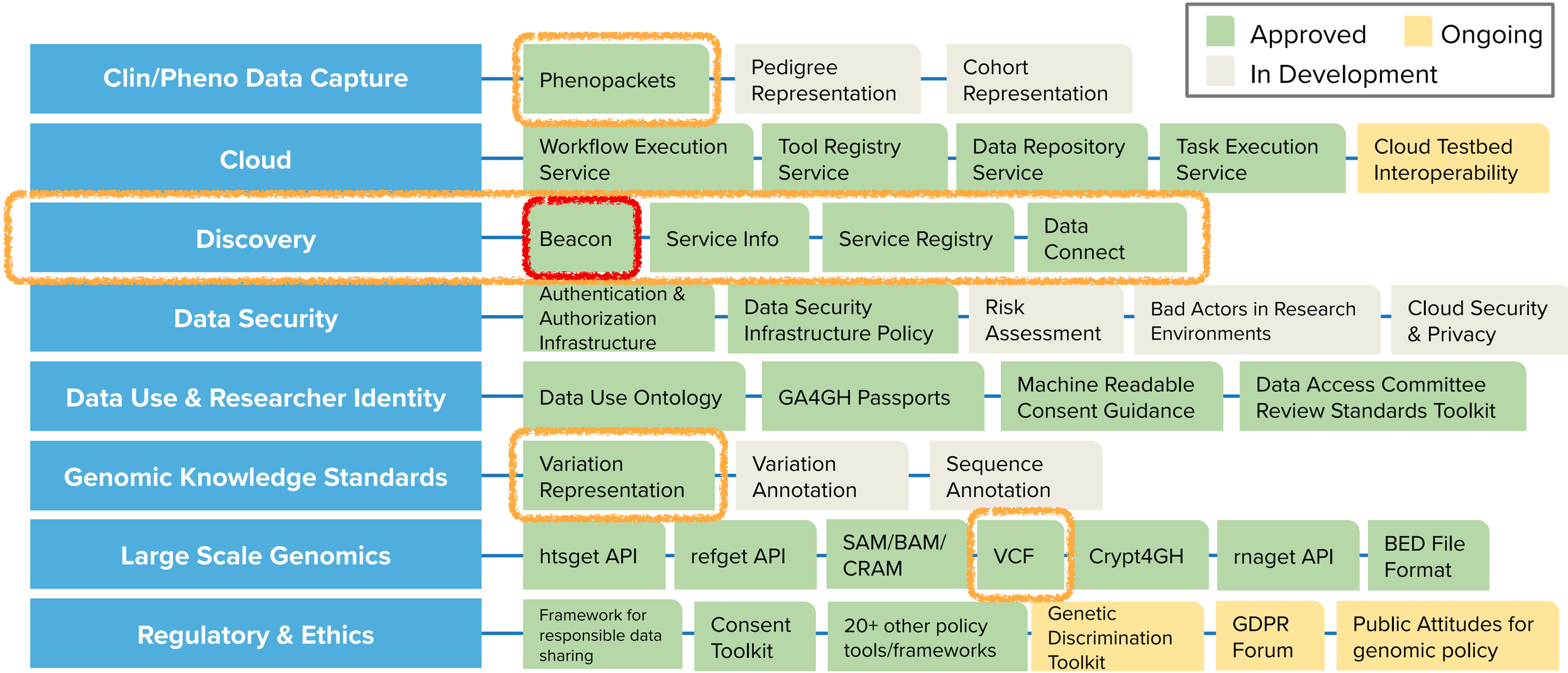
→ **UZH** and **SIB** are founding members of GA4GH



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



Overview of GA4GH standards and frameworks



GA4GH timeline

Global Alliance for Genomics & Health

GA4GH 1st Plenary London, UK

GA4GH 2nd Plenary San Diego, USA

GA4GH 3rd Plenary Leiden, Netherlands

GA4GH 4th Plenary Vancouver, Canada

GA4GH 5th Plenary Orlando, USA

GA4GH 6th Plenary Basel, Switzerland

GA4GH 7th Plenary Boston, USA

GA4GH 8th Plenary Virtual

GA4GH 9th Plenary Virtual

GA4GH 10th Plenary Barcelona, Spain

GA4GH 11th Plenary San Francisco, USA

2013 — 2014 — 2021 — 2022 — 2023 — 2024

Beacon Proposal and Conceptual Discussions

Pre-launch

73 partners sign a letter of intent to form an alliance

Building momentum

Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

Formal launch of GA4GH

Published *Framework for Responsible Sharing of Genomic and Health-Related Data*

Formed four working groups

Developed three demonstration projects

GA4GH Connect

GA4GH CONNECT

Launch of **GA4GH Connect** and Strategic Roadmap

Formation of new organizational structure consisting of eight Work Streams and over twenty Driver Projects

Gap analysis

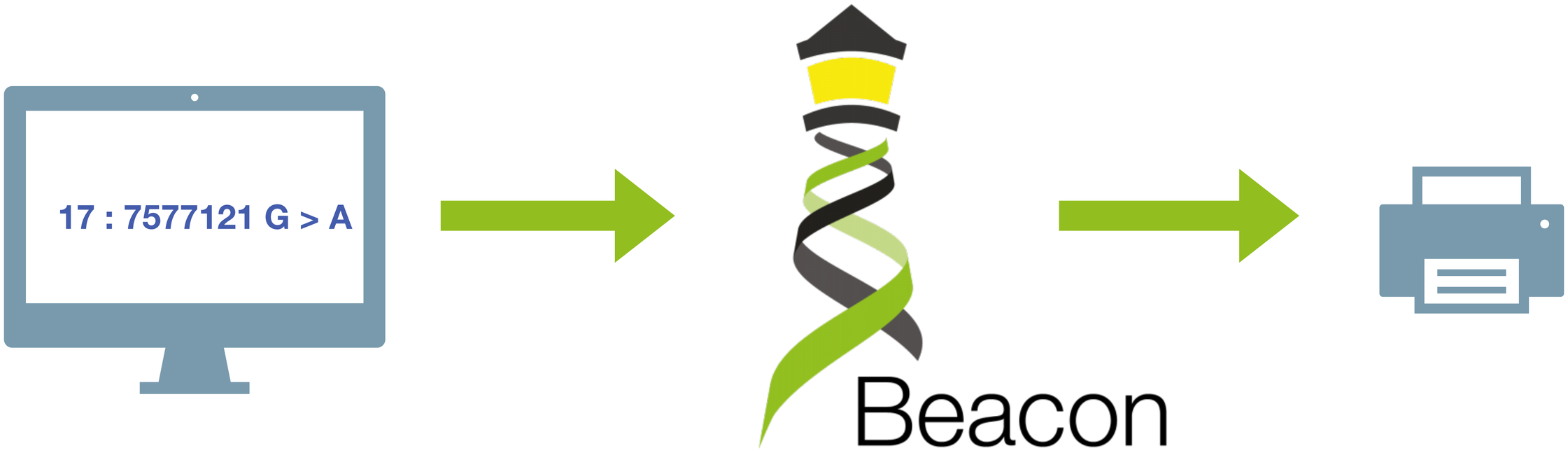
Gap analysis identifies three community imperatives

- Interoperability and alignment
- Implementation support
- Engaging with healthcar and clinical standards

Strategic Refresh

Strategic refresh introduces updates to GA4GH to better meet the three community imperatives

	Beacon v1 Development	Beacon v2 Development	Related ...
2014	**GA4GH founding event; Jim Ostell proposes Beacon concept including "more features ... version 2"**		
2015	• beacon-network.org aggregator created by DNAstack		• ELIXIR starts Beacon project support
2016	• Beacon v0.3 release • work on queries for structural variants (brackets for fuzzy start and end parameters...)	• Beacon+ concept implemented @ **progenetix.org** • concepts from GA4GH Metadata (ontologies…) • entity-scoped query parameters ("individual.age")	
2017	• OpenAPI implementation • integrating **CNV parameters** (e.g. "startMin, statMax")	• Beacon+ demos "**handover**" concept	• GA4GH re-structuring (workstreams...) • Beacon part of Discovery WS
2018	• Beacon v0.4 release in January; feature release for GA4GH approval process • **GA4GH Beacon v1 approved** at Oct plenary	• Beacon hackathon Stockholm; settling on **filters** • Barcelona goes Zurich developers meeting • Beacon API v2 Kick off • adopting "handover" concept	• new Beacon website (March)
2019	• ELIXIR Beacon Network		• Beacon publication at Nature Biotechnology
2020		• "**Scouts**" teams working on different aspects - filters, genomic variants, compliance ...	
2021		• **framework** + **models** concept implemented • range and bracket queries, variant length • starting of GA4GH review process	• Phenopackets v2 approved
2022		• changes in default model, aligning with Phenopackets and VRS • unified beacon-v2 code & docs repository • **Beacon v2 approved** at April GA4GH Connect	• *docs.genomebeacons.org*



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

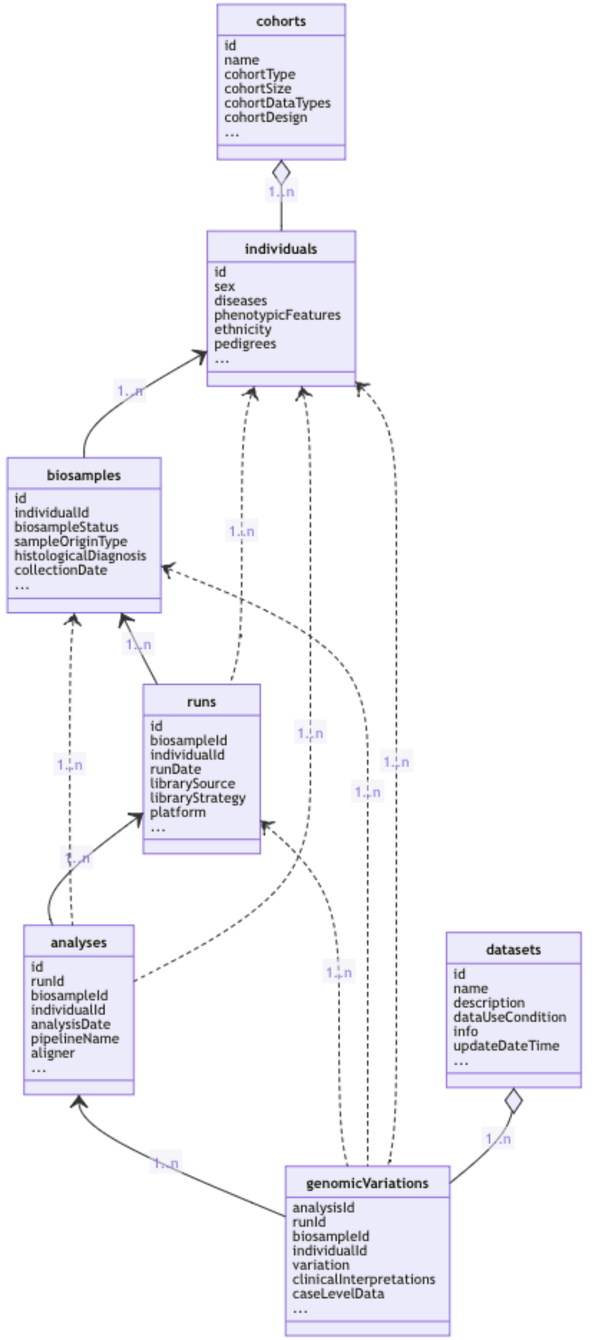
YES | **NO** | **\0**

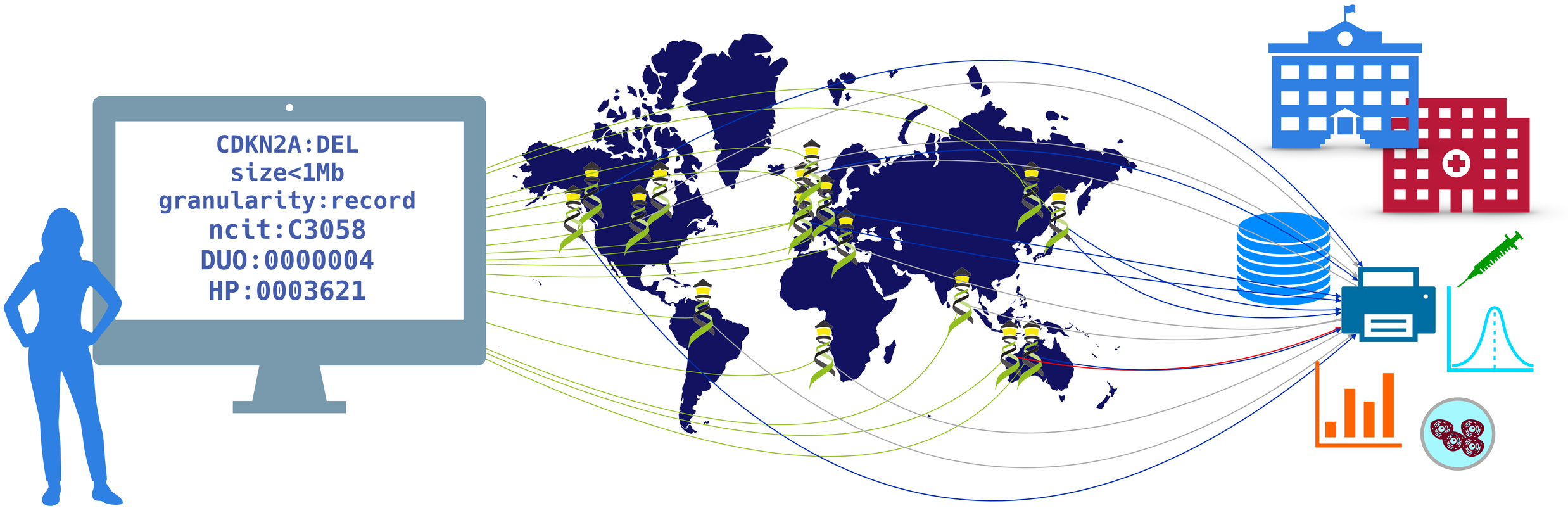
A ***Beacon*** answers a query for a specific genome variant against individual or aggregate genome collections

YES | **NO** | \0

Beacon Default v2 Model

- The Beacon ***framework*** describes the overall structure of the API requests, responses, parameters, the common components, etc.
- Beacon ***models*** describe the set of concepts included in a Beacon, like individual or biosample, and also the relationships between them.
- Besides logical concepts, the Beacon ***models*** represent the schemas for data delivery in “record” granularity
- Beacon explicitly allows the use of *other models* besides its *version specific default*.
- Adherence to a shared ***model*** empowers federation
- Use of the ***framework*** w/ different models extends adoption





Can you provide data about focal deletions in CDKN2A in Glioblastomas from juvenile patients with unrestricted access?

But... Can you really do that with the Beacon default model? Let's explore...

Request Components

Deparsing the Beacon v2 Example

- query against genomic variations, no matter how they are stored
- copy number deletion, as indicated through the VCF symbolic allele `DEL` expression

  ➡ preferably replaced by VRS v2 term
- a combination of `genId` (server side gene data) OR
- a range query and `variantMaxLength`, or positional (`start`, `end`)
- a filter for the Glioblastoma diagnosis, as NCIT term NCIT:C3058
- as an HPO term for "juvenile" HP:0003621
- full data access as per DUO:0000004

Request Components

Deparsing the Beacon v2 Example

- Where can these parameters be applied in the current default model?
  - variant parameters `/g_variants`
  - histologicalDiagnosis `/biosamples`
  - dataUseConditions `/dataset`
  - ageOfOnset term `/individual`
- Such a request requires the application of filtering terms to other entities than the ones indicated by their entry type
- Variant parameters always need **aggregation**
- ... this does not prohibit simple implementations

Standards Development & Implementation: CNV Terms

in computational (file/schema) formats

```
- EFO:0030064
- EFO:0030067
    |- EFO:0030068
    \- EFO:0020073
        \- EFO:0030069
- EFO:0030070
    |- EFO:0030071
    \- EFO:0030072
```

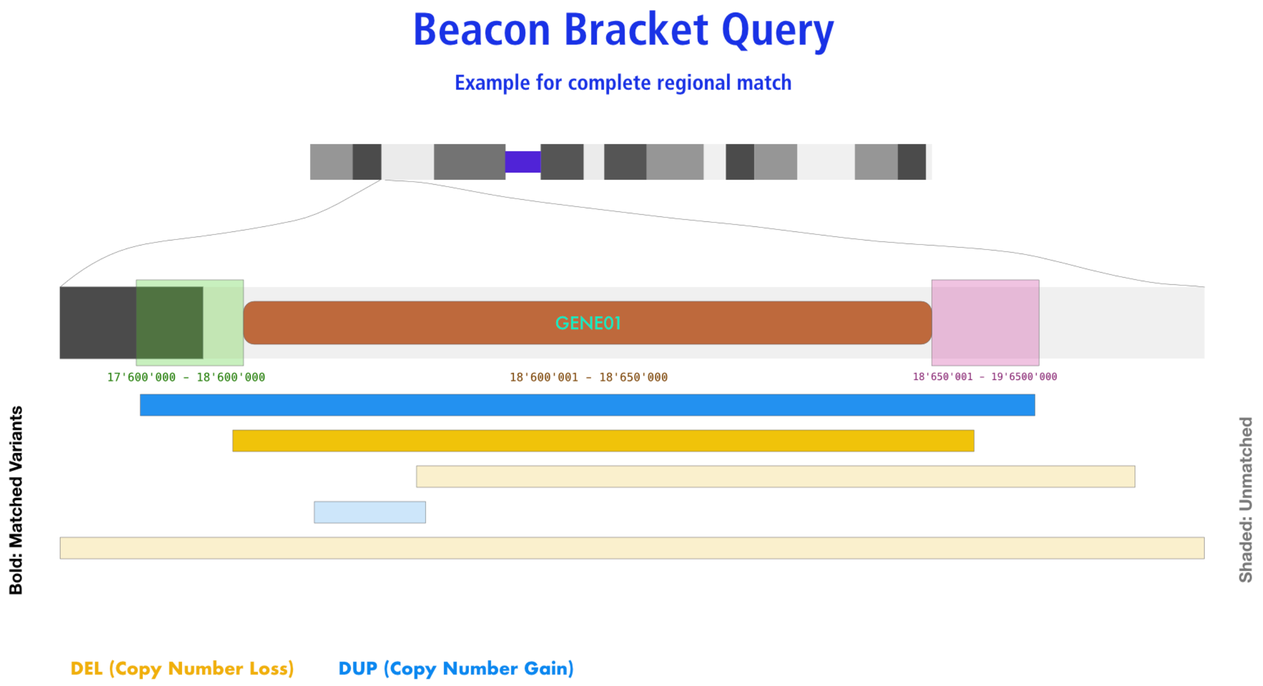
- Several competing concepts when annotating "systemic" variants such as CNV
- A vocabulary reflecting current consensus about CNV "level status types served as basis for copyNumberVariation class in GA4GH VRS
- Progenetix provides a use case for CNV profiles of ~240'000 cancer CNV profiles encoded in VRS v2 and Beacon API support

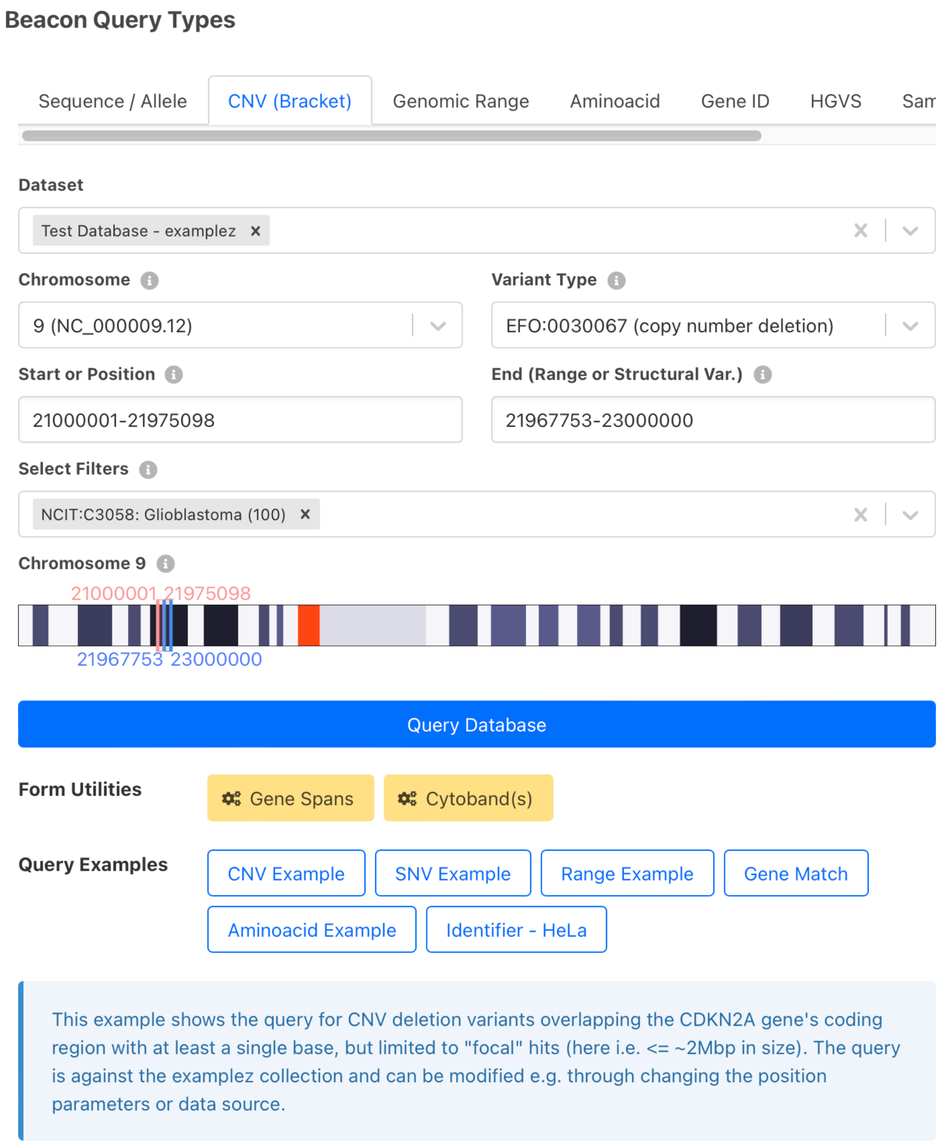
GA4GH VRS v2	Beacon v2	VCF v4.4	SO
gain (EFO:0030070)	DUP or **EFO:0030070**	DUP SVCLAIM=D	SO:0001742 copy_number_gain
low-level gain (EFO:0030071)	DUP or **EFO:0030071**	DUP SVCLAIM=D	SO:0001742 copy_number_gain
high-level gain (EFO:0030072)	DUP or **EFO:0030072**	DUP SVCLAIM=D	SO:0001742 copy_number_gain
high-level gain (EFO:0030072)	DUP or EFO:0030073 focal genome amplification	DUP SVCLAIM=D	SO:0001742 copy_number_gain
loss (EFO:0030067)	DEL or **EFO:0030067**	DEL SVCLAIM=D	SO:0001743 copy_number_loss
low-level loss (EFO:0030068)	DEL or **EFO:0030068**	DEL SVCLAIM=D	SO:0001743 copy_number_loss
high-level loss (EFO:0020073)	DEL or **EFO:0020073**	DEL SVCLAIM=D	SO:0001743 copy_number_loss
complete genomic loss (EFO:0030069)	DEL or **EFO:0030069**	DEL SVCLAIM=D	SO:0001743 copy_number_loss

Variation Queries

Bracket ("CNV") Query

- defined through the use of 2 start, 2 end
- any contiguous variant...





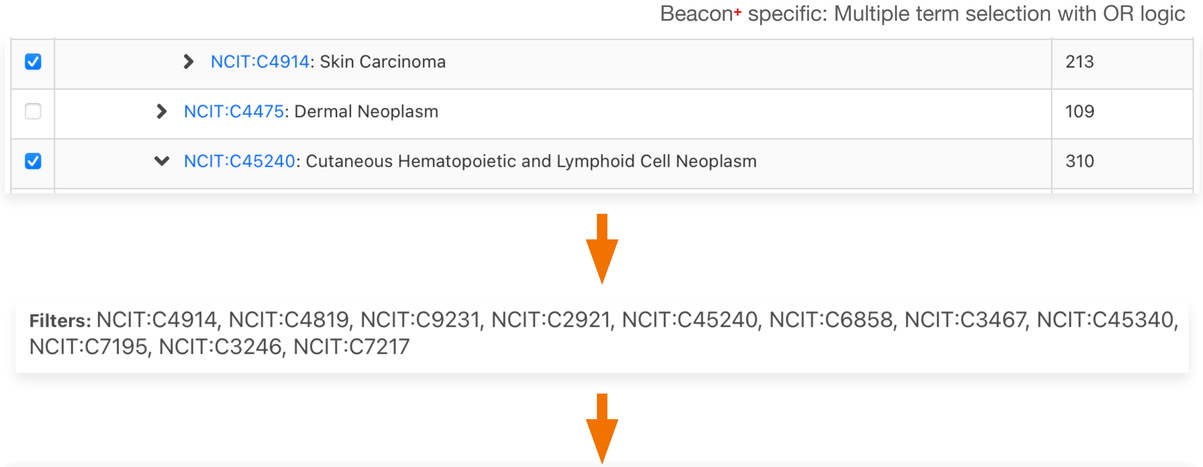
Beacon v2 Filters

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

- Beacon v2 relies heavily on "filters"
  - ontology term / CURIE
  - alphanumeric
  - custom
- 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

Beacon+ specific: Multiple term selection with OR logic



DEMONSTRATOR – Genomic queries

Female patients with focal high-level deletion in the CDKN2A locus in Glioblastoma tumors

```
GET /beacon/individuals/
 ?variantType=EFO%3A0020073
 &referenceName=NC_000009.12
 &start=21000000,21975098
 &end=21967753,23000000
 &filters=NCIT%3AC3058,NCIT%3AC16576
```

Server-side intersect of

- variants from genomic bracket query
- histology from biosample (NCIT:C3058)
- sex from individual (NCIT:C16576)

ELIXIR Beacon Network

Begriffsbestimmung

The right expressions help to conceptualize...

- **Beacon**: The protocol/API, with framework and default model
- **beacon**: Implementation of Beacon
  - ➡ using the Beacon v2 framework & supporting at minimum boolean responses
  - ➡ suggested support of Beacon v2 default model but can choose other
- Beacon **Aggregator**: service distributes queries to beacons and aggregates responses into a single Beacon response
  - ➡ potential to liftover genomes, remap filtering terms, translate between protocol versions...
  - ➡ entry point to or potentially itself node in a ...
- Beacon **Network**: Set of beacons wit shared entry point for distributed queries and aggregated response delivery
  - ➡ "true" beacon networks should have managed aspects - scope, term use...
  - ➡ networks may combine mixes of internal (protected, rich data, additional extensions...) and external interfaces

WP4 Beacon Network

Leads:

Jordi Rambla (ES)
Michael Baudis (CH)
Jaakko Leinonen (FI)

Objectives

This Work Package will

- Maintain an **operational Beacon Network service**, with ELIXIR lead participation, including a service level target for availability and user service and incident response.
- Deliver transparent and responsive governance structures that appropriately represent stakeholders, delineate strategy, and **react to user feedback**.
- Provide an approach to change management which ensures **ongoing development** of the service through other activities is integrated into the live service with minimum impact on existing users and dependencies.

ELIXIR Beacon Network Plans & Challenges

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

- ELIXIR Beacon Network has been *prototyped* in a previous IS
- This CoS will:
  - Contribute to the costs of running a service in production
  - Solidify the governing bodies
    - Strategic
    - Operational
  - While keeping the running costs to a minimum
  - Allowing for take over by other partners if ever necessary
- Challenges
  - Relies on partner willingness to run the service
    - Only a part of the operational costs are covered
    - Additional funds should be guaranteed by the partners
  - Funding for development or **software maintenance** is not included, but these activities are **clearly necessary**

cnag
European Genomic Data Infrastructure
progenetix
NBIS
GHGA
THE UNIVERSITY OF MELBOURNE
health RI
enabling data driven health & life sciences
cancercelllines
UNIVERSITY OF LEICESTER
BioData.pt
elixir

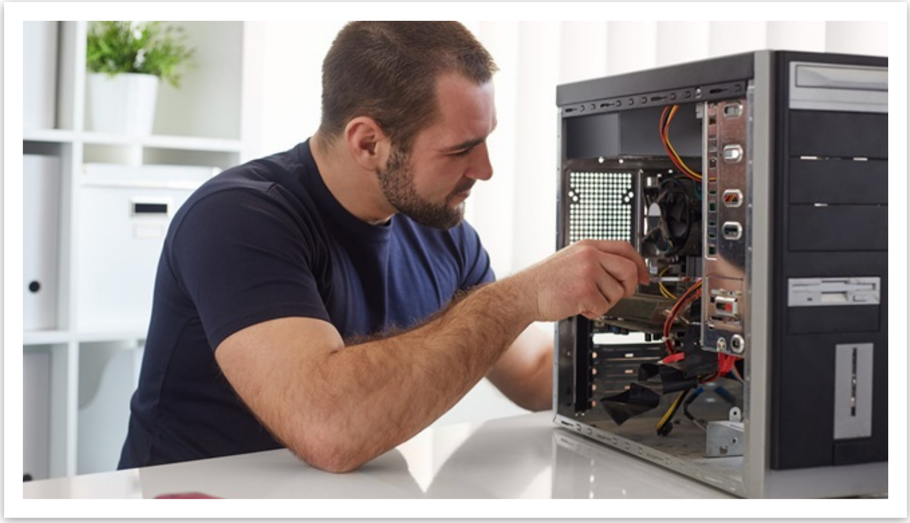
ELIXIR Beacon Network Summary

- ELIXIR Beacon Network in alignment with core stakeholders and external contributors to provide a working service with maximum capability profile
  - as GA4GH standard external input but also requirements to be accommodated
- importance of partner engagement since stakeholders (e.g. GDI …) currently lack ***live*** multi-node data scenarios but have a projected need
- high relevance of further ***protocol development*** to accommodate use cases **within** protocol specifications
  - RNAseq, pathogens, imaging …
  - massive opportunity to network people and data across domains through shared framework
- limited resources for "engage and ingest" work which is relevant for the global reach and - conversely - recognition and use of this ELIXIR standard and service

Beacon is a core ELIXIR asset. The Beacon Network provides a unique infrastructure service and is an attractor for cross-domain engagement

Beacon v2 deployment

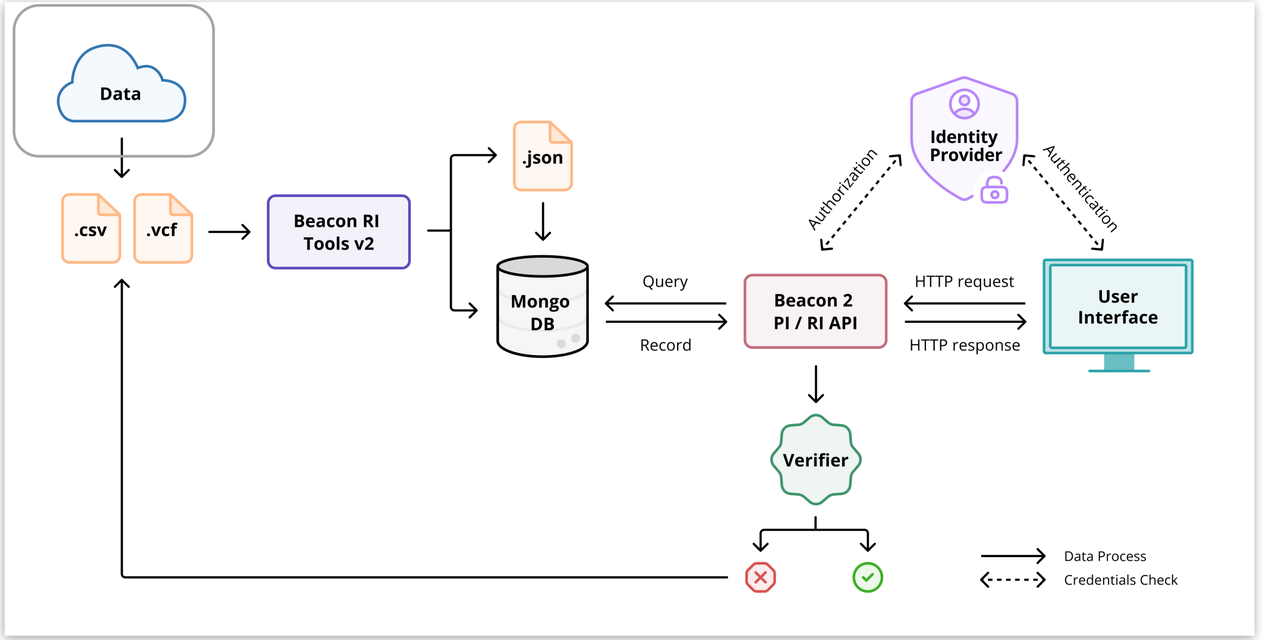
Build it yourself



GitHub

Beacon v2 API
https://github.com/ga4gh-beacon/beacon-v2

Toolkit for production environments



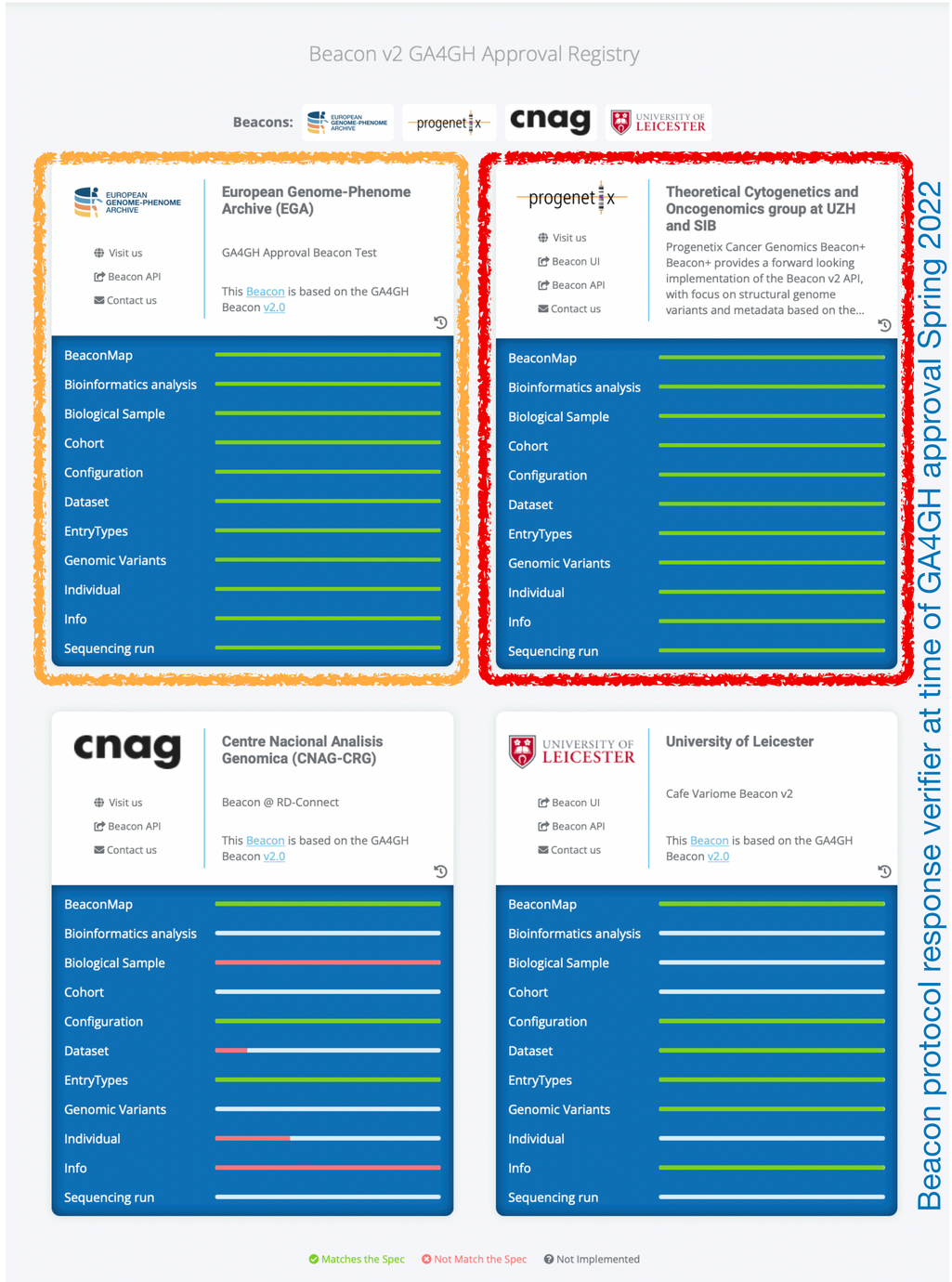
Beacon v2 Production Implementation (released Oct 2024)
https://github.com/ga4gh-beacon/beacon-v2

https://www.choice.com.au/electronics-and-technology/computers/desktop-and-laptop-computers/articles/is-building-your-own-pc-worth-it

bycon Beacon

Implementation driven standards development

- Progenetix' Beacon+ has served as implementation driver since 2016
- the *bycon* package is used to prototype advanced Beacon features such as
  - ➡ structural variant queries
  - ➡ data handovers
  - ➡ Phenopackets integration
  - ➡ variant co-occurrences
  - ➡ ...



Global Alliance for Genomics & Health

progenetix

elixir

bycon based Beacon+ Stack

progenetix

- *collations* contain pre-computed data (e.g. CNV frequencies, statistics) and information for all grouping entity instances and correspond to **filter values**
  - `pubmed:10027410, NCIT:C3222, pgx:cohort-TCGA, pgx:icdom-94703…`
  - precomputed frequencies per collection informative e.g. in form autofills
- *querybuffer* stores id values of all entities matched by a query and provides the corresponding `accessid` for **handover** generation
- complete query aggregation; i.e. individual queries are run against the corresponding entities and ids are intersected
  - ➡ retrieval of any entity, e.g. all individuals which have queried variants analyzed on a given platform
  - ➡ allows multi-variant queries, i.e. all bio samples or individuals which had matches of all of the individual variant queries

React

APACHE HTTP SERVER PROJECT

python™

mongoDB

Entity collections

Utility collections

github.com/progenetix/bycon GitHub

Beacon as a global standard

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

Beacon Scouts

Real-world needs

Cancer

Common diseases

Rare Diseases

…

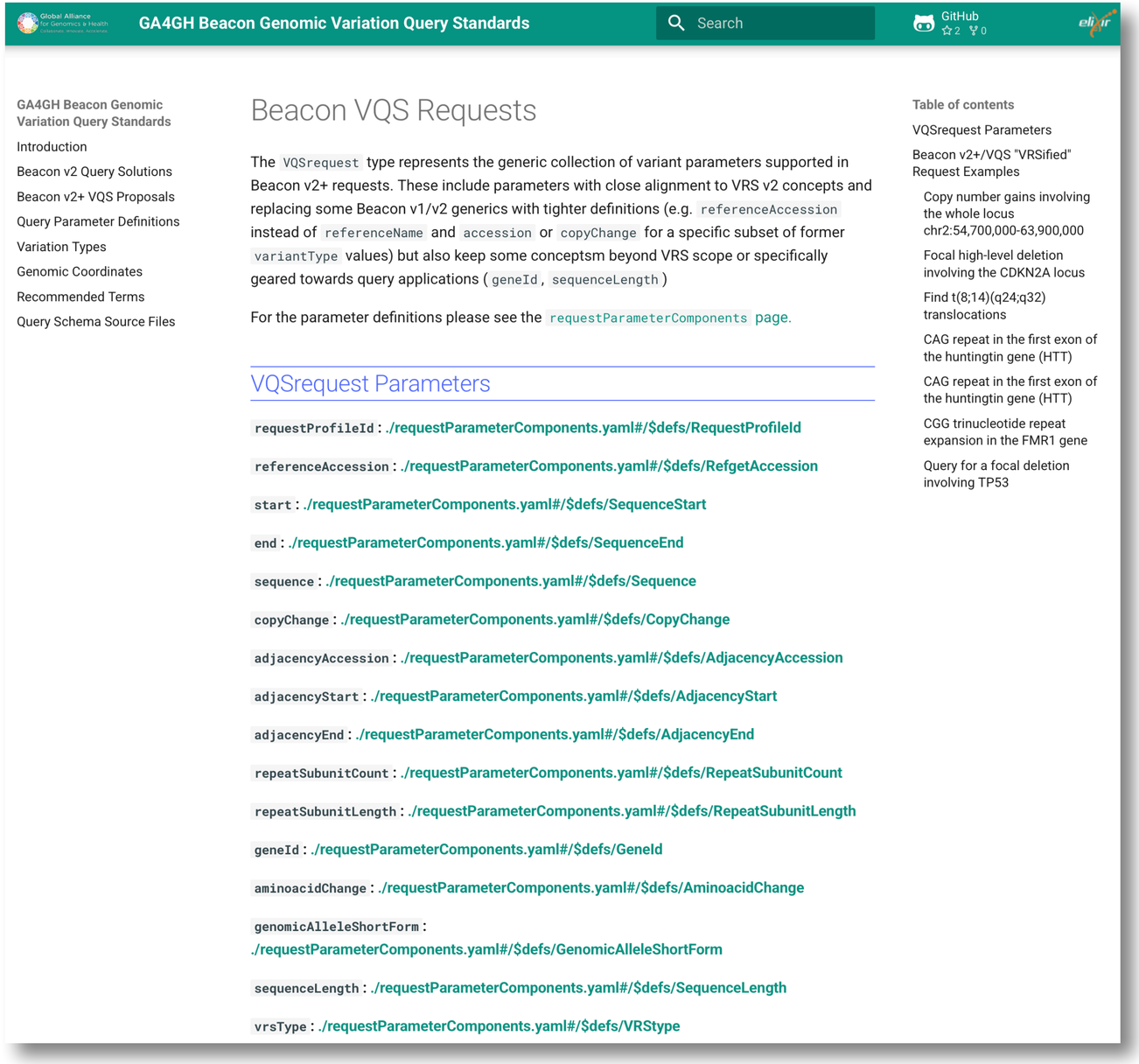
- **Beacon Filters** – improve current filter solutions
- **Beacon Cohorts** – develop aggregated request and response (e.g. counts by sex and age)
- **Beacon Variants** – expand specification to cover new use cases and typed queries
- **Beacon Dev** – improve API (cleaning code, GitHub issues)
- **Beacon Matchmaking** – implementation in matchmaking use cases

EUROPEAN GENOME-PHENOME ARCHIVE

Beacon Scouts

Finding the Paths to Beacon's Future

- Genomic Variation Scouts
  - ➡ extension to the query model based on assessed needs
    - fusions/breakpoints, cytogenetic annotations, repeats, categorical variants...
  - ➡ adoption of evolving VRS... standards for variant representation
    - adjacency, repeats…
    - re-use of parameters where clear (e.g. **`sequenceLength`** instead of **`variantMinLength`** + **`variantMaxLength`**)



https://genomebeacons.org/variant-query-types/variant-scouts-home/

VQS - Variant Query Standard

VRS aligned typed queries

- Typed queries
  - ➡ query schemas with defined set of (required and optional) parameters
    - can be verified
    - profile ids can be advertised by beacons
- VRS aligned
  - ➡ explicit reference to VRS types
  - ➡ … but differ in (some) parameter use since query NE representation
- Expanding library
  - ➡ adjacency, repeats...

```yaml
VQScopyChangeRequest:
  description: |-
    A typical Beacon v2.n request for copy number vari
    approximate positions for CNV start and end regior
    `Range` type. The `copyChange` parameter indicates
    genomic copy number (pls. refer to the class defir
  type: object
  properties:
    requestProfile:
      const: VQScopyChangeRequest
    referenceAccession:
      $ref: "./requestParameterComponents.yaml#/$defs/
    startRange:
      $ref: "./requestParameterComponents.yaml#/$defs/
    endRange:
      $ref: "./requestParameterComponents.yaml#/$defs/
    copyChange:
      $ref: "./requestParameterComponents.yaml#/$defs/
    sequenceLength:
      $ref: "./requestParameterComponents.yaml#/$defs/
    vrsType:
      const: CopyNumberChange
  required:
    - requestProfile
    - referenceAccession
    - startRange
    - endRange
    - copyChange
```

```yaml
requestProfile: VQScopyChangeRequest
referenceAccession: refseq:NC_000002.12
start:
  - 21000001
  - 21975098
end:
  - 21967753
  - 23000000
copyChange: EFO:0020073
vrsType: CopyNumberChange
```

```yaml
requestProfile: VQSadjacencyRequest
referenceAccession: refseq:NC_000008.11
start: 116700000
end: 145138636
adjacencyAccession: refseq:NC_000014.9
adjacencyStart: 89300000
adjacencyEnd: 107043718
vrsType: Adjacency
```

```yaml
requestProfile: VQSsequenceRepeatRequest
geneId: HTT
repeatSubunitLength: 3
sequenceLength:
  - 105
  - 750
vrsType: ReferenceLengthExpression
```

https://genomebeacons.org/variant-query-types/variant-scouts-home/

Aggregation / Summaries

Data Summaries

- Aggregated results offer a way for beacons to
  - de-couple information about the resource's data content from samples and individuals
  - highlight relevant features and data
  - support performant front-ends/dashboards
- *Distributions* for one and two-dimensional counts or (potentially) value representations
  - value selection, binning/bucketing trough parameters
- While custom dashboards are standard feature of genomic data resources, Beacon protocol support empowers

Federation => Aggregation^2

progenetix

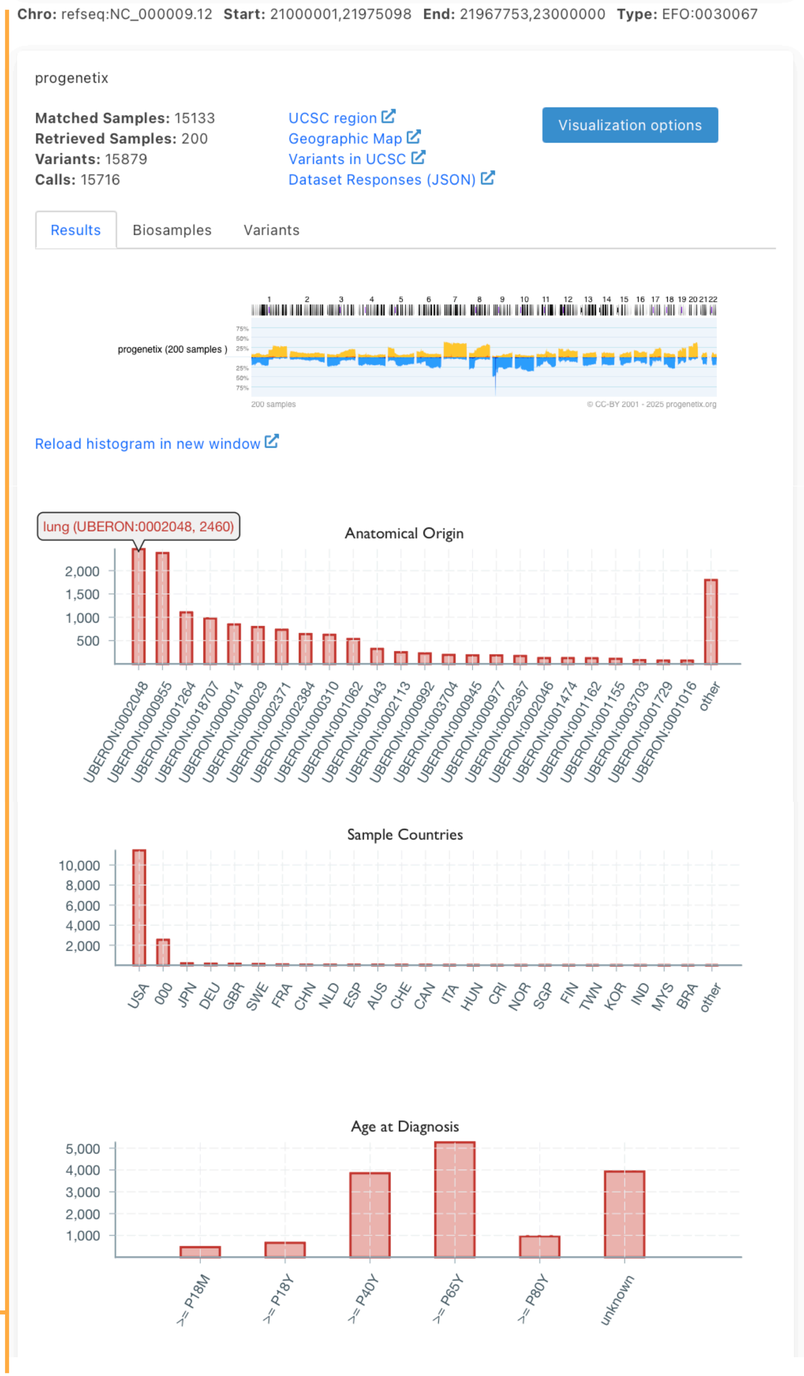
```
- concepts:
  - property: histological_diagnosis.id
    scope: biosample
  description: Histological diagnosesin matched biosamples
  distribution:
  - conceptValues:
    - id: NCIT:C132256
      label: Unspecified Tissue
    count: 18313
  - conceptValues:
    - id: NCIT:C3512
      label: Lung Adenocarcinoma
    count: 13657
  - conceptValues:
    - id: NCIT:C4017
      label: Breast Ductal Carcinoma
    count: 9634
  - conceptValues:
    - id: NCIT:C2919
      label: Prostate Adenocarcinoma
    count: 9115
...
- id: ageAtDiagnosis
  label: Age at Diagnosis
  sorted: true
  concepts:
  - property:index_disease.age
    scope: individual
    splits:
    - P0D
    - P1Y
    - P18M
    - P18Y
    - P120Y
  description: Follow-up time after diagnosis
  distribution:
  - conceptValues:
    - id: '>= P6M'
      label: '>= P6M'
    count: 3872
  - conceptValues:
    - id: '>= P1Y'
      label: '>= P1Y'
    count: 6041
...
```

Data Summaries

Representation of aggregated results in Beacon responses

- Aggregated results offer a way for beacons to
  - de-couple information about the resource's data content from samples and individuals
  - highlight relevant features and data
  - support performant front-ends/dashboards
- *Distributions* for one and two-dimensional counts or (potentially) value representations
  - value selection, binning/bucketing trough parameters
- While custom dashboards are standard feature of genomic data resources, Beacon protocol support empowers

Federation => Aggregation^2



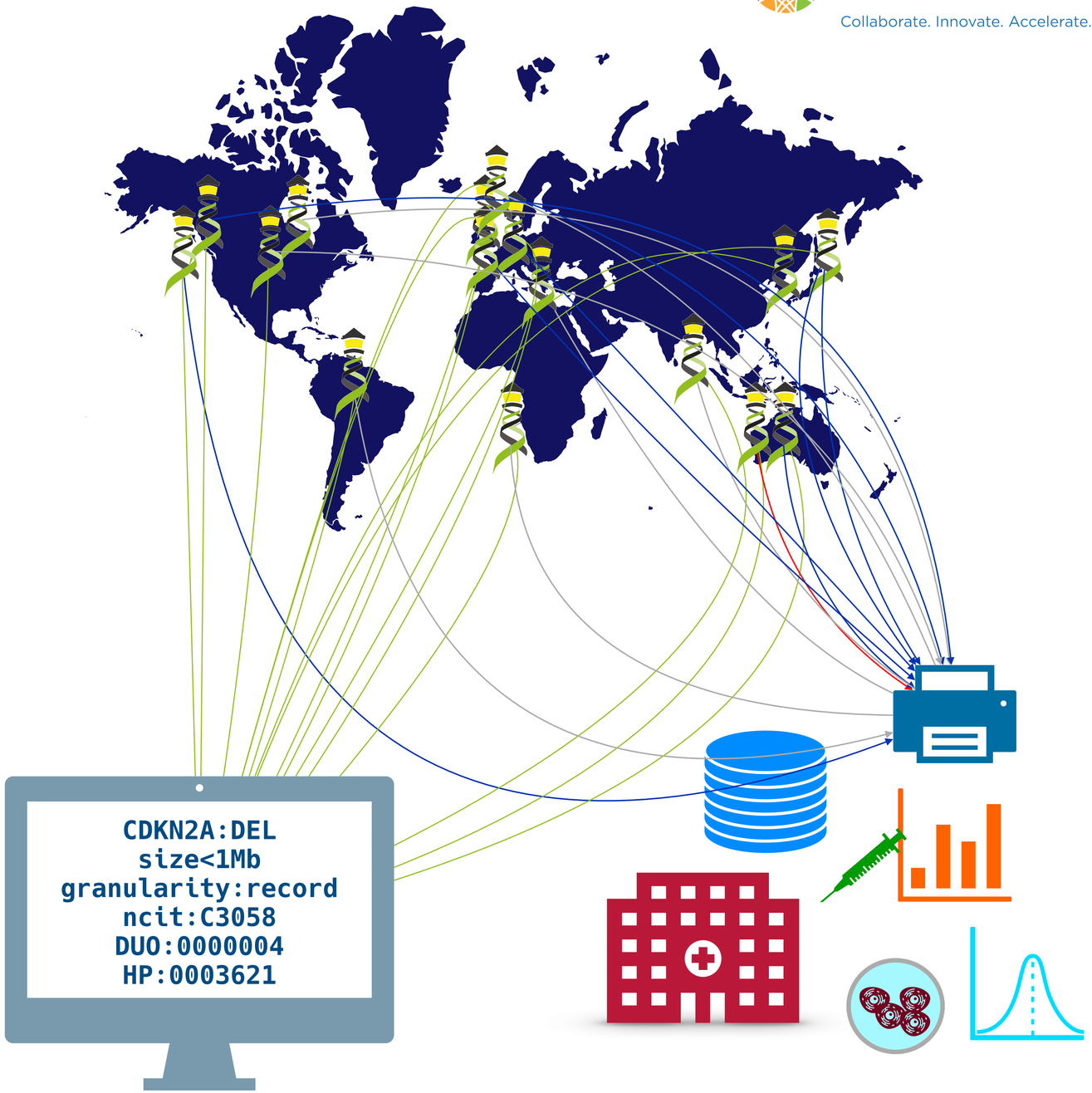
Summary aggregations implemented using a Beacon aggregated response prototype in Progenetix

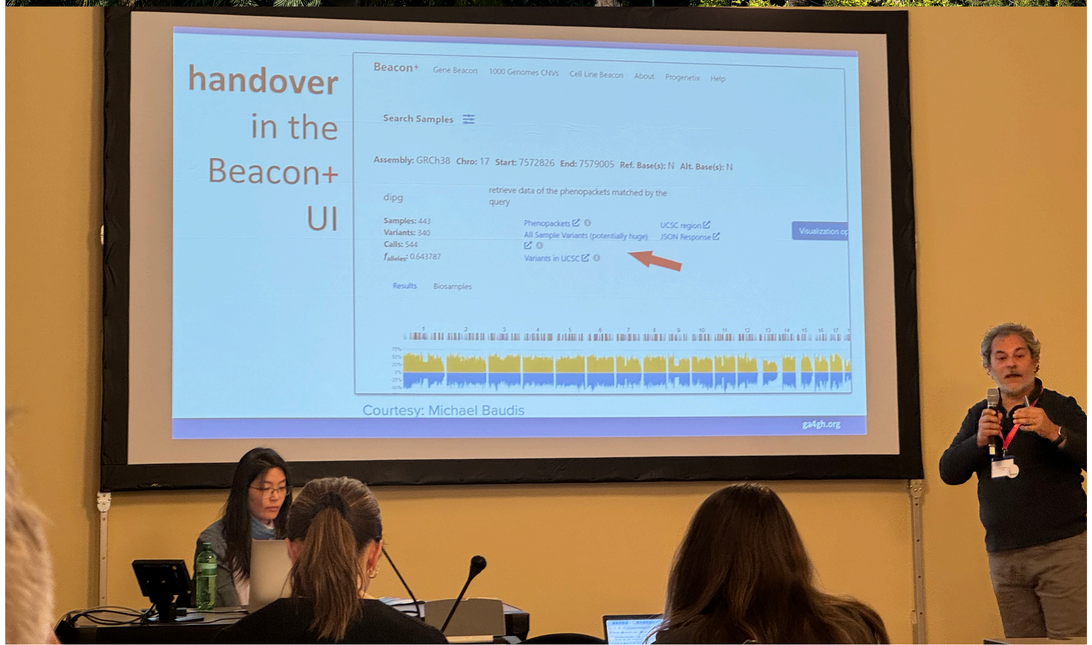
progenetix

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

BioTAPESTRY?

- host exchange visitors or project partners for
  - ➡ implementing Beacons via ***bycon***
  - ➡ porting workflows and tools to Galaxy (focus on CNV analyses)
  - ➡ text mining and metadata mappings for Progenetix
  - ➡ Beacon implementations for novel use cases
- visit for presentation, workshops, projects
  - ➡ teaching about Beacon, CNVs ..
  - ➡ group members for galaxyfying work





The Global Alliance for Genomics and Health (GA4GH) gathered for the 2024 **April Connect meeting** in Ascona, Switzerland and online from 21 to 24 April. The GA4GH Connect meetings provide an opportunity for contributors to advance the GA4GH Road Map, showcase GA4GH standards and policies in action, and gather feedback on product development and community needs. The meeting brought together 103 in-person attendees and 312 virtual attendees for updates from Work Streams and Driver Projects, breakout sessions, and themed events.

University of Zurich[UZH]
Department of Molecular Life Sciences