Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

Universität Zürich[UZH]

Federated genomic discoveries
Deploying the GA4GH Beacon protocol

Michael Baudis @ GHGA Seminar Series 2024-03-20

SIB Swiss Institute of Bioinformatics

elixir

Theoretical Cytogenetics and Oncogenomics

Cancer Genomics | Data Resources | Methods & Standards for Genomics and Personalized Health

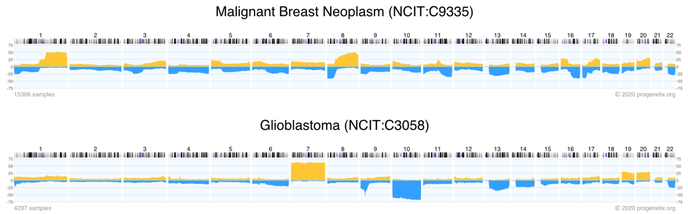
Curators

Data ~~Parasites~~

Theoretical Cytogenetics and Oncogenomics

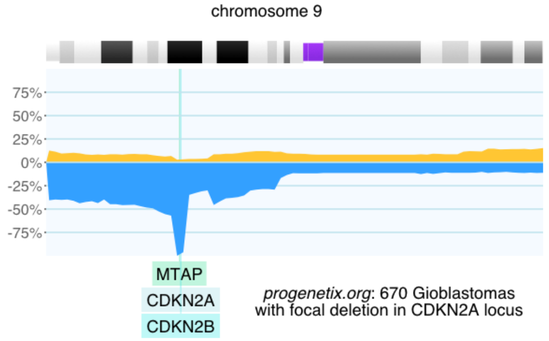
... but what does this entail @baudisgroup?

- 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"



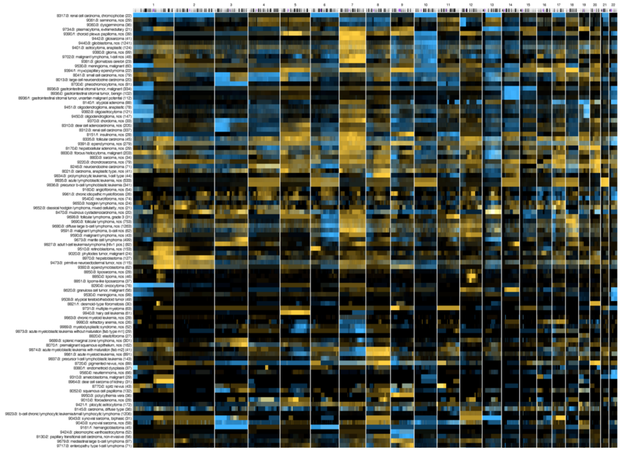
progenetix

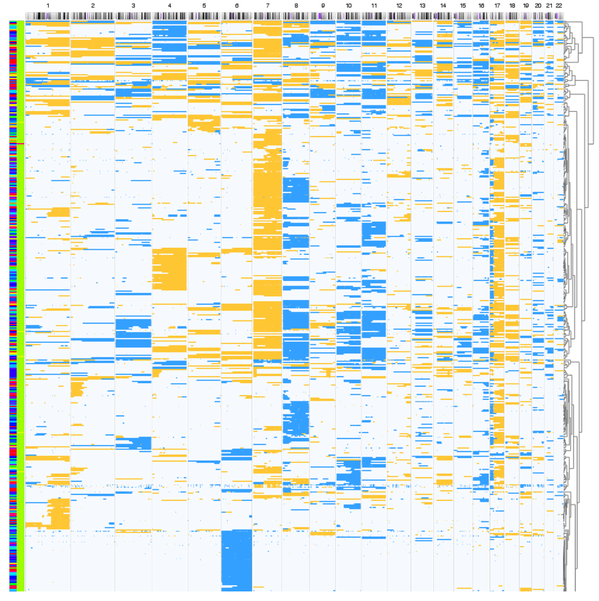
Theoretical Cytogenetics and Oncogenomics Research | Methods | Standards

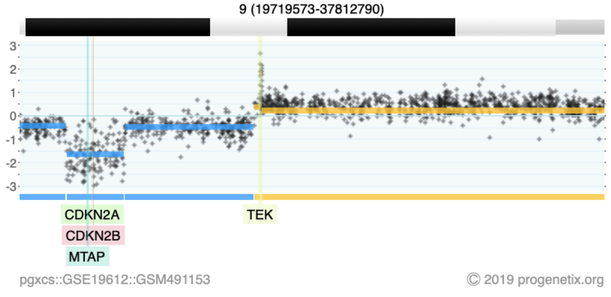


Genomic Imbalances in Cancer - Copy Number Variations (CNV)

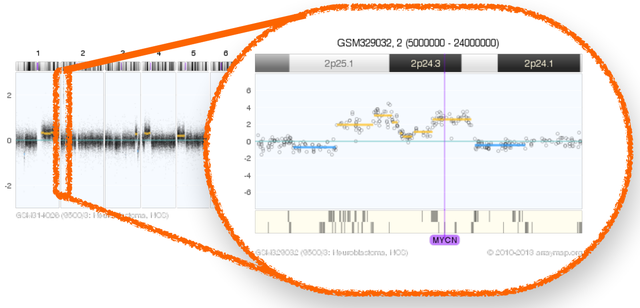
- Point mutations (insertions, deletions, substitutions)
- Chromosomal rearrangements
- **Regional Copy Number Alterations** (losses, gains)
- Epigenetic changes (e.g. DNA methylation abnormalities)







2-event, homozygous deletion in a Glioblastoma



MYCN amplification in neuroblastoma
(GSM314026, SJNB8_N cell line)

progenetix.org

Cancer Genomics Reference Resource

- ***open*** resource for oncogenomic profiles
- over **116'000 cancer CNV profiles**
- more than **800 diagnostic types**
- inclusion of reference datasets (e.g. TCGA)
- 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
- recent addition of SNV data for some series

Universität Zürich[UZH]

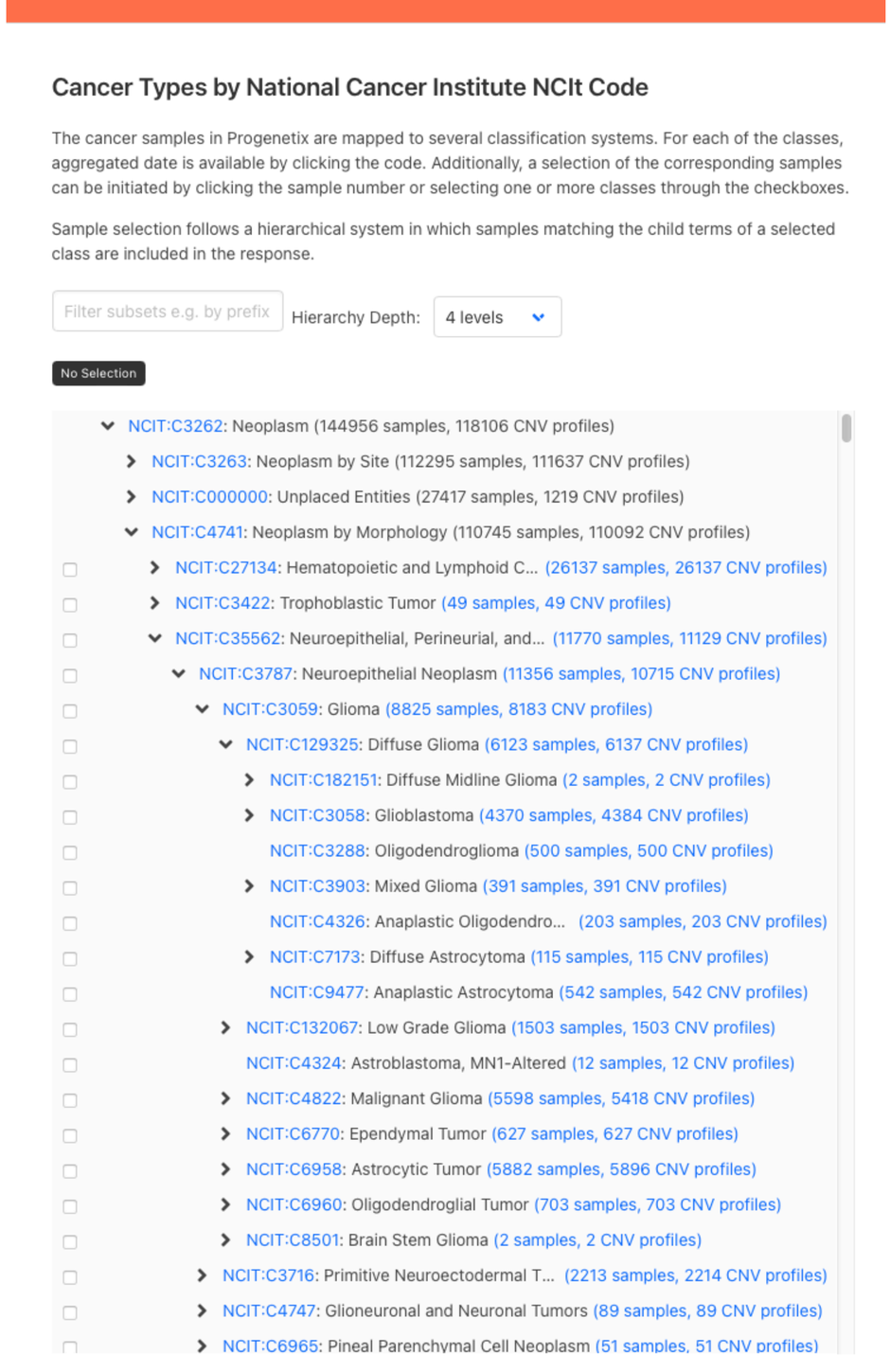
progenetix

SIB Swiss Institute of Bioinformatics

progenetix.org

Cancer Genomics Reference Resource

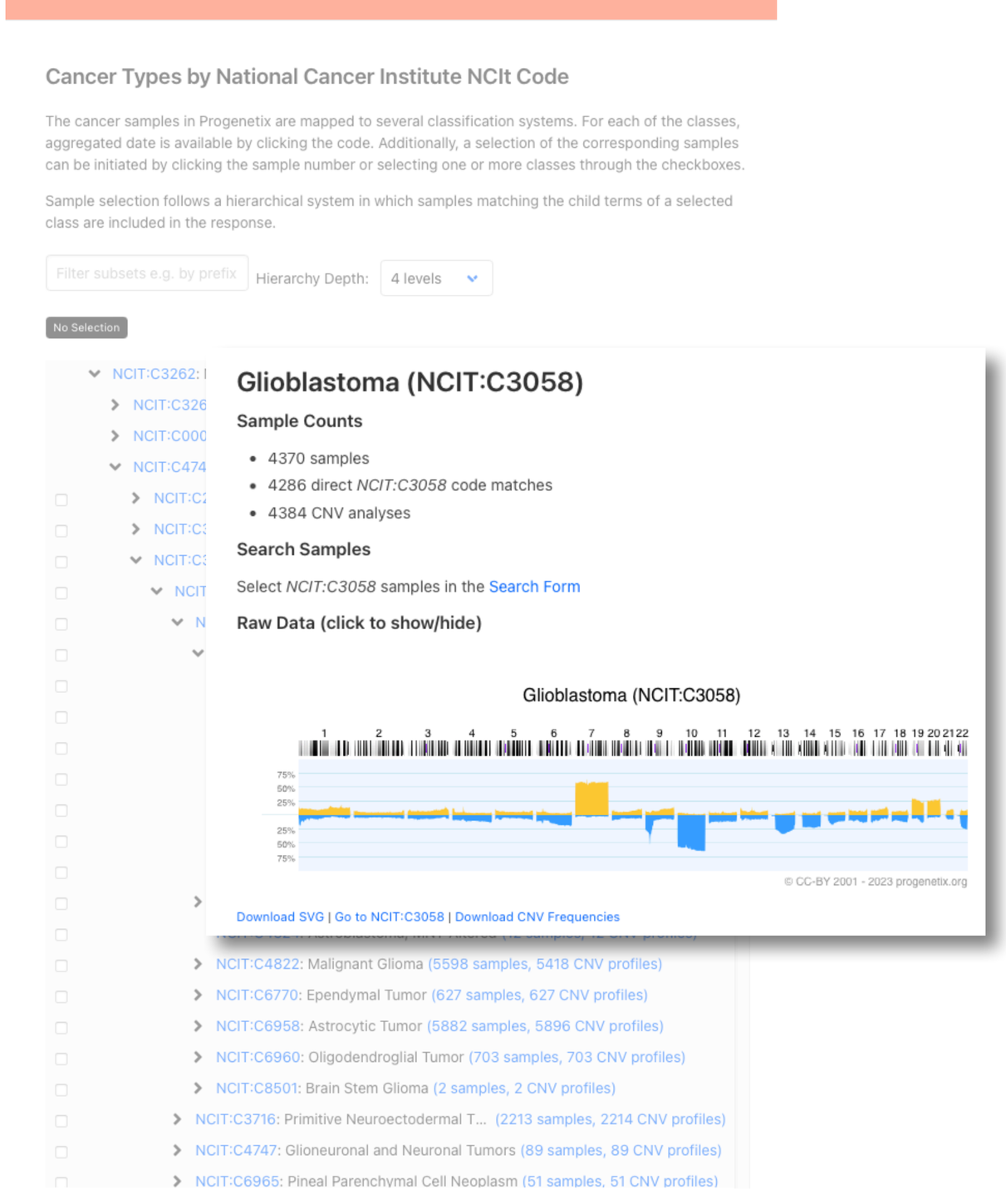
- ***open*** resource for oncogenomic profiles
- over **116'000 cancer CNV profiles**
- more than **800 diagnostic types**
- inclusion of reference datasets (e.g. TCGA)
- standardized encodings (e.g. NCIt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core clinical data (TNM, sex, survival ...)
- data mapping services
- recent addition of SNV data for some series



progenetix.org

Cancer Genomics Reference Resource

- ***open*** resource for oncogenomic profiles
- over **116'000 cancer CNV profiles**
- more than **800 diagnostic types**
- inclusion of reference datasets (e.g. TCGA)
- 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
- recent addition of SNV data for some series



Universität Zürich[UZH]

progenetix

SIB Swiss Institute of Bioinformatics

progenetix.org

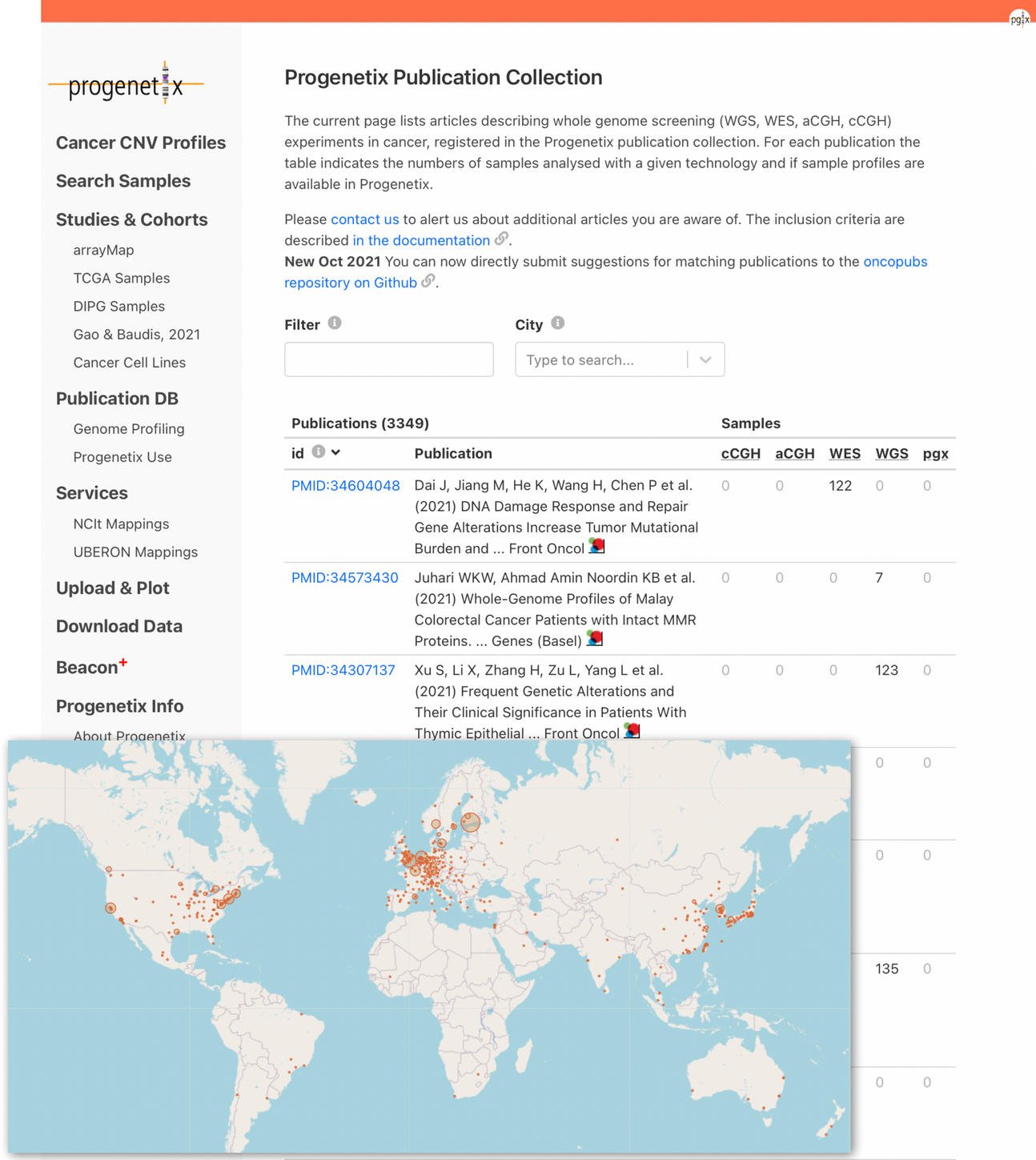
Cancer Genomics Reference Resource

- ***open*** resource for oncogenomic profiles
- over **116'000 cancer CNV profiles**
- more than **800 diagnostic types**
- inclusion of reference datasets (e.g. TCGA)
- 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
- recent addition of SNV data for some series

Universität Zürich[UZH]

progenetix

SIB Swiss Institute of Bioinformatics



Cancer Cell Lines

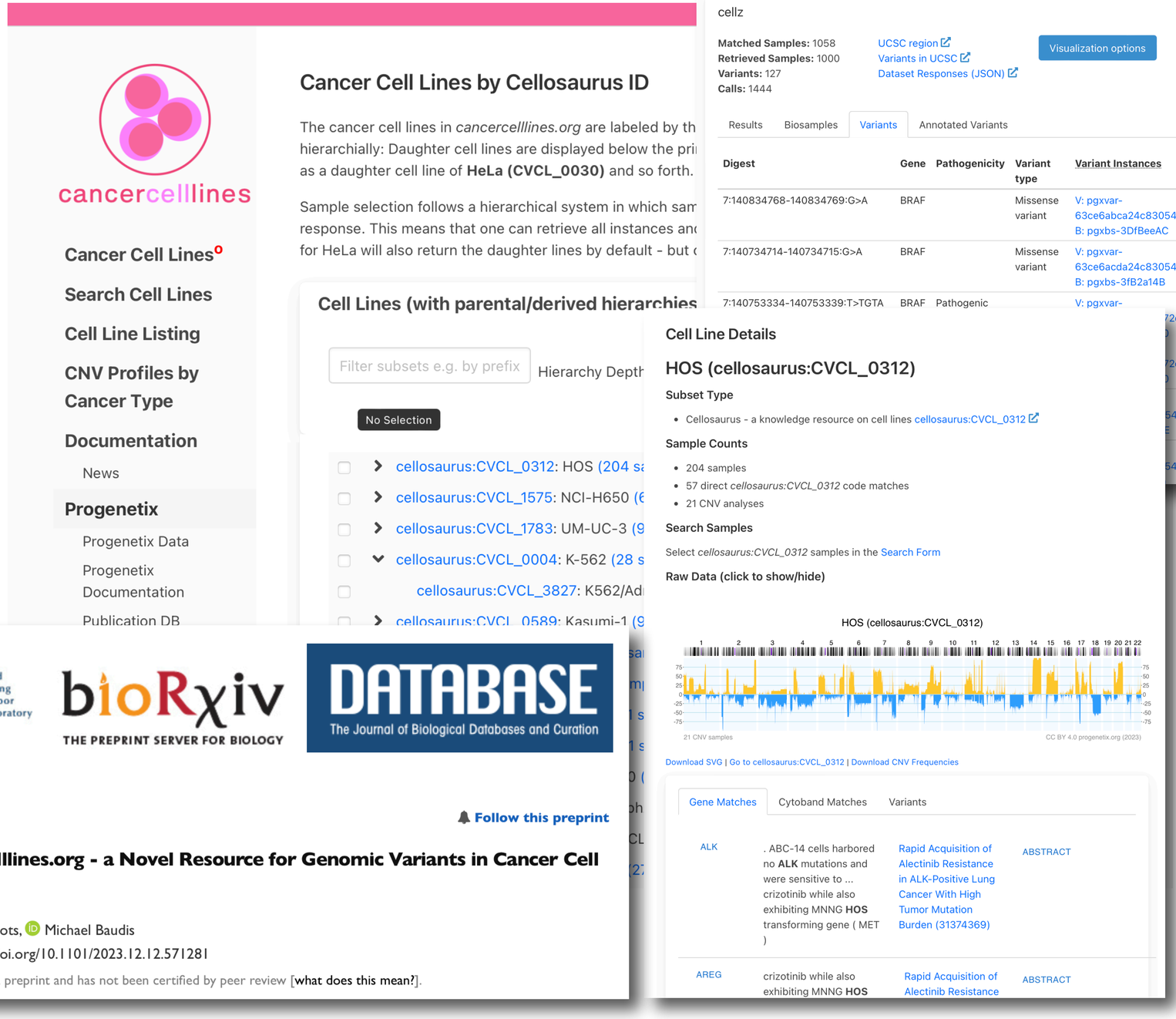
Cancer Genomics Reference Resource

- starting from >5000 cell line CNV profiles
  - 5754 samples | 2163 cell lines
  - 256 different NCIT codes
- genomic mapping of annotated variants and additional data from several resources (ClinVar, CCLE, Cellosaurus...)
  - 16178 cell lines
  - 400 different NCIT codes
- query and data delivery through Beacon v2 API

➡ integration in data federation approaches

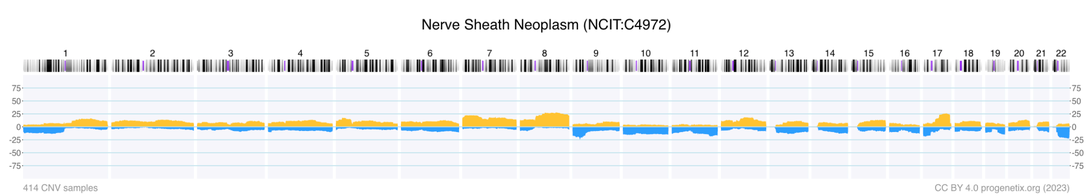
cancercelllines.org

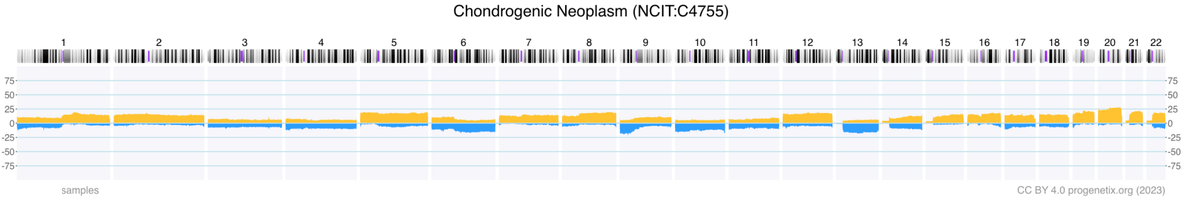
Lead: Rahel Paloots

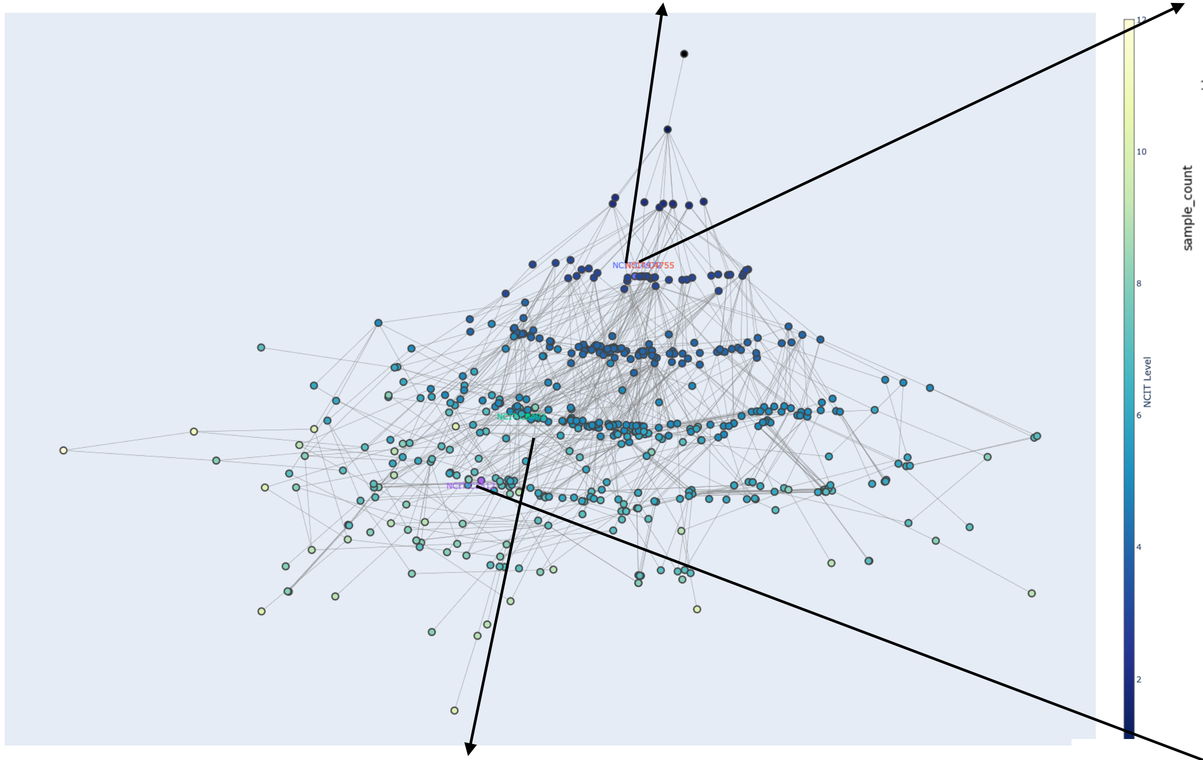


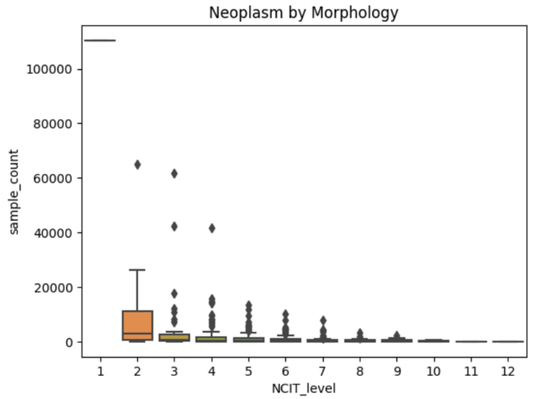
CNV profiles heterogeneity vs cancer classification

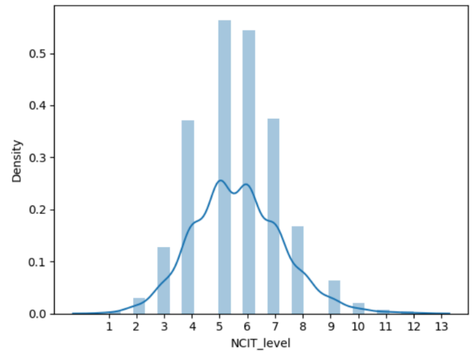
Correspondance of genomic profiles to NCIT cancer hierarchy

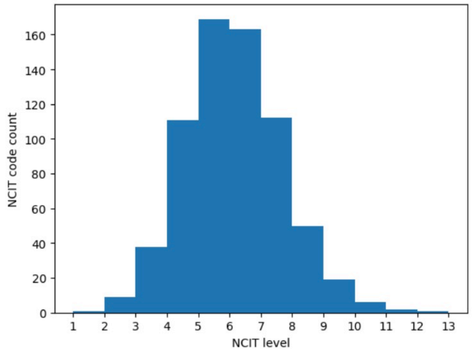


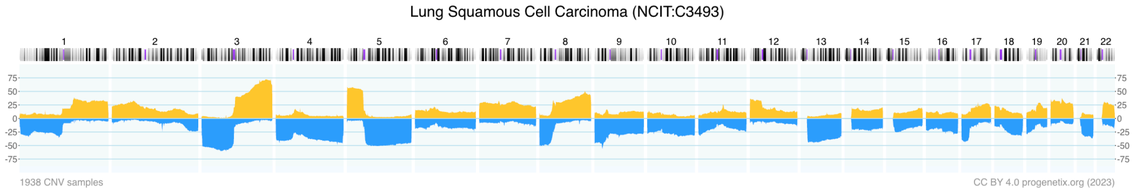


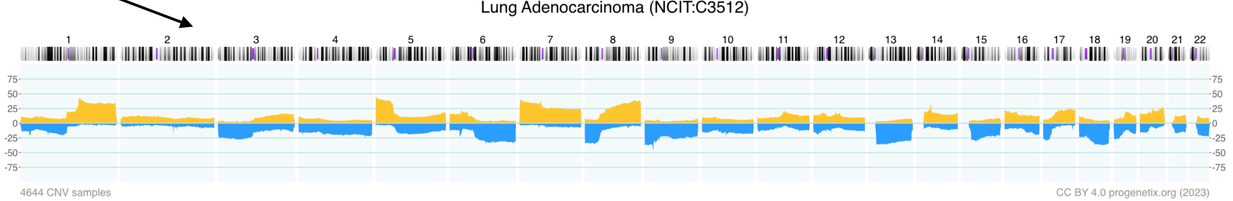








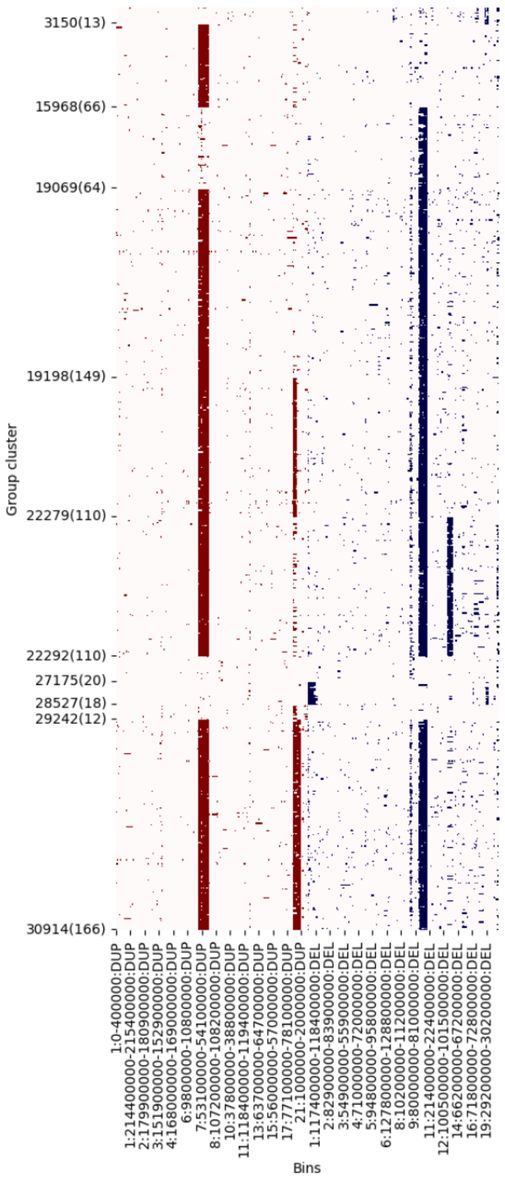


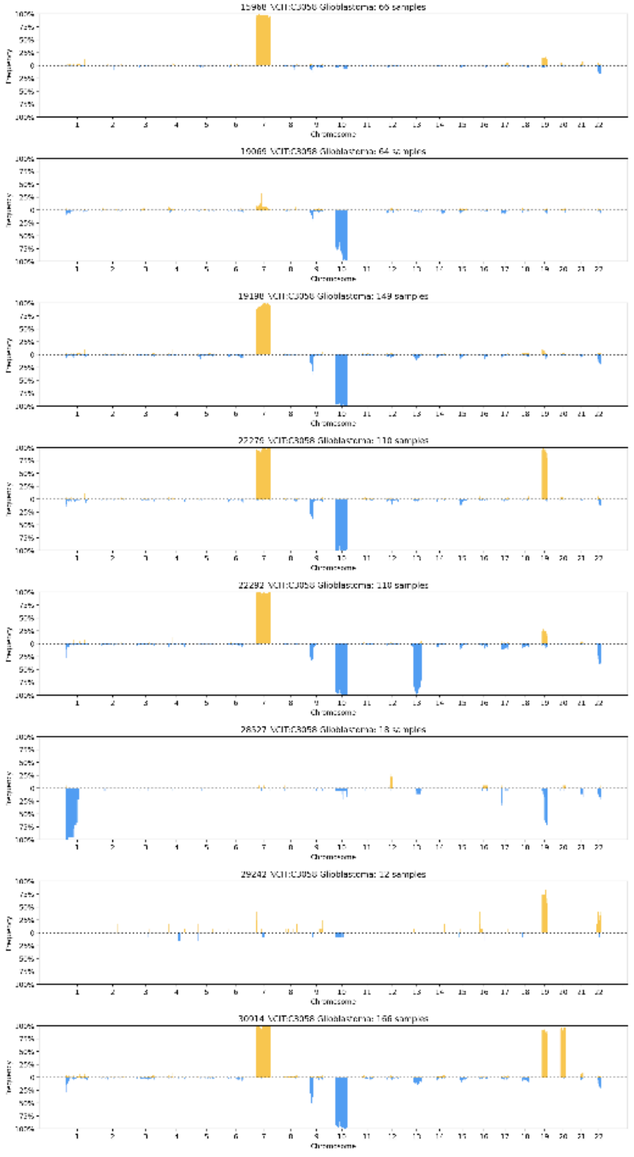


Lead: Ziying Yang

Results

Entity CNV heterogeneity: Glioblastoma



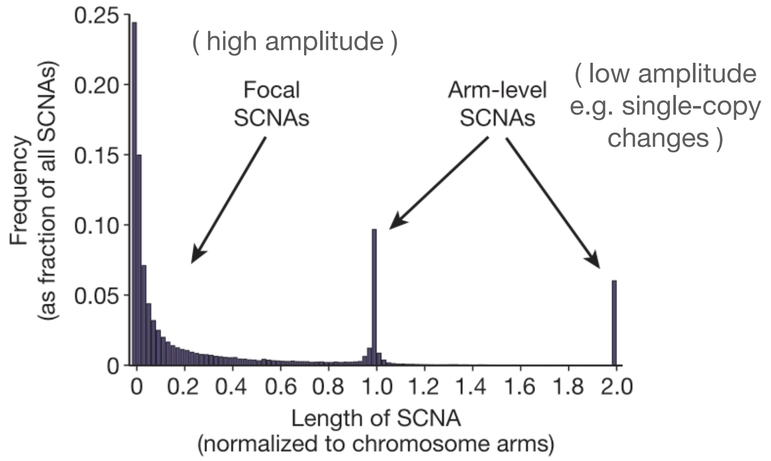


group cluster	CNV features
15968	Dup 7
19069	Del 10
19198	Dup 7, Del 10
22279	Dup 7, Del 10, Dup 19
22292	Dup 7, Del 10, Del 13
28527	Del 1p, Del 19q
29242	Dup 19
30914	Dup 7, Del 10, Dup 19, Dup 20

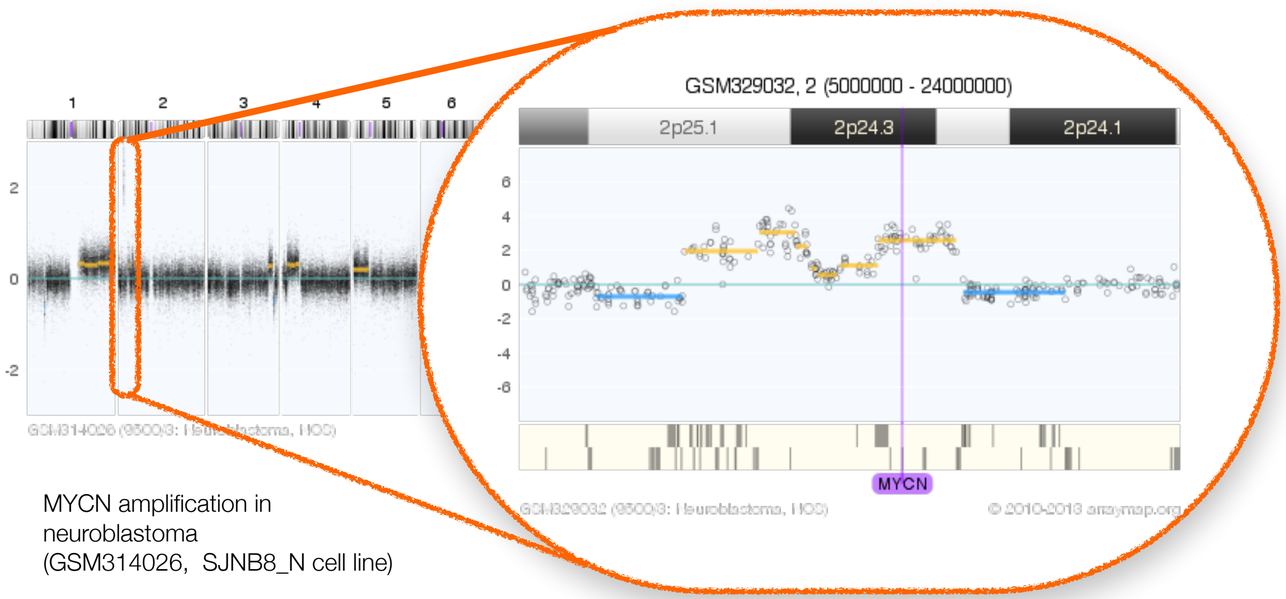
Lead: Ziying Yang

CNV Categorization

different levels of CNV



Rameen et al 2010 Nature



MYCN amplification in neuroblastoma (GSM314026, SJNB8_N cell line)

Lead: Hangjia Zhao

GA4GH Variation Representation Specification

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

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 HGVS expressions submitted to express copy number variation are interpreted to be relative copy changes.

Computational Definition

An assessment of the copy number of a Location or a Feature within a system (e.g. genome, cell, etc.) relative to a baseline ploidy.

Information Model

Some CopyNumberChange attributes are inherited from Variation.

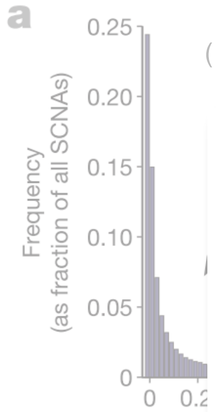
Field	Type	Limits	Description
_id	CURIE	0..1	Variation Id. MUST be unique within document.
type	string	1..1	MUST be “CopyNumberChange”
subject	Location | CURIE | Feature	1..1	A location for which the number of systemic copies is described.
copy_change	string	1..1	MUST be one of “efo:0030069” (complete 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).

CNV Categorization

different levels of CNV

GA4GH Variation Representation Specification

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



(high amplitude)

Focal

Arm-level

(low amplitude

OXFORD

Briefings in Bioinformatics, 2024, **25(2)**, 1–12

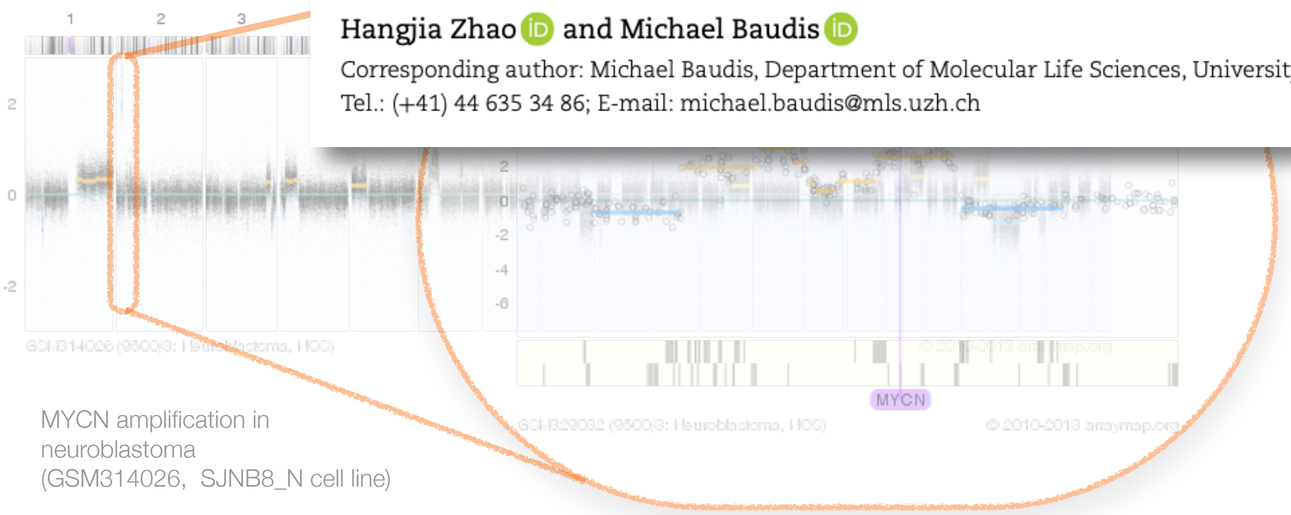
https://doi.org/10.1093/bib/bbad541

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)

CopyNumberChange

Copy Number Change captures a categorization of copies of a molecule within a system, relative to a
allers, particularly in the somatic
nd less useful in practice than
s relative statements, and many
nterpreted to be relative copy

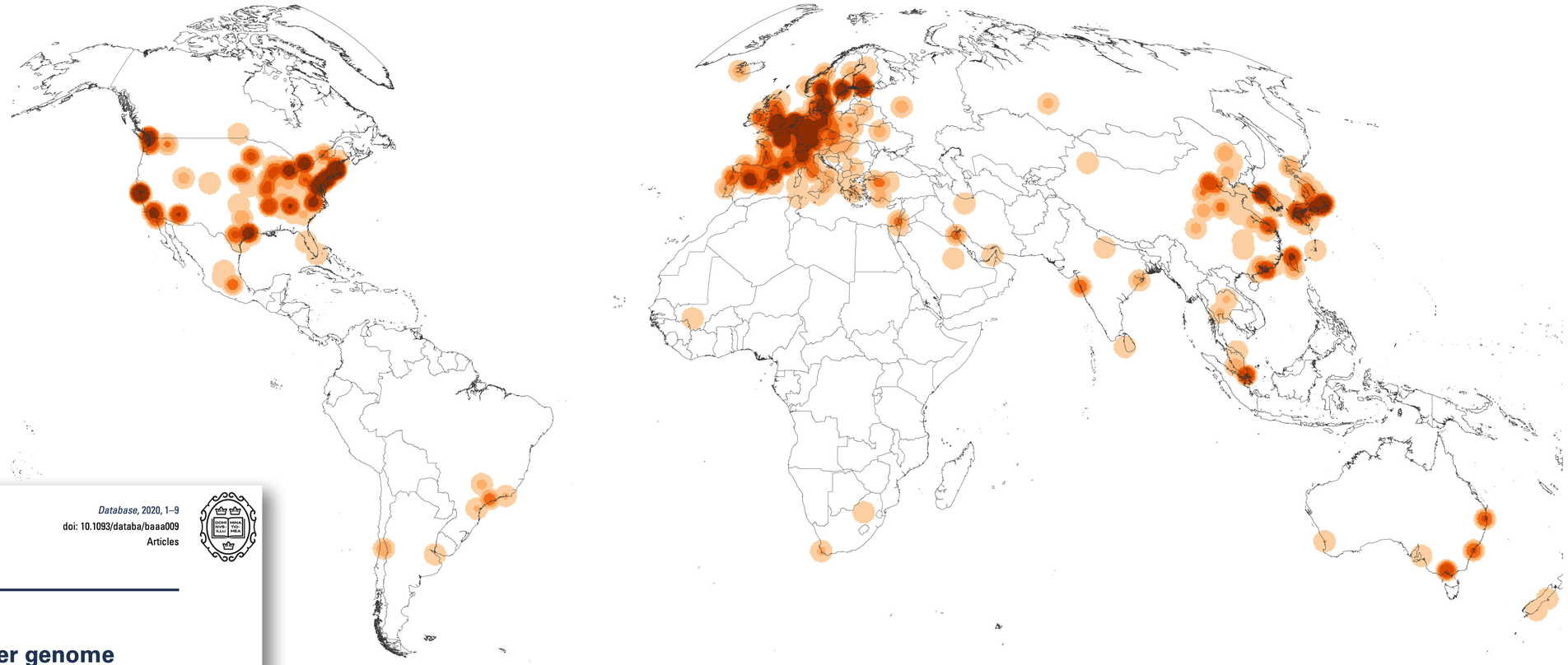
a system (e.g. genome, cell,

_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).

Lead: Hangjia Zhao

Where does Genomic Data Come From?

Geographic bias in published cancer genome profiling studies



DATABASE
The Journal of Biological Databases and Curation

Database, 2020, 1–9
doi: 10.1093/databa/baaa009
Articles

Articles

Geographic assessment of cancer genome profiling studies

Paula Carrio-Cordo[1,2], Elise Acheson[3], Qingyao Huang[1,2] and Michael Baudis[1,*]

[1]Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland [2]Swiss Institute of Bioinformatics, Zurich, Switzerland [3]Department of Geography, University of Zurich, Zurich, Switzerland

Map of the geographic distribution (by first author affiliation) of the 104'543 genomic array, 36'766 chromosomal CGH and 15'409 whole genome/exome based cancer genome datasets. The numbers are derived from the 3'240 publications registered in the Progenetix database.

Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

GENOMICS

A federated ecosystem for sharing genomic, clinical data

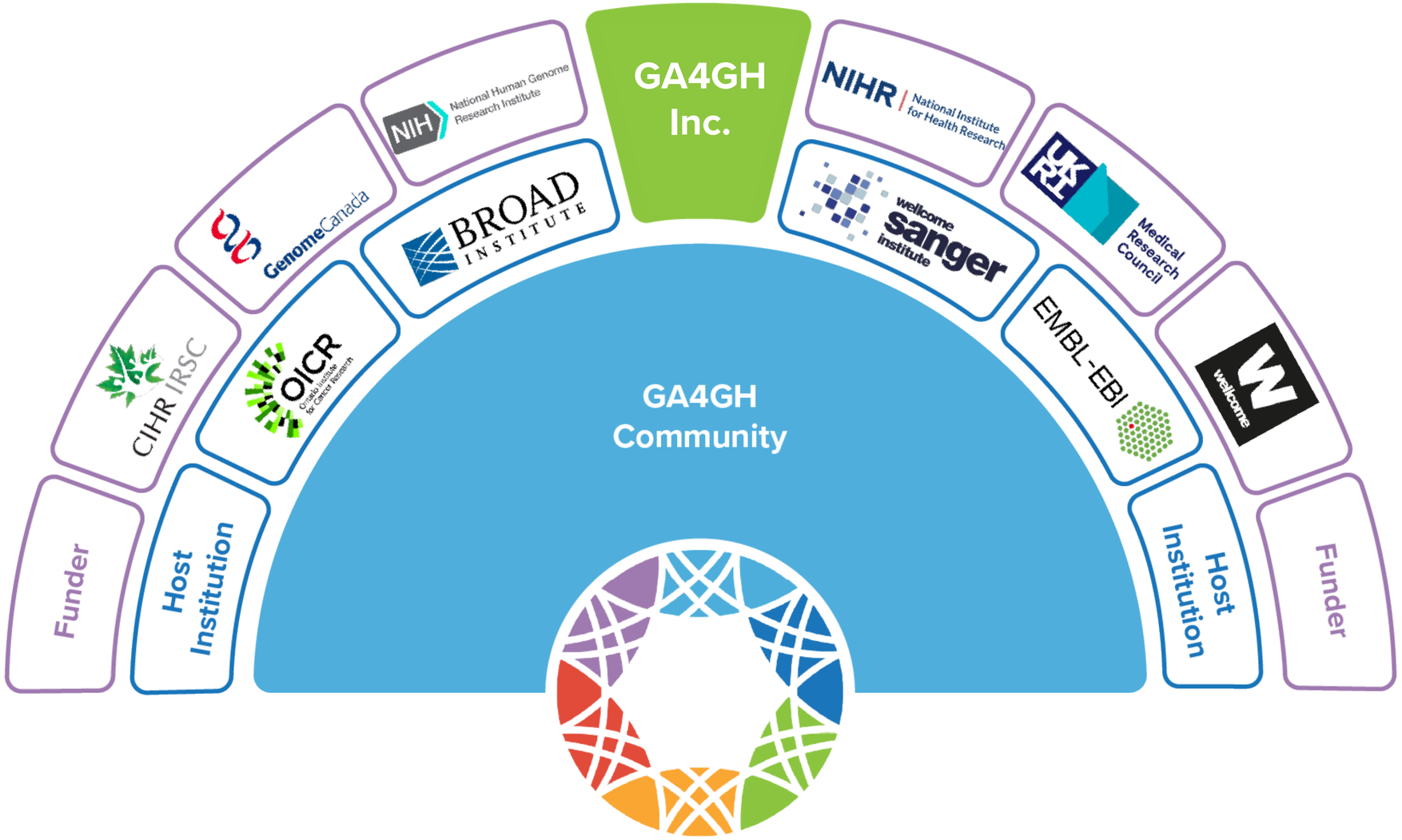
Silos of genome data collection are being transformed into seamlessly connected, independent systems

The Global Alliance for Genomics and Health*

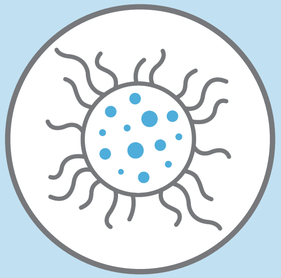
SCIENCE 10 JUNE 2016 • VOL 352 ISSUE 6291

Organization

Global Alliance
for Genomics & Health



Global Genomic Data Sharing Can…



Demonstrate patterns in health & disease

Increase statistical significance of analyses

Lead to “stronger” variant interpretations

Increase accurate diagnosis

Advance precision medicine

Different Approaches to Data Sharing

progenetix

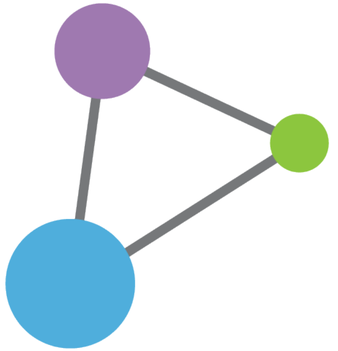
European Genome-Phenome Archive

SPHN Swiss Personalized Health Network

Centralized Genomic Knowledge Bases

Data Commons
Trusted, controlled repository of multiple datasets

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



Linkage of distributed and disparate datasets

Different Approaches to Data Sharing

progenetix

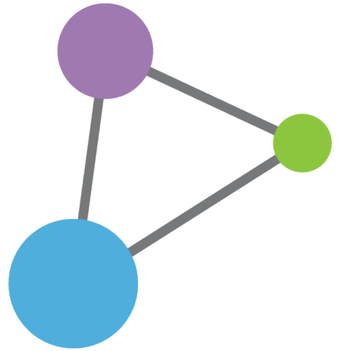
Centralized Genomic Knowledge Bases

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

Different Approaches to Data Sharing

progenetix

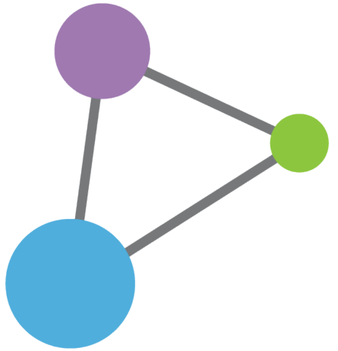
Centralized Genomic Knowledge Bases

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

The EGA

European Genome-Phenome Archive

Long term secure archive for human biomedical research sensitive data, with focus on reuse of the data for further research (or "*broad and responsible use of genomic data*")

EUROPEAN GENOME-PHENOME ARCHIVE

CRG Centre for Genomic Regulation

EMBL-EBI

Global Alliance for Genomics & Health

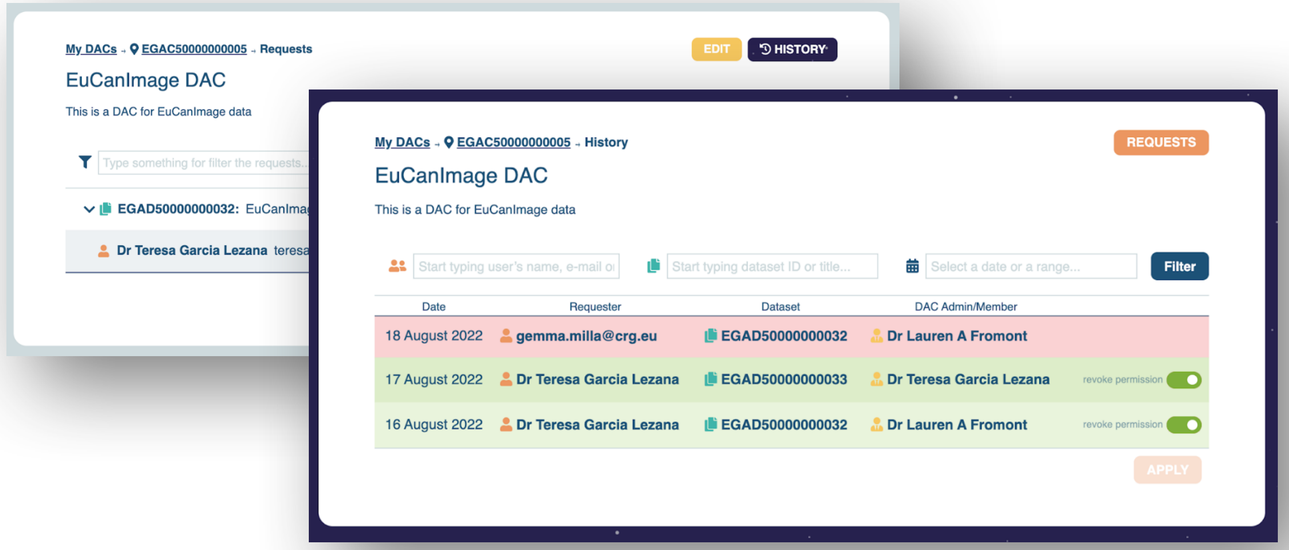
DRIVER PROJECT

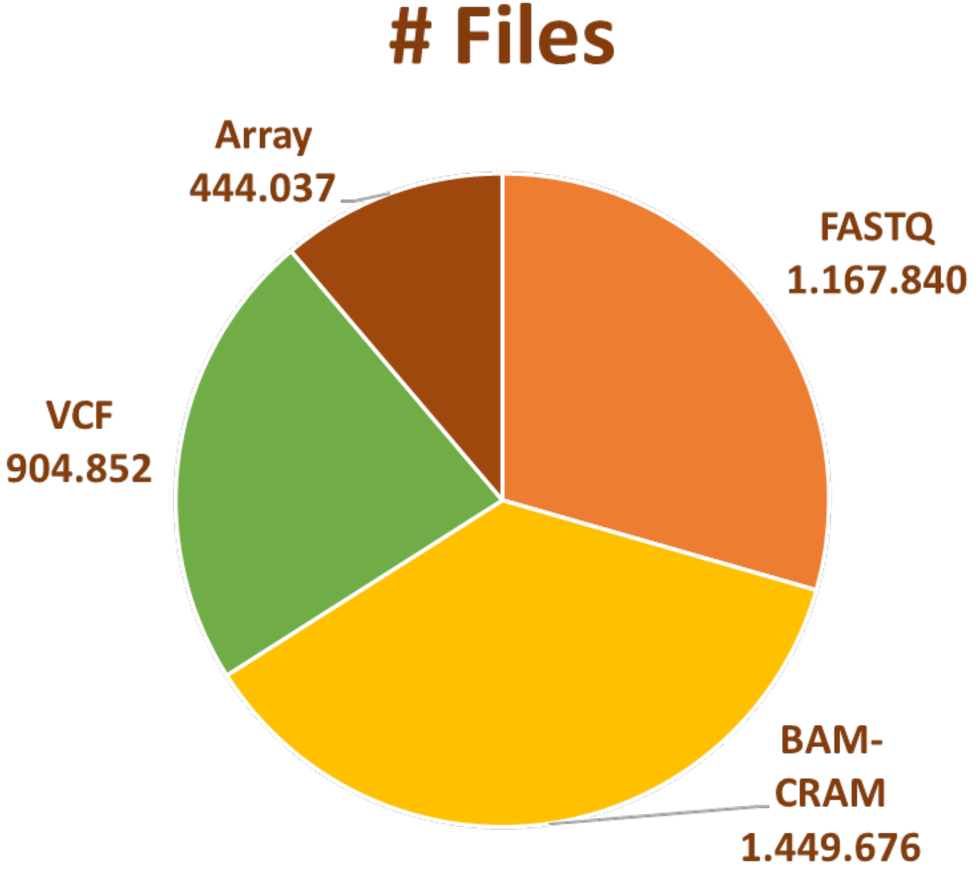
Slide: adapted from Jordi Rambla@ GA4GH 2023

The EGA

European Genome-Phenome Archive

- EGA “owns” nothing; data controllers tell who is authorized to access ***their*** datasets
- EGA admins provide smooth “all or nothing" data sharing process





4,328 Studies released

10,470 Datasets

2,309 Data Access Committees

Slide: adapted from Jordi Rambla@ GA4GH 2023

Different Approaches to Data Sharing

progenetix

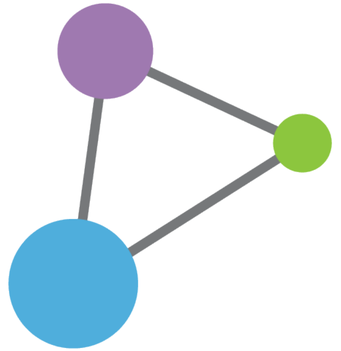
Centralized Genomic Knowledge Bases

European Genome-Phenome Archive

Data Commons
Trusted, controlled repository of multiple datasets

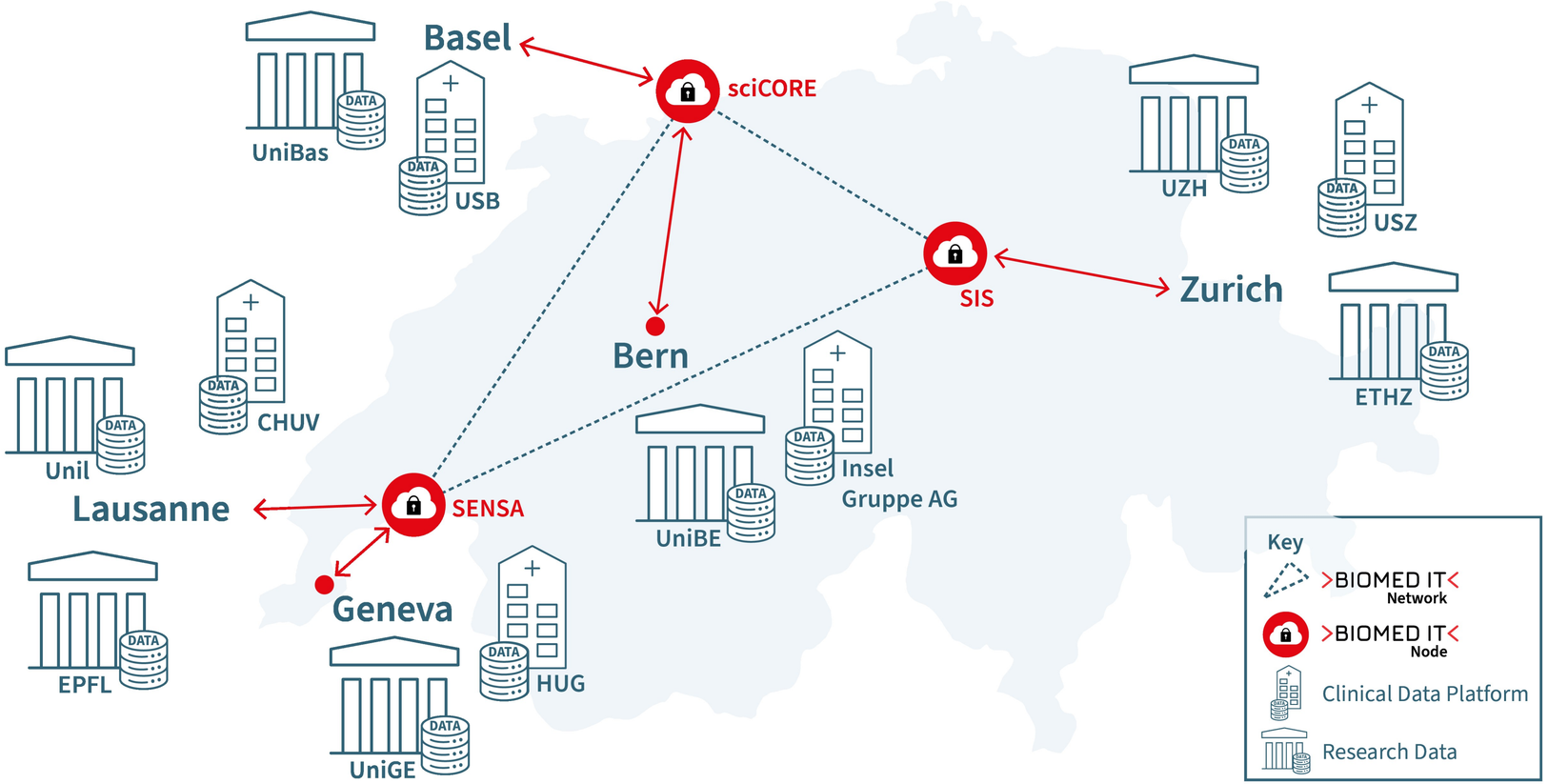
SPHN Swiss Personalized Health Network

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



Linkage of distributed and disparate datasets

The Swiss Personalized Health Network



Personalized Health Informatics Group

SPHN Data Coordination Center (DCC)
BioMedIT Network

Strategic Focus Area
Personalized Health and Related Technologies

Different Approaches to Data Sharing

progenetix

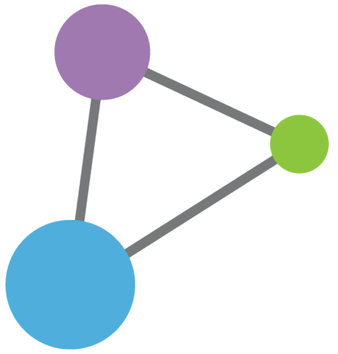
European Genome-Phenome Archive

SPHN Swiss Personalized Health Network

Centralized Genomic Knowledge Bases

Data Commons
Trusted, controlled repository of multiple datasets

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



Linkage of distributed and disparate datasets

Federation

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

INFORMATICS

Beacon v2 and Beacon networks: federated data discovery n biome

Jordi Rambla[1,2] | Michael Baudis[3] | Roberto Ariosa[1] | Tim Beck[4] |
Lauren A. Fromont[1] | Arcadi Navarro[1,5,6,7] | Rahel Paloots[3] |
Manuel Rueda[1] | Gary Saunders[8] | Babita Singh[1] | John D. Spalding[9] |
Juha Törnroos[9] | Claudia Vasallo[1] | Colin D. Veal[4] | Anthony J. Brookes[4]

Cell Genomics

CellPress
OPEN ACCESS

Commentary

International federation of genomic medicine databases using GA4GH standards

Adrian Thorogood,[1,2,*] Heidi L. Rehm,[3,4] Peter Goodhand,[5,6] Angela J.H. Page,[4,5] Yann Joly,[2] Michael Baudis,[7]
Jordi Rambla,[8,9] Arcadi Navarro,[8,10,11,12] Tommi H. Nyronen,[13,14] Mikael Linden,[13,14] Edward S. Dove,[15] Marc Fiume,[16]
Michael Brudno,[17] Melissa S. Cline,[18] and Ewan Birney[19]

Cell Genomics

OPEN ACCESS

Perspective

GA4GH: International policies and standards for data sharing across genomic research and healthcare

Heidi L. Rehm,[1,2,47] Angela J.H. Page,[1,3,*] Lindsay Smith,[3,4] Jeremy B. Adams,[3,4] Gil Alterovitz,[5,47] Lawrence J. Babb,[1]
Maxmillian P. Barkley,[6] Michael Baudis,[7,8] Michael J.S. Beauvais,[3,9] Tim Beck,[10] Jacques S. Beckmann,[11]
Sergi Beltran,[12,13,14] David Bernick,[1] Alexander Bernier,[9] James K. Bonfield,[15] Tiffany F. Boughtwood,[16,17]
Guillaume Bourque,[9,18] Sarion R. Bowers,[15] Anthony J. Brookes,[10] Michael Brudno,[18,19,20,21,38] Matthew H. Brush,[22]
David Bujold,[9,18,38] Tony Burdett,[23] Orion J. Buske,[24] Moran N. Cabili,[1] Daniel L. Cameron,[25,26] Robert J. Carroll,[27]
Esmeralda Casas-Silva,[123] Debyani Chakravarty,[29] Bimal P. Chaudhari,[30,31] Shu Hui Chen,[32] J. Michael Cherry,[33]
Justina Chung,[3,4] Melissa Cline,[34] Hayley L. Clissold,[15] Robert M. Cook-Deegan,[35] Mélanie Courtot,[23]
Fiona Cunningham,[23] Miro Cupak,[6] Robert M. Davies,[15] Danielle Denisko,[19] Megan J. Doerr,[36] Lena I. Dolman,[19]

(Author list continued on next page)

Cell Genomics

OP

Technology

The GA4GH Variation Representation Specification A computational framework for variation representation and federated identification

Alex H. Wagner,[1,2,25,*] Lawrence Babb,[3,*] Gil Alterovitz,[4,5] Michael Baudis,[6] Matthew Brush,[7] Daniel L. Cameron,[8,9]
Melissa Cline,[10] Malachi Griffith,[11] Obi L. Griffith,[11] Sarah E. Hunt,[12] David Kreda,[13] Jennifer M. Lee,[14] Stephanie Li,[15]
Javier Lopez,[16] Eric Moyer,[17] Tristan Nelson,[18] Ronak Y. Patel,[19] Kevin Riehle,[19] Peter N. Robinson,[20]
Shawn Rynearson,[21] Helen Schuilenburg,[12] Kirill Tsukanov,[12] Brian Walsh,[7] Melissa Konopko,[15] Heidi L. Rehm,[3,22]
Andrew D. Yates,[12] Robert R. Freimuth,[23] and Reece K. Hart[3,24,*]

A New Paradigm for Data Sharing

FROM

TO

Data Copying

Data Visiting

A New Paradigm for Data Sharing

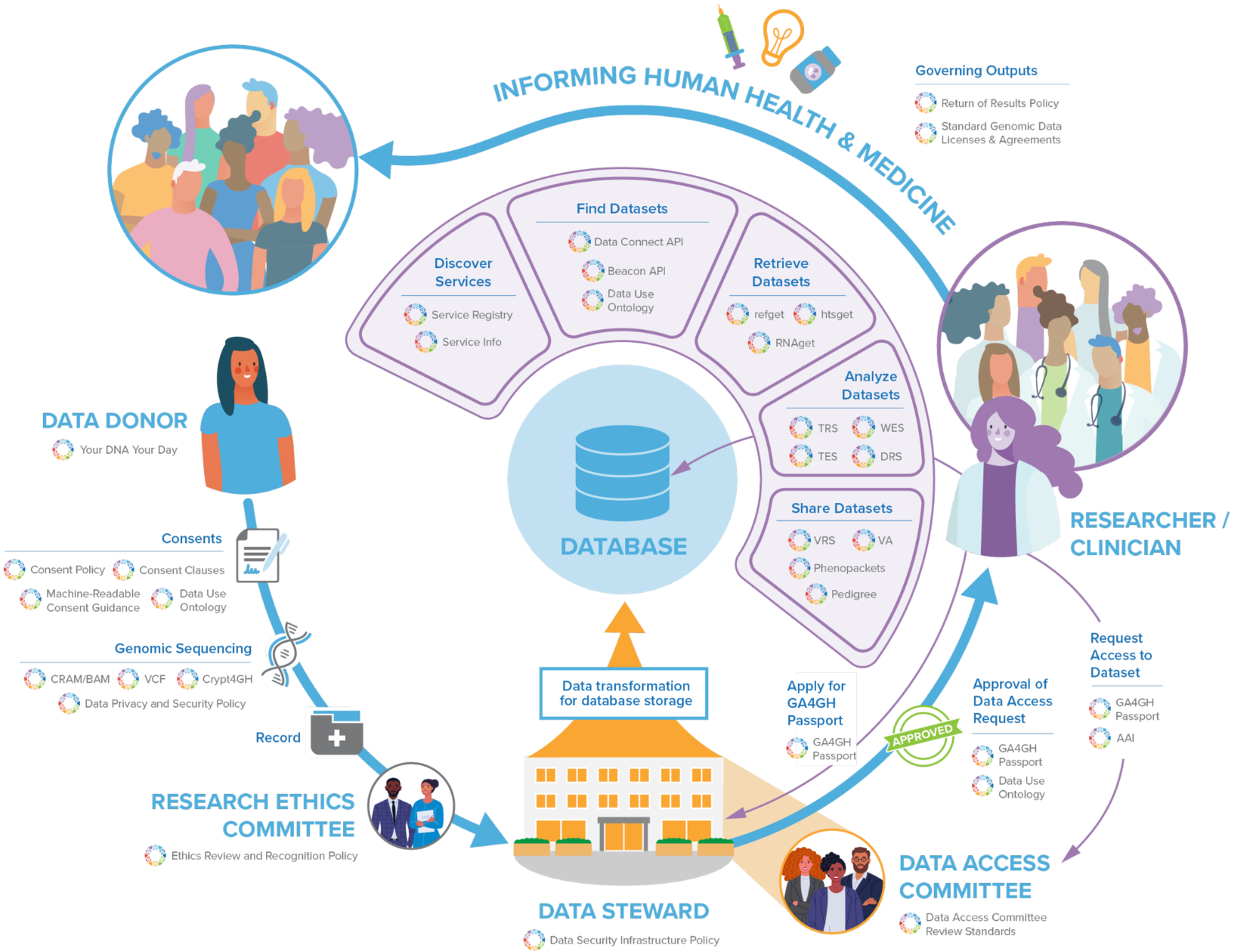
FROM

STANDARDS

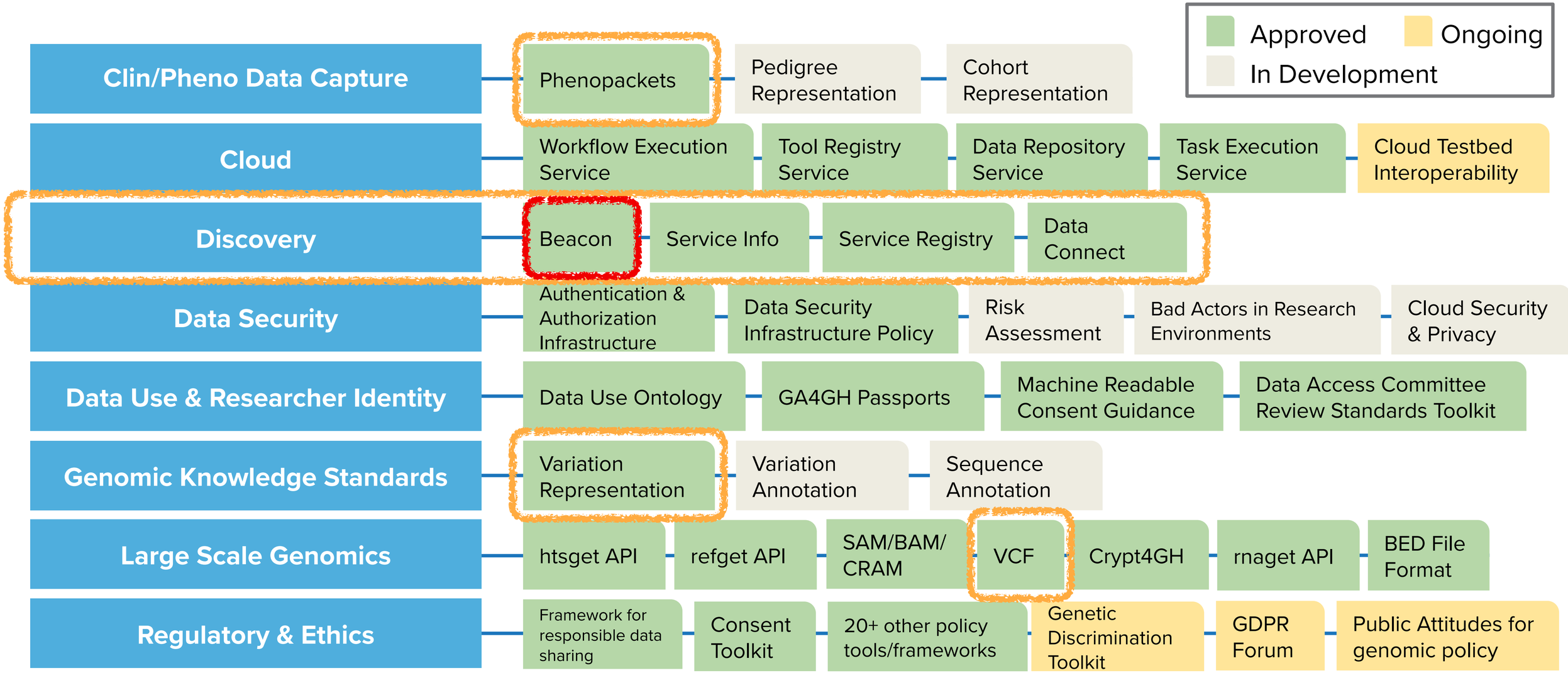
TO

Data Copying

Data Visiting


INFORMING HUMAN HEALTH & MEDICINE
Governing Outputs
Return of Results Policy
Standard Genomic Data Licenses & Agreements
Find Datasets
Data Connect API
Beacon API
Data Use Ontology
Discover Services
Service Registry
Service Info
Retrieve Datasets
refget
htsget
RNAget
Analyze Datasets
TRS
WES
TES
DRS
Share Datasets
VRS
VA
Phenopackets
Pedigree
DATABASE
DATA DONOR
Your DNA Your Day
RESEARCHER / CLINICIAN
Consents
Consent Policy
Consent Clauses
Machine-Readable Consent Guidance
Data Use Ontology
Genomic Sequencing
CRAM/BAM
VCF
Crypt4GH
Data Privacy and Security Policy
Record
Data transformation for database storage
Apply for GA4GH Passport
GA4GH Passport
APPROVED
Approval of Data Access Request
GA4GH Passport
Data Use Ontology
Request Access to Dataset
GA4GH Passport
AAI
RESEARCH ETHICS COMMITTEE
Ethics Review and Recognition Policy
DATA STEWARD
Data Security Infrastructure Policy
DATA ACCESS COMMITTEE
Data Access Committee Review Standards

Overview of GA4GH standards and frameworks



Phenopackets v2

Phenopackets is a standard schema for sharing phenotypic information.

Approved: June 24, 2021

Example Users

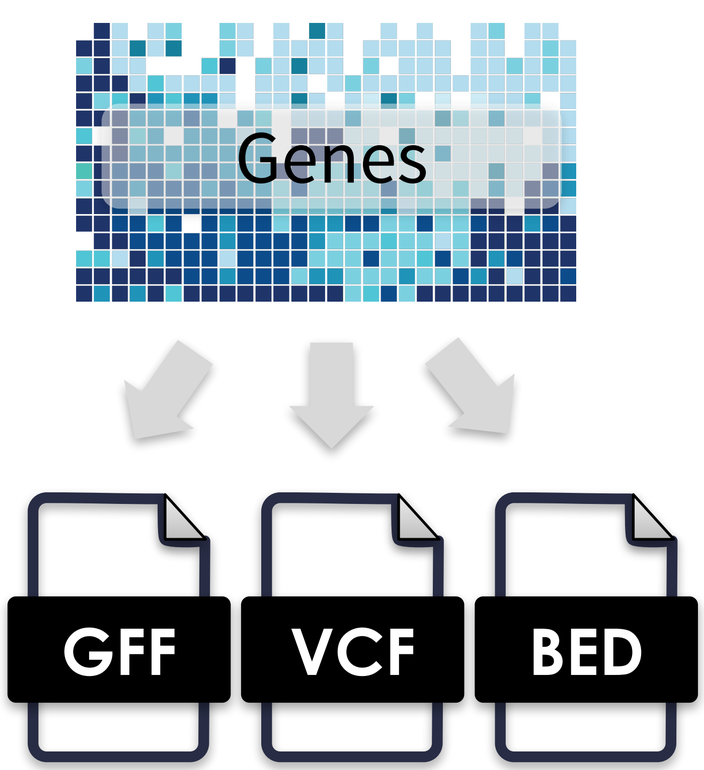
Cafe Variome

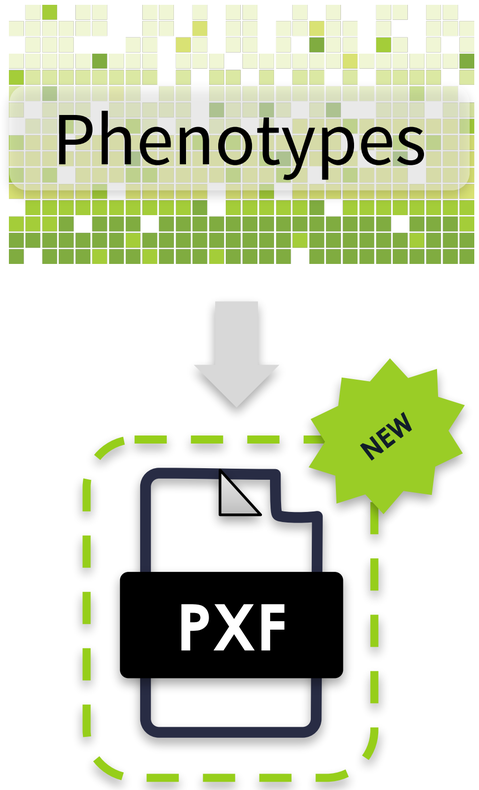
NIH

AMED

PHENOTIPS™

RD Connect





VCF/BCF

The Variant Call Format (VCF) specifies the format of a text file used in bioinformatics for storing gene sequence variations. The Binary Call Format (BCF) is the Binary equivalent, smaller and more efficient to process.

Software Libraries: htslib | htsjdk

Tools: Samtools | BCFtools

Databases: European Variation Archive (EVA) | dbGAP | dbSNP | 1000 Genomes Projects / IGSR

Genome Browsers: ENSEMBL | JBrowse | UCSC Genome Browser

Example Users

All of Us RESEARCH PROGRAM

BROAD INSTITUTE

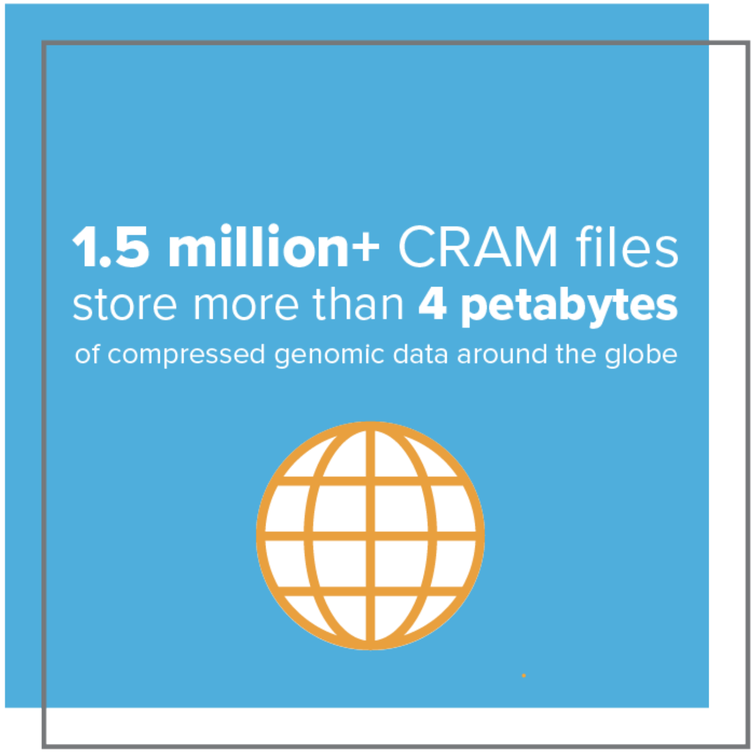
elixir

EMBL-EBI

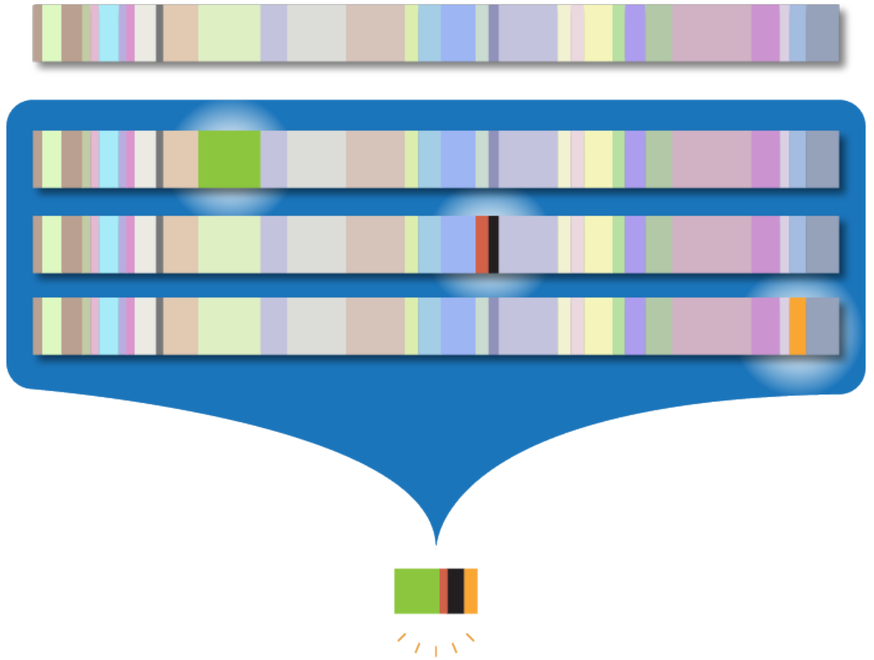
Genomics england

CRAM

CRAM is a file format for storing compressed genomic data. To make files small and efficient, the algorithm compresses information by only storing the parts that are different from the reference human genome.



CRAM compresses data by only storing the difference.



Genomics England implements GA4GH API to provide secure access to genomic data for the NHS

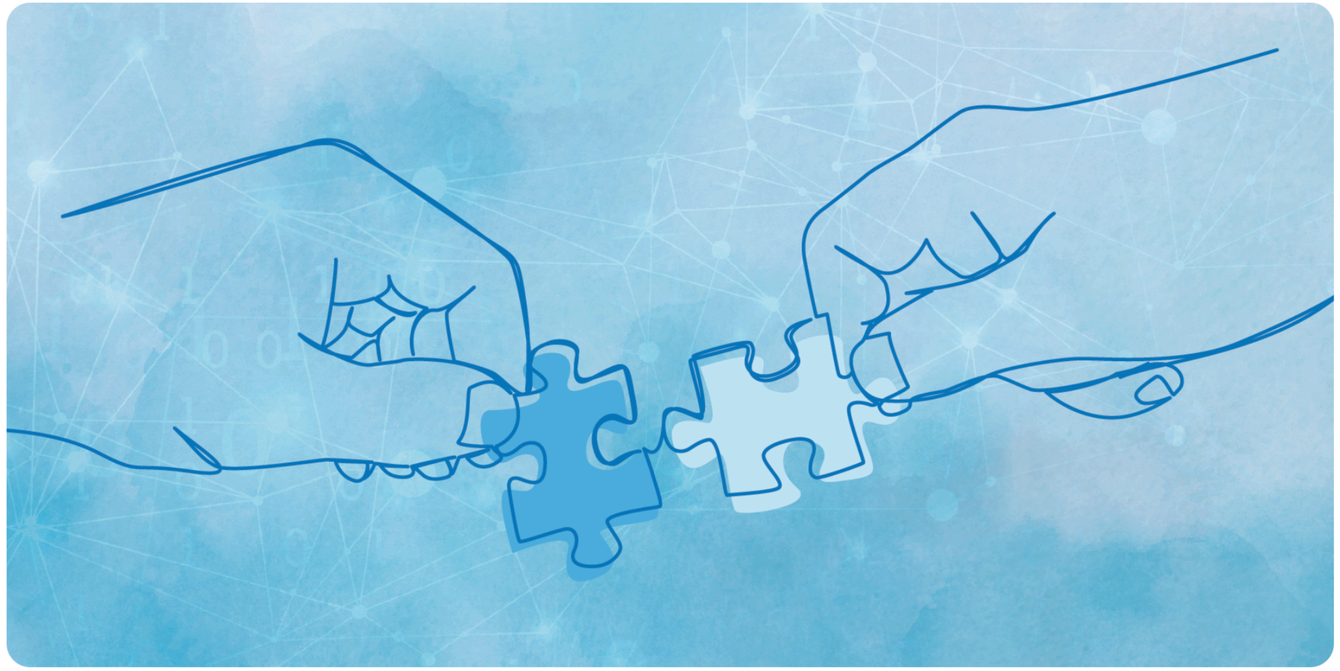
Genomics England has implemented the standard GA4GH API hts
Genomes Program and the Genomic Medicine Service.

News

14 Feb 2024

NIH and GA4GH commit to ongoing collaboration

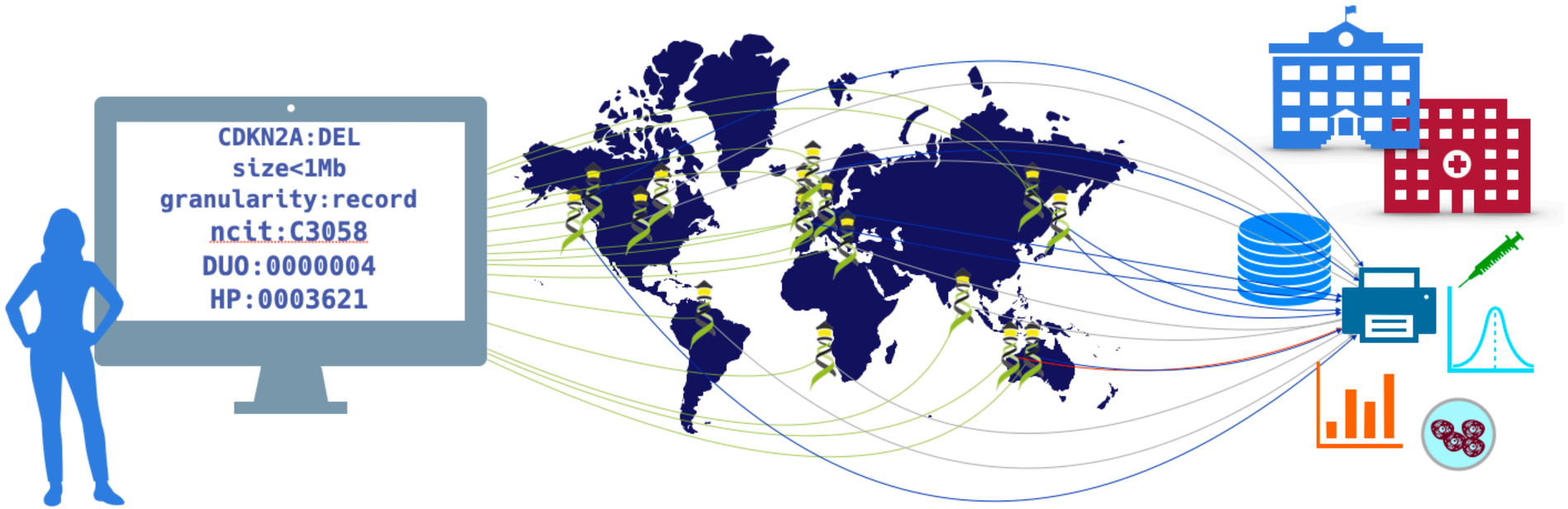
NIH and GA4GH strengthen their partnership to expand responsible data use for the benefit of human health through a Memorandum of Agreement.



The United States National Institutes of Health (NIH) Office of Data Science Strategy (ODSS) and the Global Alliance for Genomics and Health (GA4GH) have announced a strategic collaboration in the form of a Memorandum of Agreement. This partnership aims to bolster the development of technology standards, tools, and policy frameworks to support responsible sharing of genomic and related health data on a global scale.

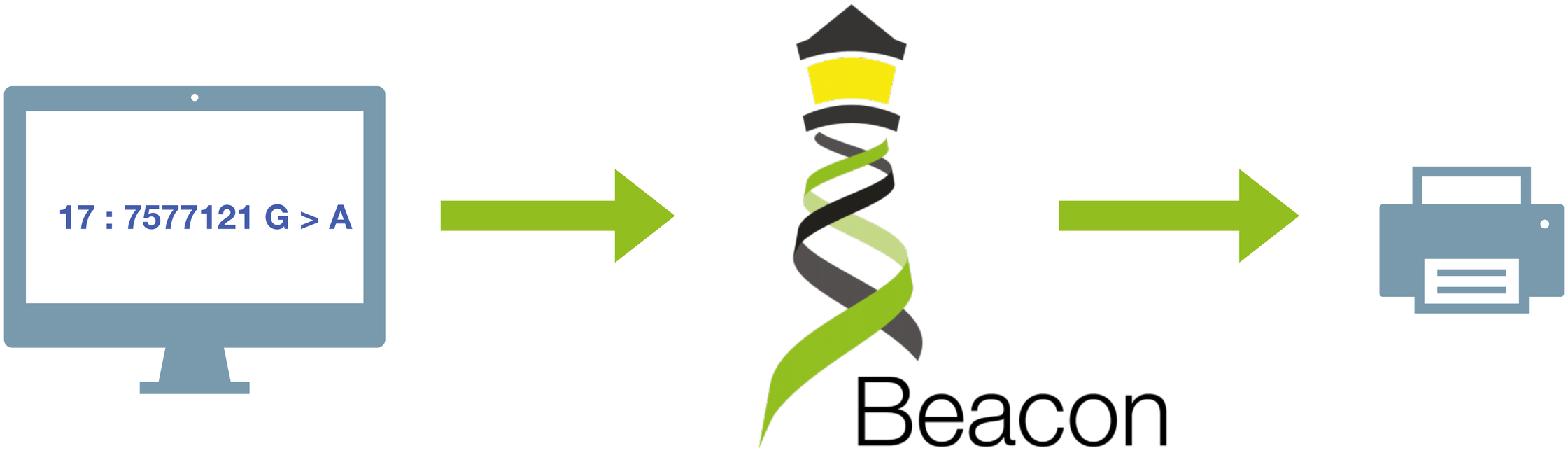
Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.



The GA4GH Beacon Protocol

Federating Genomic Discoveries



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**.

Global Alliance “Beacon” - Jim Ostell, NCBI, March 7, 2014

Global Alliance for Genomics & Health

Introduction

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

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

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

Utility

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

A number of more useful first versions have been suggested.

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

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

Implementation

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

Beacon 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 on 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

- new Beacon website (March)

2019

- ELIXIR Beacon Network

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

- Beacon publication at Nature Biotechnology

2020

- "Scouts" teams working on different aspects - filters, genomic variants, compliance ...
- discussions w/ clinical stakeholders

2021

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

- Phenopackets v2 approved

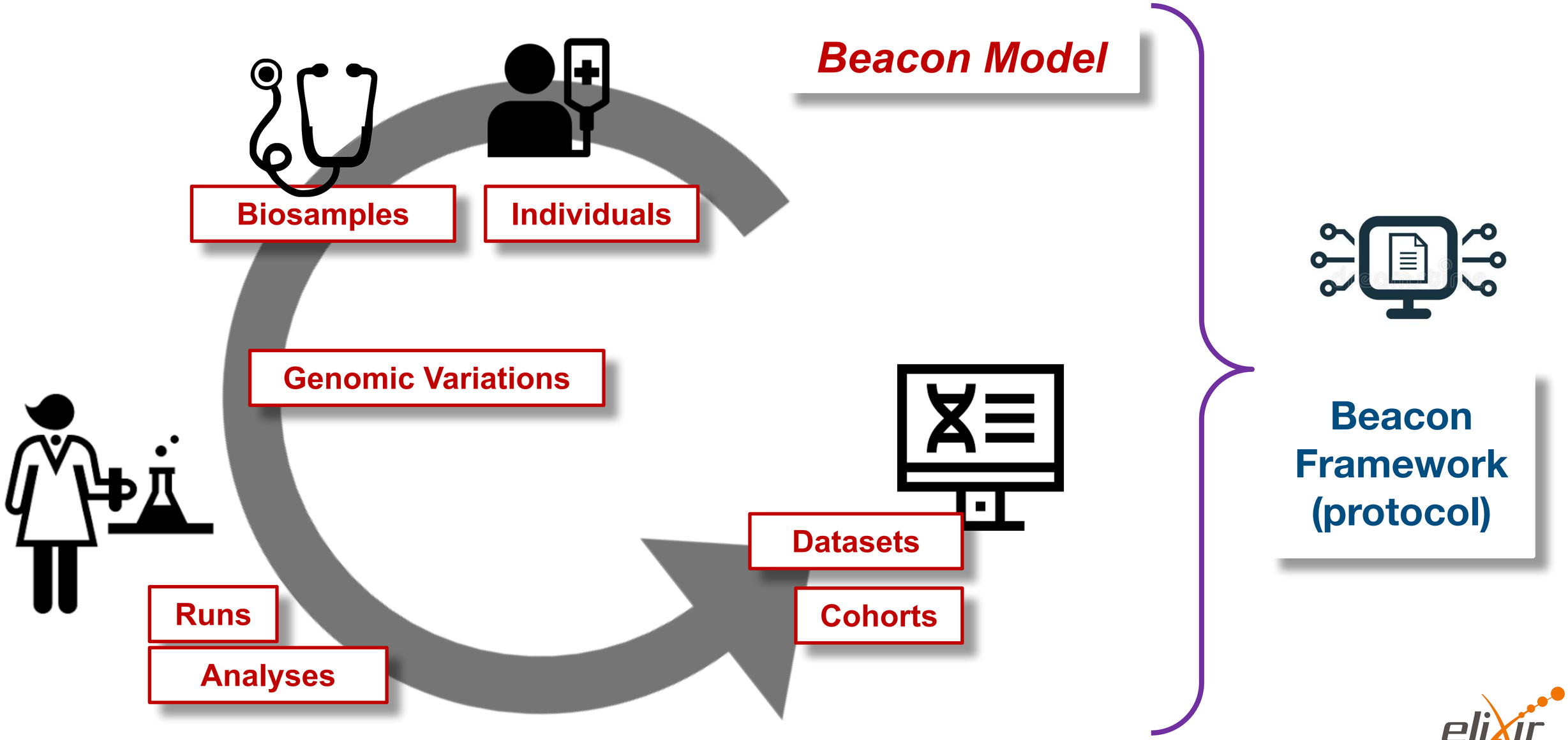
2022

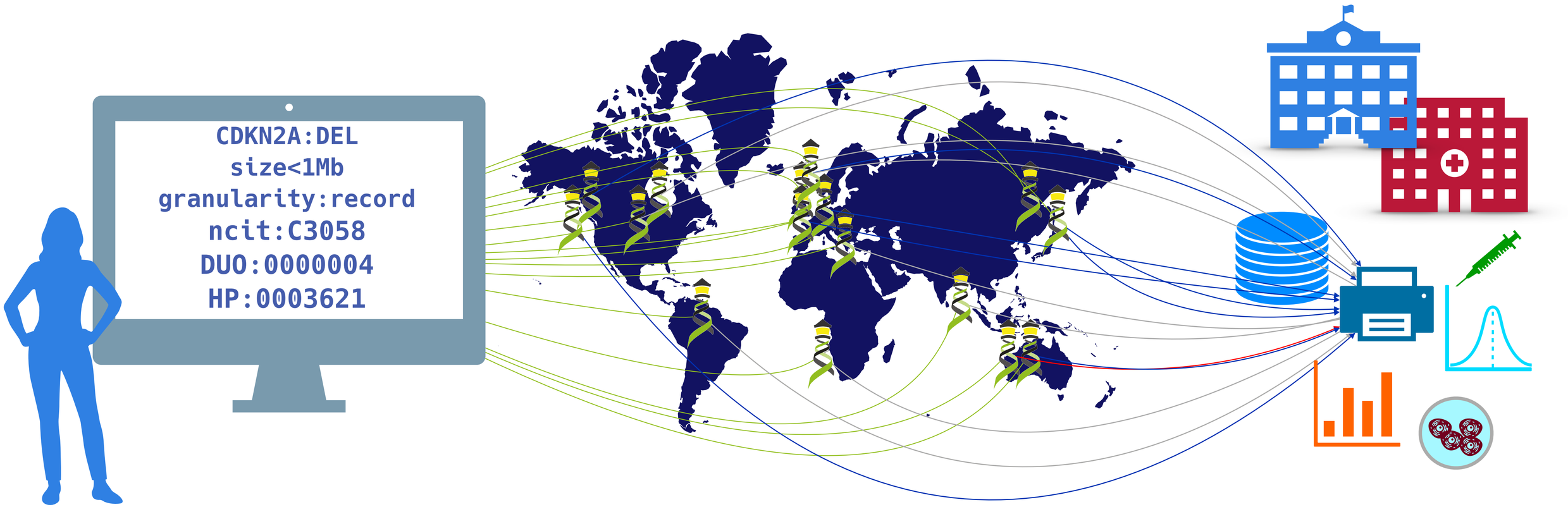
- further changes esp. in default model, aligning with Phenopackets and VRS
- unified beacon-v2 code & docs repository
- Beacon v2 approved at Apr GA4GH Connect

- *docs.genomebeacons.org*

Beacon v2

docs.genomebeacons.org





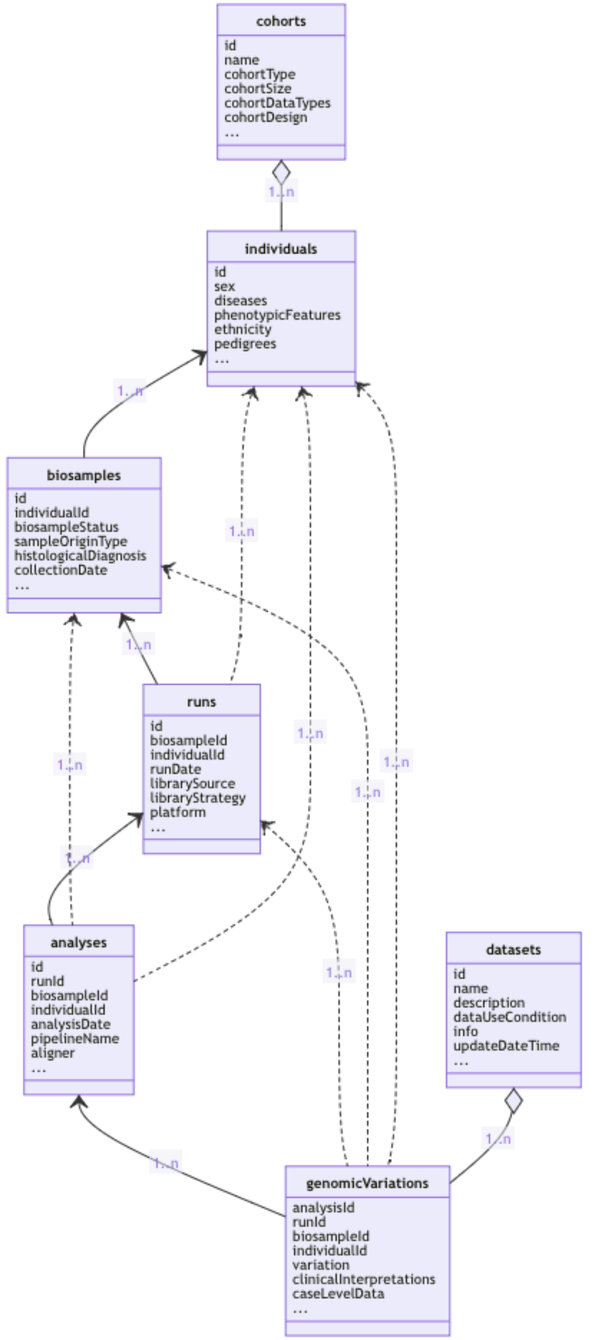
Can you provide data about focal deletions in CDKN2A in Glioblastomas from juvenile patients with unrestricted access?

Beacon *v2* API

The Beacon API v2 represents a simple but powerful **genomics API** for ***federated*** data discovery and retrieval

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

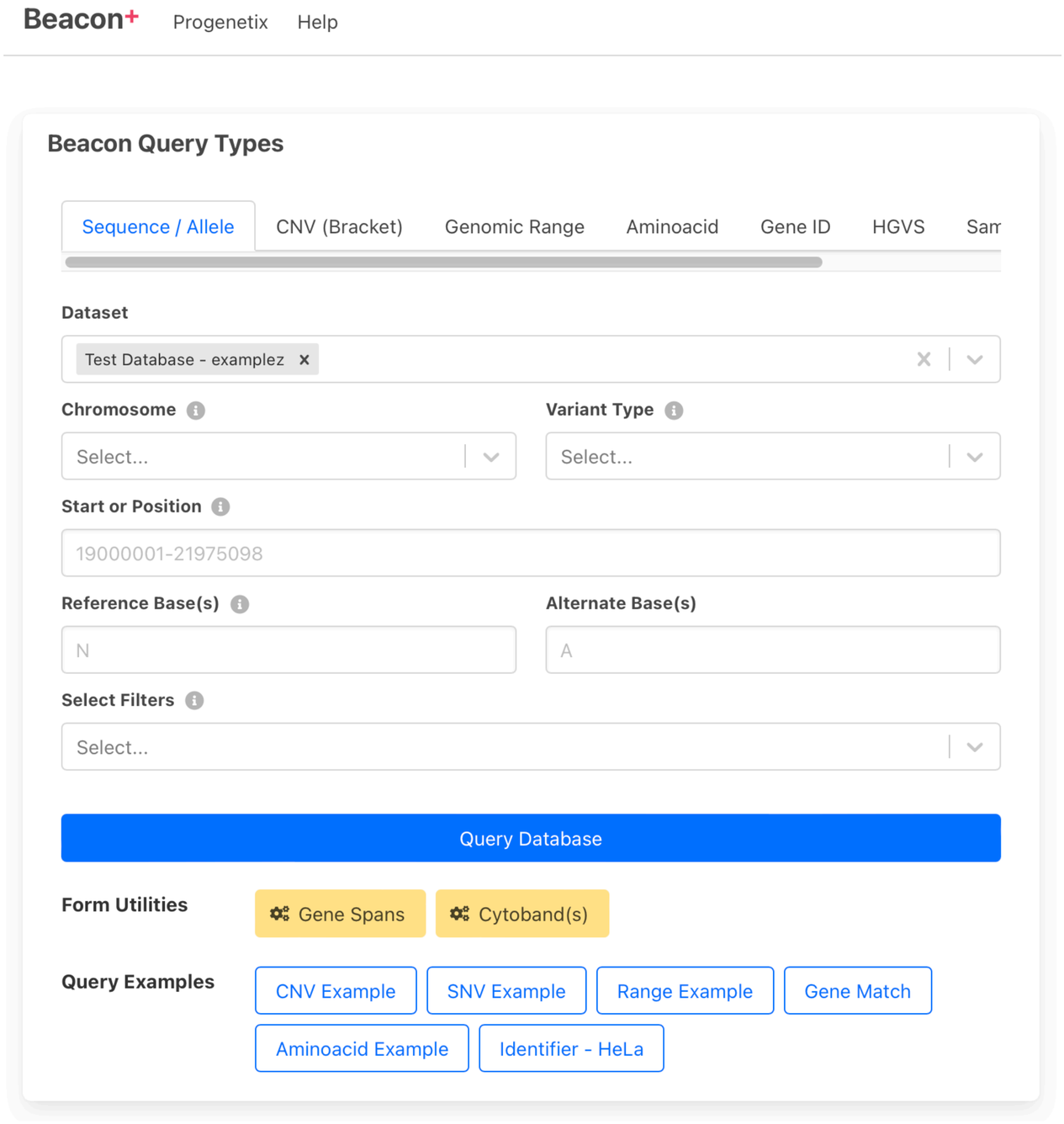


Beacon Queries

Implementation of Current Options

- (so far) the Beacon model does not define explicit query types
- disambiguation of parameters is left to implementers
- implicit query types:
  - ➡ allele/sequence query
  - ➡ range query, w/ or w/o additional parameters
  - ➡ bracket query (e.g. sized CNVs)
  - ➡ aminoacid, HGVS, gene

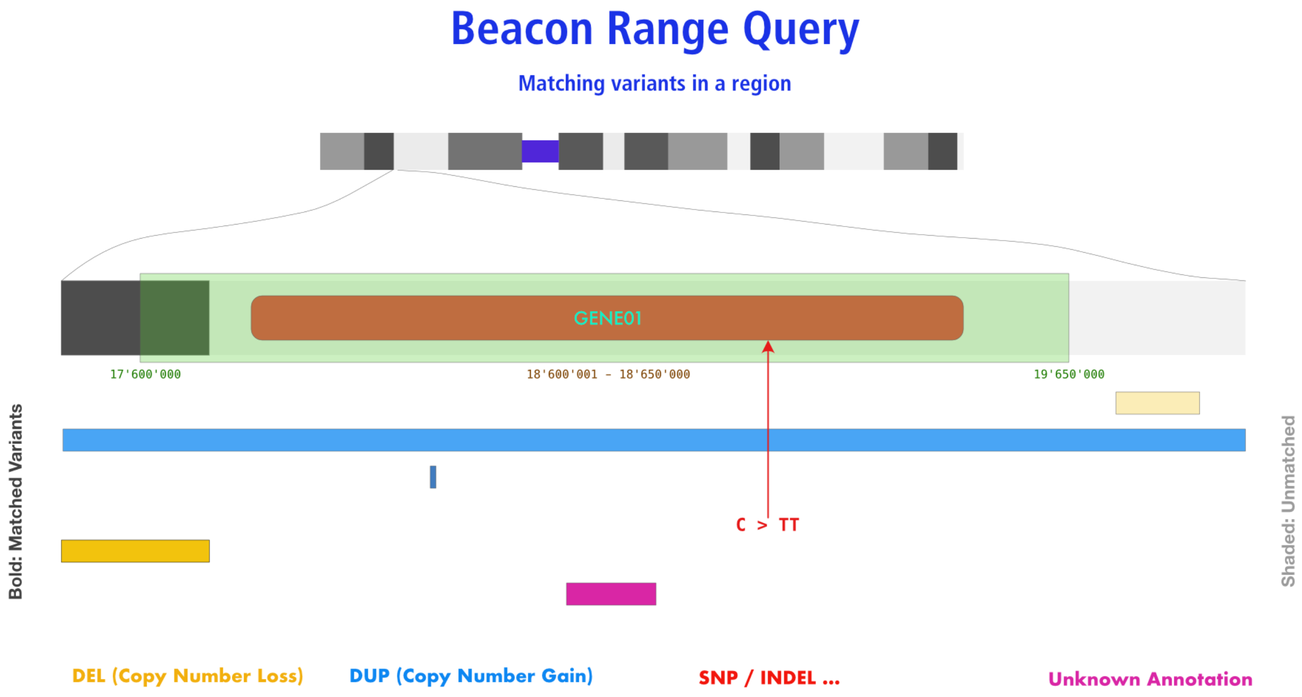
beaconplus.progenetix.org

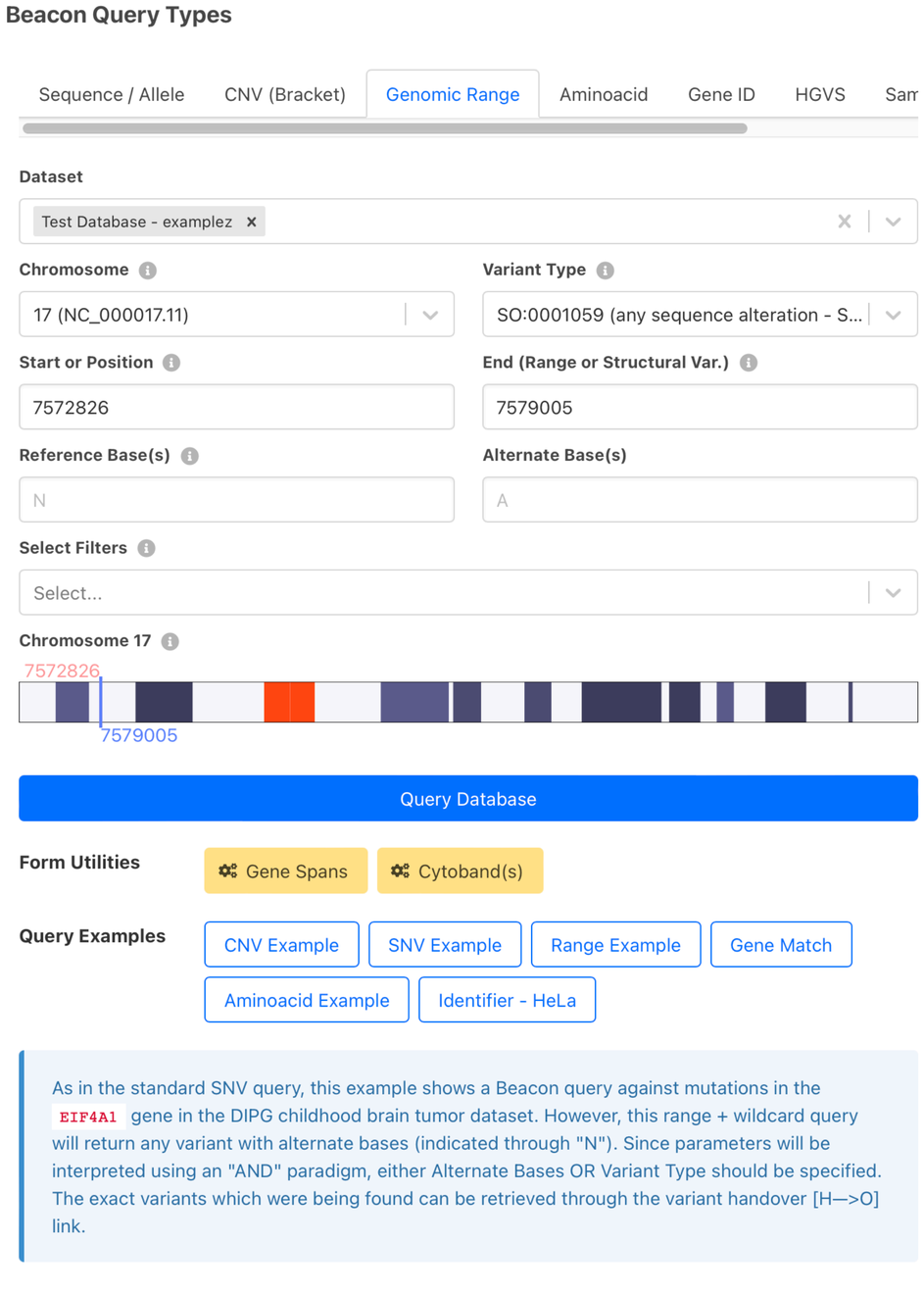


Beacon Queries

Range ("anything goes") Request

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

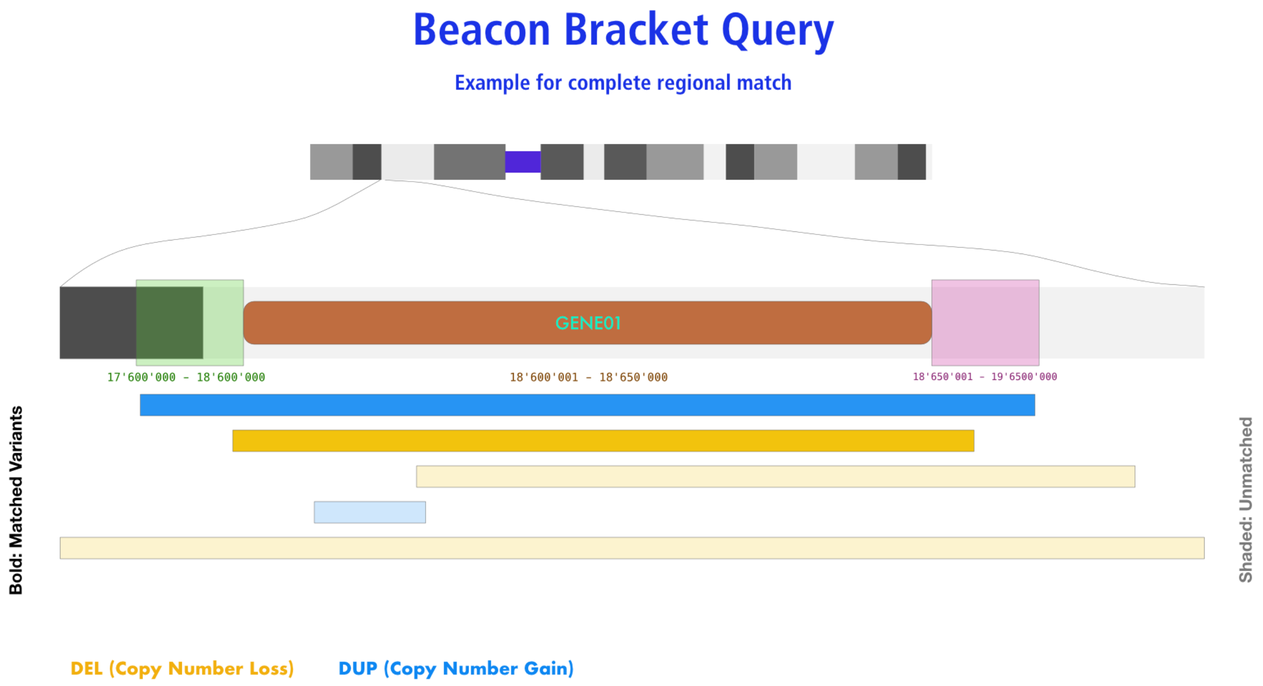


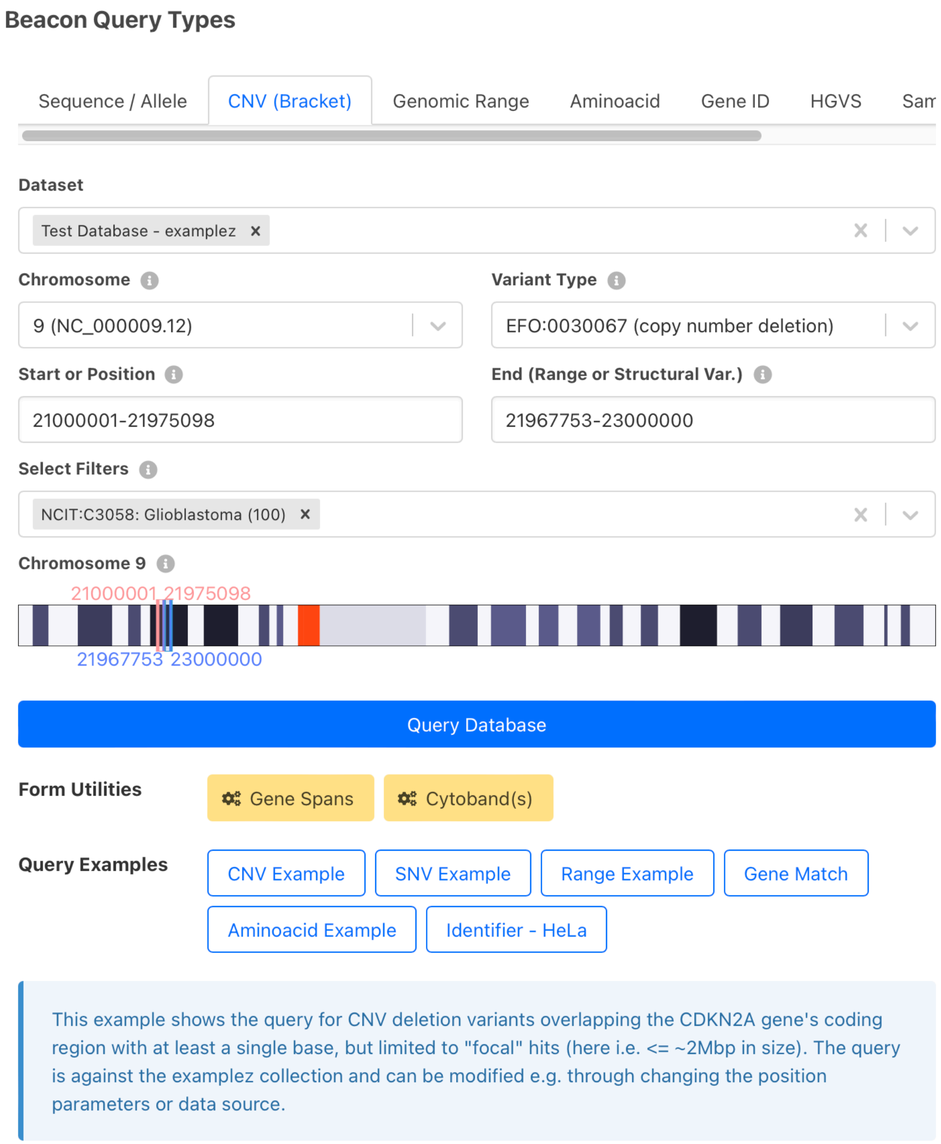


Beacon Queries

Bracket ("CNV") Query

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





Beacon v2 Filters

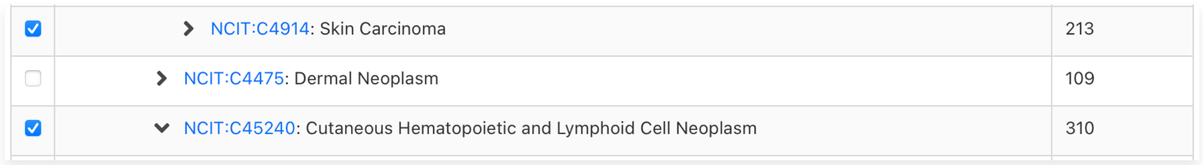
Example: Use of hierarchical classification systems (here NCIt neoplasm core)

- Beacon v2 "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



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
Calls: 0
Samples: 523

$f_{alleles}$: 0

Callsets Variants | Legacy Interface | UCSC region

Show JSON Response

Results | Biosamples

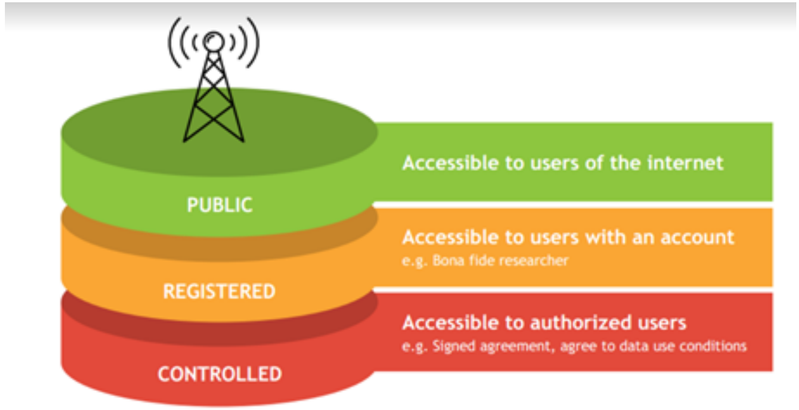
Id	Description	Classifications	Identifiers	DEL	DUP	CNV
PGX_AM_BS_MCC01	Merkel cell carcinoma	icdot-C44.9 Skin, NOS icdom-82473 Merkel cell carcinoma NCIT:C9231 Merkel Cell Carcinoma	PMID:9537255	0.116	0.104	0.22
PGX_AM_BS_MCC02	Merkel cell carcinoma	icdot-C44.9 Skin, NOS icdom-82473 Merkel cell carcinoma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

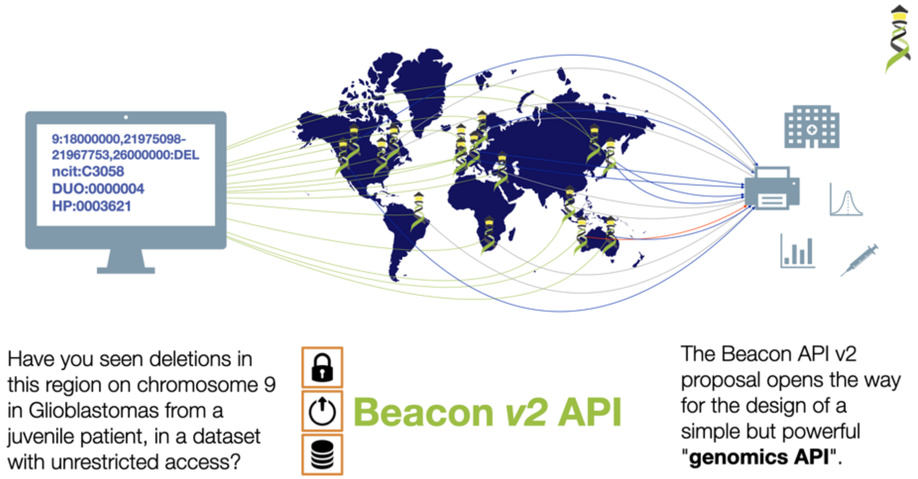
Page **1 of 105**

Beacon API v2

The Beacon API can be implemented as a web-accessible service that users may query for information about a specific allele.

Approved: April 21, 2022





Example Users

EMBL-EBI

Australian Genomics

SciLifeLab

elixir

UNIVERSITY OF CALIFORNIA SANTA CRUZ

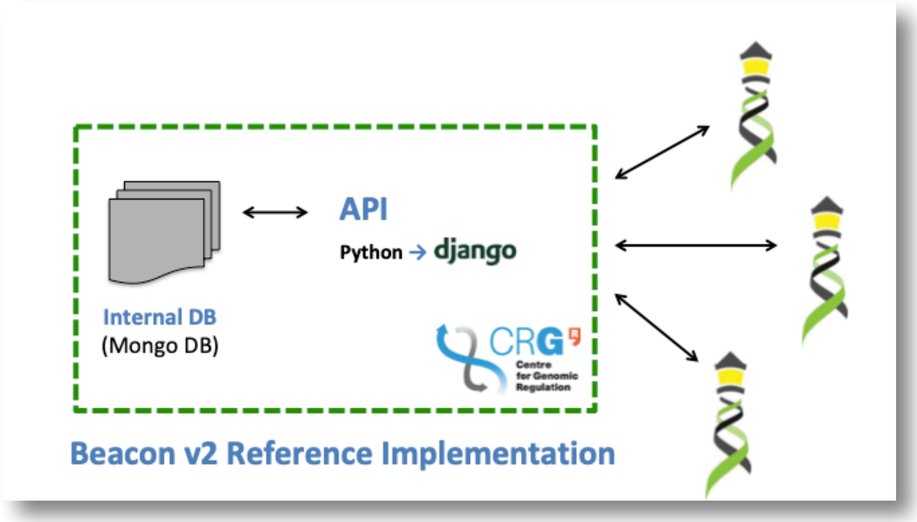
BROAD INSTITUTE

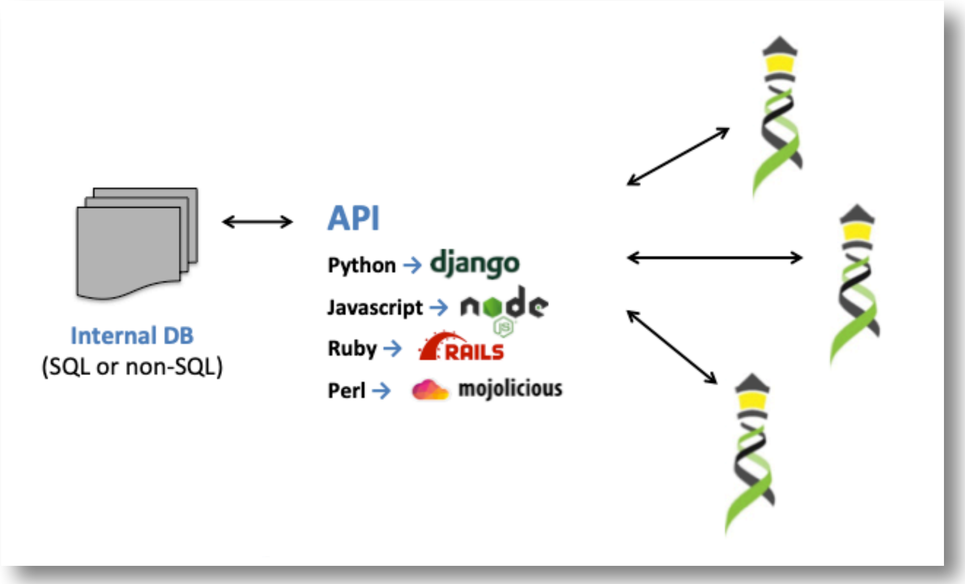
EUROPEAN GENOME-PHENOME ARCHIVE

International Cancer Genome Consortium

Implementing Beacon v2

... its just code ¯_(ツ)_/¯





Progenetix Stack

progenetix

- JavaScript front-end is populated for query results using asynchronous access to multiple handover objects
  - biosamples and variants tables, CNV histogram, UCSC .bed loader, .pgxseg variant downloads…
- the complete middleware / CGI stack is provided through the ***bycon*** package
  - schemas, query stack, data transformation (e.g. Phenopackets generation)…
- data collections mostly correspond to the main Beacon default model entities
  - no separate *runs* collection; integrated w/ analyses
  - *variants* are stored per observation instance

React

APACHE HTTP SERVER PROJECT

python

mongoDB

- *collations* contain pre-computed data (e.g. CNV frequencies, statistics) and information for all grouping entity instances and correspond to **filter values**
  - PMID:10027410, NCIT:C3222, pgx:cohort-TCGA, pgx:icdom-94703…
- *querybuffer* stores id values of all entities matched by a query and provides the corresponding access handle for **handover** generation


```
_id: ObjectId("6249bb654f8f8d67eb94953b"),
id: '876eed6-5829-4f28-b01d-9759a6f55f14',
source_collection: 'variants',
source_db: 'progenetix',
source_key: '_id',
target_collection: 'variants',
target_count: 567,
target_key: '_id',
target_values: [
  ObjectId("5bab578b727983b2e00ca99e"),
```


variants | analyses | biosamples | individuals

Entity collections

collations | geolocs | genespans | publications | qBuffer

Utility collections

The Beacon team through the ages

EUROPEAN GENOME-PHENOME ARCHIVE

CRG Centre for Genomic Regulation

Jordi Rambla
Arcadi Navarro
Roberto Ariosa
Manuel Rueda
Lauren Fromont
Mauricio Moldes
Claudia Vasallo
Babita Singh
Sabela de la Torre
Marta Ferri
Fred Haziza

CSC

Juha Törnroos
Teemu Kataja
Ilkka Lappalainen
Dylan Spalding

UNIVERSITY OF LEICESTER

Cafe Variome Central

Tony Brookes
Tim Beck
Colin Veal
Tom Shorter

Swiss Personalized Health Network (SPHN)

University of Zurich[UZH]

Michael Baudis
Rahel Paloots
Hangjia Zhao
Ziying Yang
Bo Gao
Qingyao Huang

Genomics england

Augusto Rendon
Ignacio Medina
Javier López
Jacobo Coll
Antonio Rueda

cnag
centre nacional d'anàlisi genòmica
centro nacional de análisis genómico

Sergi Beltran
Carles Hernandez

Inserm
Institut national de la santé et de la recherche médicale

David Salgado

BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Salvador Capella
Dmitry Repchevski
JM Fernández

DisGeNET

Laura Furlong
Janet Piñero

elixir

B1MG

Serena Scollen
Gary Saunders
Giselle Kerry
David Lloyd

H3Africa
Human Heredity & Health in Africa

Nicola Mulder
Mamana Mbiyavanga
Ziyaad Parker

EU Can CAN.

David Torrents

AUTISM SPEAKS

Dean Hartley

JUNTA DE ANDALUCIA
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

GEM Japan

Toshiaki Katayama

Stephane Dyke

DNASTACK

Marc Fiume
Miro Cupak

BRCA EXCHANGE

Melissa Cline

ENA

EMBL-EBI

Diana Lemos

EUROPEAN JOINT PROGRAMME RARE DISEASES

VICC Variant Interpretation for Cancer Consortium

GA4GH Phenopackets
Peter Robinson
Jules Jacobsen

GA4GH VRS
Alex Wagner
Reece Hart

Beacon PRC
Alex Wagner
Jonathan Dursi
Mamana Mbiyavanga

Alice Mann
Neerjah Skantharajah

elixir

Universität Zürich[UZH]

Global Alliance for Genomics & Health

Progenetix and GA4GH Beacon

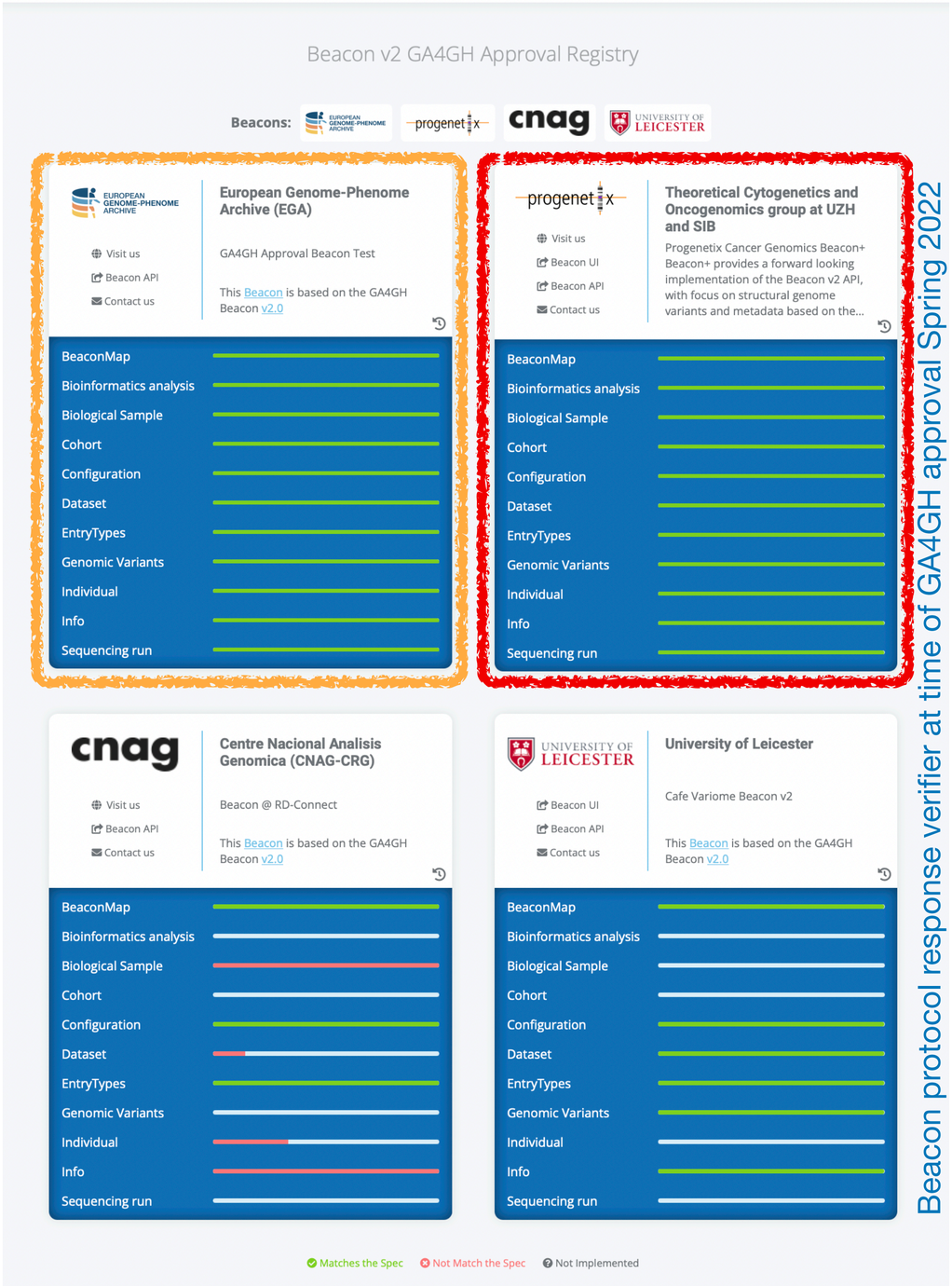
Implementation driven development of a GA4GH standard

progenetix

Progenetix & Beacon

Implementation driven standards development

- Progenetix Beacon+ has served as implementation driver since 2016
- prototyping of advanced Beacon features such as
  - ➡ structural variant queries
  - ➡ data handovers
  - ➡ Phenopackets integration



Global Alliance for Genomics & Health

progenetix

elixir

Progenetix Stack

progenetix

- JavaScript front-end is populated for query results using asynchronous access to multiple handover objects
  - biosamples and variants tables, CNV histogram, UCSC .bed loader, .pgxseg variant downloads…
- the complete middleware / CGI stack is provided through the ***bycon*** package
  - schemas, query stack, data transformation (e.g. Phenopackets generation)…
- data collections mostly correspond to the main Beacon default model entities
  - no separate *runs* collection; integrated w/ analyses
  - *variants* are stored per observation instance

React

APACHE HTTP SERVER PROJECT

python™

mongoDB

- *collations* contain pre-computed data (e.g. CNV frequencies, statistics) and information for all grouping entity instances and correspond to **filter values**
  - PMID:10027410, NCIT:C3222, pgx:cohort-TCGA, pgx:icdom-94703…
- *querybuffer* stores id values of all entities matched by a query and provides the corresponding access handle for **handover** generation

```
_id: ObjectId("6249bb654f8f8d67eb94953b"),
id: '0765ee26-5029-4f28-b01d-9759abf5bf14',
source_collection: 'variants',
source_db: 'progenetix',
source_key: '_id',
target_collection: 'variants',
target_count: 667,
target_key: '_id',
target_values: [
  ObjectId("5bab578b727983b2e00ca99e"),
  ObjectId("5bab578d727983b2e00cb505"),
```

Entity collections

Utility collections

Beacon v2 Conformity and Extensions in *bycon*

Putting the + into Beacon ...

- support & use of standard Beacon v2 PUT & GET variant queries, filters and meta parameters
  - ➡ variant parameters, geneId, lengths, EFO & VCF CNV types, pagination
  - ➡ widespread, self-scoping filter use for bio-, technical- and and id parameters with switch for descending terms use (globally or per term if using POST)
- **extensive use of handovers**
  - ➡ asynchronous delivery of e.g. variant and sample data, data plots
- + optional use of OR logic for filter combinations (global)
- + extension of query parameters
  - ➡ **geographic queries** incl. $geonear and use of GeoJSON in schemas
- ┐(ﾟ▽ﾟ)┌ no implementation of authentication on this open dataset

bycon **provides a number of additional services and output formats which are initiated over the / services path or provided as request parameters and are not considered Beacon extensions (though they follow the syntax where possible).**

progenetix

Beacon+: Phenopackets

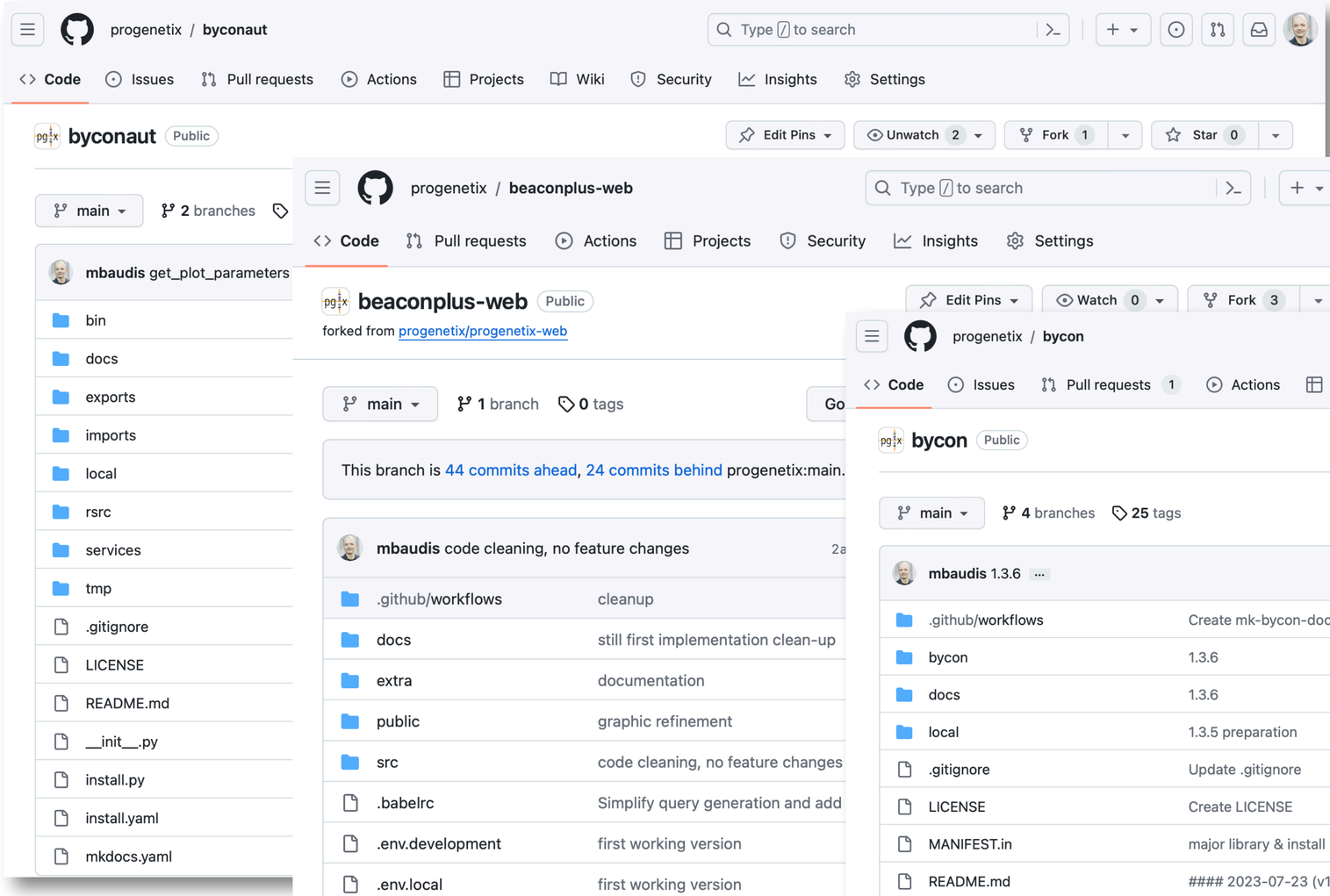
Testing alternative response schemas...

http://progenetix.org/beacon/biosamples/pgxbs-kftvhyvb/phenopackets

- 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+ this is done through *ad hoc* handover URIs

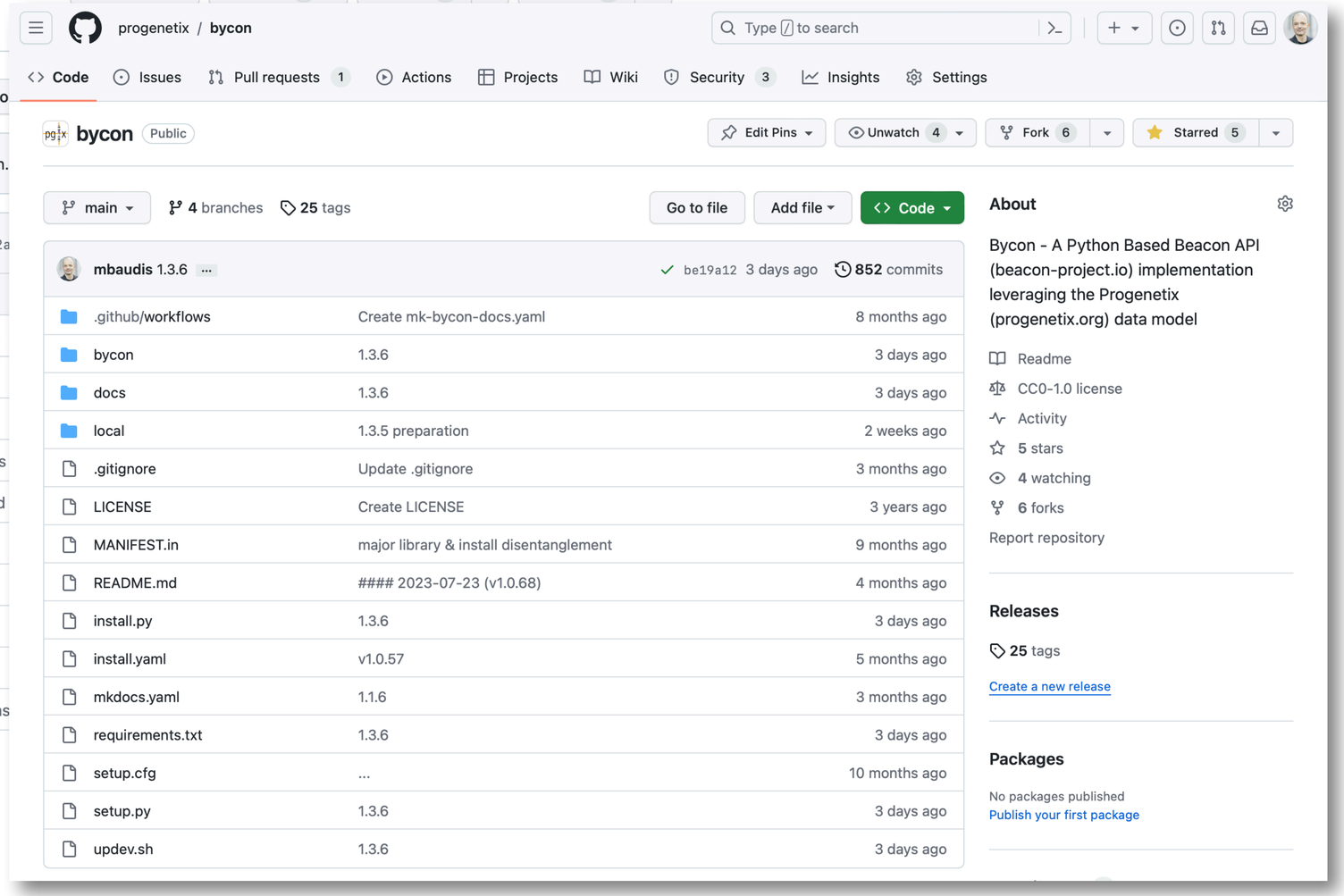

```
"id": "pgxpxf-kftx3tl5",
"metaData": {
  "phenopacketSchemaVersion": "v2",
  "resources": [
    {
      "id": "NCIT",
      "iriPrefix": "http://purl.obolibrary.org/obo/NCIT_",
      "name": "NCIt Plus Neoplasm Core",
      "namespacePrefix": "NCIT",
      "url": "http://purl.obolibrary.org/obo/ncit/neoplasm-core.c
      "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/pgxbs-kftvhyvb/variants/?output=pgxseg"
      }
    ],
    "histologicalDiagnosis": {
      "id": "NCIT:C3099",
      "label": "Hepatocellular Carcinoma"
    },
    "id": "pgxbs-kftvhyvb",
    "individualId": "pgxind-kftx3tl5",
    "pathologicalStage": {
      "id": "NCIT:C27966",
      "label": "Stage I"
    },
    "sampledTissue": {
      "id": "UBERON:0002107",
      "label": "liver"
    },
    "timeOfCollection": {
      "age": "P48Y9M26D"
    },
```



bycon.progenetix.org
github.com/progenetix/bycon/

beaconplus.progenetix.org
.../progenetix/beaconplus-web/



bycon.progenetix.org
github.com/progenetix/bycon/

pgxRpi

An interface API for analyzing Progenetix CNV data in R using the Beacon+ API

GitHub: https://github.com/progenetix/pgxRpi

Bioconductor

README.md

pgxRpi

Welcome to our R wrapper package for Progenetix REST API that leverages the capabilities of Beacon v2 specification. Please note that a stable internet connection is required for the query functionality. This package is aimed to simplify the process of accessing oncogenomic data from Progenetix database.

You can install this package from GitHub using:

```
install.packages("devtools")
devtools::install_github("progenetix/pgxRpi")
```

For accessing metadata of biosamples/individuals, or learning more about filters, get started from the vignette Introduction_1_loadmetadata.

For accessing CNV variant data, get started from this vignette Introduction_2_loadvariants.

For accessing CNV frequency data, get started from this vignette Introduction_3_loadfrequency.

For processing local pgxseg files, get started from this vignette Introduction_4_process_pgxseg.

If you encounter problems, try to reinstall the latest version. If reinstallation doesn't help, please contact us.

pgxRpi

platforms all | rank 2218 / 2221 | support 0 / 0 | in Bioc devel only
build ok | updated < 1 month | dependencies 144

DOI: 10.18129/B9.bioc.pgxRpi
This is the **development** version of pgxRpi; to use it, please install the devel version of Bioconductor.

R wrapper for Progenetix

Bioconductor version: Development (3.19)

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]

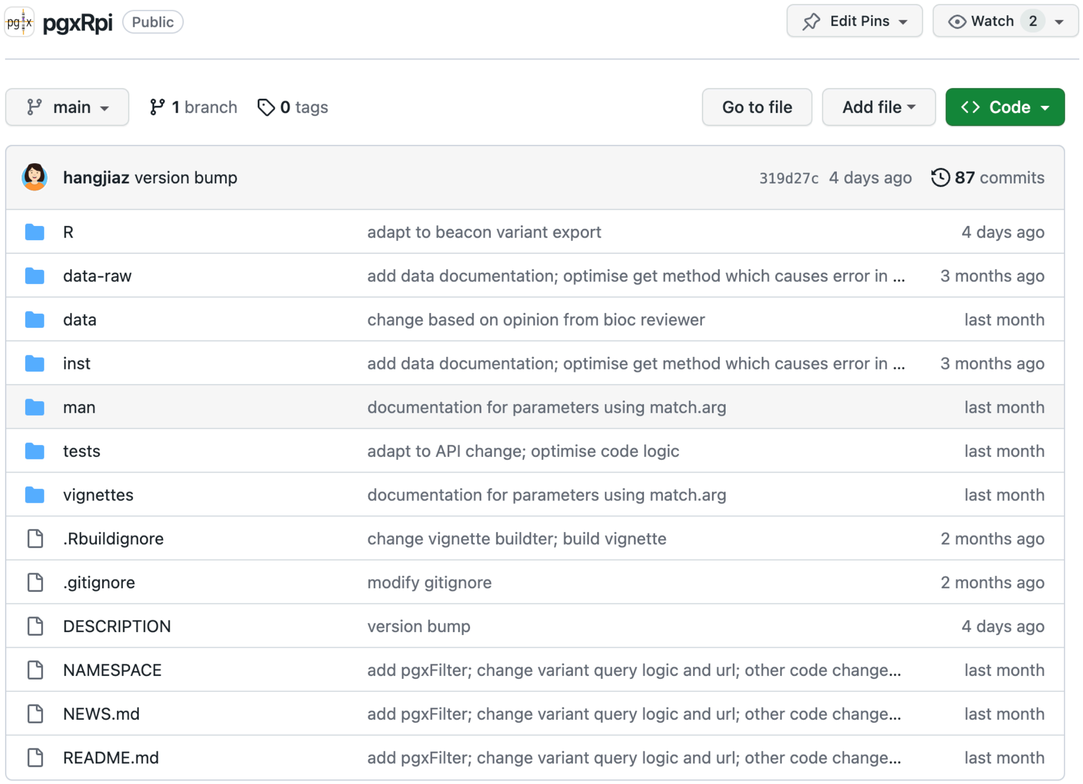
Maintainer: Hangjia Zhao <hangjia.zhao at uzh.ch>

Citation (from within R, enter `citation("pgxRpi")`):

Zhao H, Baudis M (2023). *pgxRpi: R wrapper for Progenetix*. doi:10.18129/B9.bioc.pgxRpi, R package version 0.99.9, https://bioconductor.org/packages/pgxRpi.

pgxRpi

An interface API for analyzing Progenetix CNV data in R using the Beacon+ API



pgxRpi Public

Edit Pins | Watch 2

main | 1 branch | 0 tags | Go to file | Add file | Code

hangjiaz version bump 319d27c 4 days ago 87 commits

R	adapt to beacon variant export	4 days ago
data-raw	add data documentation; optimise get method which causes error in ...	3 months ago
data	change based on opinion from bioc reviewer	last month
inst	add data documentation; optimise get method which causes error in ...	3 months ago
man	documentation for parameters using match.arg	last month
tests	adapt to API change; optimise code logic	last month
vignettes	documentation for parameters using match.arg	last month
.Rbuildignore	change vignette buildter; build vignette	2 months ago
.gitignore	modify gitignore	2 months ago
DESCRIPTION	version bump	4 days ago
NAMESPACE	add pgxFilter; change variant query logic and url; other code change...	last month
NEWS.md	add pgxFilter; change variant query logic and url; other code change...	last month
README.md	add pgxFilter; change variant query logic and url; other code change...	last month

2 Retrieve meatdata of samples

2.1 Relevant parameters

type, filters, filterLogic, individual_id, biosample_id, codematches, limit, skip

2.2 Search by filters

Filters are a significant enhancement to the Beacon query API, providing a mechanism for specifying rules to select records based on their field values. To learn more about how to utilize filters in Progenetix, please refer to the documentation.

The `pgxFilter` function helps access available filters used in Progenetix. Here is the example use:

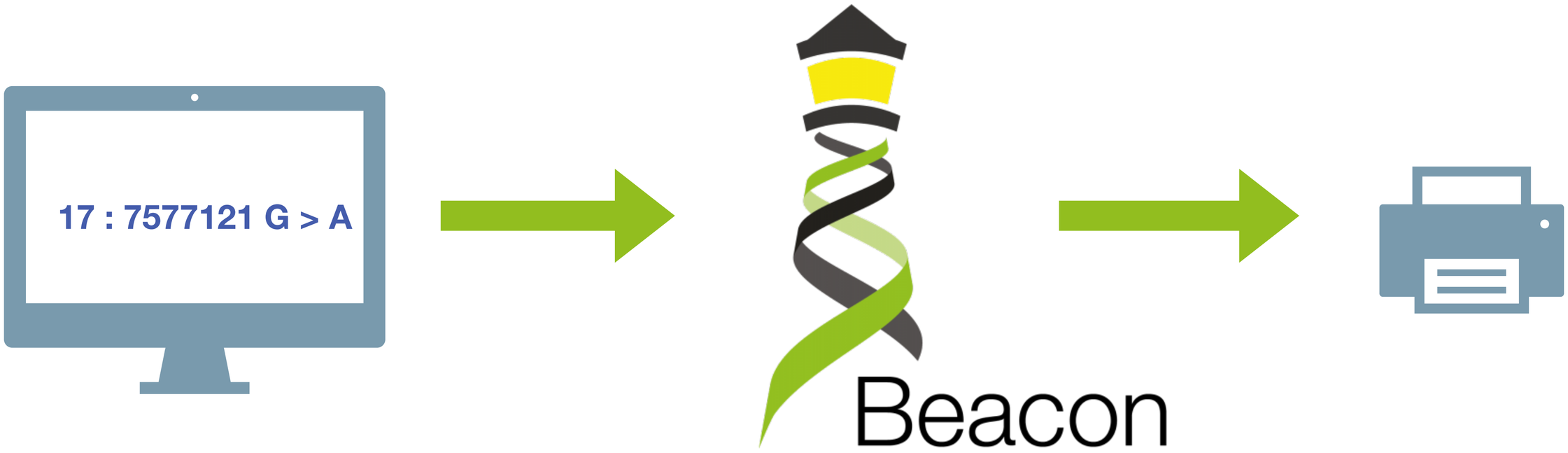
```
# access all filters
all_filters <- pgxFilter()
# get all prefix
all_prefix <- pgxFilter(return_all_prefix = TRUE)
# access specific filters based on prefix
ncit_filters <- pgxFilter(prefix="NCIT")
head(ncit_filters)
#> [1] "NCIT:C28076" "NCIT:C18000" "NCIT:C14158" "NCIT:C14161" "NCIT:C28077"
#> [6] "NCIT:C28078"
```

The following query is designed to retrieve metadata in Progenetix related to all samples of lung adenocarcinoma, utilizing a specific type of filter based on an NCIt code as an ontology identifier.

```
biosamples <- pgxLoader(type="biosample", filters = "NCIT:C3512")
# data looks like this
biosamples[c(1700:1705),]
#>       biosample_id group_id group_label   individual_id    callset_ids
#> 1700 pgxbs-kftvjjhx       NA          NA pgxind-kftx5fyd pgxcs-kftwjevi
#> 1701 pgxbs-kftvjjhz       NA          NA pgxind-kftx5fyf pgxcs-kftwjew0
#> 1702 pgxbs-kftvjji1       NA          NA pgxind-kftx5fyh pgxcs-kftwjewi
#> 1703 pgxbs-kftvjjn2       NA          NA pgxind-kftx5g4r pgxcs-kftwjg5r
#> 1704 pgxbs-kftvjjn4       NA          NA pgxind-kftx5g4t pgxcs-kftwjg6q
#> 1705 pgxbs-kftvjjn5       NA          NA pgxind-kftx5g4v pgxcs-kftwjg78
```

Genomic Data & Privacy

How about Beacon?



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

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

Genome *Beacons* Compromise Security?

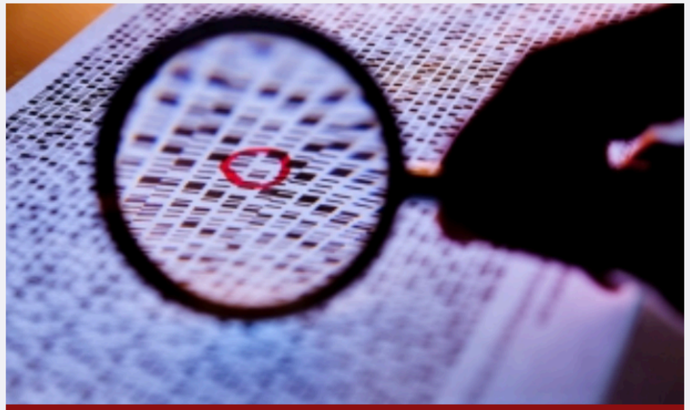
Querying for thousands of specific SNV occurrences in a genomic data pool can identify individuals in an anonymized genomic data collection

Stanford researchers identify potential security hole in genomic data-sharing network

Hackers with access to a person's genome might find out if that genome is in an international network of disease databases.

OCT 29
2015

Sharing genomic information among researchers is critical to the advance of biomedical research. Yet genomic data contains identifiable information and, in the wrong hands, poses a risk to individual privacy. If someone had access to your genome sequence — either directly from your saliva or other tissues, or from a popular genomic information service — they could check to see if you appear in a database of people with certain medical conditions, such as heart disease, lung cancer or autism.



Stanford researchers are working with the Global Alliance for Genomics and Health to make genomic information in the Beacon Project more secure.
Science photo/Shutterstock

Work by a pair of researchers at the Stanford University School of Medicine makes that genomic data more secure. Suyash Shringarpure, PhD, a postdoctoral scholar in genetics, and Carlos Bustamante, PhD, a professor of genetics, have demonstrated a technique for hacking a network of global genomic databases and how to prevent it. They are working with investigators from the Global Alliance for Genomics and Health on implementing preventive measures.

The work, published Oct. 29 in *The American Journal of Human Genetics,* also bears importantly on the larger question of how to analyze mixtures of genomes, such as those from different people at a crime scene.

IDENTIFICATION OF INDIVIDUALS FROM MIXED COLLECTIONS USING RARE ALLELES

The American Journal of Human Genetics 97, 631–646, November 5, 2015

Privacy Risks from Genomic Data-Sharing Beacons

Suyash S. Shringarpure[1,*] and Carlos D. Bustamante[1,*]

The human genetics community needs robust protocols that enable secure sharing of genomic data from participants in genetic research. Beacons are web servers that answer allele-presence queries—such as "Do you have a genome that has a specific nucleotide (e.g., A) at a specific genomic position (e.g., position 11,272 on chromosome 1)?"—with either "yes" or "no." Here, we show that individuals in a beacon are susceptible to re-identification even if the only data shared include presence or absence information about alleles in a beacon. Specifically, we propose a likelihood-ratio test of whether a given individual is present in a given genetic beacon. Our test is not dependent on allele frequencies and is the most powerful test for a specified false-positive rate. Through simulations, we showed that in a beacon with 1,000 individuals, re-identification is possible with just 5,000 queries. Relatives can also be identified in the beacon. Re-identification is possible even in the presence of sequencing errors and variant-calling differences. In a beacon constructed with 65 European individuals from the 1000 Genomes Project, we demonstrated that it is possible to detect membership in the beacon with just 250 SNPs. With just 1,000 SNP queries, we were able to detect the presence of an individual genome from the Personal Genome Project in an existing beacon. Our results show that beacons can disclose membership and implied phenotypic information about participants and do not protect privacy a priori. We discuss risk mitigation through policies and standards such as not allowing anonymous pings of genetic beacons and requiring minimum beacon sizes.

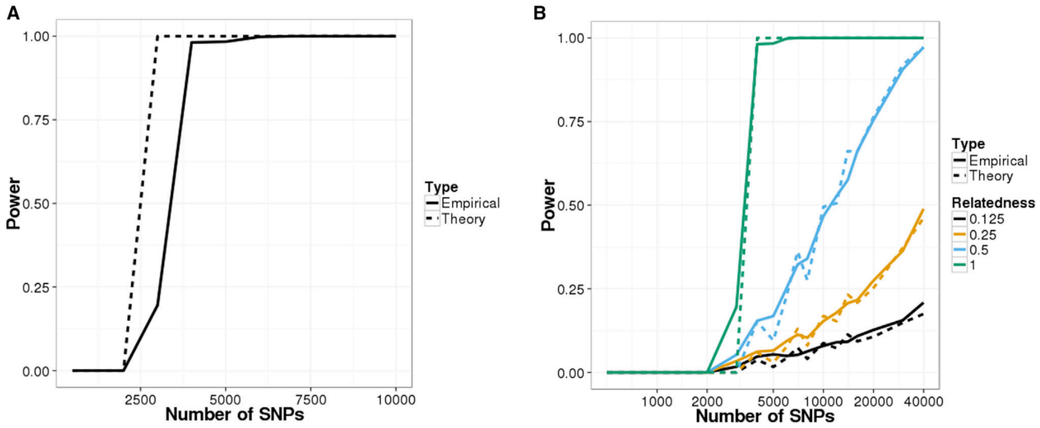


Figure 1. Power of Re-identification Attacks on Beacons Constructed with Simulated Data
Power curves for the likelihood-ratio test (LRT) on (A) a simulated beacon with 1,000 individuals and (B) detecting relatives in the simulated beacon. The false-positive rate was set to 0.05 for all scenarios.

- rare allelic variants can be used to identify an individual (or her relatives) in a genome collection without having access to individual datasets
- however, such an approach requires previous knowledge about the individual's SNPs

Making Beacons Biomedical - Beacon v2

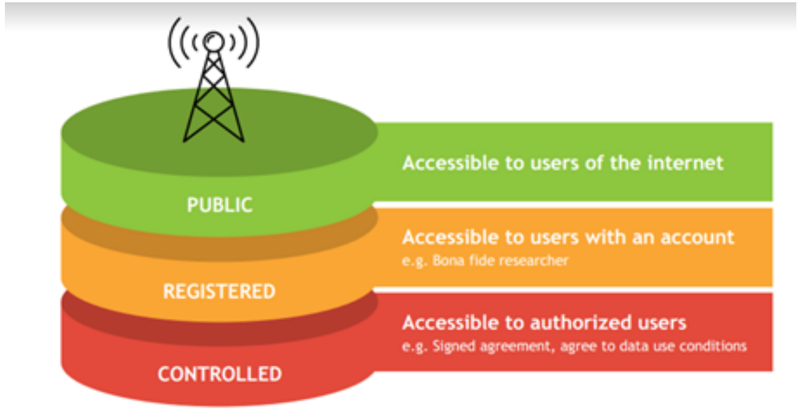
- Scoping queries through "biodata" parameters
- Extending the queries towards clinically ubiquitous variant formats
  - cytogenetic annotations, named variants, variant effects
- Beacon queries as entry for **data delivery**
  - Beacon v2 permissive to respond with variety of data types
    - Phenopackets, biosample data, cohort information ...
  - handover to stream and download using htsget, VCF, EHRs
- Interacting with EHR standards
  - FHIR translations for queries and handover ...
- Beacons as part of local, secure environments
- Authentication to enable non-aggregate, patient derived datasets
  - ELIXIR AAI with compatibility to other providers (OAuth...)

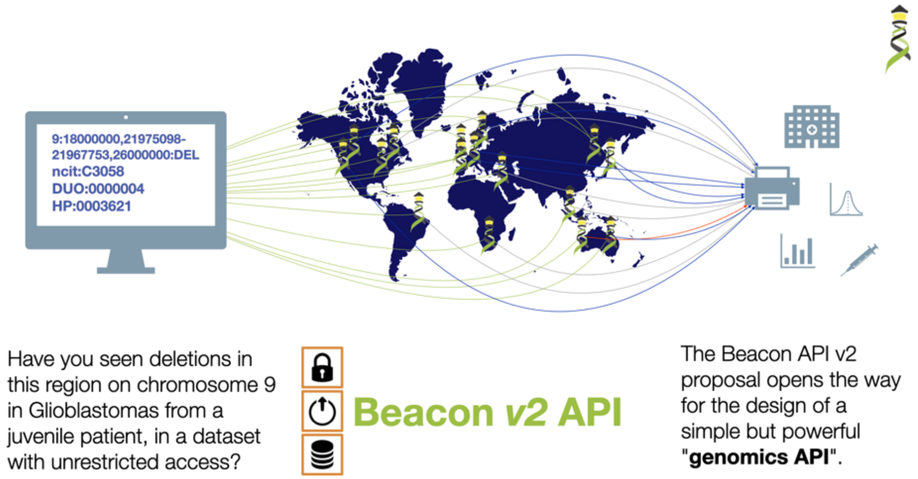
Definitely breaks the "Relative Security by Design" Concept!

Beacon API v2

The Beacon API can be implemented as a web-accessible service that users may query for information about a specific allele.

Approved: April 21, 2022





Example Users

EMBL-EBI

Australian Genomics

SciLifeLab

elixir

UNIVERSITY OF CALIFORNIA SANTA CRUZ

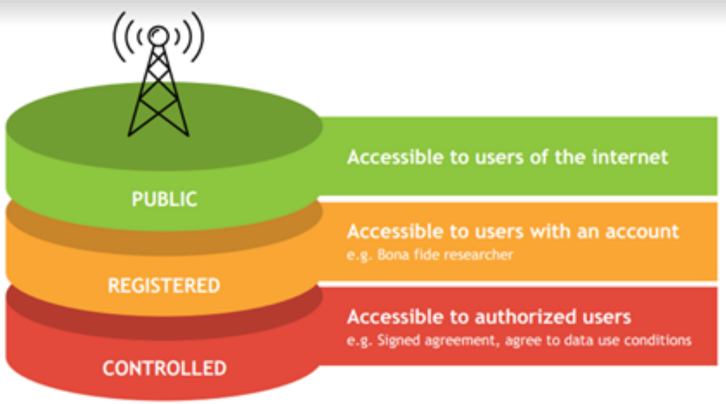
BROAD INSTITUTE

EUROPEAN GENOME-PHENOME ARCHIVE

International Cancer Genome Consortium

Beacon Security

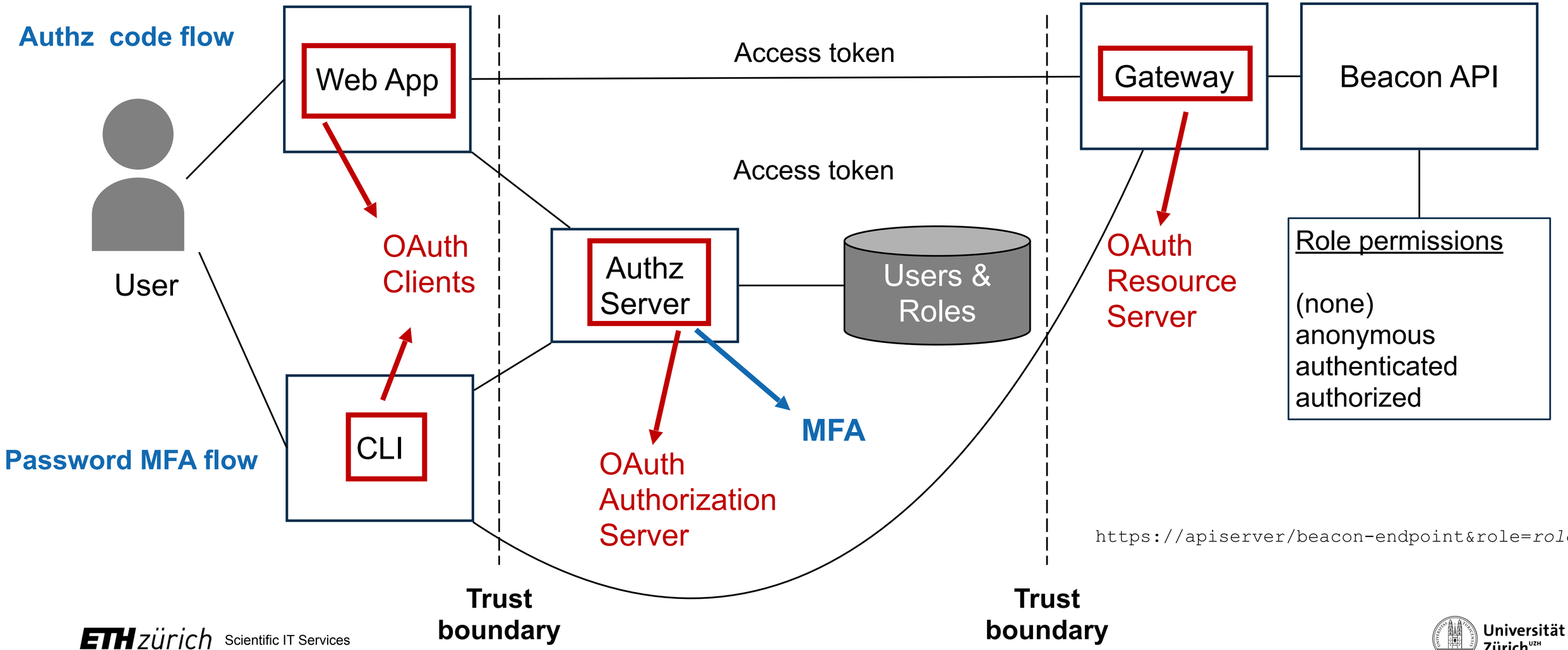
Security by Design ... if Implemented in the Environment



- the beacon API specification does not implement explicit security (e.g. checking user authentication and authorization)
- the framework implements different levels of response granularity which can be mapped to authorization levels (**boolean** / **count** / **record** level responses)
- implementations can have beacons running in secure environments with a **gatekeeper** service managing authentication and autorization levels, and potentially can filter responses for escalated levels
- the backend can implement additional access reduction, on a user <-> dataset level if needed

Architecture

Running the *bycon* stack in a secure environment



Architecture

Running the *bycon* stack in a secure environment

- The ***Beacon API*** implementation stack (e.g. bycon) is authentication procedure agnostic; i.e. it just accepts that a user has been authenticated and passed the general authorization gatekeeping
- The ***Beacon API*** server and the ***Gateway*** reside in a single VM, with only the ***Gateway***'s port exposed (with TLS). Beacon's port is not exposed by the VM and can only be reached through the ***Gateway***
- The ***Authentication Server*** can run on the same or separate VM; needs a database with user accounts.
- The ***Web Client*** can be in the same VM or a separate one.
- Separate ***Gateways*** (e.g. university firewall vs. public) can be configured to modify different roles, e.g. the public gateway may turn registered roles into anonymous, regardless of whether the user has registered status
- Users can write their own clients (web / command line) which are registered with the ***Authorization Server*** and are issued with a Client ID and Client Secret to use against the ***Authorization Server***.

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

What Can You Do?

- implement procedures and standards supporting **data discovery** (FAIR principles) and federation approaches using Beacon
- promote 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

➡ **Collaborate!**

Events

21 Apr 2024

GA4GH Ascona Connect

REGISTER FOR THIS EVENT

This hybrid working meeting aims to support GA4GH contributors in advancing product development and gathering feedback on needs.

Image summary: Join us for GA4GH Connect from 21 to 24 April 2024.

https://www.ga4gh.org/event/ga4gh-connect-2/

Friday, 19 April 2024

10:00 - 14:15	GA4GH Connect satellite workshop: Beacon implementers			

Sunday, 21 April 2024

10:30 - 12:00	VCF v4.5: scalability and methylation	GA4GH Implementation Forum (GIF)	REWS genera	
12:00 - 13:00	Lunch			
13:00 - 14:30	Driver Project workshop: data harmonisation	Opportunities and obligations in conducting responsible genomics research: role-based perspectives	GA4GH Infect Community of	
14:30 - 15:00	Break			
15:00 - 16:30	Federated variant level matching	Data Model and Schema Consensus (DaMaSC) session 1	GKS Connect 2024 release workshop	Rare data

Monday, 22 April 2024

09:00 - 10:30	Opening session (plenary session)			
10:30 - 12:00	Phenopackets for querying Beacon	Driver Project workshop: artificial intelligence and machine learning	Ethi final Acc Clau	
12:00 - 13:00	Lunch			
13:00 - 14:30	Welcome to the Cat-VRS	Experiments Metadata Standard	GA4	
14:30 - 15:00	Break			
15:00 - 16:30	Passports in production	Driver Project workshop: data discovery	Command line interfa for GA4GH environments enhanc with confidential computing	

Tuesday, 23 April 2024

09:00 - 10:30	Around the world in one query	Unveiling GA4GH's five-year Equity, Diversity and Inclusion (EDI) Strategic Plan	
09:00 - 12:00	DRS v1.5: key features and plans for 2024		
10:30 - 12:00	Technical Alignment Subcommittee (TASC) meeting	Beacon cohorts and aggregator	
12:00 - 13:00	Lunch		
13:00 - 14:30	Beacon variants	Data visiting	Sequence Annotation (SA) model hackathon (v0.2)
14:30 - 15:00	Break		
18:00	Cocktail reception at Hotel Eden Roc		

Wednesday, 24 April 2024

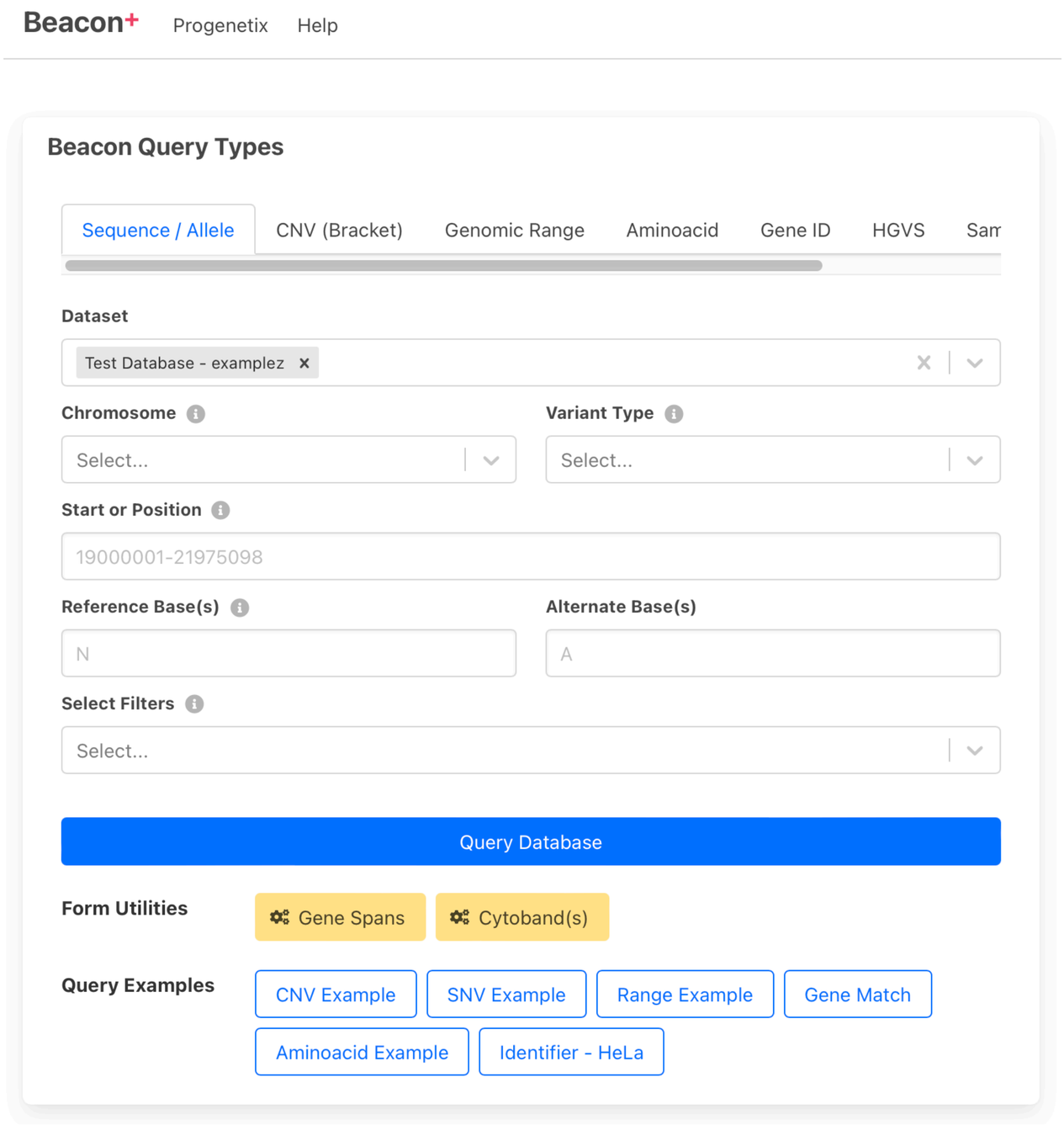
09:00 - 10:30	Implementation of Beacon in cancer use cases		
09:00 - 12:00	GKS implementation forum		
10:30 - 12:00	Beacon resolved: a session for developement	AAI and Passports "what's next?" unconference sessions	
12:00 - 13:00	Lunch		
13:00 - 14:30	Federated cohort building	GKS road map: scope and priorities	Crypt4gh, workflows, and key management
14:30 - 15:00	Break		
15:00 - 16:30	Data Model and Schema Consensus (DaMaSC) session 2	Beacon filter solutions	Driver Project workshop: multi-site geography collaboration and governance

Beacon Queries

Missing or ill defined options

- Translocations
  - in principle possible (start bracket with "referenceName" and end bracket with "mateName" but not yet documented / battle tested)
- Functional elements?
- Cytoband queries?
- Exon hits beyond specifying individual ones by sequence
- Tandem dups ...

➡ **Beacon & Genomic Variations Scout Team**

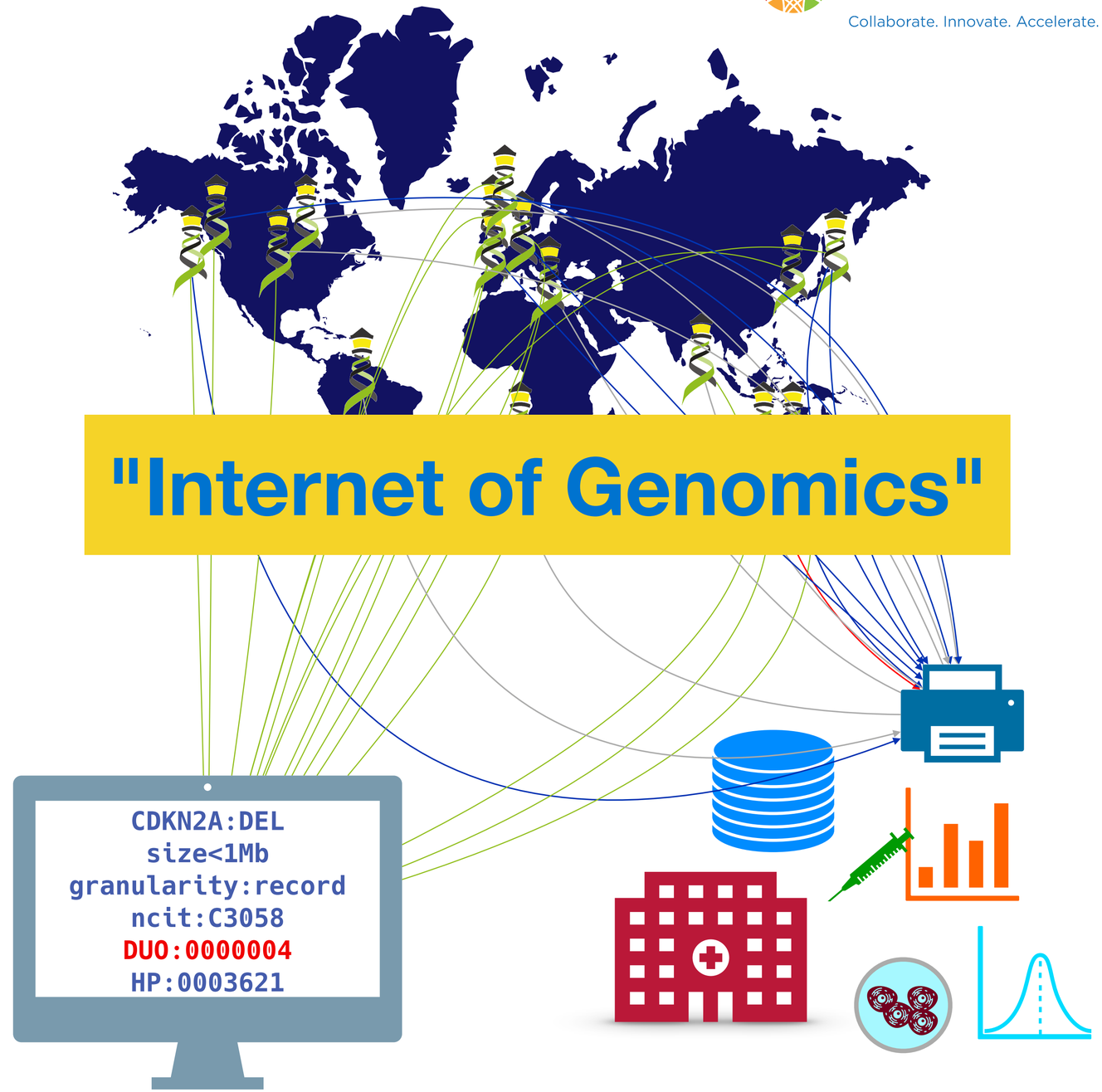


Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

What Can You Do?

- implement procedures and standards supporting **data discovery** (FAIR principles) and federation approaches using Beacon
- promote 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

➡ **Collaborate!**



Universität Zürich[UZH]

SIB Swiss Institute of Bioinformatics

elixir

Universität Zürich[UZH]

Swiss Institute of Bioinformatics