



# Driving oncogenomic and CNV reference resources through Beacon ✓

*Data Discovery :: Data Sharing :: Analysis Support*

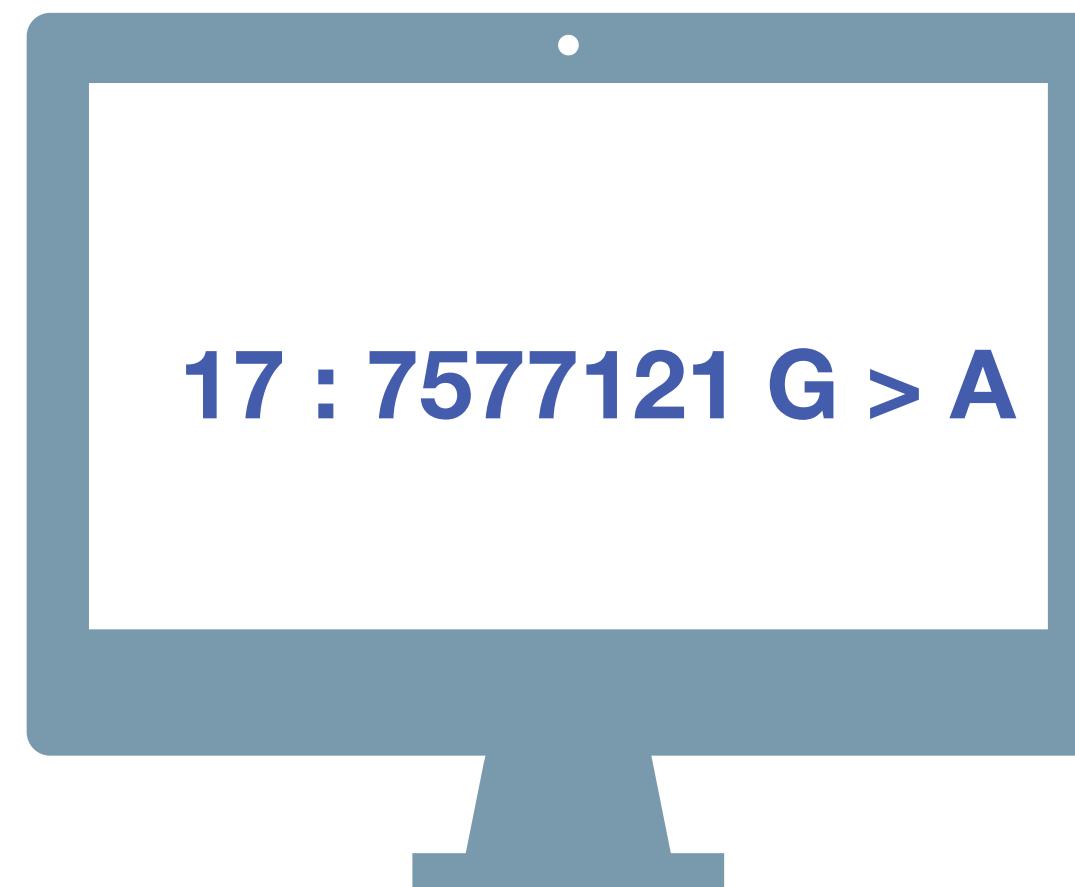


Global Alliance  
for Genomics & Health



University of  
Zurich UZH

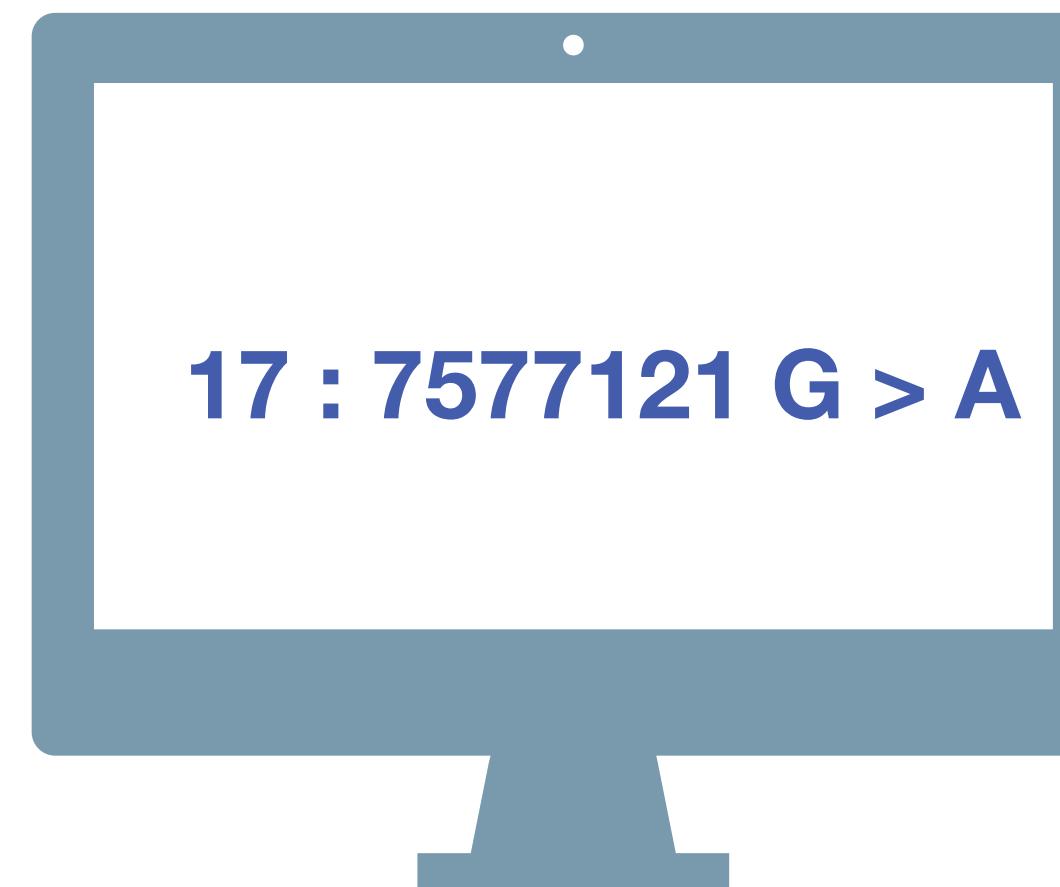
Michael Baudis :: ELIXIR AHM :: 2025-06-04



# Beacon

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

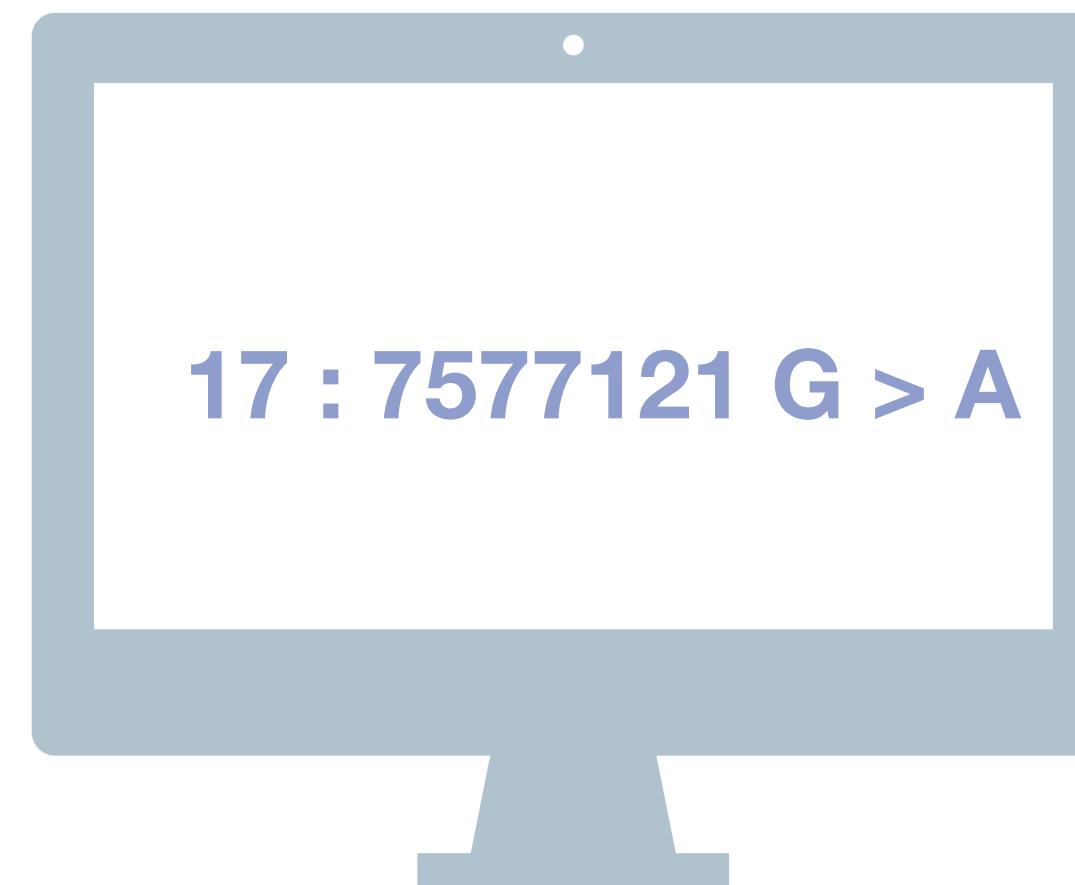
**YES | NO | \0**



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

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

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



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

**YES | NO | \0**

## Why

... 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 ... research for ... ***signs of willing participants in far reaching data sharing***, ... it has remained a dark and quiet place. [This] challenge is to 1) ***trigger the issues*** blocking groups ... in way that isn’t masked by the ... complexities ... 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**. ... provide a ***low bar for ... real ... engagement***. ... there is **some utility** in ...locating a rare allele in your data ... A number of more useful first versions have been suggested:

1. Provide ***frequencies of all alleles*** at that point
2. .. alleles seen in a gene ***region***
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 ... [potentially including a] ***“phone home” response*** ...

## Beacon v1 Development

2014

GA4GH founding event; Jim Ostell proposes Beacon concept including "more features ... version 2"

2015

- beacon-network.org aggregator created by DNAstack

2016

- Beacon v0.3 release
- work on queries for structural variants (brackets for fuzzy start and end parameters...)

2017

- OpenAPI implementation
- integrating **CNV parameters** (e.g. "startMin, statMax")
- Beacon v0.4 release in January; feature release for GA4GH approval process
- **GA4GH Beacon v1 approved** at Oct plenary

2019

- ELIXIR Beacon Network

2020



2021

## Beacon v2 Development

- Beacon+ concept implemented @ [progenetix.org](#)
- concepts from GA4GH Metadata (ontologies...)
- entity-scoped query parameters ("individual.age")

- Beacon+ demos "handover" concept

- Beacon hackathon Stockholm; settling on **filters**
- Barcelona goes Zurich developers meeting
- Beacon API v2 Kick off
- adopting "handover" concept

- "**Scouts**" teams working on different aspects - filters, genomic variants, compliance ...

- **framework + models** concept implemented
- range and bracket queries, variant length
- starting of GA4GH review process

- changes in default model, aligning with Phenopackets and VRS
- unified beacon-v2 code & docs repository
- **Beacon v2 approved** at April GA4GH Connect

2022

## Related ...

- ELIXIR starts Beacon project support

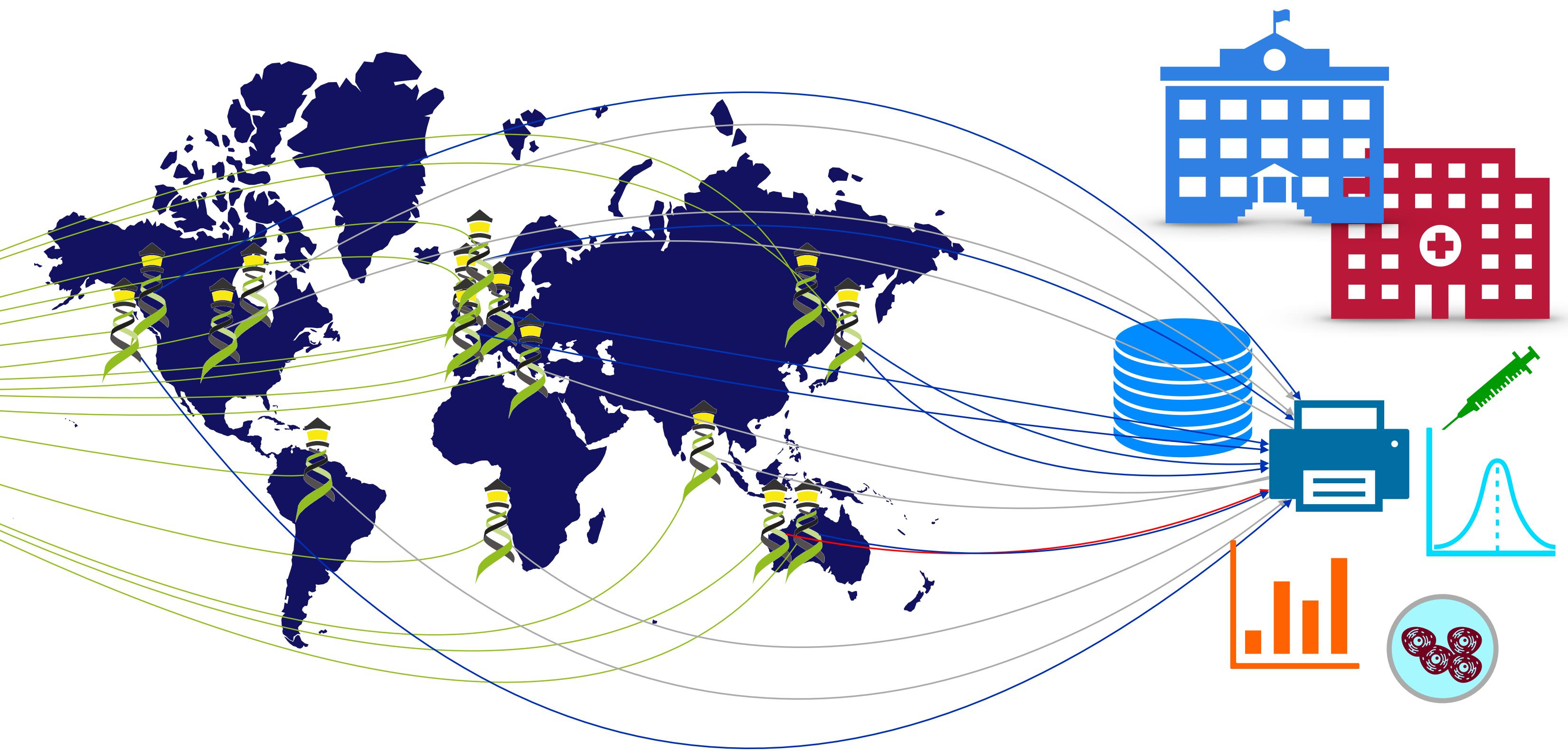
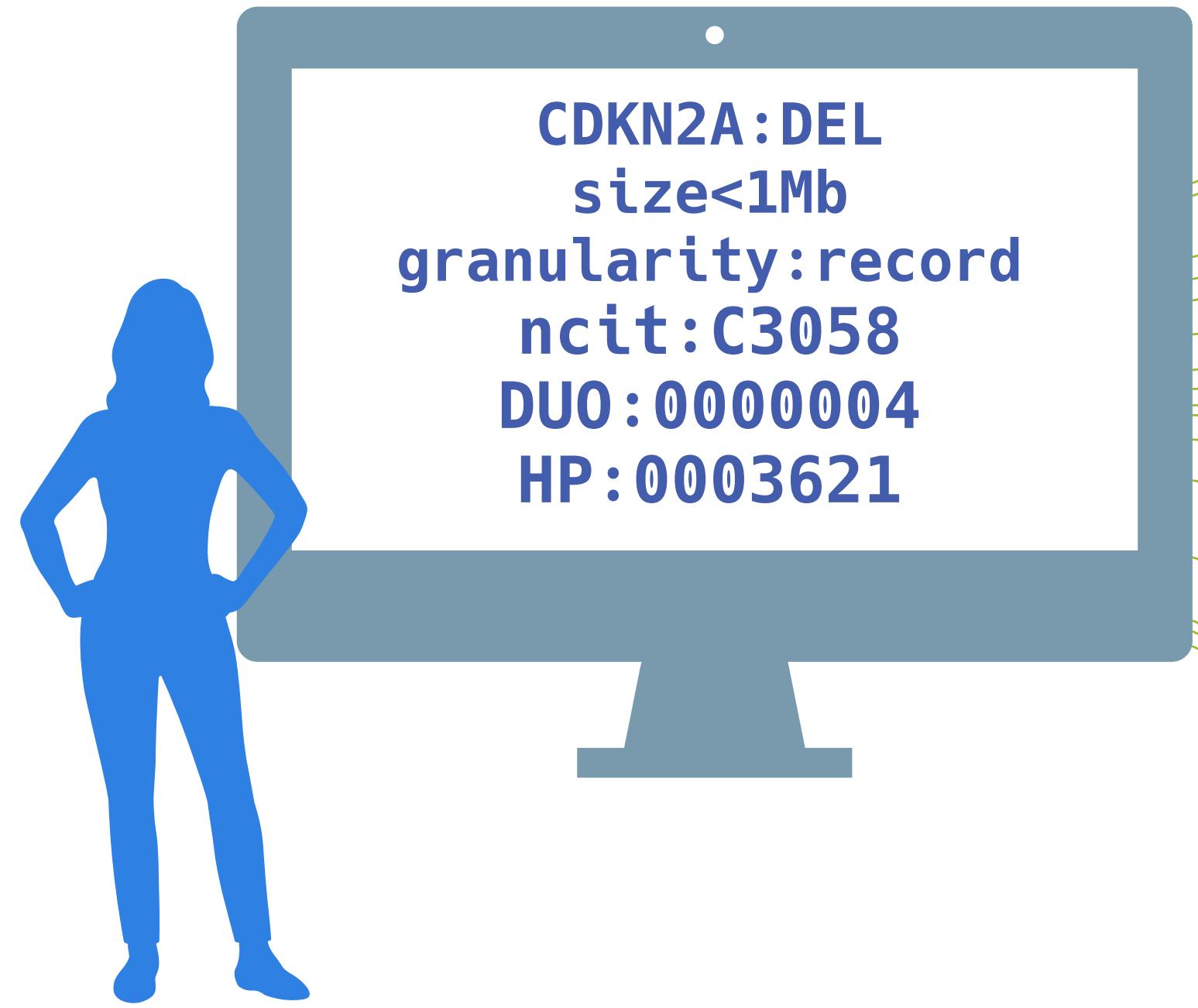
- GA4GH re-structuring (workstreams...)
- Beacon part of Discovery WS

- new Beacon website (March)

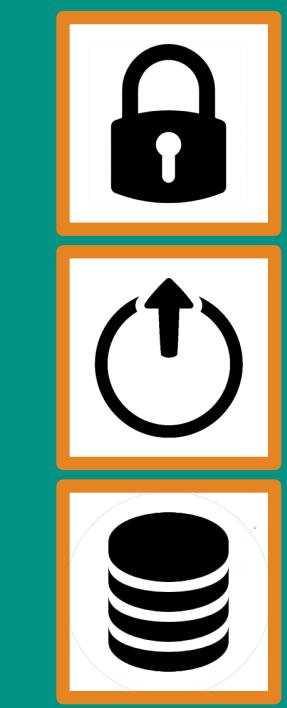
- Beacon publication at Nature Biotechnology

- Phenopackets v2 approved

- [docs.genomebeacons.org](#)



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

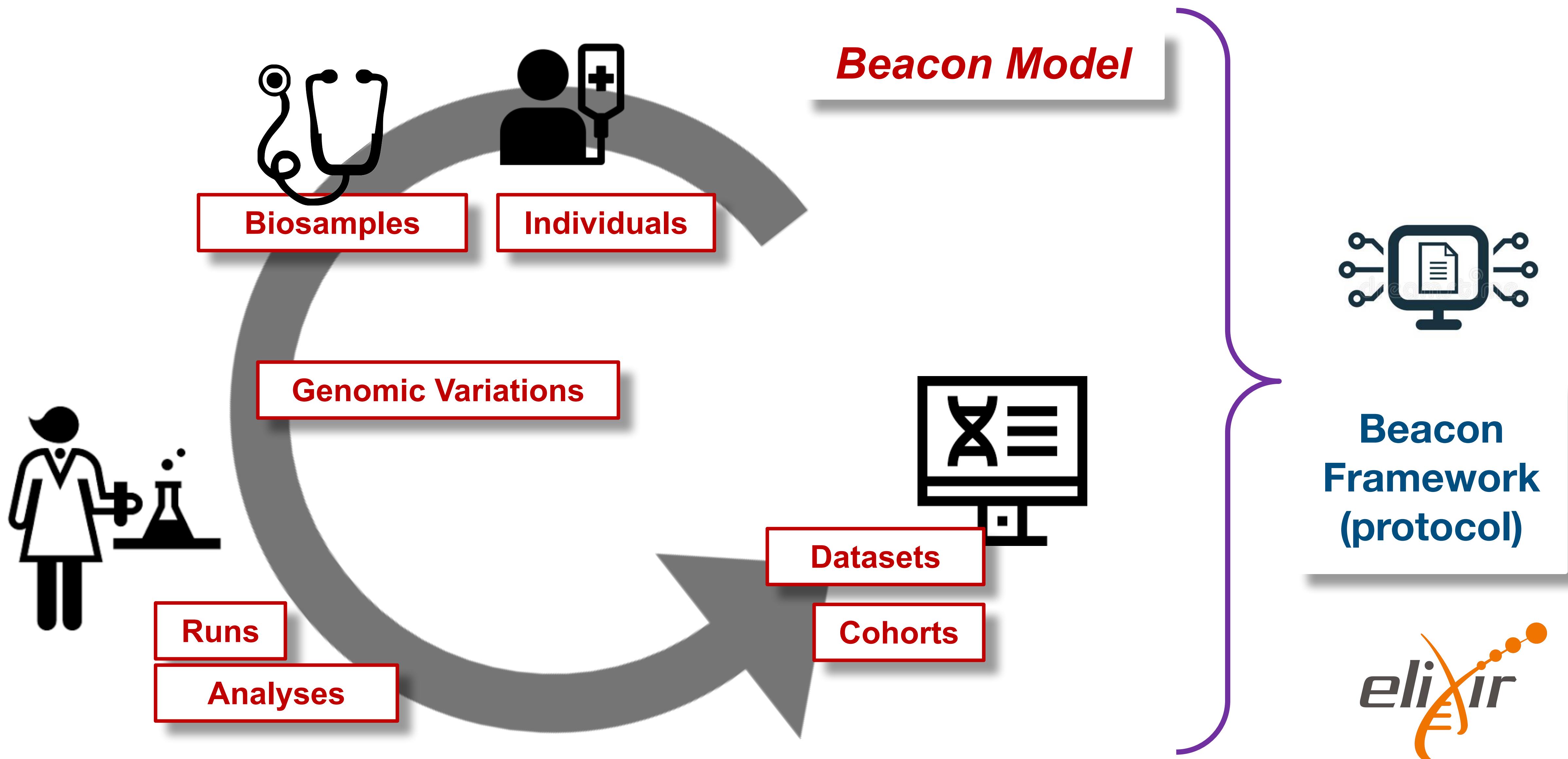


# Beacon API

The Beacon API represents a simple but powerful **genomics API** for **federated** data discovery and retrieval

# Beacon v2

docs.genomebeacons.org

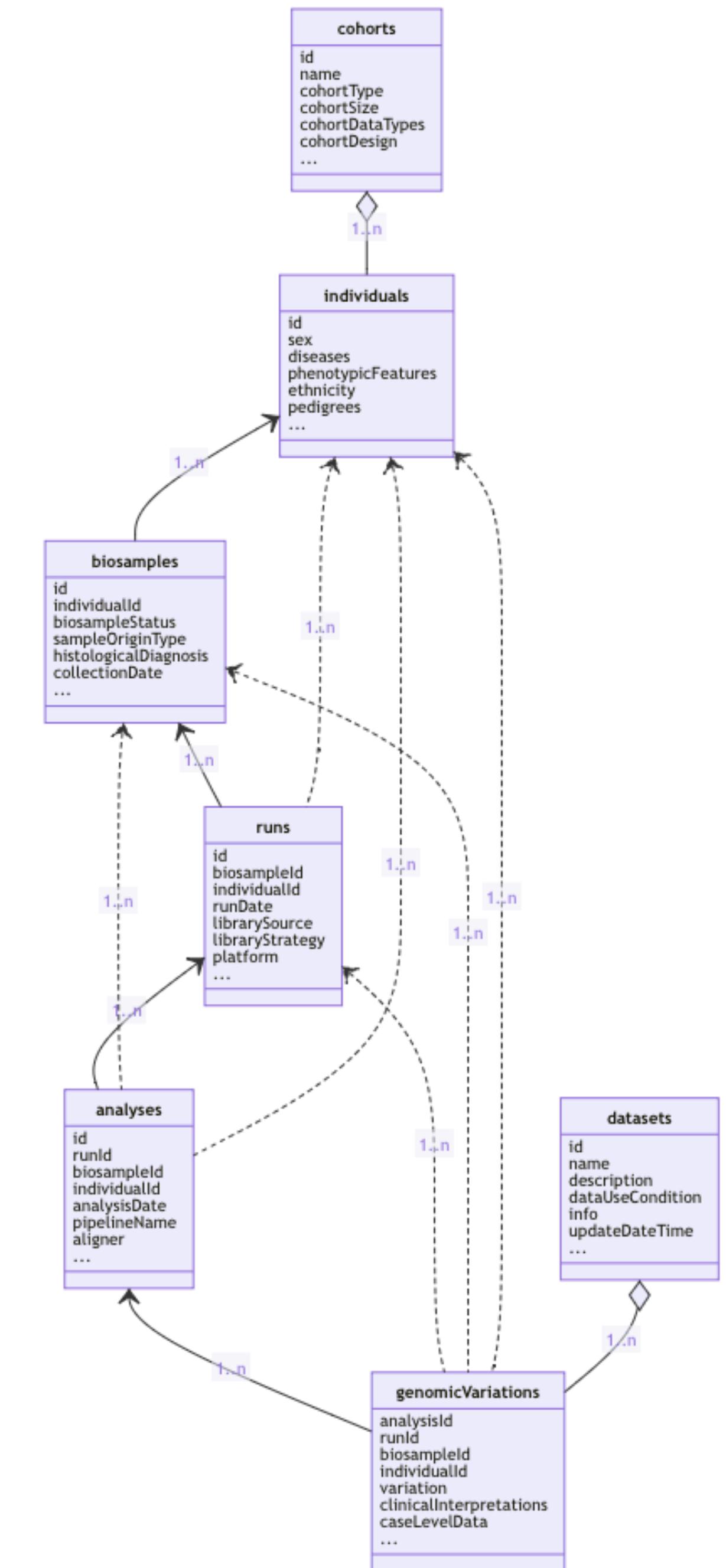


**Beacon  
Framework  
(protocol)**



# 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



# 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 with 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

# 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
  - ...

Beacon protocol response verifier at time of GA4GH approval Spring 2022

Beacon v2 GA4GH Approval Registry

Beacons:    

Category	EGA	progenetix	Theoretical Cytogenetics and Oncogenomics group at UZH and SIB
BeaconMap	Green	Green	Green
Bioinformatics analysis	Green	Green	Green
Biological Sample	Green	Green	Green
Cohort	Green	Green	Green
Configuration	Green	Green	Green
Dataset	Green	Green	Green
EntryTypes	Green	Green	Green
Genomic Variants	Green	Green	Green
Individual	Green	Green	Green
Info	Green	Green	Green
Sequencing run	Green	Green	Green

Category	cnag	University of Leicester
BeaconMap	Green	Green
Bioinformatics analysis	White	White
Biological Sample	Red	White
Cohort	White	White
Configuration	Green	White
Dataset	Red	White
EntryTypes	Green	White
Genomic Variants	White	White
Individual	Red	White
Info	White	White
Sequencing run	White	White

Green: Matches the Spec, Red: Not Match the Spec, White: Not Implemented



# progenetix.org

## Cancer Genomics Reference Resource

- **open** resource for oncogenomic profiles
- over **150'000 cancer CNV profiles**
- SNV data for some series (e.g. TCGA)
- more than **900 diagnostic types**
- inclusion of reference datasets (e.g. TCGA)
- standardized encodings (e.g. NCIIt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core clinical data (TNM, sex, survival ...)
- data mapping services



### Cancer CNV Profiles

ICD-O Morphologies  
ICD-O Organ Sites  
Cancer Cell Lines  
Clinical Categories

### Search Samples

arrayMap  
TCGA Samples  
1000 Genomes  
Reference Samples  
DIPG Samples  
cBioPortal Studies  
Gao & Baudis, 2021

### Publication DB

Genome Profiling  
Progenetix Use

### Services

NCIt Mappings  
UBERON Mappings

### Upload & Plot

### Beacon<sup>+</sup>

### Documentation

News  
Downloads & Use  
Cases  
Sevices & API

### Baudisgroup @ UZH

## Cancer genome data @ progenetix.org

The Progenetix database provides an overview of mutation data in cancer, with a focus on copy number abnormalities (CNV / CNA), for all types of human malignancies. The data is based on *individual sample data* from currently **142063** samples.

### Floor of the Mouth Neoplasm (NCIT:C4401)



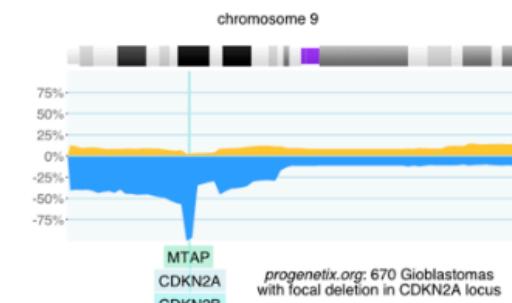
[Download SVG](#) | [Go to NCIT:C4401](#) | [Download CNV Frequencies](#)

Example for aggregated CNV data in 126 samples in Floor of the Mouth Neoplasm.  
Here the frequency of regional **copy number gains** and **losses** are displayed for all 22 autosomes.

### Progenetix Use Cases

#### Local CNV Frequencies

A typical use case on Progenetix is the search for local copy number aberrations - e.g. involving a gene - and the exploration of cancer types with these CNVs. The [\[ Search Page \]](#) provides example use cases for designing queries. Results contain basic statistics as well as visualization and download options.



progenetix.org: 670 Glioblastomas with local deletion in CDKN2A locus

#### Cancer CNV Profiles

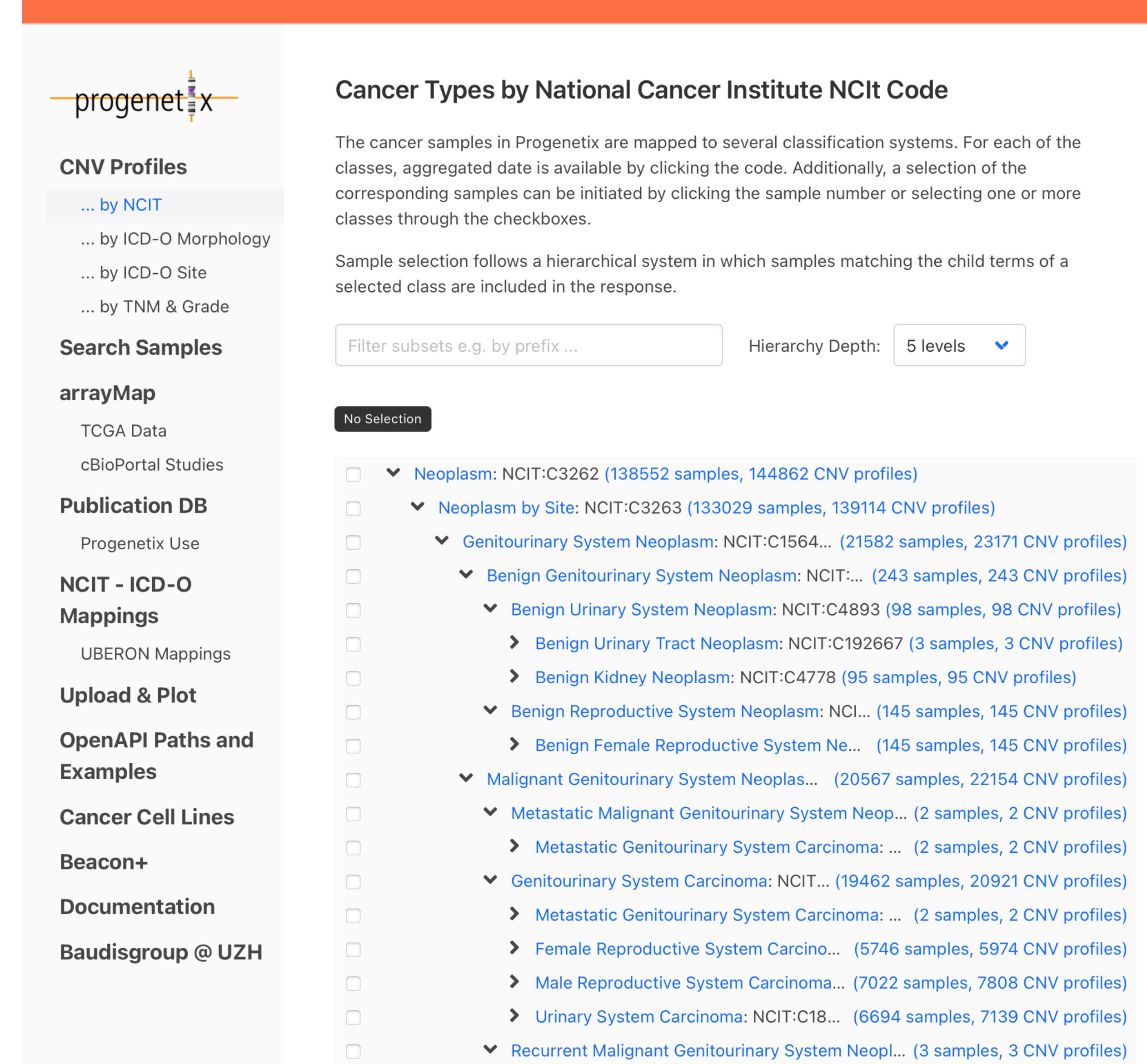
The progenetix resource contains data of **834** different cancer types (NCIt neoplasm classification), mapped to a variety of biological and technical categories. Frequency profiles of regional genomic gains and losses for all categories (diagnostic entity, publication, cohort ...) can be accessed through the [\[ Cancer Types \]](#) page with direct visualization and options for sample retrieval and plotting options.

#### Cancer Genomics Publications

Through the [\[ Publications \]](#) page Progenetix provides **4164** annotated references to research articles from cancer genome screening experiments (WGS, WES, aCGH, cCGH). The numbers of analyzed samples and possible availability in the Progenetix sample collection are indicated.

## Cancer Genomics Reference Resource

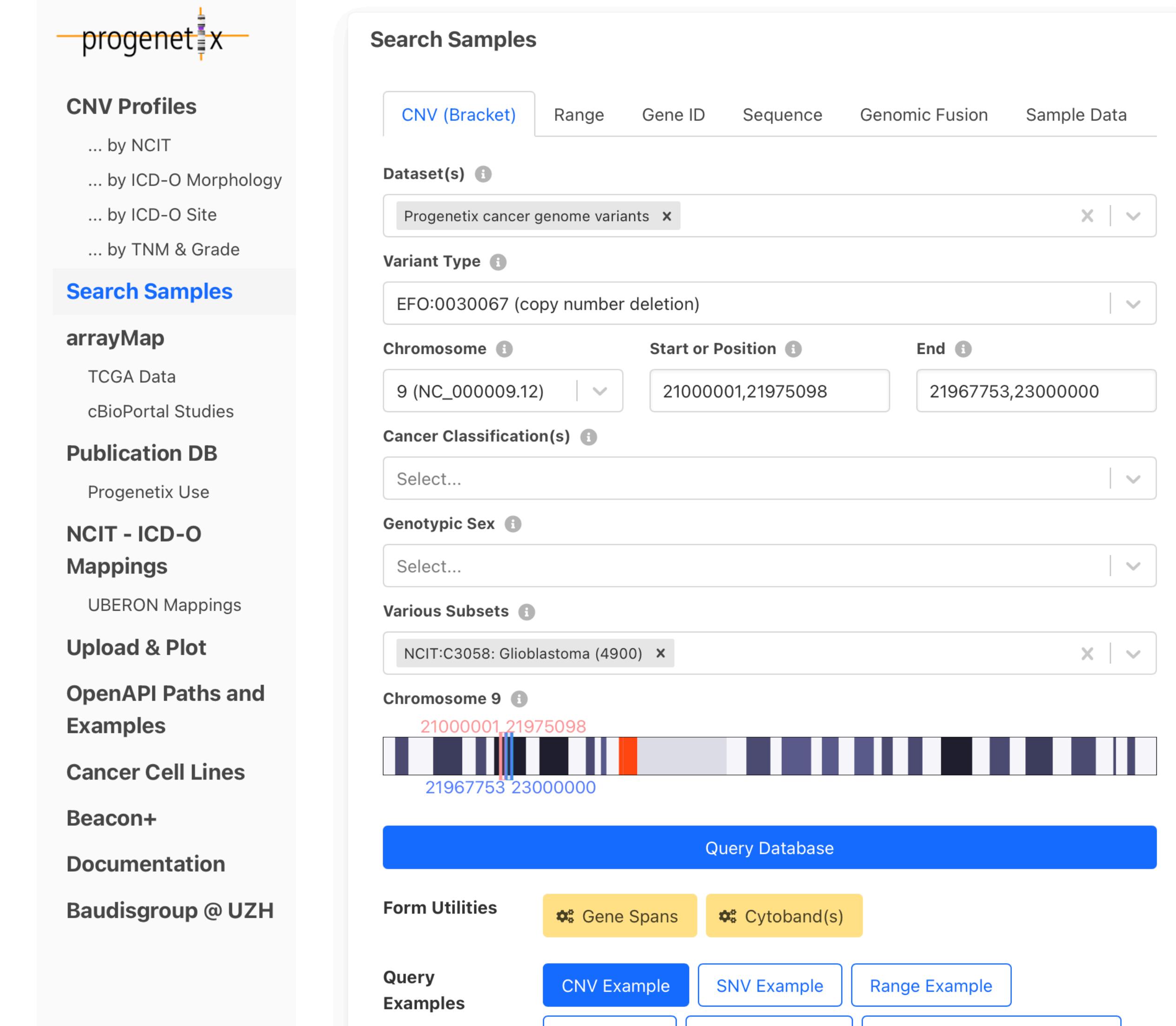
- **open** resource for oncogenomic profiles
- over **150'000 cancer CNV profiles**
- SNV data for some series (e.g. TCGA)
- more than **900 diagnostic types**
- inclusion of reference datasets (e.g. TCGA)
- standardized encodings (e.g. NCI<sup>t</sup>, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core clinical data (TNM, sex, survival ...)
- data mapping services



The screenshot shows the Progenetix interface with a sidebar on the left containing links like 'CNV Profiles', 'Search Samples', 'arrayMap', 'Publication DB', 'NCIT - ICD-O Mappings', 'Upload & Plot', 'OpenAPI Paths and Examples', 'Cancer Cell Lines', 'Beacon+', 'Documentation', and 'Baudisgroup @ UZH'. The main area is titled 'Cancer Types by National Cancer Institute NCI Code' and contains a detailed hierarchical tree of cancer types. A search bar at the top says 'Filter subsets e.g. by prefix ...' and a dropdown says 'Hierarchy Depth: 5 levels'. The tree starts with 'Neoplasm' (138552 samples, 144862 CNV profiles), which branches into 'Neoplasm by Site' (133029 samples, 139114 CNV profiles), then 'Genitourinary System Neoplasm', 'Benign Genitourinary System Neoplasm' (243 samples, 243 CNV profiles), 'Benign Urinary System Neoplasm' (98 samples, 98 CNV profiles), 'Benign Urinary Tract Neoplasm' (3 samples, 3 CNV profiles), 'Benign Kidney Neoplasm' (95 samples, 95 CNV profiles), 'Benign Reproductive System Neoplasm' (145 samples, 145 CNV profiles), 'Benign Female Reproductive System Neoplasm' (145 samples, 145 CNV profiles), 'Malignant Genitourinary System Neoplasms' (20567 samples, 22154 CNV profiles), 'Metastatic Malignant Genitourinary System Neoplasms' (2 samples, 2 CNV profiles), 'Metastatic Genitourinary System Carcinoma' (2 samples, 2 CNV profiles), 'Genitourinary System Carcinoma' (19462 samples, 20921 CNV profiles), 'Metastatic Genitourinary System Carcinoma' (2 samples, 2 CNV profiles), 'Female Reproductive System Carcinoma' (5746 samples, 5974 CNV profiles), 'Male Reproductive System Carcinoma' (7022 samples, 7808 CNV profiles), 'Urinary System Carcinoma' (6694 samples, 7139 CNV profiles), and 'Recurrent Malignant Genitourinary System Neoplasms' (3 samples, 3 CNV profiles).

## Cancer Genomics Reference Resource

- **open** resource for oncogenomic profiles
- over **150'000 cancer CNV profiles**
- SNV data for some series (e.g. TCGA)
- more than **900 diagnostic types**
- inclusion of reference datasets (e.g. TCGA)
- standardized encodings (e.g. NCIIt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core clinical data (TNM, sex, survival ...)
- data mapping services

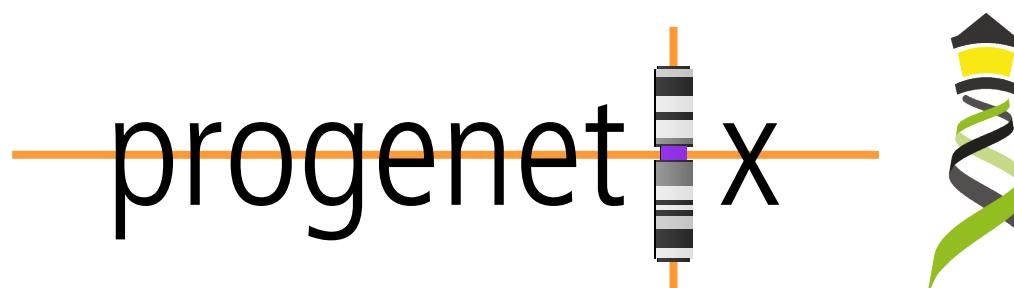


The screenshot shows the Progenetix search interface. At the top, there's a navigation bar with links like "Search Samples", "CNV Profiles", "arrayMap", "Publication DB", "NCIT - ICD-O Mappings", "Upload & Plot", "OpenAPI Paths and Examples", "Cancer Cell Lines", "Beacon+", "Documentation", and "Baudisgroup @ UZH". Below the navigation bar is a search form titled "Search Samples". It includes fields for "Dataset(s)" (set to "Progenetix cancer genome variants"), "Variant Type" (set to "EFO:0030067 (copy number deletion)"), "Chromosome" (set to "9 (NC\_000009.12)"), "Start or Position" (set to "21000001,21975098"), "End" (set to "21967753,23000000"), "Cancer Classification(s)", "Genotypic Sex", and "Various Subsets" (set to "NCIT:C3058: Glioblastoma (4900)"). A blue button at the bottom right says "Query Database". Below the search form is a "Form Utilities" section with buttons for "Gene Spans" and "Cytoband(s)". At the bottom, there's a "Query Examples" section with buttons for "CNV Example", "SNV Example", and "Range Example".

# Beacon v2 Filters

# **Example: Use of hierarchical classification systems (here NCI 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

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

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

progenetix

Variants: 0     $f_{alleles}$ : 0    Callsets Variants ↗    UCSC region ↗    Calls: 0    Legacy Interface ↗    Samples: 523    [Show JSON Response](#)

Results    [Biosamples](#)

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

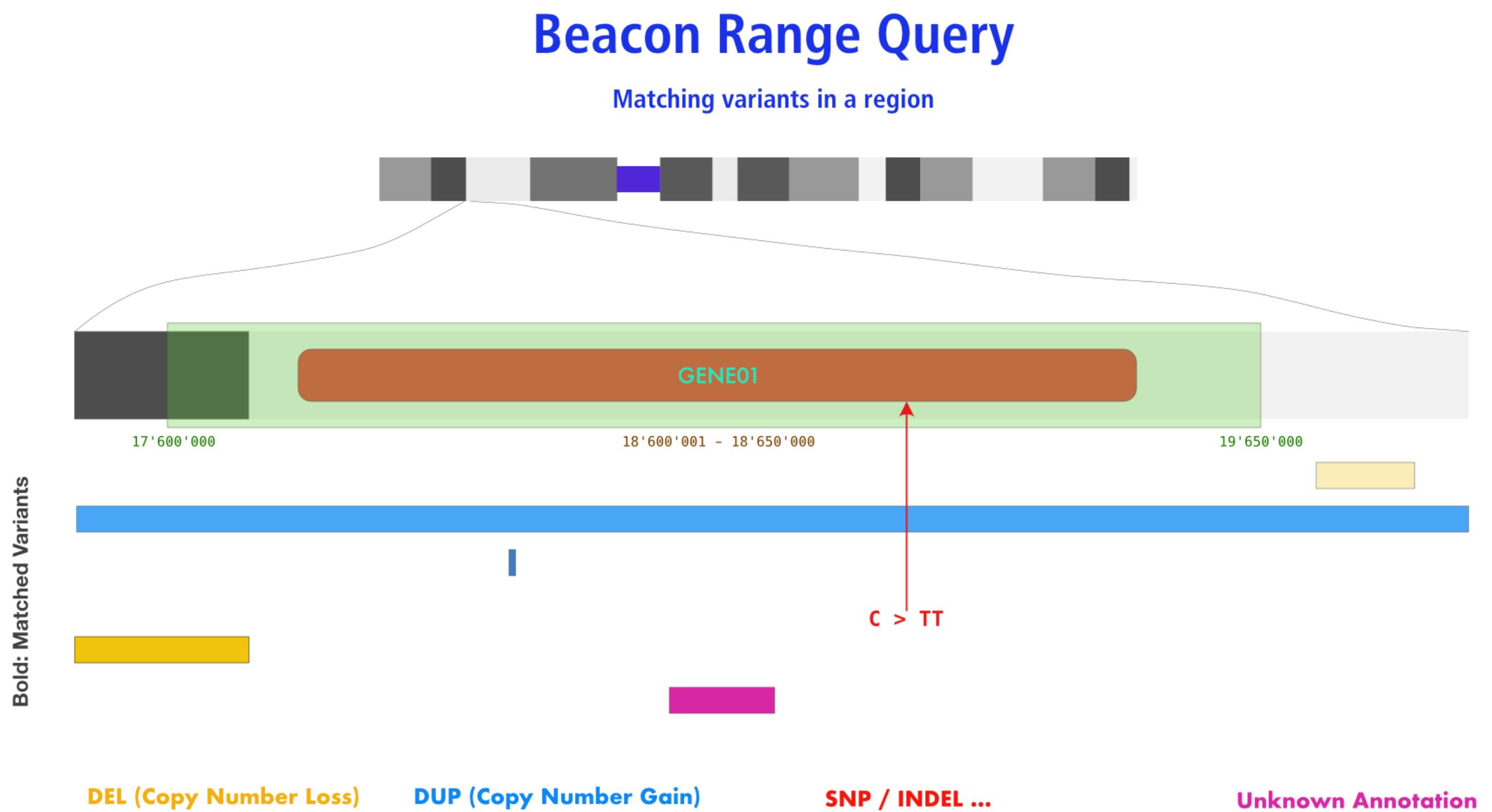
« < > »

Page 1 of 105

# Variation Queries

## Range ("anything goes") Request

- defined through the use of 1 start, 1 end
- any variant... but can be limited by type etc.



### Beacon Query Types

Sequence / Allele   CNV (Bracket)   **Genomic Range**   Aminoacid   Gene ID   HGVS   Sam

Dataset: Test Database - examplez

Chromosome: 17 (NC\_000017.11)

Variant Type: SO:0001059 (any sequence alteration - S...)

Start or Position: 7572826

End (Range or Structural Var.): 7579005

Reference Base(s): N

Alternate Base(s): A

Select Filters: Chromosome 17

Query Database

Form Utilities: Gene Spans, Cytoband(s)

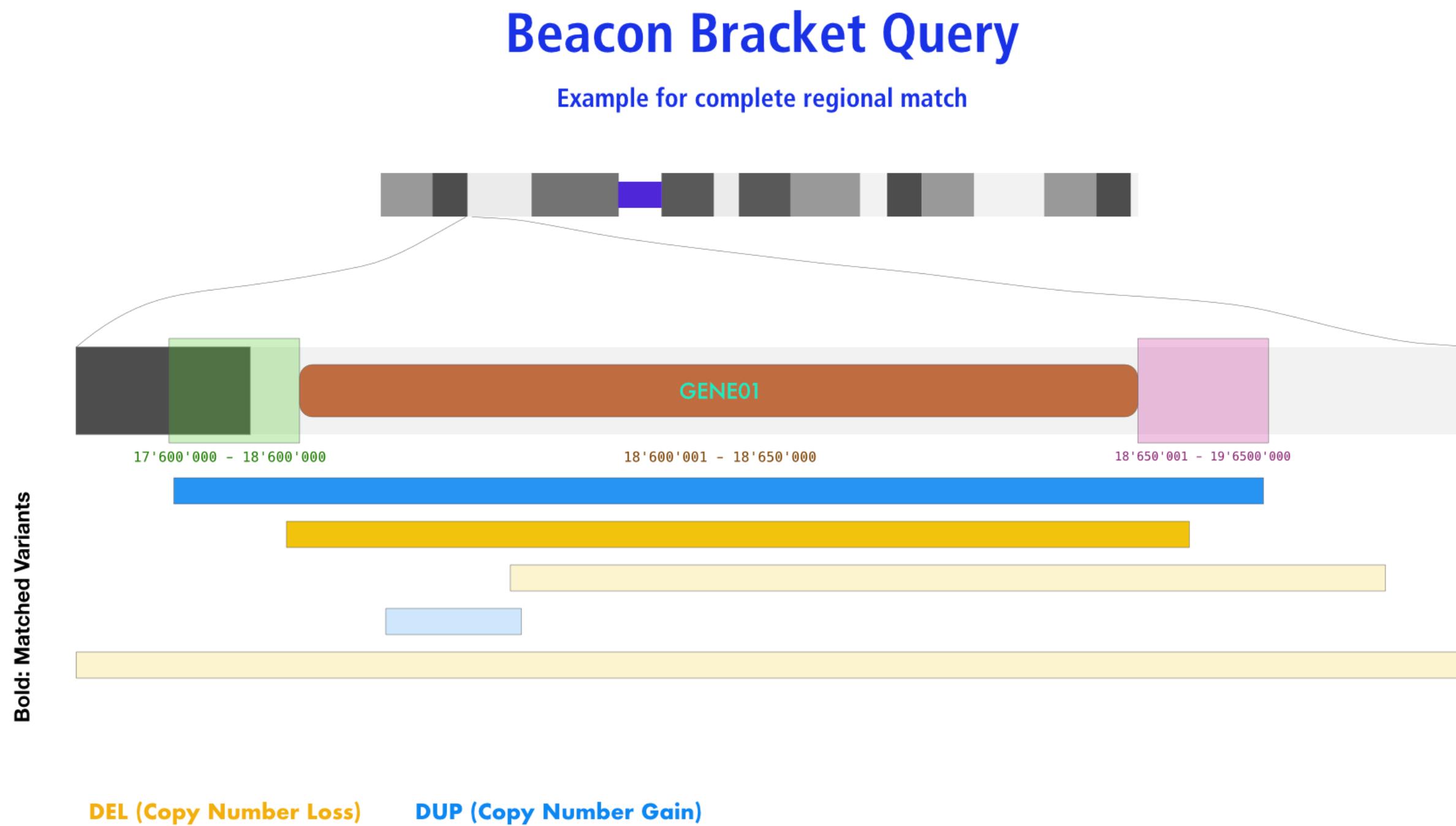
Query Examples: CNV Example, SNV Example, Range Example, Gene Match, Aminoacid Example, Identifier - HeLa

As in the standard SNV query, this example shows a Beacon query against mutations in the EIF4A1 gene in the DIPG childhood brain tumor dataset. However, this range + wildcard query will return any variant with alternate bases (indicated through "N"). Since parameters will be interpreted using an "AND" paradigm, either Alternate Bases OR Variant Type should be specified. The exact variants which were being found can be retrieved through the variant handover [H->O] link.

# Variation Queries

## Bracket ("CNV") Query

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



### Beacon Query Types

Sequence / Allele   **CNV (Bracket)**   Genomic Range   Aminoacid   Gene ID   HGVS   Sam

#### Dataset

Test Database - examplez X | ▼

#### Chromosome

9 (NC\_000009.12) | ▼

#### Variant Type

EFO:0030067 (copy number deletion) | ▼

#### Start or Position

21000001-21975098

#### End (Range or Structural Var.)

21967753-23000000

#### Select Filters

NCIT:C3058: Glioblastoma (100) X | ▼

#### Chromosome 9

21000001-21975098



### Query Database

#### Form Utilities

Gene Spans

Cytoband(s)

#### Query Examples

[CNV Example](#)

[SNV Example](#)

[Range Example](#)

[Gene Match](#)

[Aminoacid Example](#)

[Identifier - HeLa](#)

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.  $\leq \sim 2\text{Mbp}$  in size). The query is against the examplez collection and can be modified e.g. through changing the position parameters or data source.

# Website populated by asynchronous retrieval of Beacon query results using handovers

 [Edit Query](#)

**CNV Profiles**  
... by NCIT  
... by ICD-O Morphology  
... by ICD-O Site  
... by TNM & Grade

**Search Samples**

**arrayMap**  
TCGA Data  
cBioPortal Studies

**Publication DB**  
Progenetix Use

**NCIT - ICD-O Mappings**  
UBERON Mappings

**Upload & Plot**

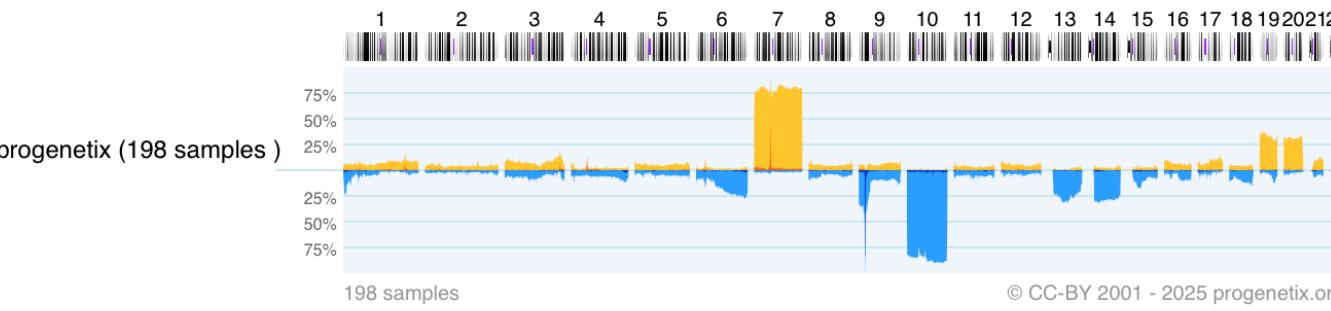
**OpenAPI Paths and Examples**

**Cancer Cell Lines**

**progenetix**

Matched Samples: 969      UCSC region   
Retrieved Samples: 200      Geographic Map   
Variants: 984      Variants in UCSC   
Calls: 976      Dataset Responses (JSON) 

[Results](#) [Biosamples](#) [Variants](#)



[Reload histogram in new window](#) 

Matched Subset Codes	Subset Samples	Matched Samples	Subset Match Frequencies
pgx:icdot-C71.4	4	1	0.250
pgx:icdot-C71.1	14	1	0.071
pgx:icdom-94403	4816	200	0.042
NCIT:C3058	4900	200	0.041
pgx:icdot-C71.9	13758	192	0.014
pgx:icdot-C71.0	1714	6	0.004

**progenetix Data Downloads**

**Download Sample Data (TSV)**  
Part1  Part2  Part3  Part4  Part5 

**Download Sample Data (JSON)**  
Part1  Part2  Part3  Part4  Part5 

**Download Variants (Beacon VRS)**  
Part1  Part2  Part3  Part4  Part5 

**Download Variants (VCF)**

Results			
Biosample	Dx Classifications	Identifiers	Variants
pgxbs-kftvl1hz	pgx:icdom-94403 Glioblastoma, NOS pgx:icdot-C71.9 Brain, NOS NCIT:C3058 Glioblastoma	pubmed:28481359 Zehir A, Benayed R et al. (2017): Mutational landscape of metastatic cancer revealed... cbiportal:msk_impact_2017	
pgxbs-kftvl7f4	pgx:icdom-94403 Glioblastoma, NOS pgx:icdot-C71.9 Brain, NOS NCIT:C3058 Glioblastoma	pubmed:28481359 Zehir A, Benayed R et al. (2017): Mutational landscape of metastatic cancer revealed... cbiportal:msk_impact_2017	
pgxbs-kftvhm6s	pgx:icdom-94403 Glioblastoma, NOS pgx:icdot-C71.9 Brain, NOS NCIT:C3058 Glioblastoma	pgx:TCGA-GBM Glioblastoma Multiforme 18772890 Cancer Genome Atlas Research Network. (2008): Comprehensive genomic characterization defines human glioblastoma...	
Biosamples			
Digest	Gene	Pathogenicity	Variant type
9:21626201- 21981584:EFO_0030068			CopyNumberChange V: pgxvar- 665749ab2d6be9a260e55de8 A: pgxcs-kftwnmzs B: pgxbs-kftvjywz I: pgxind-kftx5yjj
9:21846286- 22201587:EFO_0030068			CopyNumberChange V: pgxvar- 6656fc5fbe3f6845a3555b82 A: pgxcs-kftw53z6 B: pgxbs-kftvi872 I: pgxind-kftx3t9l
9:21949762- 22004847:EFO_0020073			CopyNumberChange V: pgxvar- 6657226e8f6b96158261aa6 A: pgxcs-kftw3vh5 B: pgxbs-kftvi4e1 I: pgxind-kftx3vx4
9:21164528- 21990552:EFO_0020073			CopyNumberChange V: pgxvar- 66572dfe2d6be9a260e3d189 A: pgxcs-kftw95rl B: pgxbs-kftvilhz I: pgxind-kftx4au8

# Standards Development & Implementation: CNV Terms

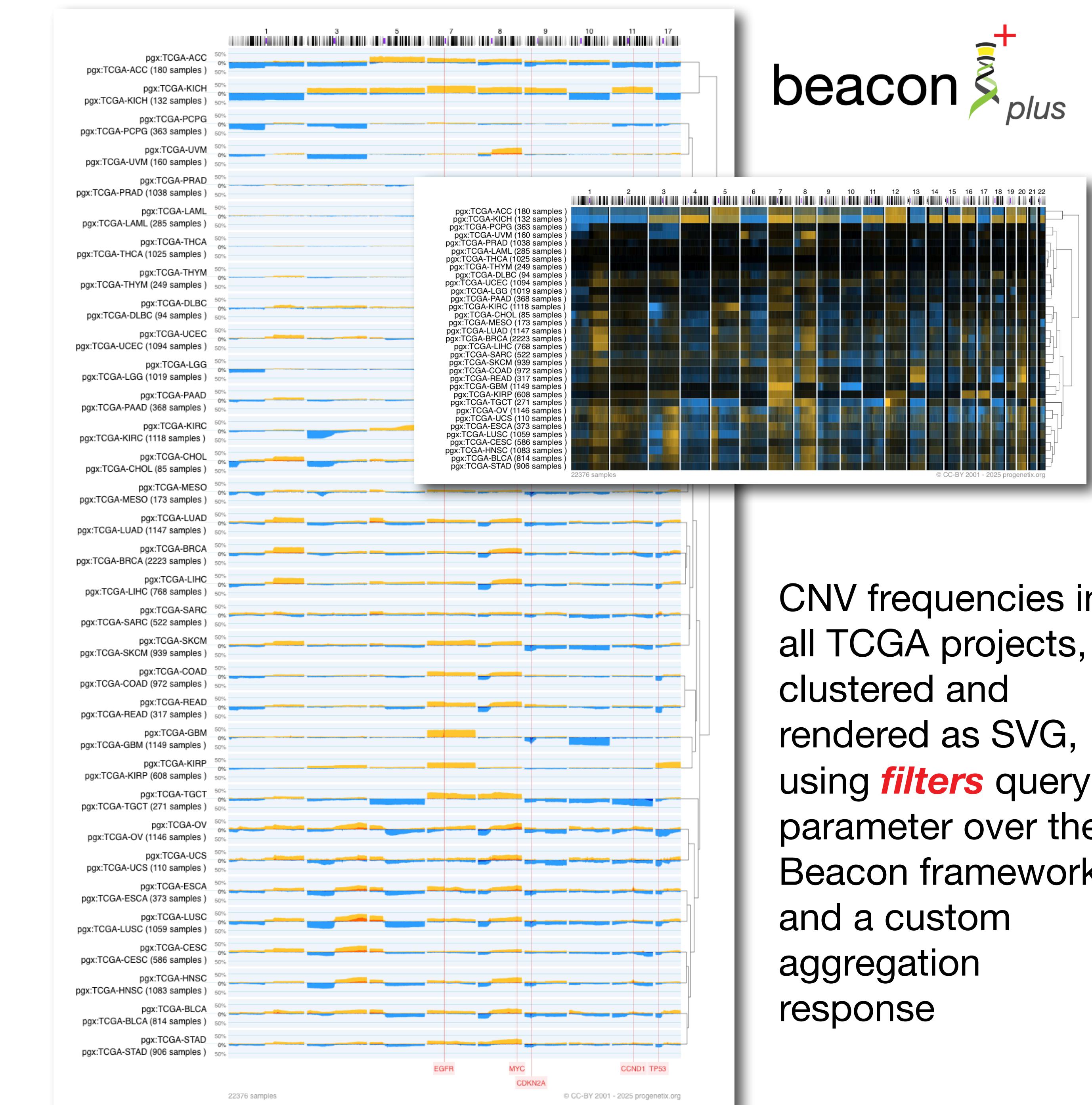
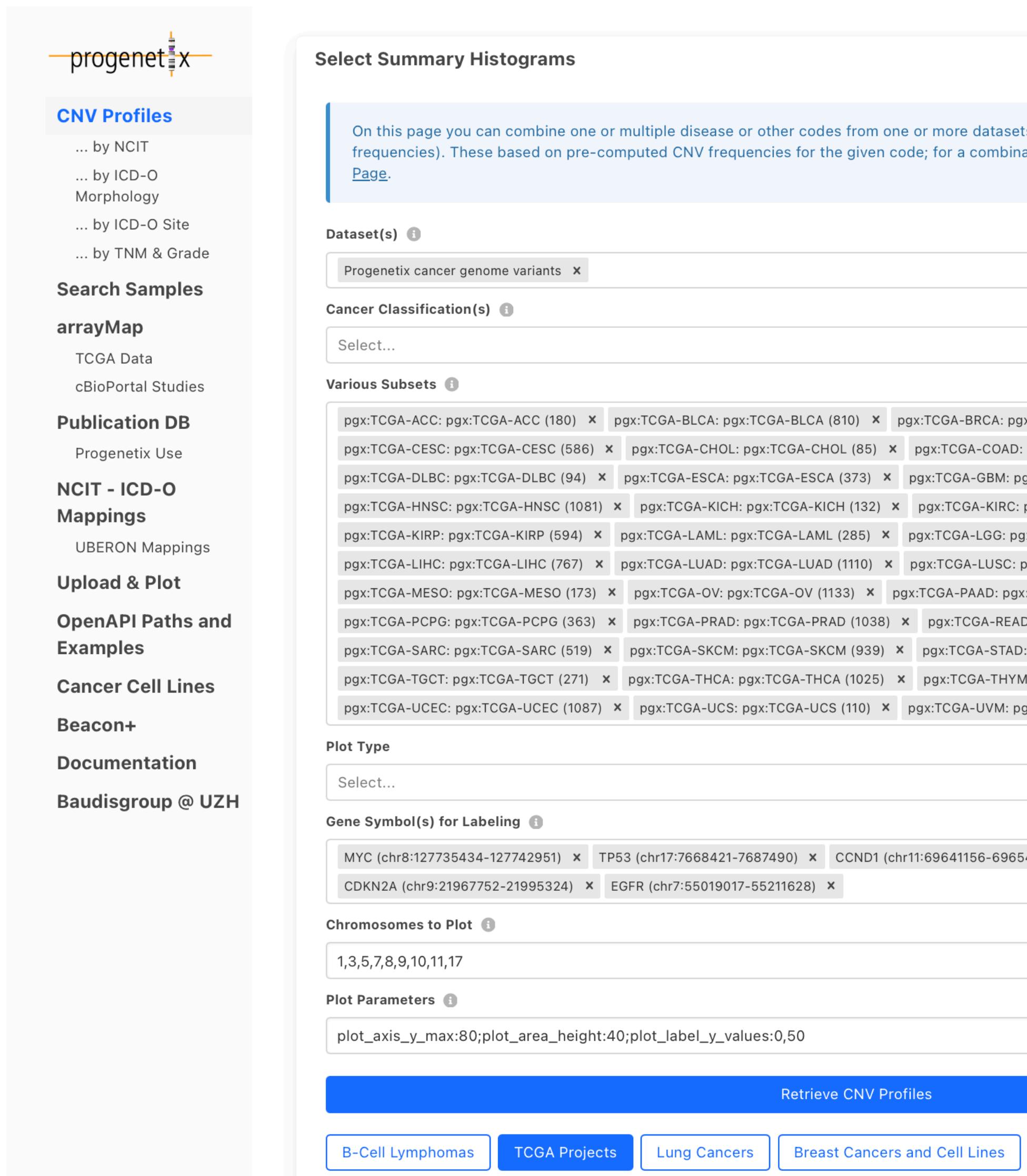
## in computational (file/schema) formats

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

<b>GA4GH VRS1.3+</b>	<b>Beacon v2</b>	<b>VCF v4.4</b>	<b>SO</b>
<b>EFO:0030070</b> gain	DUP or <b>EFO:0030070</b>	DUP SVCLAIM=D	SO:0001742 copy_number_gain
<b>EFO:0030071</b> low-level gain	DUP or <b>EFO:0030071</b>	DUP SVCLAIM=D	SO:0001742 copy_number_gain
<b>EFO:0030072</b> high-level gain	DUP or <b>EFO:0030072</b>	DUP SVCLAIM=D	SO:0001742 copy_number_gain
<b>EFO:0030072</b> high-level gain	DUP or EFO:0030073	DUP SVCLAIM=D	SO:0001742 copy_number_gain
<b>EFO:0030067</b> loss	DEL or <b>EFO:0030067</b>	DEL SVCLAIM=D	SO:0001743 copy_number_loss
<b>EFO:0030068</b> low-level loss	DEL or <b>EFO:0030068</b>	DEL SVCLAIM=D	SO:0001743 copy_number_loss
<b>EFO:0020073</b> high-level loss	DEL or <b>EFO:0020073</b>	DEL SVCLAIM=D	SO:0001743 copy_number_loss
<b>EFO:0030069</b> complete genomic loss	DEL or <b>EFO:0030069</b>	DEL SVCLAIM=D	SO:0001743 copy_number_loss

# Pushing the envelope...

# Custom Beacon aggregation response for displaying CNV frequencies



CNV frequencies in all TCGA projects, clustered and rendered as SVG, using ***filters*** query parameter over the Beacon framework and a custom aggregation response

# Beacon<sup>+</sup>: Phenopackets

## Testing alternative response schemas...

<https://progenetix.org/beacon/phenopackets/pgxind-kftx26j0>

- the v2 default schemas are mostly aligned w/ Phenopackets v2
- creating phenopackets can be done mostly by re-wrapping of Beacon entities (individual, biosample)
- variants can be included through file resource URLs; in Beacon<sup>+</sup> this is done through *ad hoc* handover URIs

```

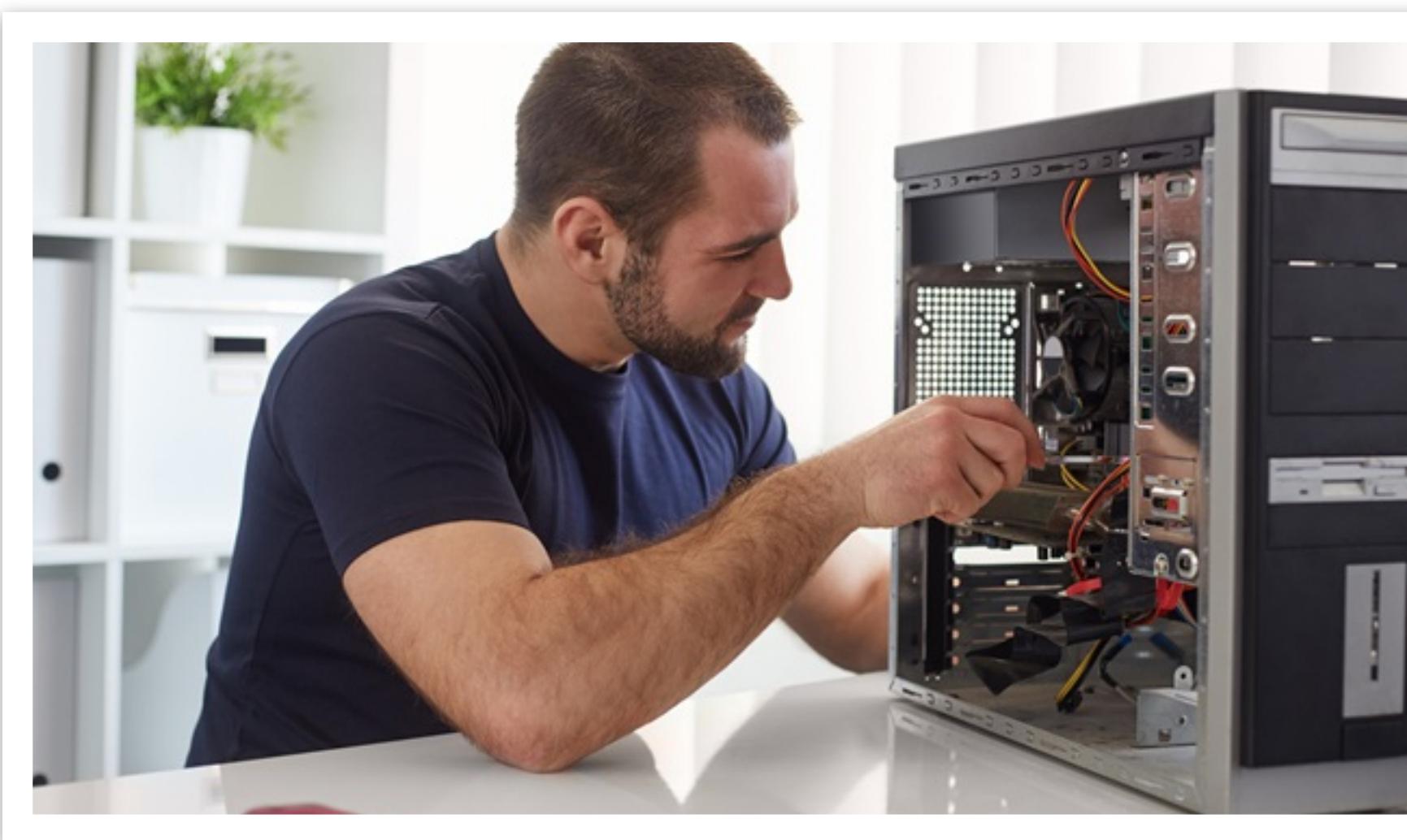
    "id": "pgxpxf-kftx3tl5",
    "metaData": {
      "phenopacketSchemaVersion": "v2",
      "resources": [
        {
          "id": "NCIT",
          "iriPrefix": "http://purl.obolibrary.org/obo/NCIT_",
          "name": "NCIt Plus Neoplasm Core",
          "namespacePrefix": "NCIT",
          "url": "http://purl.obolibrary.org/obo/ncit/neoplasm-core.c",
          "version": "2022-04-01"
        },
        ...
      ],
      "subject": {
        "dataUseConditions": {
          "id": "DUO:0000004",
          "label": "no restriction"
        },
        "diseases": [
          {
            "clinicalTnmFinding": [],
            "diseaseCode": {
              "id": "NCIT:C3099",
              "label": "Hepatocellular Carcinoma"
            },
            "onset": {
              "age": "P48Y9M26D"
            },
            "stage": {
              "id": "NCIT:C27966",
              "label": "Stage I"
            }
          }
        ],
        "id": "pgxind-kftx3tl5",
        "sex": {
          "id": "PATO:0020001",
          "label": "male genotypic sex"
        },
        "updated": "2018-12-04 14:53:11.674000",
        "vitalStatus": {
          "status": "UNKNOWN_STATUS"
        }
      }
    },
    "biosamples": [
      {
        "biosampleStatus": {
          "id": "EFO:0009656",
          "label": "neoplastic sample"
        },
        "dataUseConditions": {
          "id": "DUO:0000004",
          "label": "no restriction"
        },
        "description": "Primary Tumor",
        "externalReferences": [
          {
            "id": "pgx:TCGA-0004d251-3f70-4395-b175-c94c2f5b1b81",
            "label": "TCGA case_id"
          },
          {
            "id": "pgx:TCGA-TCGA-DD-AAVP",
            "label": "TCGA submitter_id"
          },
          {
            "id": "pgx:TCGA-9259e9ee-7279-4b62-8512-509cb705029c",
            "label": "TCGA sample_id"
          },
          {
            "id": "pgx:TCGA-LIHC",
            "label": "TCGA LIHC project"
          }
        ],
        "files": [
          {
            "fileAttributes": {
              "fileFormat": "pgxseg",
              "genomeAssembly": "GRCh38"
            },
            "uri": "https://progenetix.org/beacon/biosamples/pgxbss-kftvhvzb/variants/?output=pgxseg"
          }
        ],
        "histologicalDiagnosis": {
          "id": "NCIT:C3099",
          "label": "Hepatocellular Carcinoma"
        },
        "id": "pgxbss-kftvhvzb",
        "individualId": "pgxind-kftx3tl5",
        "pathologicalStage": {
          "id": "NCIT:C27966",
          "label": "Stage I"
        },
        "sampledTissue": {
          "id": "UBERON:0002107",
          "label": "liver"
        },
        "timeOfCollection": {
          "age": "P48Y9M26D"
        }
      }
    ]
  }
}

```



# Beacon v2 deployment

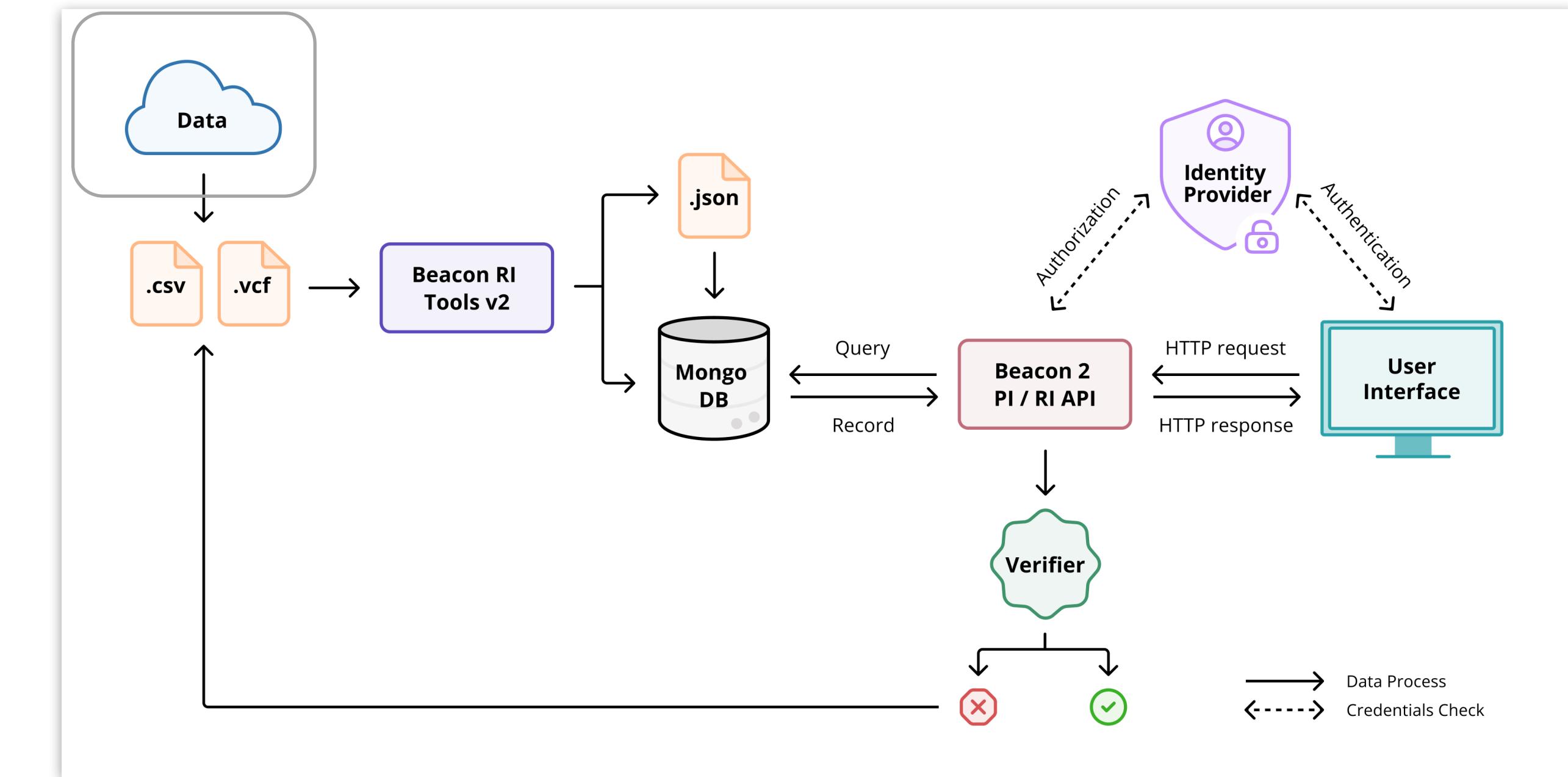
## Build it yourself



### Beacon v2 API

<https://github.com/ga4gh-beacon/beacon-v2>

## Toolkit for production environments



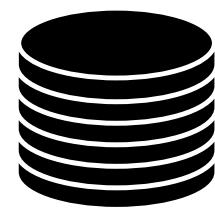
### Beacon v2 Production Implementation (released Oct 2024)

<https://github.com/ga4gh-beacon/beacon-v2>

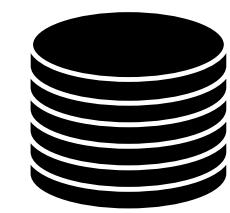
# *bycon* based Beacon+ Stack

progenetix

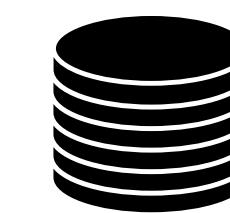
- *collations* contain pre-computed data (e.g. CNV frequencies, statistics) and information for all grouping entity instances and correspond to **filter values**
  - ▶ [pubmed:10027410](#), [NCIT:C3222](#), [pgx:cohort-TCGA](#), [pgx:icdom-94703...](#)
  - ▶ precomputed frequencies per collection informative e.g. in form autfills
- *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



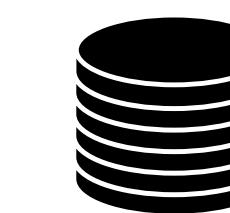
variants



analyses



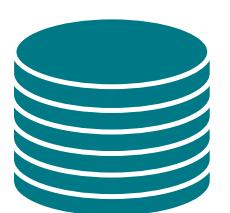
biosamples



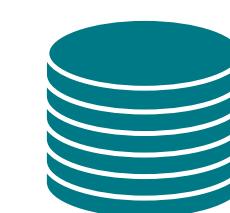
individuals



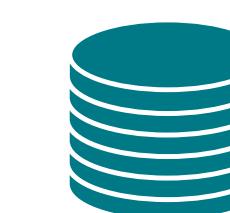
collations



geolocs



genespans



qBuffer

Entity collections

Utility collections

[github.com/progenetix/bycon](https://github.com/progenetix/bycon)





# front-end showcases query strategies

**Beacon Search Demonstrator**

This search form shows parameter combinations and examples for different Beacon search patterns. Please be aware that search types and examples are *independent* of each other, so not all combinations are automatically adjusted.

Additionally, the search options here might extend the latest stable version of the Beacon API in a sense of "implementation driven development" but are supported through this version of the [bycon](#) library.

**Search Samples**

**Compare CNV Profiles**

**CNV Profiles by Cancer Type**

NCIT Neoplasia Codes  
ICD-O Morphologies  
ICD-O Organ Sites  
TNM & Grade

**OpenAPI Paths and Examples**

**Documentation**

**Progenetix**

**Baudisgroup @ UZH**

**Search Samples**

**CNV (Bracket)** Range Gene ID Sequence Genomic Fusion Sample Data

**Dataset(s)** Test Database - examplez

**Variant Type** EFO:0030067 (copy number deletion)

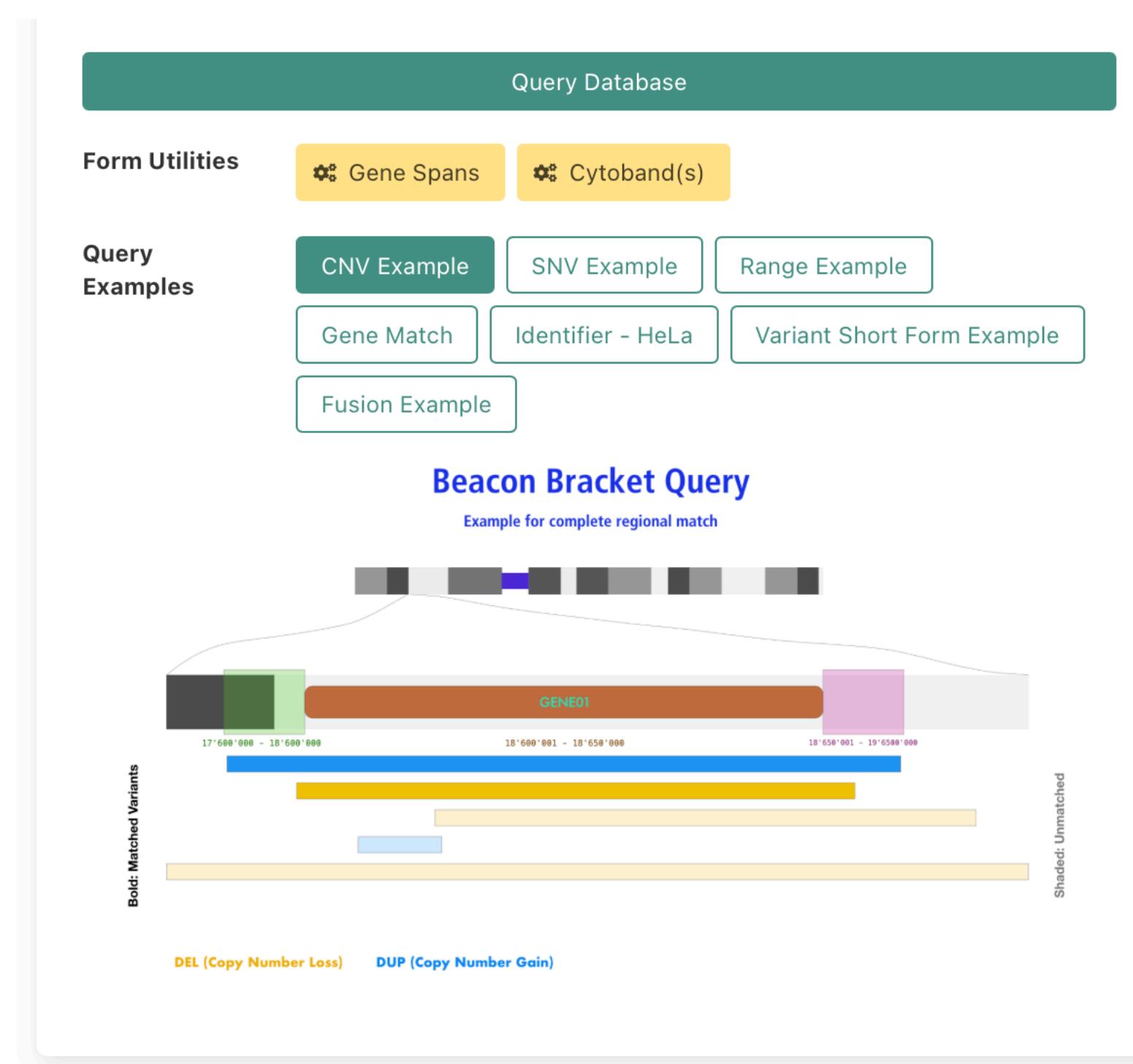
**Chromosome** 9 (NC\_000009.12) **Start or Position** 21000001,21975098 **End (Range or Structural Var.)** 21967753,23000000

**Cancer Classification(s)** Select...

**Genotypic Sex** Select...

**Various Subsets** NCIT:C3058: Glioblastoma (28)

Chromosome 9 21000001,21975098 21967753,23000000



- React based website included in *bycon*
- exposes Beacon query options
- package provided examples can be extended by local ones

# Progenetix Cancer Genomics Beacon+

/api

Beacon+ provides a forward looking implementation of the Beacon v2 API, with focus on structural genome variants and metadata based on the cancer and reference genome profiling data in the Progenetix resource (<https://progenetix.org>).

This page presents a prototype for an OpenAPI (Swagger) definition for the [GA4GH Beacon API](#). The definitions are generated from the `entity_defaults` and `argument_definitions` in the [bycon project](#). The complete. Please be aware that the whole capabilities of the project cannot be represented solely through the OpenAPI definitions and also involve features such as filtering terms logic and result aggregation entities. Additionally, the bycon project implements a number of data services beyond Beacon standards which again are only partially covered here.

## bycon and Data Aggregation

The Beacon standard implements a REST style syntax - e.g. consistent id-based document retrieval for entities indicated through their framework provide full data aggregation; i.e. queries with parameters against **any** of the main data entities (g\_variants, runs, analyses, intersection of the query results at the level of the response entity).

[Beacon v2 API] | [Example: Bv2 CNV / bracket] | [Proposal: VQS CNV / bracket] | [Proposal: VQS CNV by Gene ID] | [Proposal: VQS CNV by Reference Name]

Contact the developer

Servers

<https://progenetix.org>

## Beacon

[GET /beacon/info](#) Get info entries

[GET /beacon/datasets](#) Get dataset entries

[GET /beacon/cohorts](#) Get cohort entries

[GET /beacon/g\\_variants](#) Get genomicVariant entries

[GET /beacon/g\\_variants/{id}](#) Get genomicVariant entries

[GET /beacon/g\\_variants/{id}/analyses](#) Get analysis entries

[GET /beacon/g\\_variants/{id}/biosamples](#) Get biosample entries

[GET /beacon/g\\_variants/{id}/individuals](#) Get individual entries

[GET /beacon/analyses](#) Get analysis entries

[GET /beacon/analyses/{id}](#) Get analysis entries

[GET /beacon/analyses/{id}/g\\_variants](#) Get genomicVariant entries

[GET /beacon/analyses/{id}/biosamples](#) Get biosample entries

## Bv2minimalAlleleRequest

[GET /beacon/g\\_variants](#) Get genomicVariant entries

### Parameters

Name	Description
filters	array<string> (query)
referenceName	string (query)
start	array<integer> (query)
end	array<integer> (query)
variantType	string (query)
requestedGranularity	string (query)

## Bv2cnvbracketquery

[GET /beacon/g\\_variants](#) Get genomicVariant entries

Get genomicVariant entries

### Parameters

Name	Description
filters	array<string> (query)
referenceName	string (query)
start	array<integer> (query)
end	array<integer> (query)
variantType	string (query)
requestedGranularity	string (query)

## VQSadjacencyRequest

[GET /beacon/g\\_variants](#) Get genomicVariant entries

[GET /beacon/analyses](#) Get analysis entries

[GET /beacon/biosamples](#) Get biosample entries

Get biosample entries

### Parameters

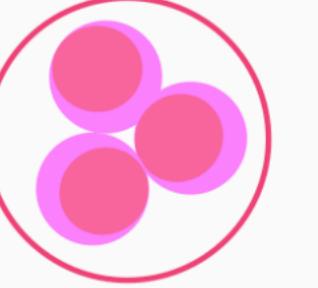
Name	Description
filters	array<string> (query)
referenceAccession	string (query)
breakpointRange	array<integer> (query)
adjacencyAccession	string (query)
adjacencyRange	array<integer> (query)
vrsType	string (query)

# Cancer Cell Lines

&&

# refCNV

cancercelllines.org  
refcnv.org



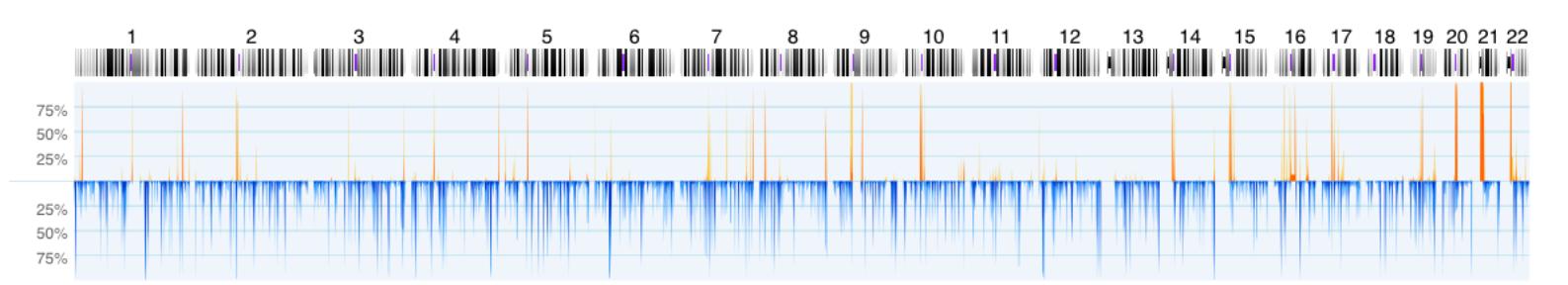
**cancercelllines**

Genomic Copy Number Variation (CNV) data from reference samples

**Under Construction**

This site is currently under construction, with contributions by groups from the University of Zurich and Erasmus MC. Neither data content nor representation have been finalized. PLEASE DO NOT USE FOR ANY RESEARCH OR REFERENCE PURPOSES!

Frequency profiles of regional genomic gains and losses for all categories (diagnostic entity, publication, cohort ...) can be accessed through the respective Cancer Types pages with visualization and sample retrieval options. Below is a typical example of the aggregated CNV data in 3201 samples of the 1000 Genomes Dragen CNV analysis set. The frequency of regional **copy number gains (high level)** and **losses (high level)** displayed for the 22 autosomes as occurrence of any of these CNVs in the 1Mb binned intervals.



Download SVG | Go to DRAGEN-CNV | Download CNV Frequencies

The repository contains CNV tracks for many of the 1000 Genomes samples, analyzed by different platforms or data pipelines and therefore allows to compare private analysis data to results from these different call sets, to avoid interpretation biases from using reference data with a different analysis profile from the one used in your study. The plot below shows analysis specific CNV tracks for chromosome 13 in the HG01572 sample from the 1000 Genomes set, for several calling pipelines.



Please be aware that the small size of most CNVs is not correctly represented at this zoom level (overplotting due to limited resolution).

Assembly: GRCh38 Chro: NC\_000007.14 Start: 140713328 End: 140924929

Type: SNV

cellz

Matched Samples: 1058  
Retrieved Samples: 1000  
Variants: 127  
Calls: 1444

UCSC region ↗  
Variants in UCSC ↗  
Dataset Responses (JSON) ↗

Visualization options

Results Biosamples Variants Annotated Variants

Digest	Gene	Pathogenicity	Variant type	Variant Instances
7:140834768-140834769:G>A	BRAF		Missense variant	V: pgxvar-63ce6abca24c83054b B: pgxbs-3fB2a14B
7:140734714-140734715:G>A	BRAF		Missense variant	V: pgxvar-63ce6acda24c83054b B: pgxbs-3fB2a14B
7:140753334-140753339:T>TGTA	BRAF	Pathogenic		V: pgxvar-

## Hierarchies

### Cell Line Details

#### HOS (cellosaurus:CVCL\_0312)

##### Subset Type

- Cellosaurus - a knowledge resource on cell lines cellosaurus:CVCL\_0312 ↗

##### Sample Counts

204 samples

57 direct cellosaurus:CVCL\_0312 code matches  
21 CNV analyses

##### Search Samples

Select cellosaurus:CVCL\_0312 samples in the Search Form

##### Raw Data (click to show/hide)

28 samples

562/Ad...

umi-1 (9)

97 (2 samples)

(11 samples)

:U-1 (1 sample)

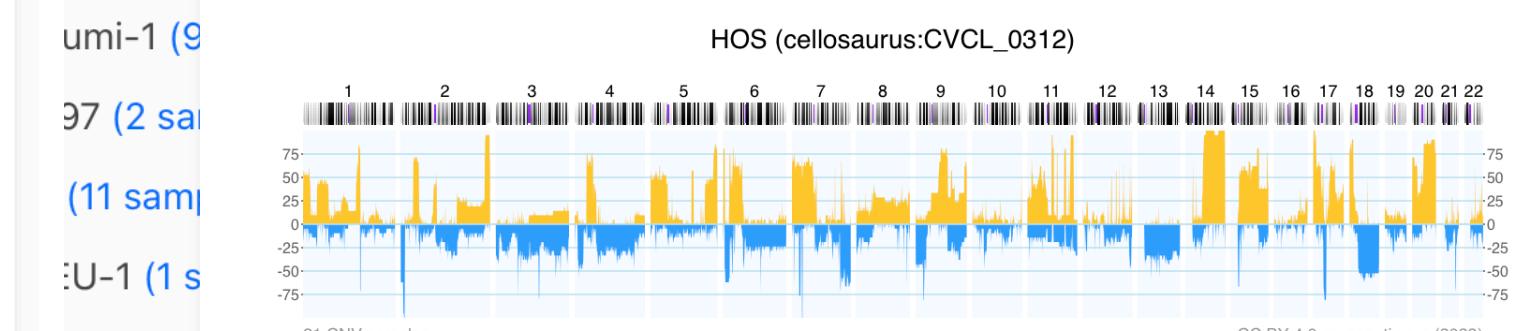
M-3 (1 sample)

JOI-90 (1 sample)

Reh/Eph (1 sample)

NSU-CL (1 sample)

C827 (2 samples)



Download SVG | Go to cellosaurus:CVCL\_0312 | Download CNV Frequencies

Gene Matches	Cytoband Matches	Variants
ALK	. ABC-14 cells harbored no ALK mutations and were sensitive to ... crizotinib while also exhibiting MNNG HOS transforming gene ( MET )	Rapid Acquisition of Alectinib Resistance in ALK-Positive Lung Cancer With High Tumor Mutation Burden (31374369) ABSTRACT
AREG	crizotinib while also exhibiting MNNG HOS	Rapid Acquisition of Alectinib Resistance ABSTRACT



## bycon Documentation

Documentation Home

Recent Changes

Setup &amp; Maintenance

Installation

Importing Data

Housekeeping

Beacon API

Beacon API

Services API

API Parameters

Front End

Code Repositories

bycon

Progenetix Front End

More Info

Progenetix Site

baudisgroup@UZH

Beacon Documentation

## Changes &amp; To Do

## Changes Tracker

While changes are documented for individual point versions we actually do not push releases out for all of them; they serve more as internal development milestones.

## 2025-05-15: (v2.4.3 "Bologna")

- expanded `NCITsex` ontology to have hierarchical terms with the current NCIT terms at the tip of the branches
  - e.g. `pgx:sex => pgx:sex-female => PATO:0020001 => NCIT:C16576`
  - allows for query expansion & use of alternate terms (e.g. PATO)
  - not strictly correct since the NCIT terms are for "any description of biological sex or gender", whereas PATO is for genotypic sex; so might be flipped later w/ annotations in the database switched accordingly (this was the original state but Beacon docs used NCIT ...)
- changed `byconautServiceResponse` to `byconServiceResponse`
- added a new subset / cancer type histogram multi-selection to the `beaconplusWeb` front-end (at [beaconplus.progenetix.org/subsetsSearch/](https://beaconplus.progenetix.org/subsetsSearch/))

## Table of contents

Changes Tracker

2025-05-15: (v2.4.3 "Bologna")

2025-05-02 (v2.4.2)

2025-04-25 (v2.4.1)

2025-04-25 (v2.4.0)  
"Cotswolds")

2025-04-15 (v2.3.1)

2025-04-04 (v2.3.0 "Logan  
Airport")

2025-03-10 (v2.2.6)

2025-03-06 (v2.2.5)

2025-03-03 (v2.2.4)

2025-02-26 (v2.2.3)

2025-02-21 (v2.2.2)

2025-02-21 (v2.2.1)

2025-02-14 (v2.2.0)

2025-02-08 (v2.1.5)

2025-01-29 (v2.1.4)

2025-01-16 (v2.1.3)

2024-12-20 (v2.1.2)

2024-12-19 (v2.1.1)

2024-12-09 (v2.1.0)

# pgxRpi: an R/Bioconductor package

## Client for Accessing Beaconized Data

- **Query and export variants**

[https://progenetix.org/beacon/biosamples/pgxbs-kftvh94d/g\\_variants](https://progenetix.org/beacon/biosamples/pgxbs-kftvh94d/g_variants)

```
> variants <- pgxLoader(type="variant",biosample_id="pgxbs-kftvh94d")
```

- **Query metadata of biosamples and individuals by filters (e.g. NCIt, PMID)**

<http://progenetix.org/services/sampletable/?filters=NCIT:C3697>

```
> biosamples <- pgxLoader(type="biosample",filters="NCIT:C3697")
```

- **Query and visualize CNV frequency by filters**

<http://www.progenetix.org/services/intervalFrequencies/?filters=NCIT:C3512>

```
> freq <- pgxLoader(type="frequency",output="pgxfreq",filter  
> pgxFreqplot(freq)
```

- **Process local .pgxseg files**

```
> info <- pgxSegprocess(file=file, show_KM_plot = T,  
return_seg = T, return_metadata = T, return_frequency = T)
```

## pgxRpi

This is the **development** version of pgxRpi; for the stable release version, see [pgxRpi](#).

### R wrapper for Progenetix

platforms all rank 2178 / 2266 support 0 / 0 in Bioc < 6 months build unknown updated < 1 month dependencies 137

DOI: [10.18129/B9.bioc.pgxRpi](https://doi.org/10.18129/B9.bioc.pgxRpi)

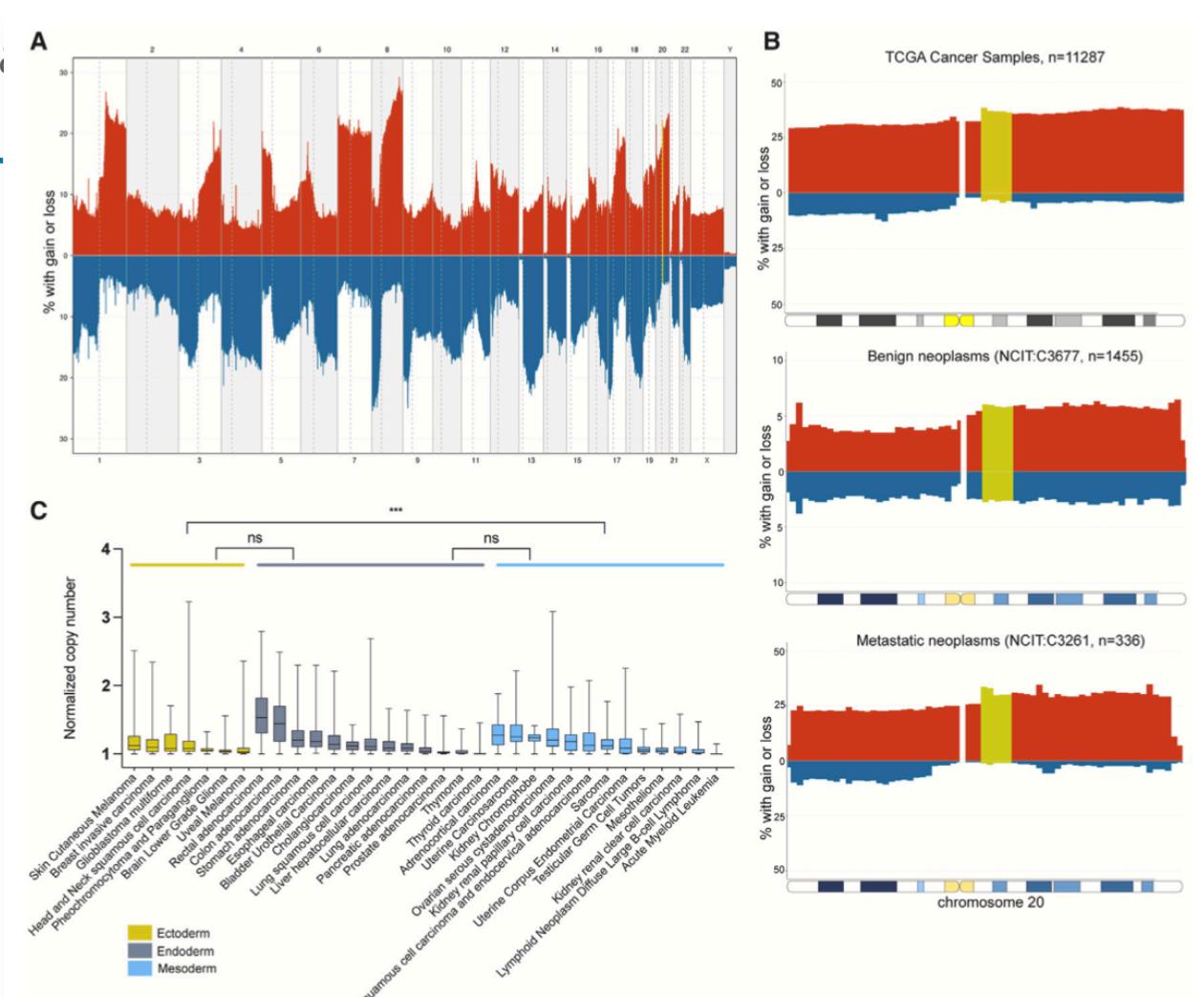
**Bioconductor version:** Development (3.20)

The package is an R wrapper for Progenetix REST API built upon the Beacon v2 protocol. Its purpose is to provide a seamless way for retrieving genomic data from Progenetix database—an open resource dedicated to curated oncogenomic profiles. Empowered by this package, users can effortlessly access and visualize data from Progenetix.

**Author:** Hangjia Zhao [aut, cre]  Michael Baudis [aut] 

**Maintainer:** Hangjia Zhao <hangjia.zhao at uzh.ch>

Use case: 2024 article using Progenetix' *pgxRpi* to retrieve & visualize 117'587 cancer CNV profiles for a study into pluripotent stem cells' genomics



## Stem Cell Reports Review



OPEN ACCESS

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

Nuša Krivec,<sup>1,2</sup> Manjusha S. Ghosh,<sup>1,2</sup> and Claudia Spits<sup>1,2,\*</sup>

<sup>1</sup>Research Group Reproduction and Genetics, Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel, Brussels, Laarbeeklaan 103, 1090 Brussels, Belgium

<sup>2</sup>These authors contributed equally

\*Correspondence: [claudia.spits@vub.be](mailto: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.

# Looking for implementers and contributors

- containerization
- data I/O ...
- standard library integration  
(VRSification of variants...)

The screenshot shows the GitHub repository page for 'bycon'. The repository is public and has 4 branches and 25 tags. The main branch is 'main'. The commit history is listed, showing contributions from 'mbaudis' for version 1.3.6. The commits include creating mk-bycon-docs.yaml, updating .gitignore, creating LICENSE, and major library & install disentanglement. Other commits mention README.md, install.py, install.yaml, mkdocs.yaml, requirements.txt, setup.cfg, setup.py, and updev.sh. The commits are dated from 3 days ago to 9 months ago.

File / Commit	Description	Date
.github/workflows	Create mk-bycon-docs.yaml	8 months ago
bycon	1.3.6	3 days ago
docs	1.3.6	3 days ago
local	1.3.5 preparation	2 weeks ago
.gitignore	Update .gitignore	3 months ago
LICENSE	Create LICENSE	3 years ago
MANIFEST.in	major library & install disentanglement	9 months ago
README.md	#### 2023-07-23 (v1.0.68)	4 months ago
install.py	1.3.6	3 days ago
install.yaml	v1.0.57	5 months ago
mkdocs.yaml	1.1.6	3 months ago
requirements.txt	1.3.6	3 days ago
setup.cfg	...	10 months ago
setup.py	1.3.6	3 days ago
updev.sh	1.3.6	3 days ago

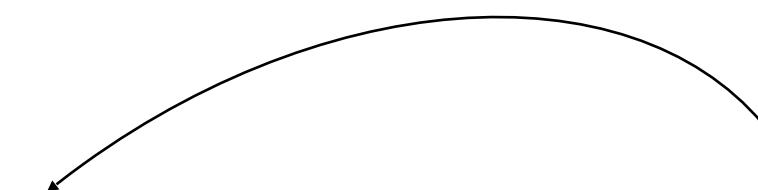
[bycon.progenetix.org](https://bycon.progenetix.org)  
[github.com/progenetix/bycon/](https://github.com/progenetix/bycon/)



## Beacon as a global standard



### 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

# 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**)

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

GA4GH Beacon Genomic Variation Query Standards

Search GitHub elixir

## Beacon VQS Requests

The `VQSRequest` type represents the generic collection of variant parameters supported in Beacon v2+ requests. These include parameters with close alignment to VRS v2 concepts and replacing some Beacon v1/v2 generics with tighter definitions (e.g. `referenceAccession` instead of `referenceName` and `accession` or `copyChange` for a specific subset of former `variantType` values) but also keep some concepts beyond VRS scope or specifically geared towards query applications (`geneId`, `sequenceLength`)

For the parameter definitions please see the [requestParameterComponents page](#).

### VQSRequest Parameters

```
requestProfileId: ./requestParameterComponents.yaml#/defs/RequestProfileId
referenceAccession: ./requestParameterComponents.yaml#/defs/RefgetAccession
start: ./requestParameterComponents.yaml#/defs/SequenceStart
end: ./requestParameterComponents.yaml#/defs/SequenceEnd
sequence: ./requestParameterComponents.yaml#/defs/Sequence
copyChange: ./requestParameterComponents.yaml#/defs/CopyChange
adjacencyAccession: ./requestParameterComponents.yaml#/defs/AdjacencyAccession
adjacencyStart: ./requestParameterComponents.yaml#/defs/AdjacencyStart
adjacencyEnd: ./requestParameterComponents.yaml#/defs/AdjacencyEnd
repeatSubunitCount: ./requestParameterComponents.yaml#/defs/RepeatSubunitCount
repeatSubunitLength: ./requestParameterComponents.yaml#/defs/RepeatSubunitLength
geneId: ./requestParameterComponents.yaml#/defs/GeneId
aminoacidChange: ./requestParameterComponents.yaml#/defs/AminoacidChange
genomicAlleleShortForm:
./requestParameterComponents.yaml#/defs/GenomicAlleleShortForm
sequenceLength: ./requestParameterComponents.yaml#/defs/SequenceLength
vrsType: ./requestParameterComponents.yaml#/defs/VRStype
```

Table of contents

- VQSRequest Parameters
- Beacon v2+/VQS "VRSified"
- Request Examples
  - Copy number gains involving the whole locus chr2:54,700,000-63,900,000
  - Focal high-level deletion involving the CDKN2A locus
  - Find t(8;14)(q24;q32) translocations
  - CAG repeat in the first exon of the huntingtin gene (HTT)
  - CAG repeat in the first exon of the huntingtin gene (HTT)
  - CGG trinucleotide repeat expansion in the FMR1 gene
  - Query for a focal deletion involving TP53

<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...

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

```
requestProfile: VQScopyChangeRequest  
referenceAccession: refseq:NC_00002.12  
start:  
  o 21000001  
  o 21975098  
end:  
  o 21967753  
  o 23000000  
copyChange: EFO:0020073  
vrsType: CopyNumberChange
```

```
requestProfile: VQSadjacencyRequest  
referenceAccession: refseq:NC_00008.11  
start: 116700000  
end: 145138636  
adjacencyAccession: refseq:NC_00014.9  
adjacencyStart: 89300000  
adjacencyEnd: 107043718  
vrsType: Adjacency
```

```
requestProfile: VQSSequenceRepeatRequest  
geneId: HTT  
repeatSubunitLength: 3  
sequenceLength:  
  o 105  
  o 750  
vrsType: ReferenceLengthExpression
```

## VQSadjacencyRequest:

description: |-

A typical Beacon v2.n request for sequence adjacency queries, e.g. for the retrieval of chromosomal translocation events or sequence fusions.

TODO: In VRS v2 there is an implicit sequence directionality from the use of either start or end parameters for either side of the adjacency. This might be problematic on the query side where in many instances just the approximate position of the (fused) breakpoints might be of interest.

This might need additional clarification (e.g. use of `startRange` or `endRange`, `adjacencyStartRange` and `adjacencyEndRange` parameters to indicate direction dependent matching).

type: object

properties:

requestProfile:

const: VQSadjacencyRequest

referenceAccession:

\$ref: "./requestParameterComponents.yaml#/defs/RefgetAccession"

sequenceRange:

\$ref: "./requestParameterComponents.yaml#/defs/Range"

adjacencyAccession:

\$ref: "./requestParameterComponents.yaml#/defs/AdjacencyAccession"

adjacencyRange:

\$ref: "./requestParameterComponents.yaml#/defs/Range"

vrsType:

const: Adjacency

required:

- requestProfile
- referenceAccession
- sequenceRange
- adjacencyAccession
- adjacencyRange
- vrsType

examples:

VQSadjacency\_01:

description: |-

### Find t(8;14)(q24;q32) translocations

#### Solution for `VQSrequest` using genomic ranges (`VQSadjacencyRequest`)

This is a query for translocations between the MYC and IgH loci, where the breakpoints are loosely defined through these well known cytogenetic bands. The query here follows the VRS adjacency model. In contrast to the VRS representational model, here:

- VRS uses an array of 2 genomic locations while Beacon names the location parameters individually (using "adjacency..." for the second partner)
- VRS explicitly encodes directionality by using either `start` or `end` position parameters (integers or ranges) while this query example creates non-directional ranges on both sides since directionality might not be known, the storage system might not encode this or all options could be of interest

request:

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

# Variant Query Standard

## VRS aligned typed queries - Open Questions...

- Parameter Zoo?

- Should we be explicit in parameters themselves

- ▶ **startRange** vs. **start** and “requires 2 pos. in context of profile”

- Level of VRSification?

- Queries don't necessarily correspond to VRS objects (polymorphic matches) - is the use of VRS vocabularies appropriate?

The screenshot shows a slide with two main parts: a JSON schema on the left and an example query on the right.

**JSON Schema (Left):**

```
VQScopyChangeRequest:
  description: |-  
    A typical Beacon v2.n request for copy number variations (CNVs) queries  
    approximate positions for CNV start and end regions through use of the  
    `Range` type. The `copyChange` parameter indicates the relative change in  
    genomic copy number (pls. refer to the class definition.)  
  type: object  
  properties:  
    requestProfile:  
      const: VQScopyChangeRequest  
    referenceAccession:  
      $ref: "./requestParameterComponents.yaml#/defs/RefgetAccession"  
    startRange:  
      $ref: "./requestParameterComponents.yaml#/defs/Range"  
    endRange:  
      $ref: "./requestParameterComponents.yaml#/defs/Range"  
    copyChange:  
      $ref: "./requestParameterComponents.yaml#/defs/CopyChange"  
    sequenceLength:  
      $ref: "./requestParameterComponents.yaml#/defs/SequenceLength"  
    vrsType:  
      const: CopyNumberChange  
  required:  
    - requestProfile  
    - referenceAccession  
    - startRange  
    - endRange  
    - copyChange
```

**Example Query (Right):**

```
requestProfile: VQScopyChangeRequest  
referenceAccession: refseq:NC_000002.12  
start:  
  ○ 21000001  
  ○ 21975098  
end:  
  ○ 21967753  
  ○ 23000000  
copyChange: EF0:0020073  
vrsType: CopyNumberChange
```

The 'startRange' field in the schema and the 'start' field in the example are both highlighted with red boxes.

# Beacon v2 Variant Requests

## Mix & Match?

- parameters allow positional and some identifier/classification based queries
  - ➡ genomic positions, sequences, variant types
  - ➡ no definition of allowed combinations so strange options possible...
    - ▶ genome assembly + versioned reference
  - ➡ patterns by convention/documentation
    - ▶ single start, end => range
    - ▶ 2 start, 2 end => bracket/CNV-style

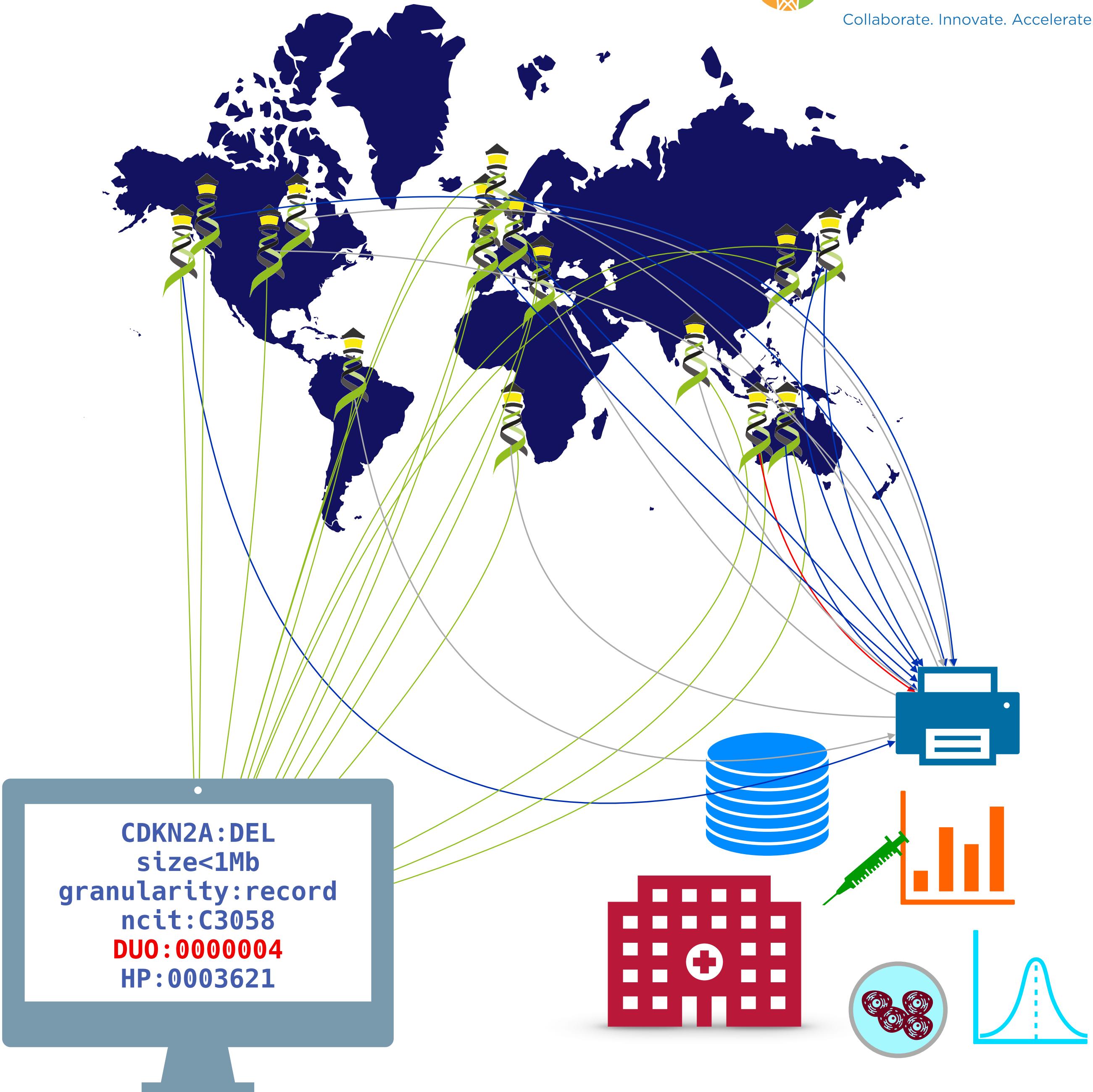
### g\_variant Parameters

```
assemblyId : ./requestParameterComponents.yaml#/defs/Assembly  
referenceName : ./requestParameterComponents.yaml#/defs/RefSeqId  
referenceBases : ./requestParameterComponents.yaml#/defs/ReferenceBases  
alternateBases : ./requestParameterComponents.yaml#/defs/AlternateBases  
variantType : ./requestParameterComponents.yaml#/defs/VariantType  
start : ./requestParameterComponents.yaml#/defs/Start  
end : ./requestParameterComponents.yaml#/defs/End  
geneId : ./requestParameterComponents.yaml#/defs/GenelId  
aminoacidChange : ./requestParameterComponents.yaml#/defs/AminoacidChange  
genomicAlleleShortForm :  
./requestParameterComponents.yaml#/defs/GenomicAlleleShortForm  
variantMinLength : ./requestParameterComponents.yaml#/defs/VariantMinLength  
variantMaxLength : ./requestParameterComponents.yaml#/defs/VariantMaxLength
```

# What Can You Do?

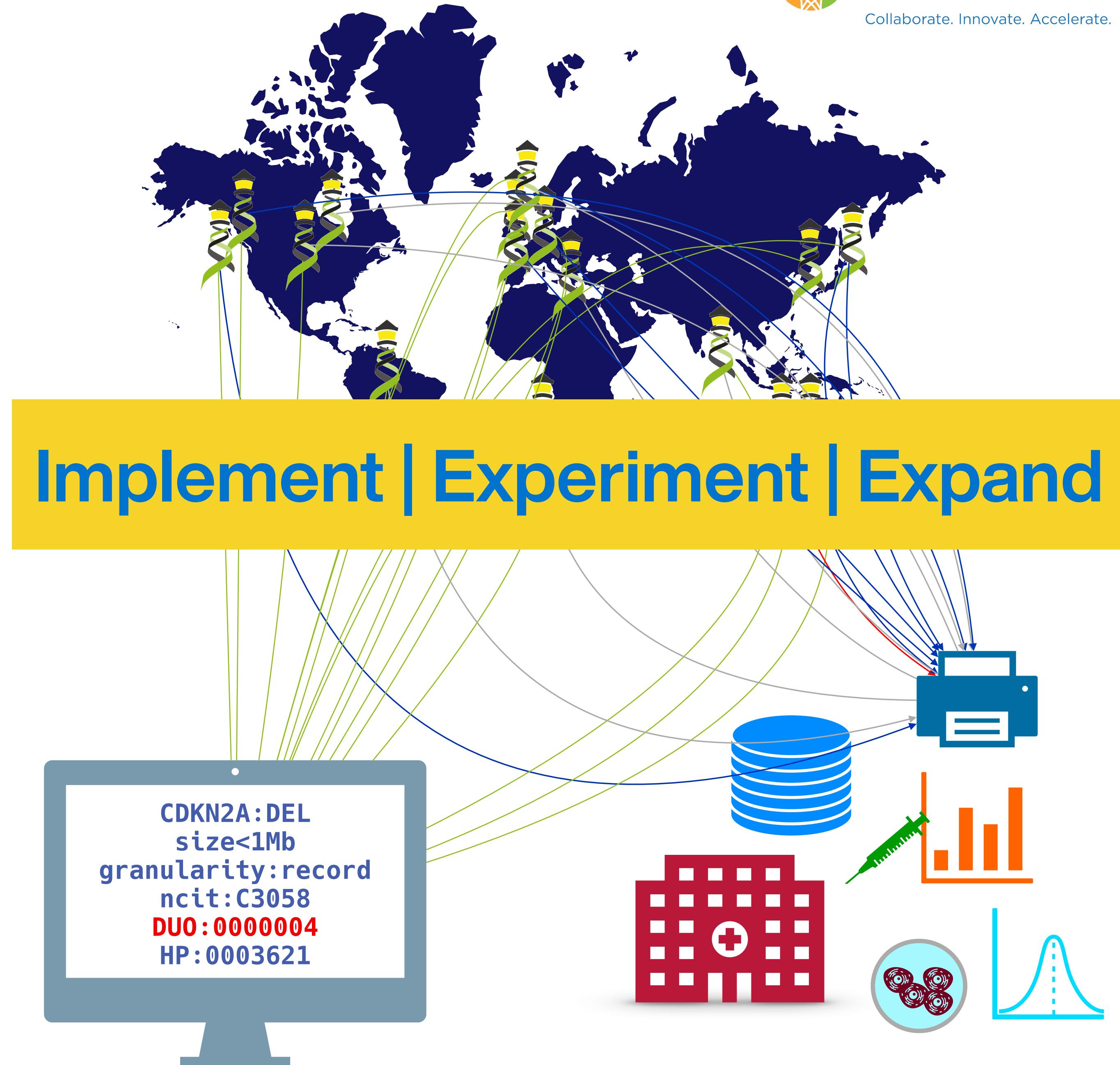
- find a way to make your (patients') **data discoverable** - through adding *at least* the relevant metadata to national or project centric repositories
- use forward looking consent and data protection models (**ORD** principle "as secure as necessary, as open as possible")
- **support** and/or get involved with international **data standards** efforts and projects
- ... **talk to us**

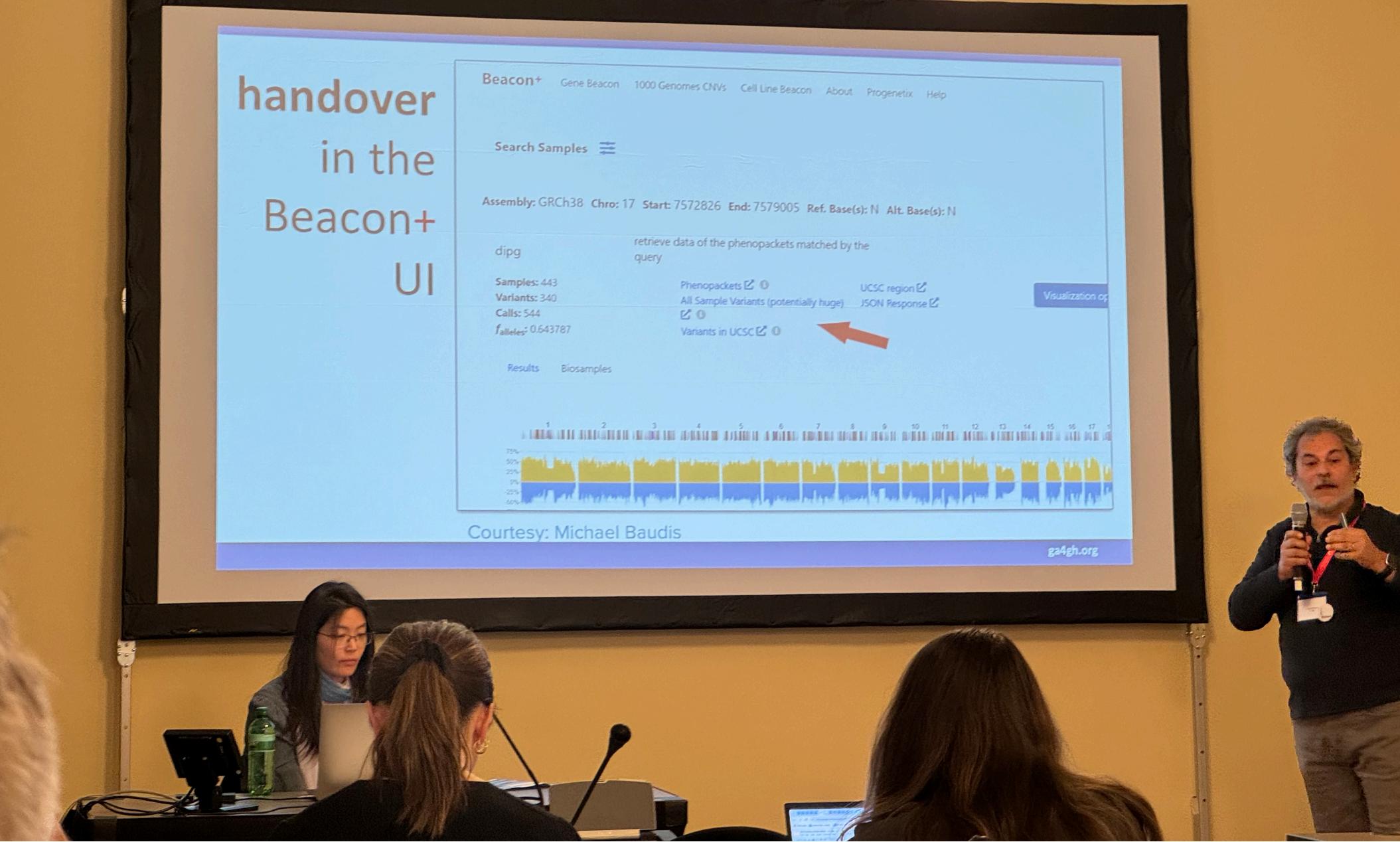
[bycon.progenetix.org](http://bycon.progenetix.org)  
[github.com/progenetix/bycon/](https://github.com/progenetix/bycon/)



# Beacon for Genomic Discovery Proxies

- Feature beacons for privacy protecting data discovery
  - privacy protection through aggregated data, cohorts
  - alternative is "**horizontal gatekeeping**": separate Beacons for **discovery** of e.g. genomic and phenotypic data and **data delivery** upon request / authentication
  - We'd love to help launching your beacon (especially as a **bycon**...)





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



Swiss Institute of  
Bioinformatics

Michael Baudis

Hangjia Zhao

Ziying Yang

Ramon Benitez

Brito

Rahel Paloots

Bo Gao

Qingyao Huang



Jordi Rambla

Arcadi Navarro

Roberto Ariosa

Manuel Rueda

Lauren Fromont

Mauricio Moldes

Liina Nagirnaja

Claudia Vasallo

Babita Singh

Sabela de la Torre

Fred Haziza



Tony Brookes

Tim Beck

Colin Veal

Tom Shorter



Juha Törnroos

Teemu Kataja

Ikkka Lappalainen

Dylan Spalding



Augusto Rendon

Ignacio Medina

Javier López

Jacobo Coll

Antonio Rueda



centre nacional d'anàlisi genòmica  
centro nacional de análisis genómico

Sergi Beltran

Carles Hernandez



Institut national  
de la santé et de la recherche médicale

David Salgado



Barcelona Supercomputing Center

Centro Nacional de Supercomputación

Salvador Capella

Dmitry Repchevski

JM Fernández



Laura Furlong

Janet Piñero



Serena Scollen

Gary Saunders

Giselle Kerry

David Lloyd



Nicola Mulder

Mamana

Mbiyavanga

Ziyaad Parker



David

Torrents



Dean Hartley



Fundación Progreso y Salud  
CONSEJERÍA DE SALUD

Joaquin Dopazo

Javier Pérez

J.L. Fernández

Gema Roldan

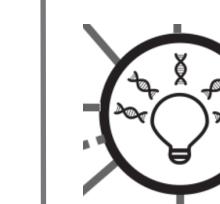


CINECA

Thomas Keane

Melanie Courtot

Jonathan Dursi

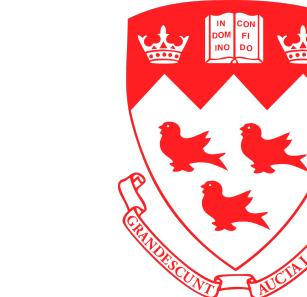


Heidi Rehm

Ben Hutton



Toshiaki Katayama



Stephane Dyke

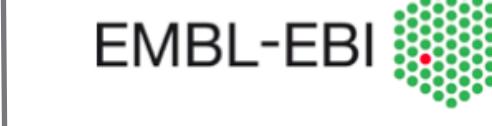


Marc Fiume

Miro Cupak



Melissa Cline



Diana Lemos



GA4GH Phenopackets

Peter Robinson

Jules Jacobsen



GA4GH VRS

Alex Wagner

Reece Hart

Beacon PRC

Alex Wagner

Jonathan Dursi

Mamana Mbiyavanga

Alice Mann

Neerjah Skantharajah



# The Beacon team through the ages