Global Alliance
for Genomics & Health

progenetix

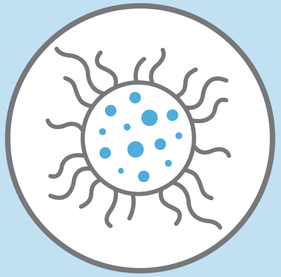
Open Reference Resource for Oncogenomic Copy Number Aberrations Implementing Emerging Global Standards

Progenetix - From Experiments to APIs

Michael Baudis | EORTC-PAMM | Firenze December 2022

Genomic Data Aggregation Can...

Global Alliance
for Genomics & Health



Demonstrate patterns in health & disease

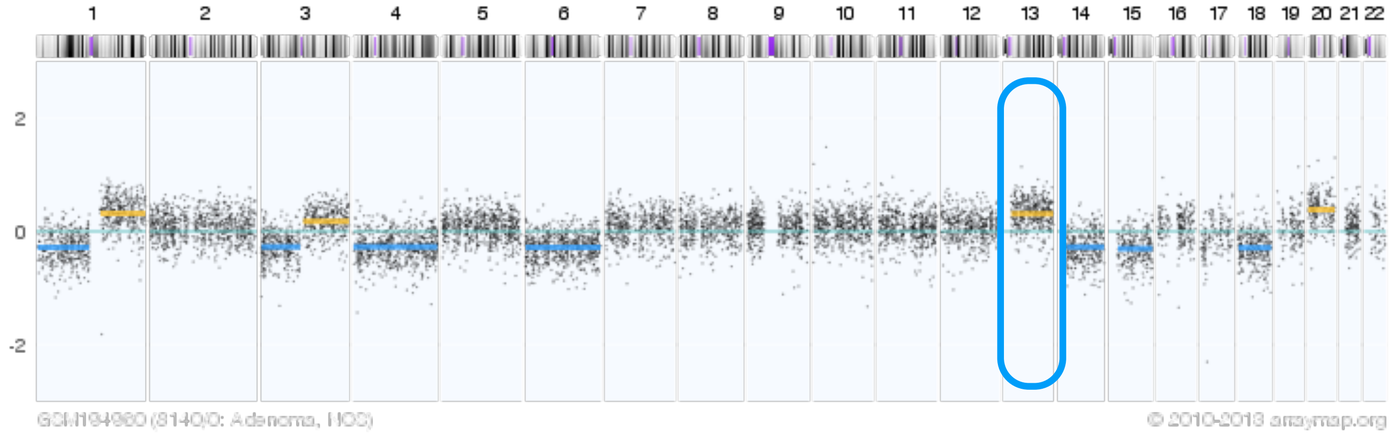
Increase statistical significance of analyses

Lead to “stronger” variant interpretations

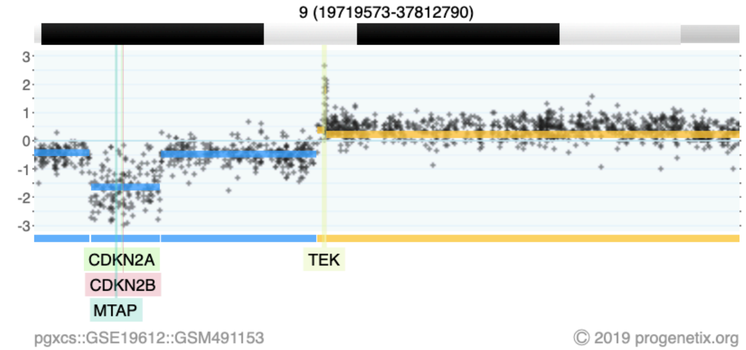
Increase accurate diagnosis

Advance precision medicine

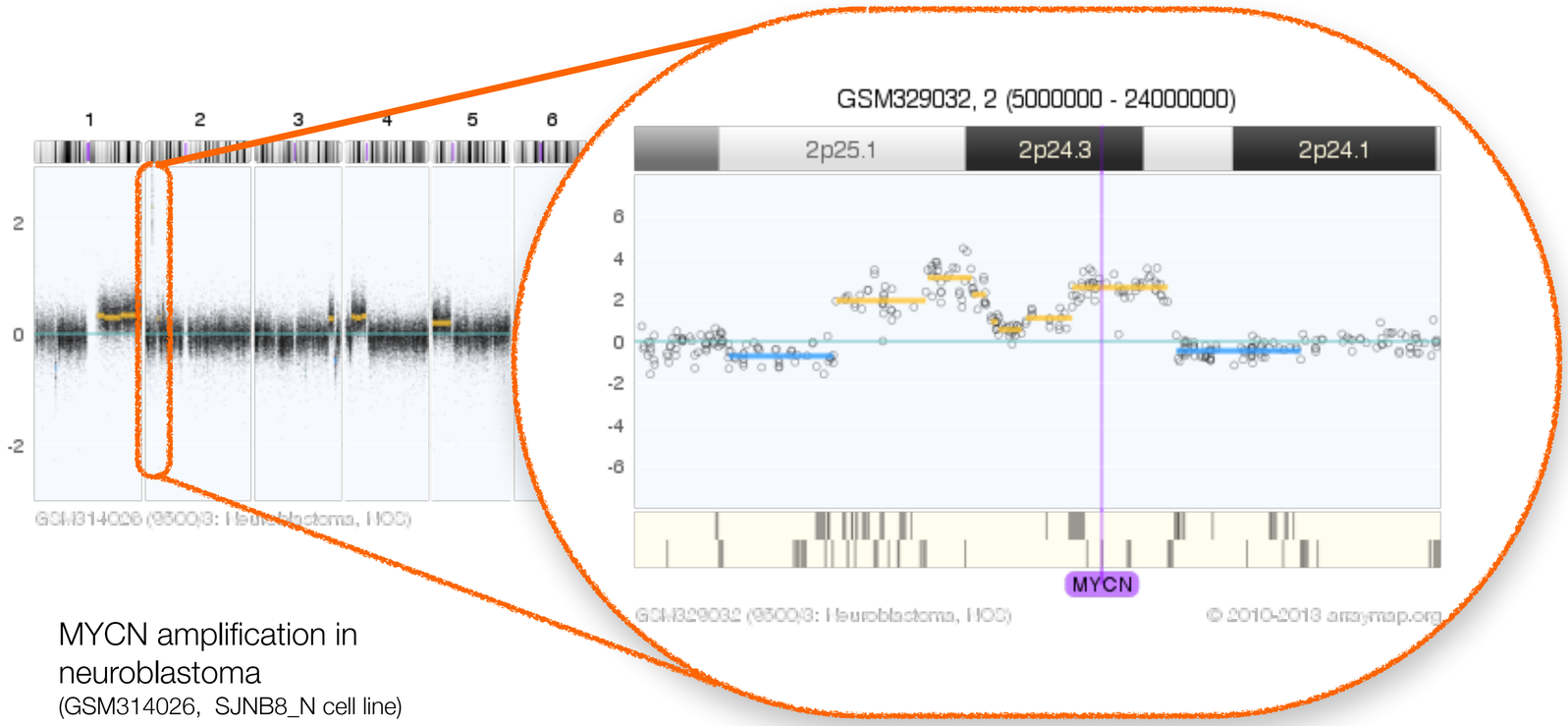
Somatic Copy Number Variations



Gain of chromosome arm 13q in colorectal carcinoma



2-event, homozygous deletion in a Glioblastoma



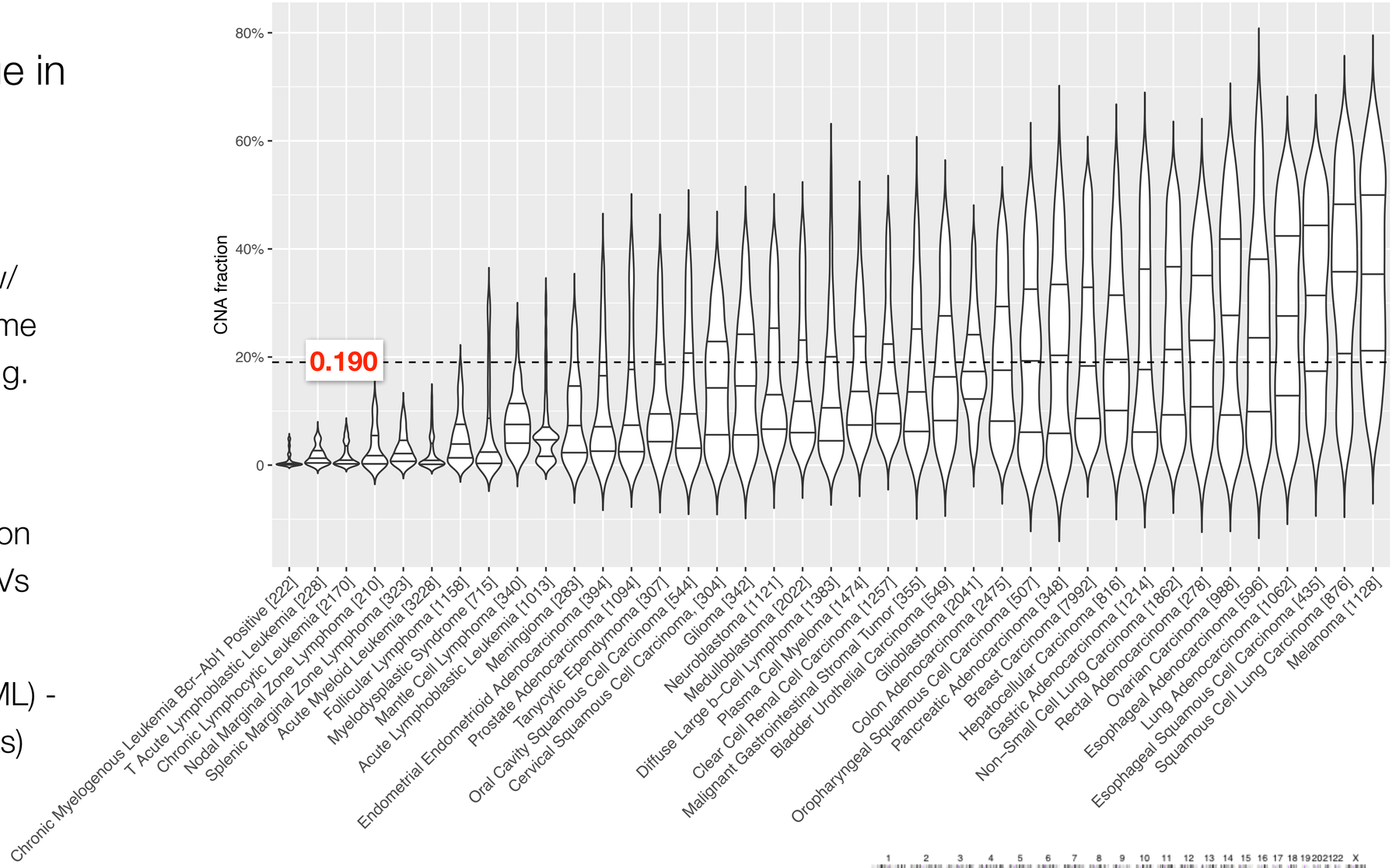
MYCN amplification in neuroblastoma
(GSM314026, SJNB8_N cell line)

low level/**high level** copy number alterations (CNAs)

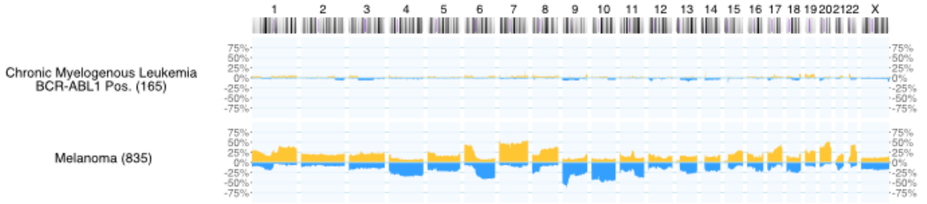
arrayMap

Genome CNV coverage in Cancer Classes

- 43654 out of 93640 CNV profiles; filtered for entities w/ >200 samples (removed some entities w/ high CNV rate, e.g. sarcoma subtypes)
- Single-sample CNV profiles were assessed for the fraction of the genome showing CNVs (relative gains, losses)
- range of medians 0.001 (CML) - 0.358 (malignant melanomas)



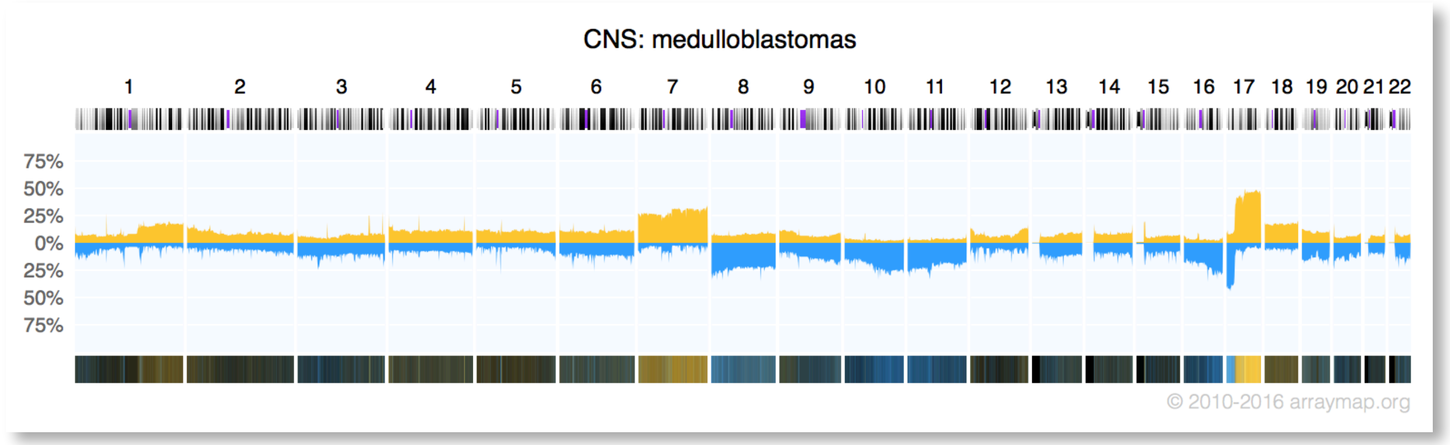
Lowest / Highest CNV fractions =>

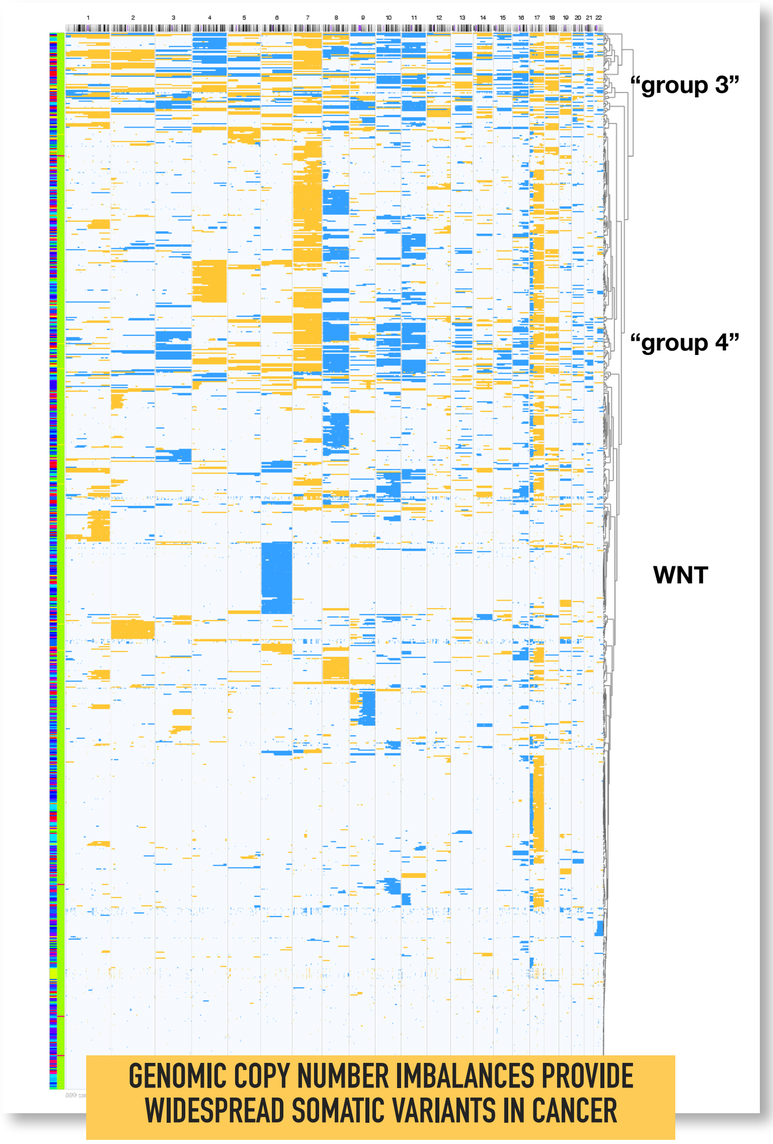


Somatic CNVs In Cancer

Recurrent mutation patterns

How can those patterns be used for classification and determination of biological mechanisms?





A genomic copy number histogram for malignant medulloblastomas, the most frequent type of pediatric brain tumors, displaying regions of genomic duplications and deletions. These can be decomposed into individual tumor profiles which segregate into several clusters of related mutation patterns with functional relevance and clinical correlation.

arrayMap

CNVs Come in a Variety of Formats

Text conversion from ISCN

- articles and supplements with **cytoband-based** *rev ish* CGH results are a great source of CNV data
- conversion by mapping cytoband locations (e.g. UCSC annotation files) to genome coordinates and assigning CNV types (enh, dim, amp are standard)

progenetix

TABLE 3. Comparison of Primary Tumors and Metastases by CGH

Case	Gain in common	Loss in common	Primary tumor only	Metastasis only
108		18		
113	7, 8q24-qter, 13q11-qter, 20q11-qter, Xq11-Xter	1p33-pter, 2p21-pter, 4q24-qter, 15q11-q15, 17p11-pter, 18		
LM	12q22-qter, 15q23-qter, 17q11-ter, 20p11-p12, 20q11-ter, 22q11-ter	1p11-p32, 1q24-31, 4, 13q11-qter, 17p11-pter, 18, 20p11-ter	11p11-pter-	12+
145	4q26-q28, 6p11-p13, 8p11-p12, 920q11-qter	1p11-pter, 4q31-qter, 6q11-qter, 8p12-pter, 11, 15q11-qter, 16q11-qter, 17p11-pter, 18, 21q11-qter	13q21-qter+, 20p11-pter-	8q11-qter+,10-, 6p21-pter-
53	7, 8q11-qter, 9q33-qter, 13q11-qter, 20p11-p12, 20q11-qter	4p13-pter, 4q21-qter, 8p12-pter, 15q14-qter, 18q11-qter, 20p12-pter	5p11-pter-, 5q13-qter-, 14q11-qter-	11+, 16p11-pter+, 17q11-qter+, 19+, 21q11-qter+, 22q11-qter+
147	7, 13q11-qter, 20q11-qter	8p21-pter, 18	4p14-pter-, 4q28-qter+, 8p11-21-, 17q11-q2+, 21q11-qter-	11q22-qter+, 16+, 1p11-33-

TABLE 1. Clinical Data

Case number	Age	Sex	Site	Stage[a]	Grade[b]	Diagnosis of metastatic disease[c]
2	40	M	Transverse colon	IV	3	Synchronous
6	79	M	Ascending colon	IV	2	Synchronous
9	73	M	Transverse colon	II	2	N/A
11	56	M	Rectosigmoid	IV	2	Metachronous
12	70	F	Sigmoid colon	IV	2	Synchronous
13	65	M	Descending colon	II	9	Synchronous
14	60	M	Rectum	III	3	Metachronous
15	51	F	Rectum	III	2	Metachronous
19	63	M	Rectosigmoid Junction	III	2	Synchronous
20	63	M	Rectum	IV	9	Metachronous
21	64	F	Sigmoid colon	IV	2	Synchronous
35	71	M	Rectum	III	9	Metachronous
49	72	M	Cecum	IV	3	Synchronous
53	72	F	Sigmoid colon	IV	2	Synchronous
104	61	M	Sigmoid colon	IV	2	Metachronous
105	58	M	Ascending colon	II	2	Metachronous
107	77	F	Cecum	IV	2	Metachronous
108	53	F	Splenic flexure	IV	2	Synchronous
112	68	M	Rectum	III	3	Synchronous
113	41	M	Splenic flexure	IV	2	Synchronous
114	49	M	Splenic flexure	IV	3	Synchronous
116	73	M	Rectosigmoid	III	9	Metachronous
120	24	F	Descending colon	IV	2	Synchronous
123	62	F	Rectum	III	2	Metachronous
124	42	M	Rectum	IV	9	Synchronous
145	70	M	Rectosigmoid	IV	2	Synchronous
147	86	F	Cecum	IV	2	Synchronous

[a]AJCC/UICC staging system (Hutter and Sobin, 1986).
[b]Grade of primary tumor: 1–3, low, moderate, high grade; 9, grading unknown.
[c]Synchronous, diagnosis of metastatic disease within 12 months following diagnosis of primary tumor; metachronous, diagnosis of metastatic disease after 12 months or later.

GENES, CHROMOSOMES & CANCER 25:82-90 (1999)

Chromosome Arm 20q Gains and Other Genomic Alterations in Colorectal Cancer Metastatic to Liver, as Analyzed by Comparative Genomic Hybridization and Fluorescence In Situ Hybridization

W. Michael Korn,[1]* Toru Yasutake,[2] Wen-Lin Kuo,[1] Robert S. Warren,[3] Colin Collins,[1] Masao Tomita,[2] Joe Gray,[1] and Frederic M. Waldman[1]

CNVs Come in a Variety of Formats: VCF

Issue 1: There are two fields to specify SV/CNV

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
1	2827694	rs2376870	CGTGGATGCGGGGAC	C	.	PASS	SVTYPE=DEL;END=28277(
2	321682	.	T	<DEL>	6	PASS	SVTYPE=DEL;END=321887
2	14477084	.	C	<DEL:ME:ALU>	12	PASS	SVTYPE=DEL;END=144773
3	9425916	.	C	<INS:ME:L1>	23	PASS	SVTYPE=INS;END=942591
3	12665100	.	A	<DUP>	14	PASS	SVTYPE=DUP;END=126862
4	18665128	.	T	<DUP:TANDEM>	11	PASS	SVTYPE=DUP;END=186652

1) Symbolic allele (SA) ↗

↖ 2) SVTYPE

- DEL Deletion relative to the reference
- INS Insertion of novel sequence relative to the reference
- DUP Region of elevated copy number relative to the reference
- INV Inversion of reference sequence
- CNV Copy number variable region (may be both deletion and duplication)
- BND Breakend

The CNV category should not be used when a more specific category can be applied. Reserved subtypes include:

- DUP:TANDEM Tandem duplication
- DEL:ME Deletion of mobile element relative to the reference
- INS:ME Insertion of a mobile element relative to the reference

- DEL: Deletion relative to the reference
- INS: Insertion of novel sequence relative to the reference
- DUP: Region of elevated copy number relative to the reference
- INV: Inversion of reference sequence
- CNV: Copy number variable region (may be both deletion and duplication)
- BND: Breakend

VCF v4.4 deprecate SVTYPE

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FOR
chrA	2	.	TGC	T	.	.	EVENT=DEL_seq	
chrA	2	.	T	<DEL>	.	.	SVLEN=2;SVCLAIM=DJ;EVENT=DEL_symbolic;END=4	
chrA	2	delbp1	T	T[chrA:5[	.	.	MATEID=delbp2;EVENT=DEL_split_bp_cn	
chrA	2	delbp2	A	]chrA:2]A	.	.	MATEID=delbp1;EVENT=DEL_split_bp_cn	
chrA	2	.	T	<DEL>	.	.	SVLEN=2;SVCLAIM=D;EVENT=DEL_split_bp_cn;END=4	
chrA	5	.	G	GAAA	.	.	EVENT=homology_seq	
chrA	5	.	G	<DUP>	.	.	SVLEN=3;CIPOS=0,5;EVENT=homology_dup	
chrA	14	.	T	<INS>	.	.	IMPRECISE;SVLEN=100;CILEN=-50,50;CIPOS=-10,10;END=14	
chrA	14	.	G	.CCCCCCG	.	.	EVENT=single_breakend	

Symbolic allele (SA)

- DEL Region of lowered copy number relative to the reference, or a deletion breakpoint
- INS Insertion of novel sequence relative to the reference
- DUP Region of elevated copy number relative to the reference, or a tandem duplication breakpoint
- INV Inversion of reference sequence
- CNV Copy number variable region (may be both deletion and duplication)

The CNV category should not be used when a more specific category can be applied.
Implementations are free to define their own subtypes. The presence of a subtype does not change either the copy number or breakpoint interpretation of a symbolic structural variant allele. The following subtypes are recommended:

- DUP:TANDEM Tandem duplication
- DEL:ME Deletion of mobile element relative to the reference
- INS:ME Insertion of a mobile element relative to the reference

Reserved specific subtypes

Note that the position of symbolic structural variant alleles is the position of the base immediately preceding the ariant.

Use Subtype to define new structural variant

- <DUP:TANDEM> precise form of duplication
- <DEL:ME:LINE>

Subtypes do not change the meaning symbolic allele.

- using genome positions (POS, INFO.END) for start, end mappings
- treatment of markers for imprecision during matching is left to the implementer
- DUP, DEL are interpreted as indicators for the type of copy number change

Progenetix Oncogenomics Resource

CNV Profiles for Most Cancer Entities from Individual Experiments

University of Zurich[UZH]

progenetix

SIB Swiss Institute of Bioinformatics

Progenetix in 2022

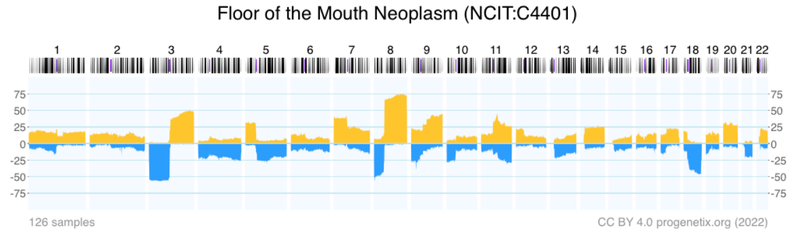
Cancer Genomics Reference Resource

- open resource for curated oncogenomic profiles
- >116'000 cancer CNV profiles, from >800 types
- majority of data from genomic arrays with ~50% overall from SNP platforms with original data re-processing
- standardized encodings (e.g. NCIt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core biosample and technical metadata where accessible (TNM, sex, survival ...)
- publication database and code mapping services

progenetix

Cancer genome data @ progenetix.org

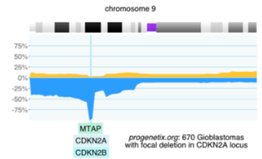
The Progenetix database provides an overview of mutation data in cancer, with a focus on copy number abnormalities (CNV / CNA), for all types of human malignancies. The data is based on *individual sample data* from currently **142063** samples.



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.

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

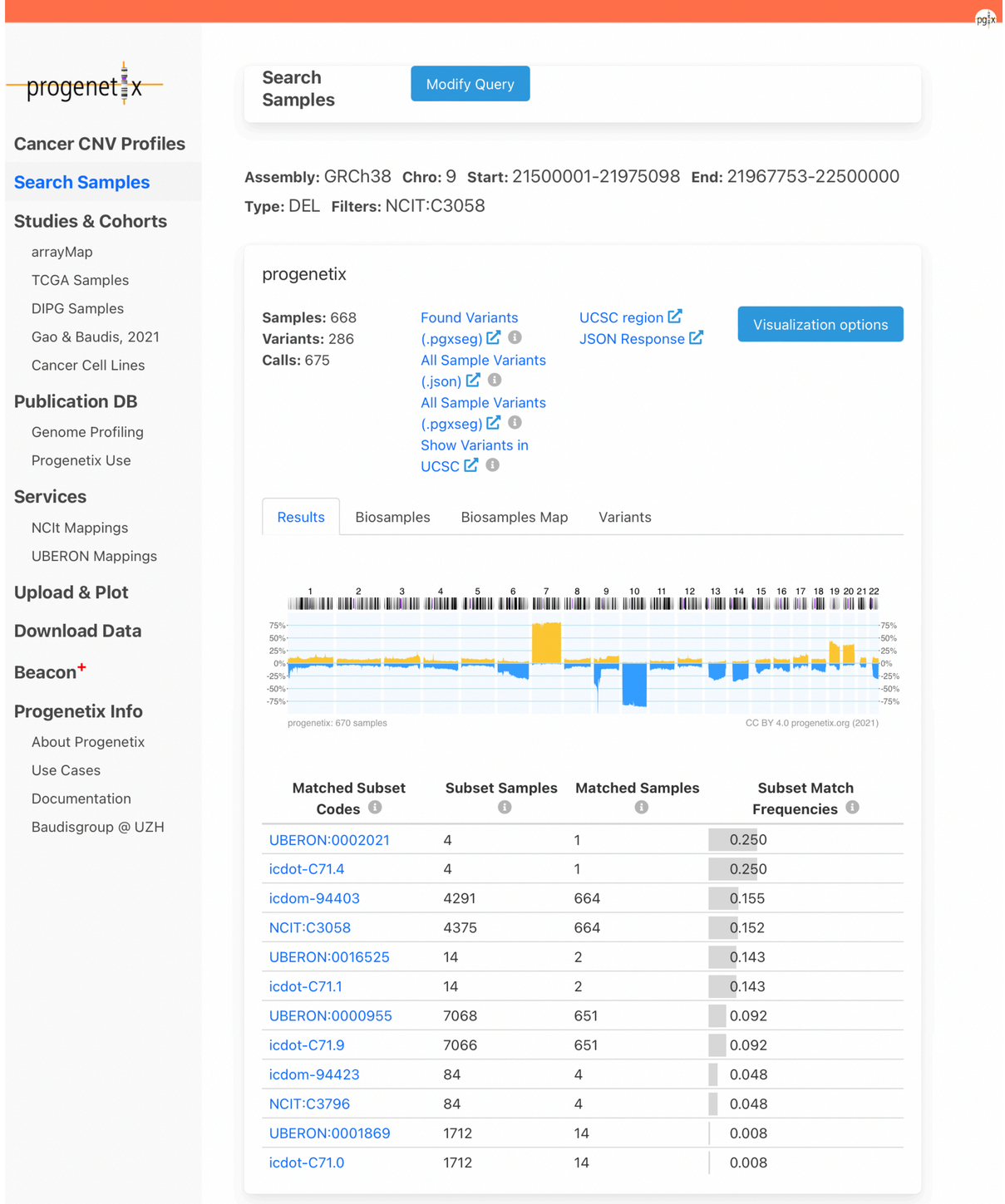
Through the [Publications] page Progenetix provides **4164** annotated references to research articles from cancer genome screening experiments (WGS, WES, aCGH, cCGH). The numbers of analyzed samples and possible availability in the Progenetix sample collection are indicated.

Progenetix in 2022

Cancer Genomics Reference Resource

- open resource for curated oncogenomic profiles
- >116'000 cancer CNV profiles, from >800 types
- majority of data from genomic arrays with ~50% overall from SNP platforms with original data re-processing
- standardized encodings (e.g. NCIt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core biosample and technical metadata where accessible (TNM, sex, survival ...)
- publication database and code mapping services

progenetix



Progenetix in 2022

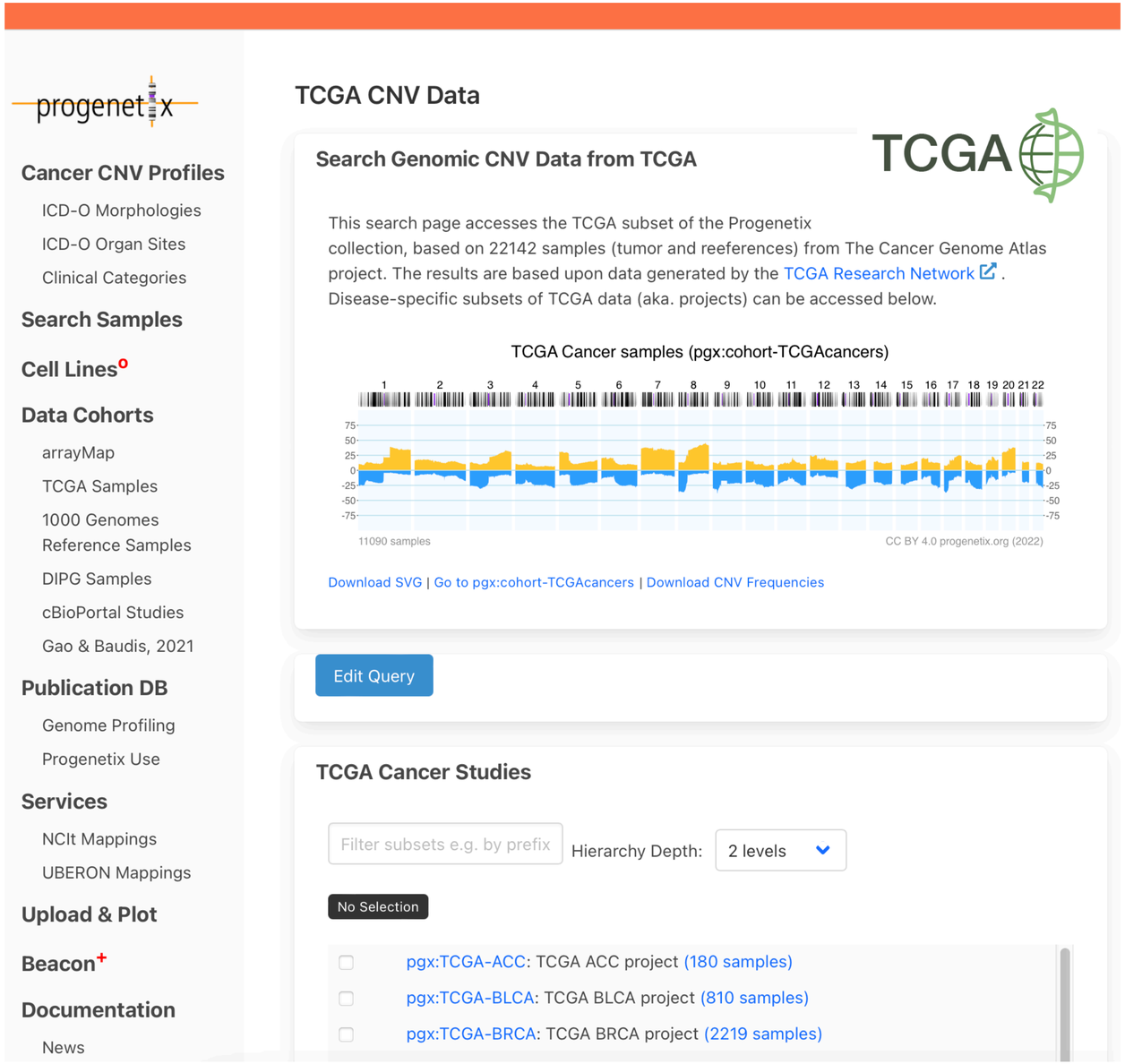
Cancer Genomics Reference Resource

- open resource for curated oncogenomic profiles
- >116'000 cancer CNV profiles, from >800 types
- majority of data from genomic arrays with ~50% overall from SNP platforms with original data re-processing
- standardized encodings (e.g. NCIt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core biosample and technical metadata where accessible (TNM, sex, survival ...)
- publication database and code mapping services

Progenetix in 2022

Cancer Genomics Reference Resource

- contains special data subsets, identified using the "cohorts" concept
  - TCGA CNV data
  - 1000Genomes germline CNVs (WGS)
  - Cancer cell line CNVs with upcoming addition of annotated SNV ... data
  - cBioPortal studies
  - ...

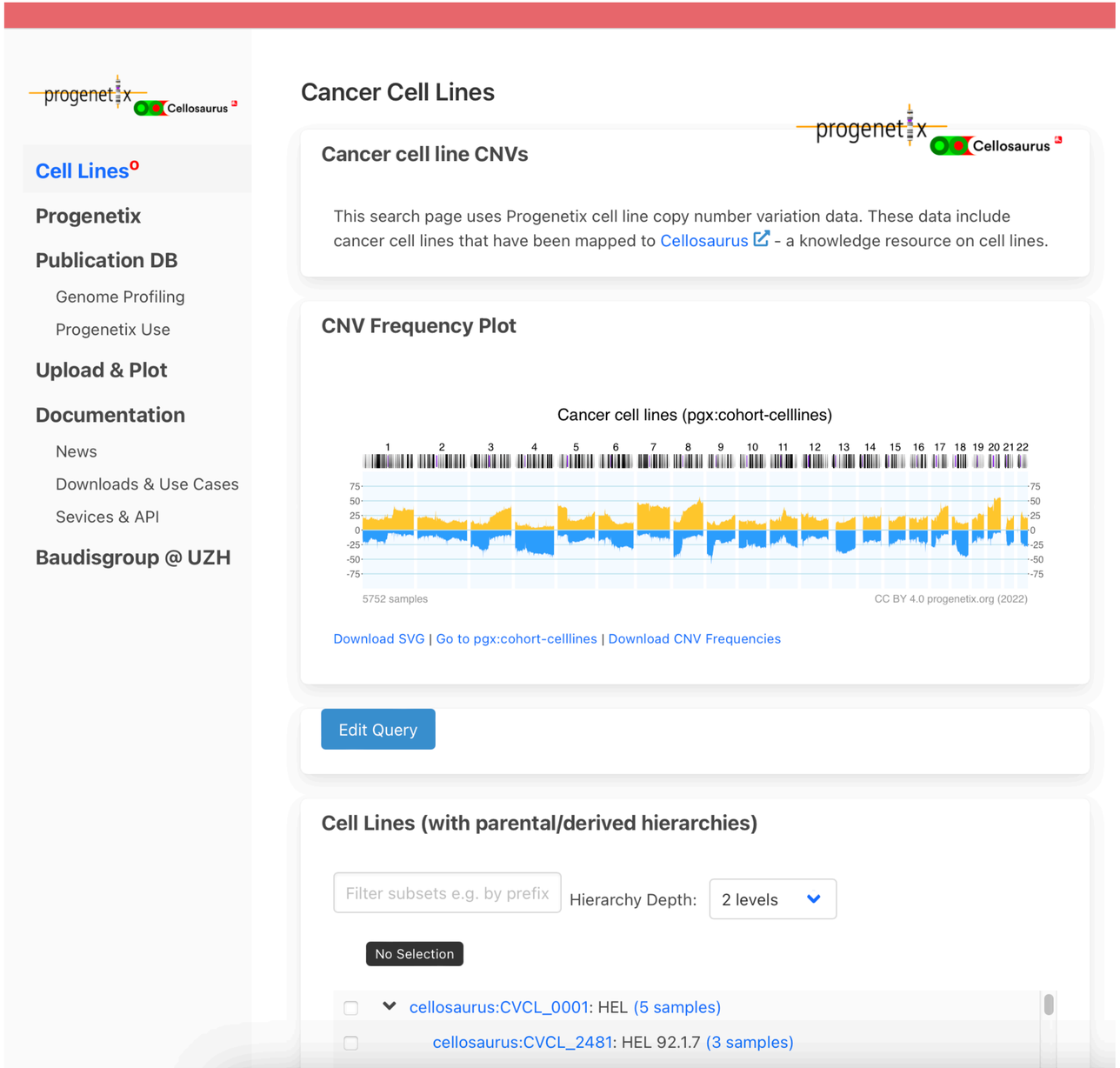


progenetix

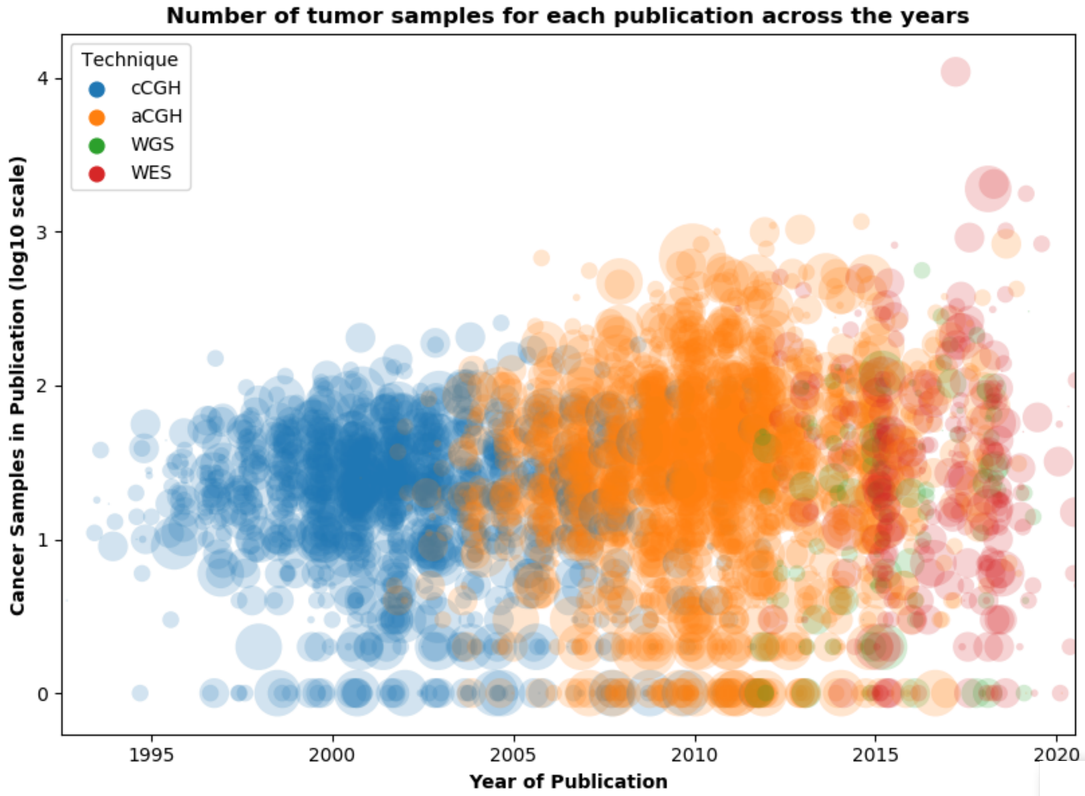
Progenetix in 2022

Cancer Genomics Reference Resource

- contains special data subsets, identified using the "cohorts" concept
  - TCGA CNV data
  - 1000Genomes germline CNVs (WGS)
  - Cancer cell line CNVs with upcoming addition of annotated SNV ... data
  - cBioPortal studies
  - ...



progenetix



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.

progenetix

progenetix

- Cancer CNV Profiles
- Search Samples
- Studies & Cohorts
  - arrayMap
  - TCGA Samples
  - DIPG Samples
  - Gao & Baudis, 2021
  - Cancer Cell Lines
- Publication DB
- Services
  - NCIt Mappings
  - UBERON Mappings
- Upload & Plot
- Download Data

Progenetix Publication Collection

The current page lists articles describing whole genome screening (WGS, WES, aCGH, cCGH) experiments in cancer, registered in the Progenetix publication collection. For each publication the table indicates the numbers of samples analysed with a given technology and if sample profiles are available in Progenetix.

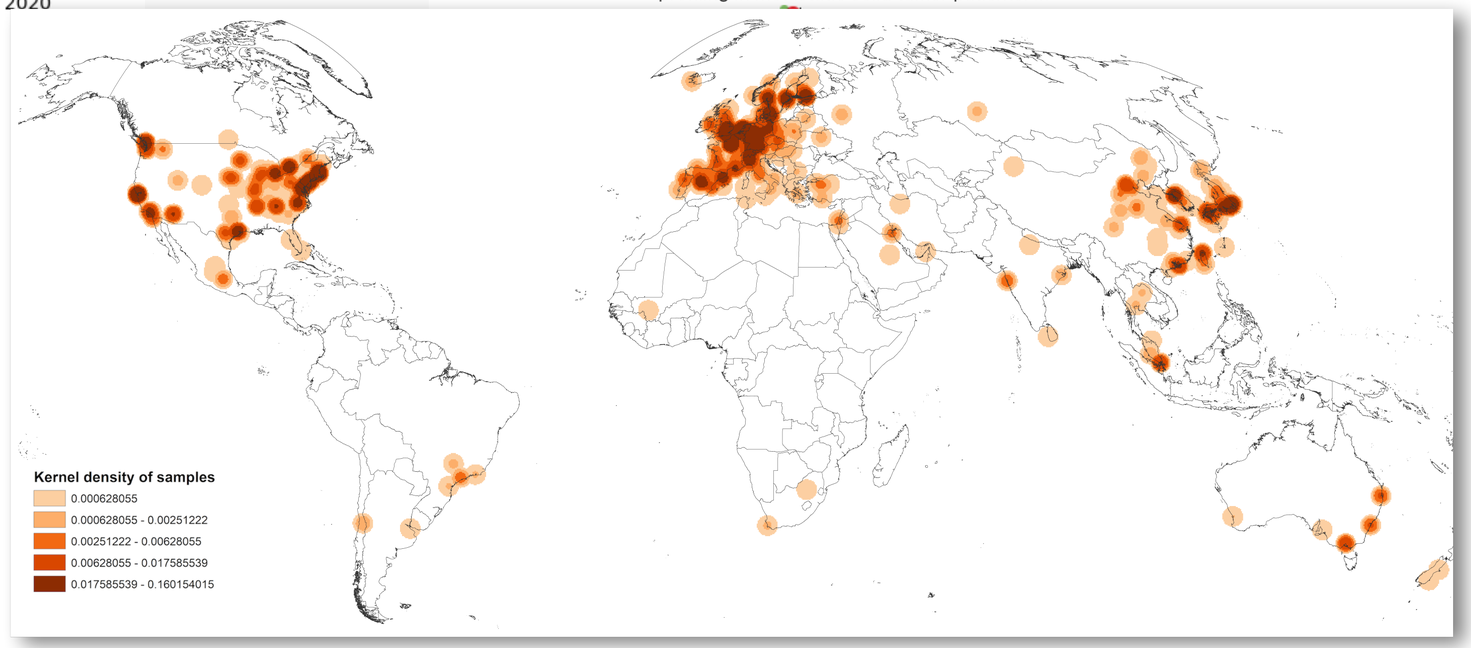
Please contact us to alert us about additional articles you are aware of. The inclusion criteria are described in the documentation.

Filter

City

Type to search...

Publications (3324)		Samples				
id	Publication	cCGH	aCGH	WES	WGS	pgx
PMID:34103027	Peng G, Chai H, Ji W, Lu Y, Wu S et al. (2021) Correlating genomic copy number alterations with clinicopathologic findings in 75 cases of ... BMC Med Genomics	0	79	0	0	0
PMID:34059130	Tsui DWY, Cheng ML, Shady M, Yang JL et al. (2021) Tumor fraction-guided cell-free DNA profiling in metastatic solid tumor patients. ...	0	0	5	113	0



200+ Genomic Data Initiatives Globally

Clinical/Genomic Medicine

Research

National

Cohorts

Since data is distributed globally, we need interoperable standards to answer research questions

Enabling genomic data sharing for the benefit of human health

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

Genomic Data Toolkit →

Regulatory & Ethics Toolkit →

Data Security Toolkit →

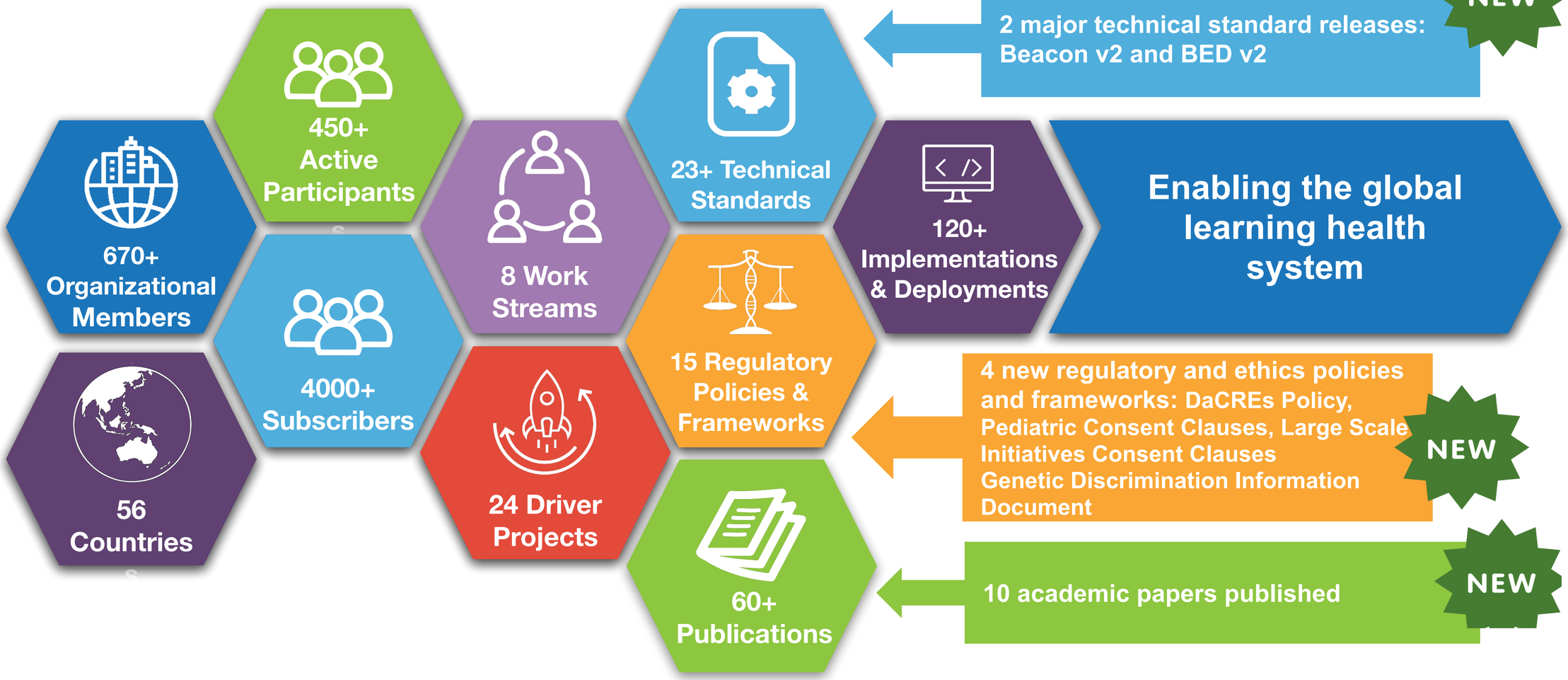
VIEW OUR LEADERSHIP

MORE ABOUT US

BECOME A MEMBER

The GA4GH ecosystem and outputs

Global Alliance for Genomics & Health



GENOMICS

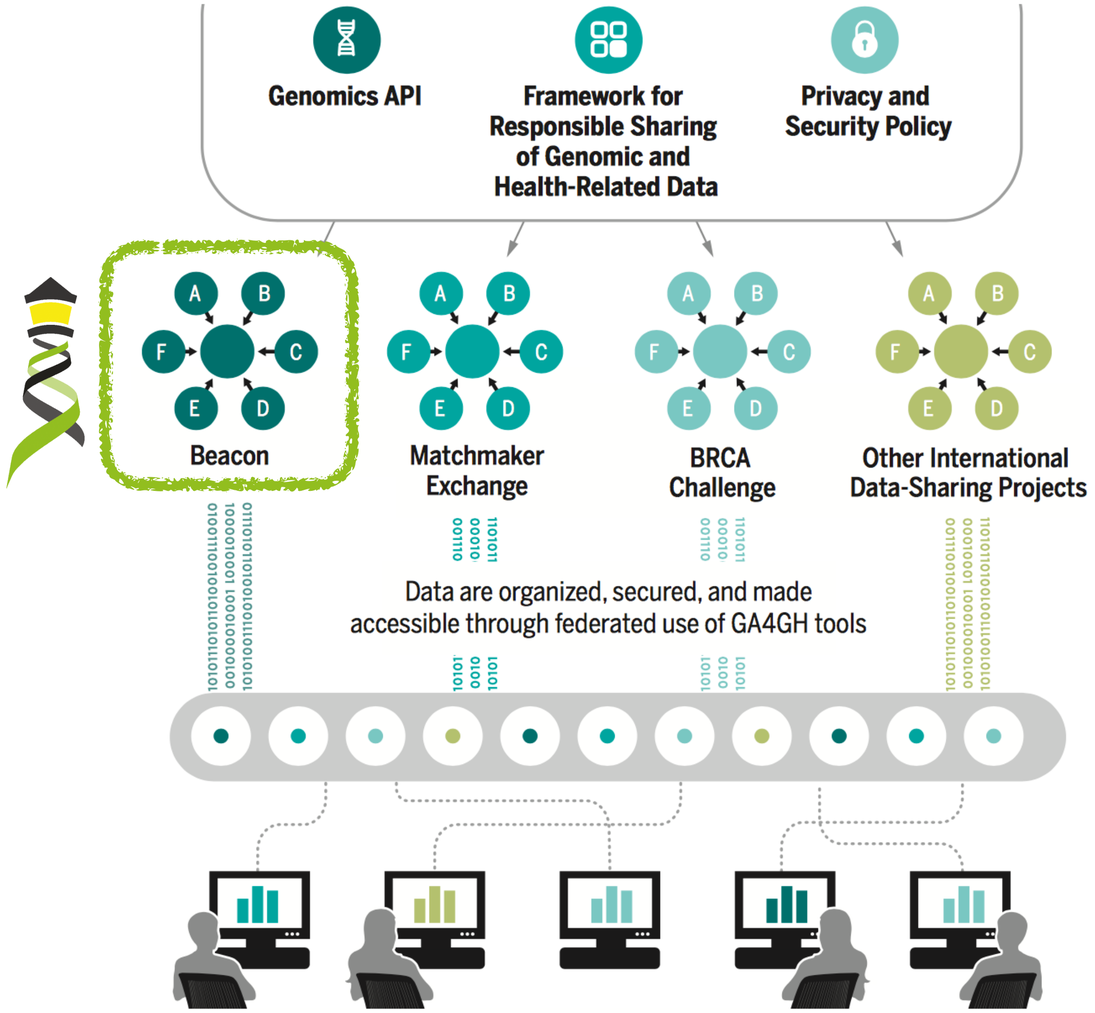
A federated ecosystem for sharing genomic, clinical data

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

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



Global Alliance for Genomics & Health

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***

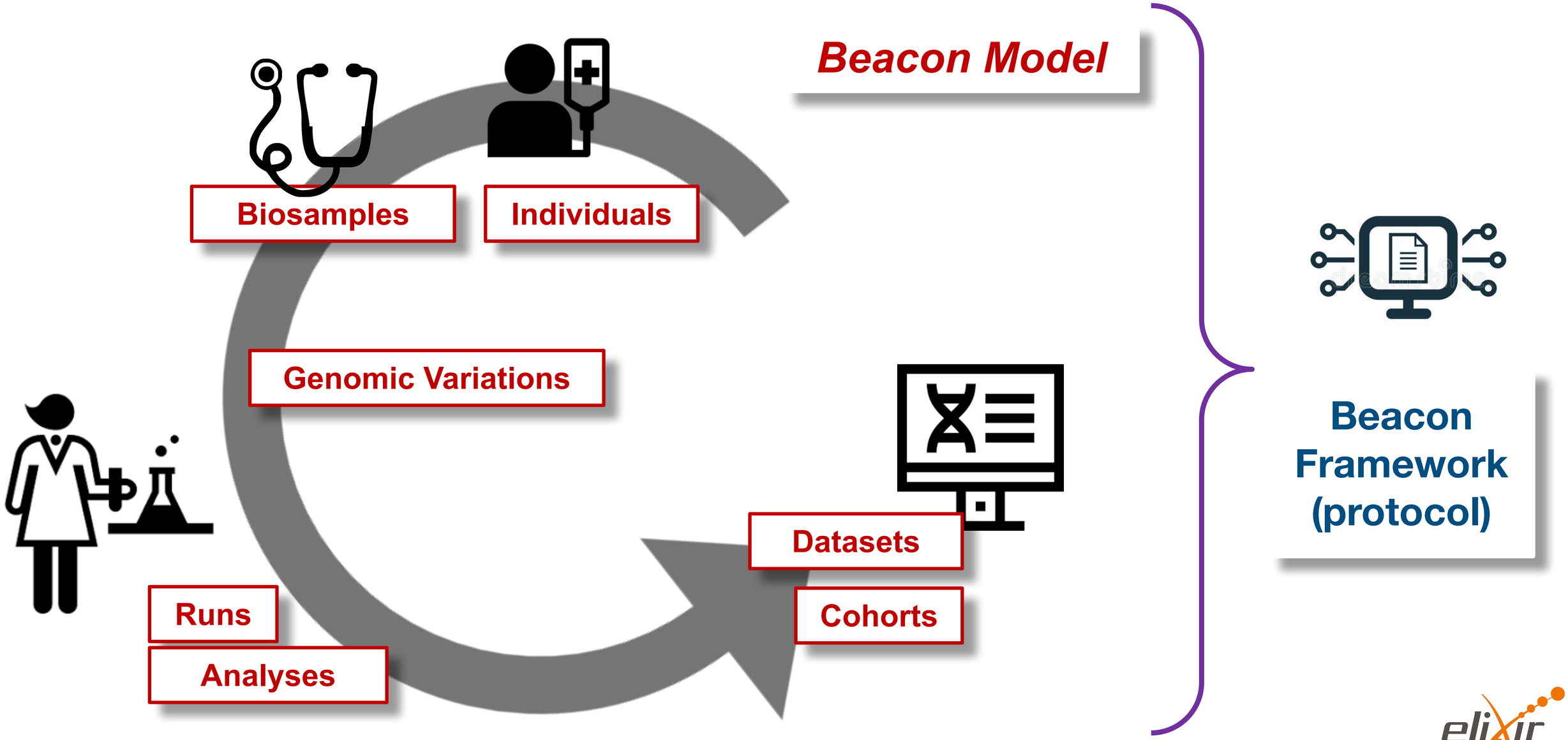
Here: The variant has been found in **few** resources, and those are from **disease** specific **collections**.

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...)

Beacon v2

docs.genomebeacons.org



Progenetix & GA4GH Beacon v2

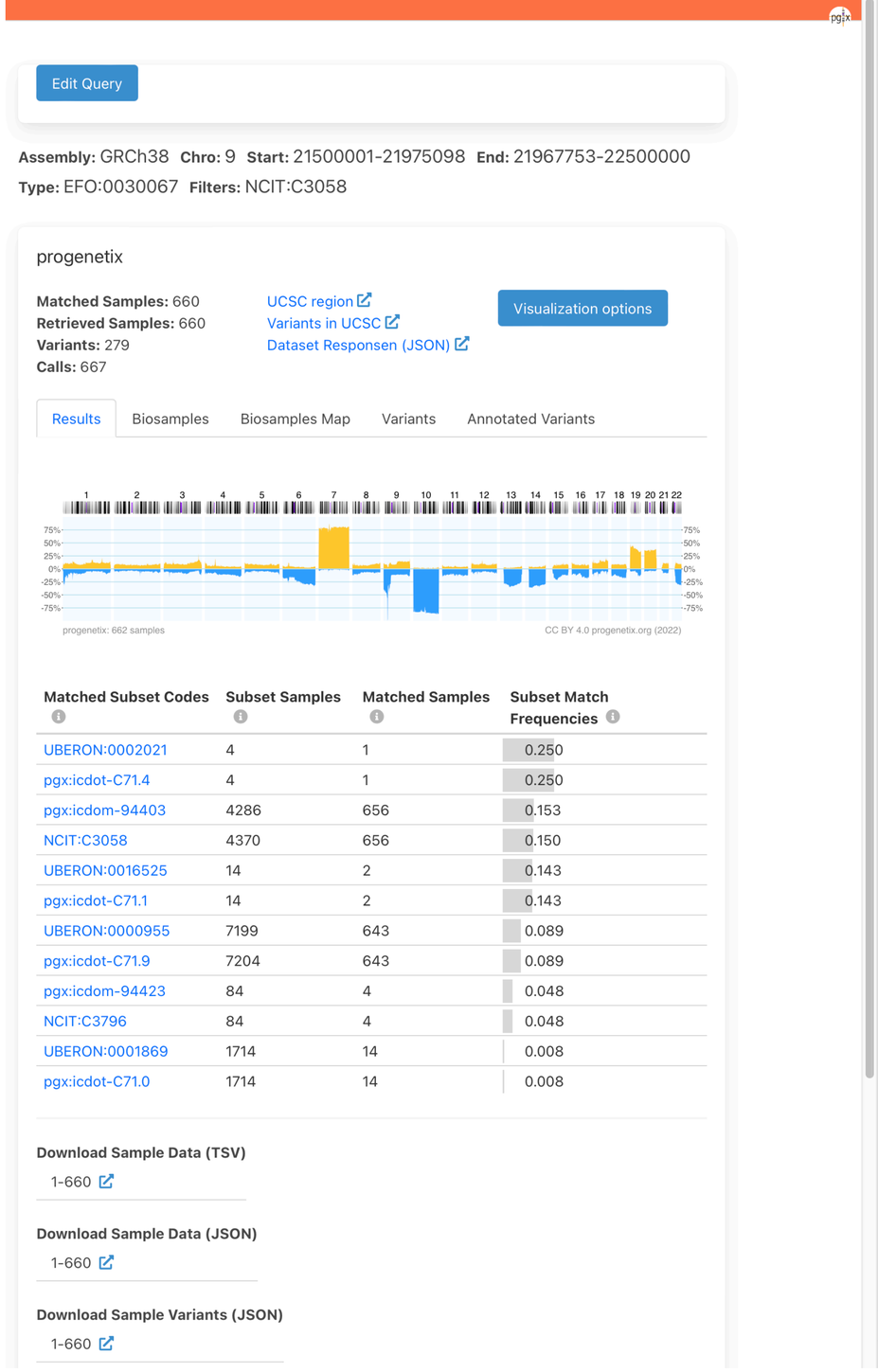
A custom "full stack" implementation of a genomics resource around Beacon data model & API

progenetix

Progenetix

Genomic resource utilizing Beacon v2 calls

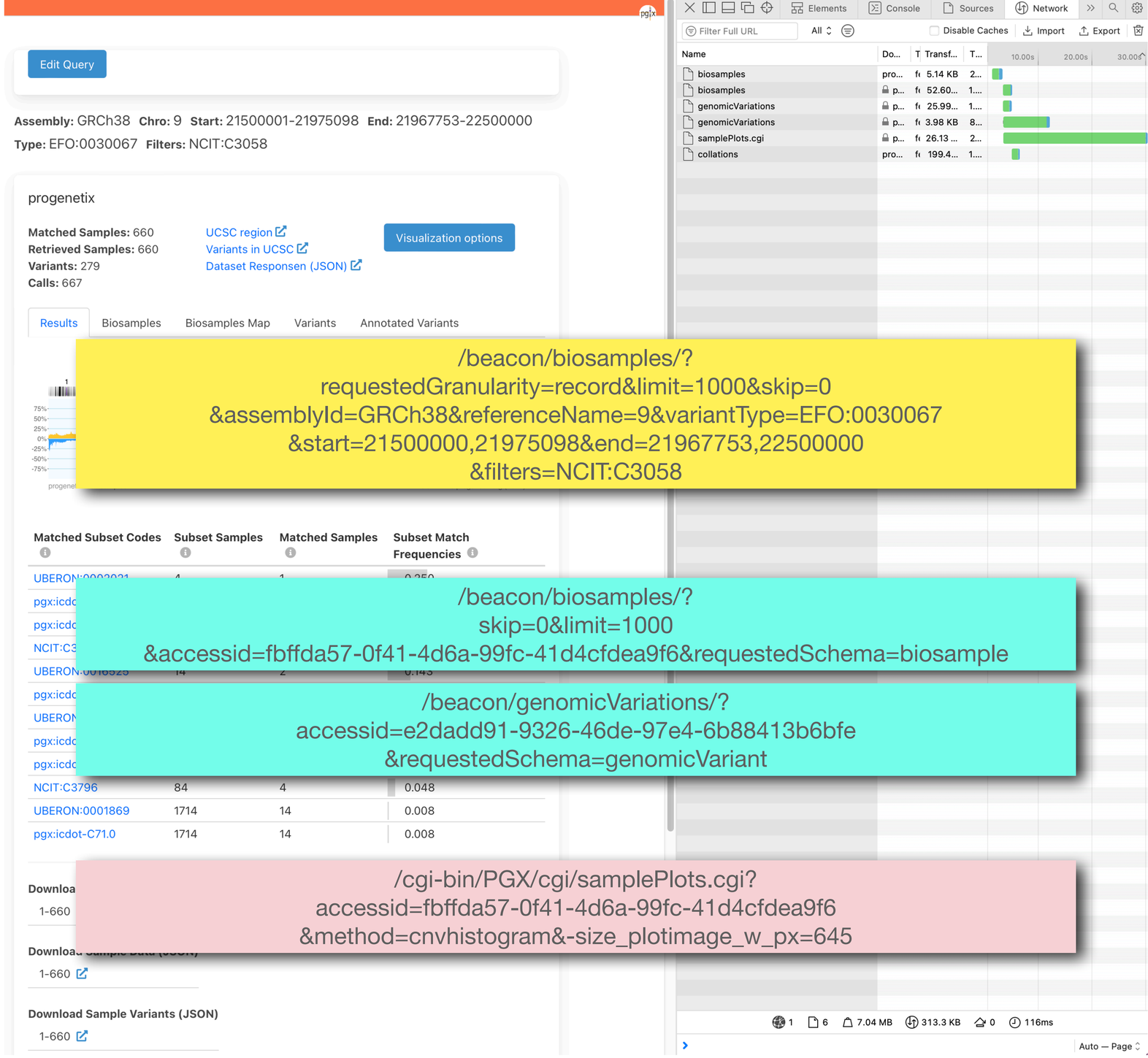
- Progenetix uses Beacon v2 queries to drive its UI
- all individuals, biosamples, variants, analyses matched by a given query are stored by their object ids
- handovers for variant purposes (e.g. to retrieve all matched variants) are returned in the original response and asynchronously retrieved by the front end app



Progenetix

Genomic resource utilizing Beacon v2 calls

- Progenetix uses Beacon v2 queries to drive its UI
- all individuals, biosamples, variants, analyses matched by a given query are stored by their object ids
- handovers for variant purposes (e.g. to retrieve all matched variants) are returned in the original response and asynchronously retrieved by the front end app



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"),
```

variants analyses biosamples individuals

Entity collections

collations geolocs genespans publications qBuffer

Utility collections

Beacon v1 => v2

Genomic variation queries

- Beacon v2 defines query schemas through JSON Schema documents for POST requests and REST paths in OpenAPI documents
- Additional variant parameters:
  - ➡ variantType, mateName (existing in v1)
  - ➡ geneId
  - ➡ variantMinLength, variantMaxLength
  - ➡ aminoacidChange
  - ➡ genomicAlleleShortForm

```
{
    "$schema":"beaconRequestBody.json",
    "meta": {
        "apiVersion": "2.0",
        "requestedSchemas": [
            {
                "entityType": "genomicVariation",
                "schema:": "https://
raw.githubusercontent.com/ga4gh-beacon/beacon-v2/
main/models/json/beacon-v2-default-model/
genomicVariations/defaultSchema.json"
            }
        ]
    },
    "query": {
        "requestParameters": {
            "g_variant": {
                "referenceName": "NC_000017.11",
                "start": [7577120],
                "referenceBases": "G",
                "alternateBases": "A"
            }
        }
    },
    "requestedGranularity": "record",
    "pagination": {
        "skip": 0,
        "limit": 5
    }
}
```

Beacon & CNVs

Open types w/ some definitions

- Beacon supports structural variant queries through the ***variantType*** parameter
- The default model does not prescribe which types can be used (but documents VCF derived DUP & DEL)
- CNV values are not (yet) supported but EFO offers common classes
- Progenetix supports EFO *relative CN terms* (but accepts & interpolates DUP & DEL)

Beacon	VCF	SO	EFO	VRS	Notes
DUP	DUP[1]	SO:0001742 copy_number_gain	EFO:0030070 copy number gain	low-level gain (implicit)	a sequence alteration whereby the copy number of a given genomic region is greater than the reference sequence
DUP	DUP[1]	SO:0001742 copy_number_gain	EFO:0030071 low-level copy number gain	low-level gain	
DUP	DUP[1]	SO:0001742 copy_number_gain	EFO:0030072 high-level copy number gain	high-level gain	commonly but not consistently used for >=5 copies on a bi-allelic genome region
DUP	DUP[1]	SO:0001742 copy_number_gain	EFO:0030073 focal genome amplification	high-level gain	commonly but not consistently used for >=5 copies on a bi-allelic genome region, of limited size (operationally max. 1-5Mb)
DEL	DEL[1]	SO:0001743 copy_number_loss	EFO:0030067 copy number loss	partial loss (implicit)	a sequence alteration whereby the copy number of a given genomic region is smaller than the reference sequence
DEL	DEL[1]	SO:0001743 copy_number_loss	EFO:0030068 low-level copy number loss	partial loss	
DEL	DEL[1]	SO:0001743 copy_number_loss	EFO:0030069 complete genomic deletion	complete loss	complete genomic deletion (e.g. homozygous deletion on a bi-allelic genome region)

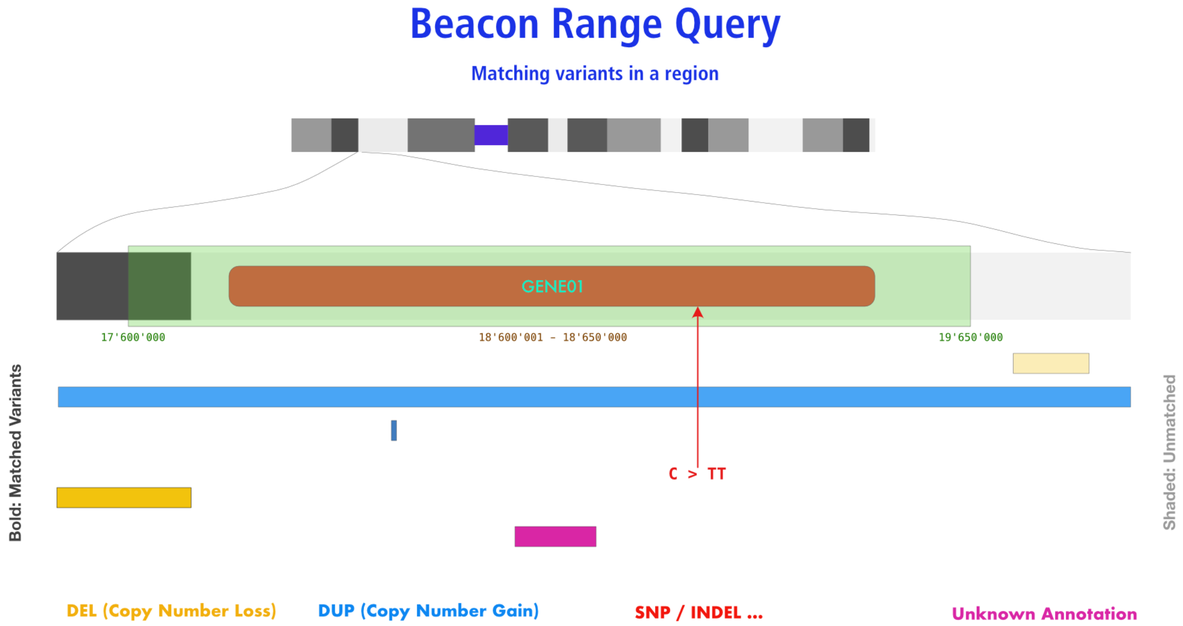
[1] VCFv4.4 introduces an SVCLAIM field to disambiguate between in situ events (such as tandem duplications; known adjacency/ break junction: SVCLAIM=J) and events where e.g. only the change in abundance / read depth (SVCLAIM=D) has been determined. Both J and D flags can be combined.

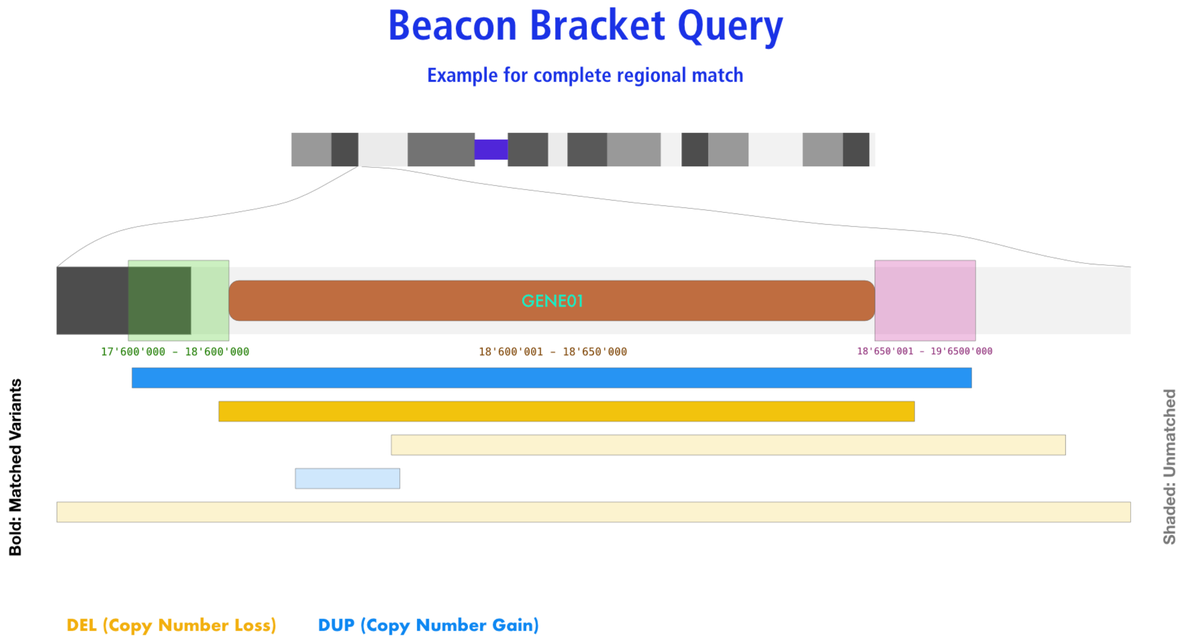
docs.genomebeacons.org/variant-queries/#term-use-comparison

Positional Queries

Going beyond single positions...

- Beacon v1 already provided support for "bracket" queries, e.g. for CNV queries - v2 improves documentation
- Use cases w/ focus on structural variants were evaluated by a Beacon "scout" team
- new "range" option
  - ➡ anything w/ overlap
  - ➡ matched variants can optionally be filtered by type, size, sequence
- query option are not hard defined but derived from parameters
  - ➡ Strong wish for defined types?



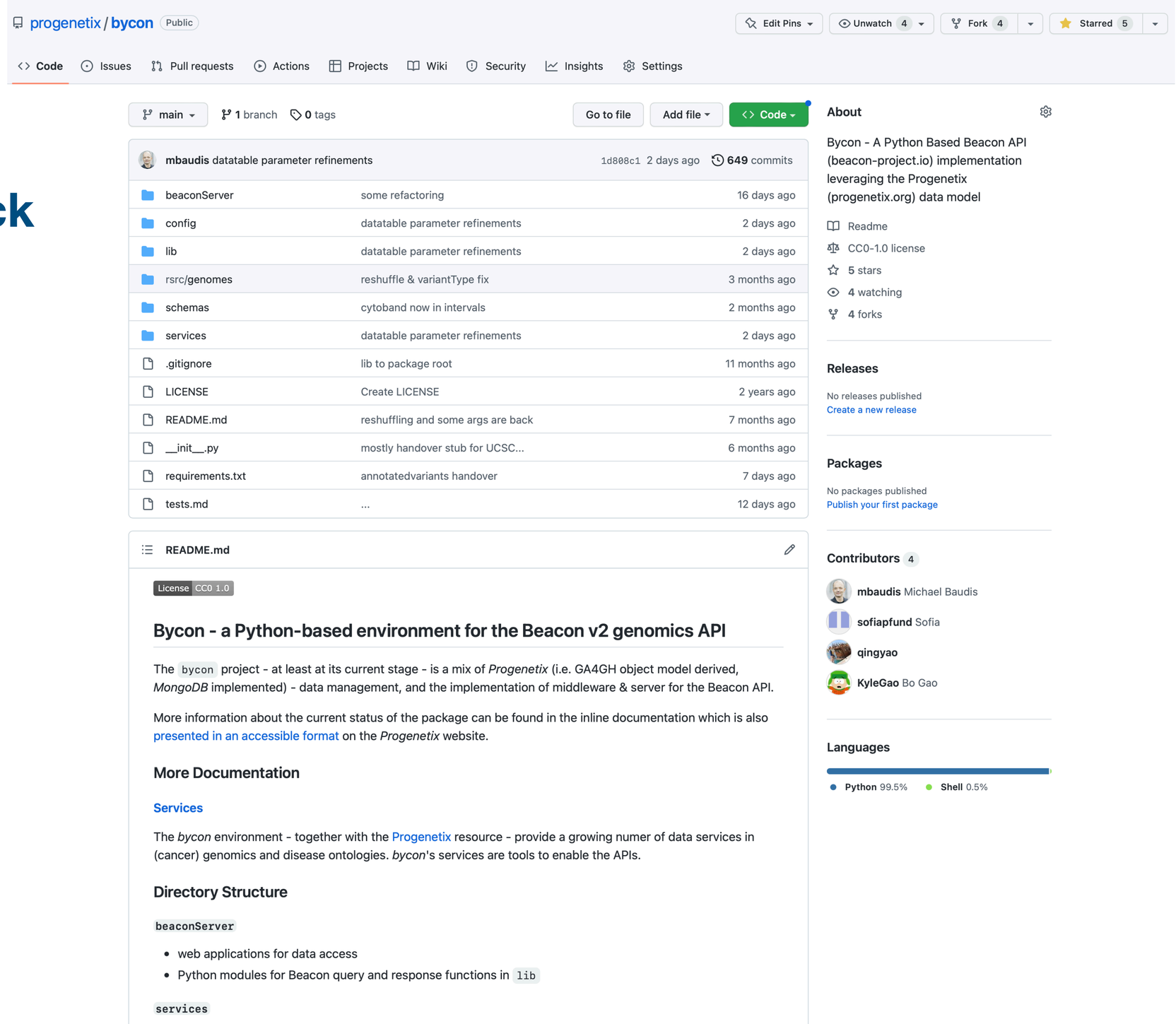


bycon

Progenetix' Beacon Stack

- Python-based software stack
- originally developed for in-house use
- public code base & increasingly sophisticated documentation

➡ Happy about adoption & contributions!



progenetix / bycon Public

Edit Pins · Unwatch 4 · Fork 4 · Starred 5

Code · Issues · Pull requests · Actions · Projects · Wiki · Security · Insights · Settings

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

mbaudis datatable parameter refinements 1d808c1 2 days ago 649 commits

beaconServer	some refactoring	16 days ago
config	datatable parameter refinements	2 days ago
lib	datatable parameter refinements	2 days ago
rsrc/genomes	reshuffle & variantType fix	3 months ago
schemas	cytoband now in intervals	2 months ago
services	datatable parameter refinements	2 days ago
.gitignore	lib to package root	11 months ago
LICENSE	Create LICENSE	2 years ago
README.md	reshuffling and some args are back	7 months ago
__init__.py	mostly handover stub for UCSC...	6 months ago
requirements.txt	annotatedvariants handover	7 days ago
tests.md	...	12 days ago

README.md

License CC0 1.0

Bycon - a Python-based environment for the Beacon v2 genomics API

The `bycon` project - at least at its current stage - is a mix of *Progenetix* (i.e. GA4GH object model derived, *MongoDB* implemented) - data management, and the implementation of middleware & server for the Beacon API.

More information about the current status of the package can be found in the inline documentation which is also presented in an accessible format on the *Progenetix* website.

More Documentation

Services

The *bycon* environment - together with the Progenetix resource - provide a growing numer of data services in (cancer) genomics and disease ontologies. *bycon*'s services are tools to enable the APIs.

Directory Structure

`beaconServer`

- web applications for data access
- Python modules for Beacon query and response functions in `lib`

`services`

About

Bycon - A Python Based Beacon API (beacon-project.io) implementation leveraging the Progenetix (progenetix.org) data model

Readme · CC0-1.0 license · 5 stars · 4 watching · 4 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Contributors 4

mbaudis Michael Baudis
sofiapfund Sofia
qingyao
KyleGao Bo Gao

Languages

Python 99.5% · Shell 0.5%

Documentation Home

Search

progenetix-web
☆3 1

Progenetix Documentation

- Documentation Home
  - Progenetix Source Code
    - bycon
    - progenetix-web
    - PGX
    - Additional Projects
- News & Changes
- Pages & Forms
- Services & API
- Use Case Examples
- Classifications, Ontologies & Standards
- Publication Collection
- Data Review
- Beacon+ & bycon
- Technical Notes
- Progenetix Data
- Baudisgroup @ UZH

Progenetix Source Code ¶

With exception of some utility scripts and external dependencies (e.g. MongoDB the software (from database interaction to website) behind Progenetix and Beac

`bycon`

- Python based service based on the GA4GH Beacon protocol
- software powering the Progenetix resource
- Beacon+ implementation(s) use the same code base

`progenetix-web`

- website for Progenetix and its Beacon+ implementations
- provides Beacon interfaces for the `bycon` server, as well as other Progenetix sevices (e.g. the publicat
- implemented as React / Next.js project
- contains this documentation tree here as `mkdocs` project, with files in the `docs` directory

Org.progenetix

Search

beacon-v2
☆2 8

Base `/biosamples`

`/BIOSAMPLES/` + QUERY

- /biosamples?filters=cellosaurus:CVCL_0004
- this example retrieves all biosamples having an annotation for the Cellosaurus *CVCL_0004* identifier (K562)

es/pgxbs-kftva5c9

a single biosample

MODE=TRUE

es?testMode=true

some random samples

- for testing API responses

`/BIOSAMPLES/{ID}/G_VARIANTS`

- /biosamples/pgxbs-kftva5c9/g_variants/
- retrieval of all variants from a single biosample

Base `/individuals`

`/INDIVIDUALS` + QUERY ¶

/individuals?filters=NCIT:C7541

Rapidly evolving documentation of both the Beacon API itself and its use and technical implementation on
docs.genomebeacons.org
docs.progenetix.org

Shoutout to Laure(e)n Fromont & Manuel Rueda for being instrumental in the Beacon v2 documentation!

Beacon API

Beacon-style JSON responses

The Progenetix resource's API utilizes the `bycon` framework for data query and delivery and represents a custom implementation of the Beacon *v2* API.

The standard format for JSON responses corresponds to a generic Beacon v2 response, with the `meta` and `response` root elements. Depending on the endpoint, the main data will be a list of objects either inside `response.results` or (mostly) in `response.resultSets.results`. Additionally, most API responses (e.g. for biosamples or variants) provide access to data using *handover* objects.

Beacon v2 Documentation

Search

beacon-v2
☆2 8

Org.progenetix

Progenetix & Beacon+

The Beacon+ implementation - developed in the Python & MongoDB based `bycon` project - implements an expanding set of Beacon v2 paths for the Progenetix resource.

Scoped responses from query object

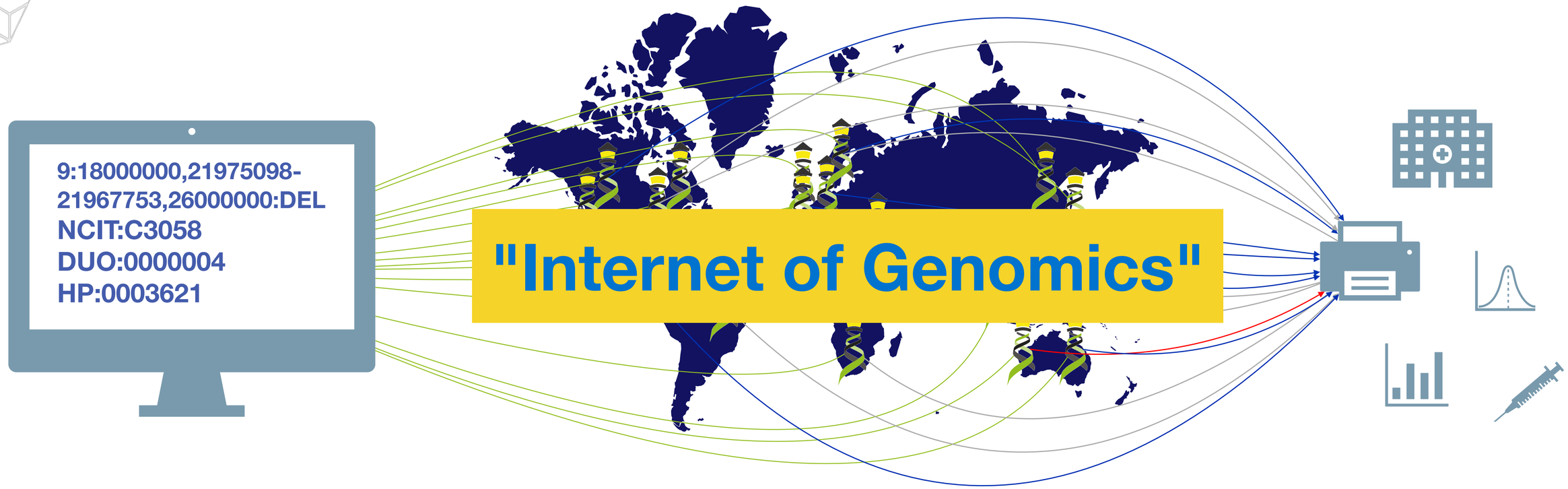
In queries with a complete `beaconRequestBody` the type of the delivered data is independent of the path and determined in the `requestedSchemas`. So far, Beacon+ will compare the first of those to its supported responses and provide the results accordingly; it doesn't matter if the endpoint was `/beacon/biosamples/` or `/beacon/variants/` etc.

Below is an example for the standard test "small deletion CNVs in the CDKN2A locus, in gliomas" Progenetix test query, here responding with the matched variants. Exchanging the `entityType` entry to

- `{ "entityType": "biosample", "schema:": "https://progenetix.org/services/schemas/Biosample/"}`

would change this to a biosample response. The example ccan be tested by POSTing this as `application/json` to `http://progenetix.org/beacon/variants/` or `http://progenetix.org/beacon/biosamples/`.

```
{
    "$schema":"beaconRequestBody.json",
    "meta": {
        "apiVersion": "2.0",
        "requestedSchemas": [
            {
                "entityType": "genomicVariant",
                "schema:": "https://progenetix.org/services/schemas/genomicVariant"
            }
        ]
    },
    "query": {
        "requestParameters": {
```



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

Beacon *v2* API

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

Progenetix & Beacon v2+

A cancer genomics reference resource powered by GA4GH standards

Copy Number Variations (CNV / CNA)

➡ complex, exciting and still poorly understood area in cancer and rare diseases

Progenetix

➡ largest public resource for CNV in cancers (w/ increasing germline data)

Global Alliance for Genomics and Health (GA4GH)

➡ policies & technical standards with focus on ***genomic data sharing***

Federated Data Access

➡ genomic complexity requires ***data access beyond individual resources***

Beacon v2

➡ main GA4GH **data discovery** and sharing protocol, developed with support from the European bioinformatics organization **ELIXIR**

progenetix

Progenetix & Beacon v2+

A cancer genomics reference resource powered by GA4GH standards

Copy Number Variations (CNV / CNA)

➡ complex, exciting and still poorly understood area in cancer and rare diseases

Progenetix serves as a testbed for the early implementation of GA4GH standards such as Beacon extensions, Phenopackets and VRS

➡ policies & technical standards with focus on ***genomic data sharing***

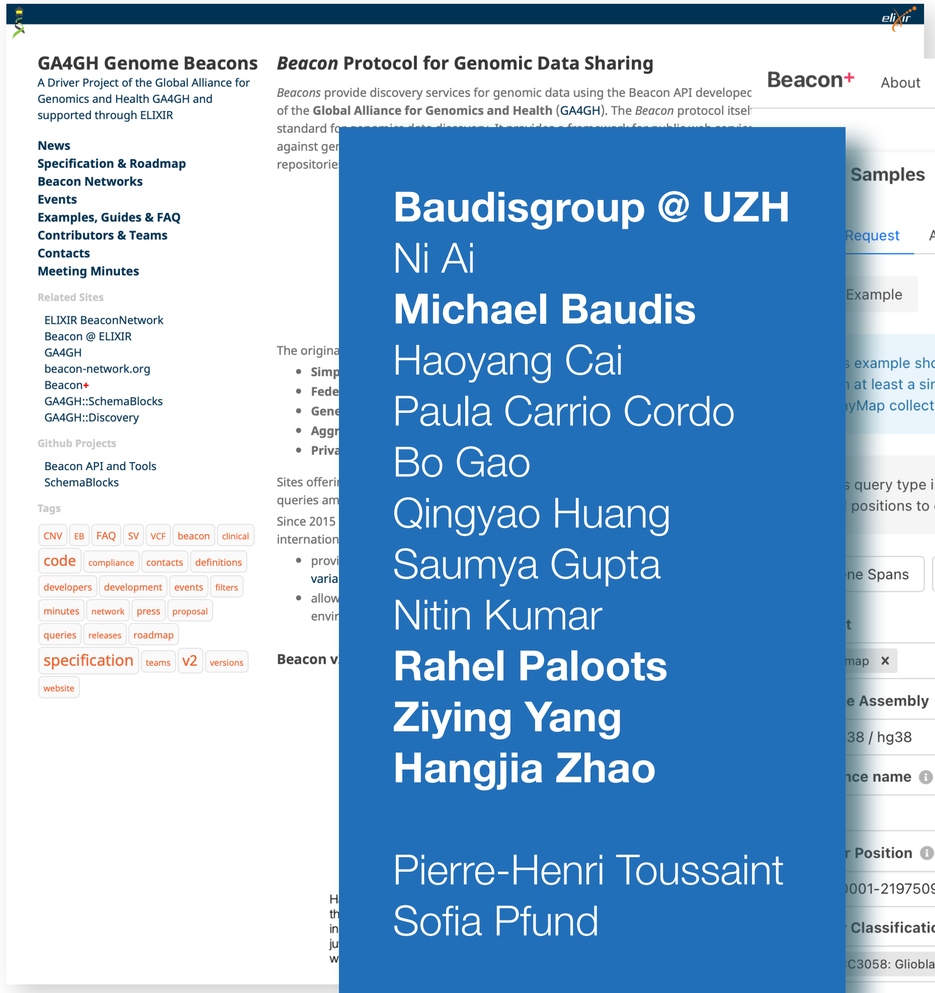
Federated Data Access

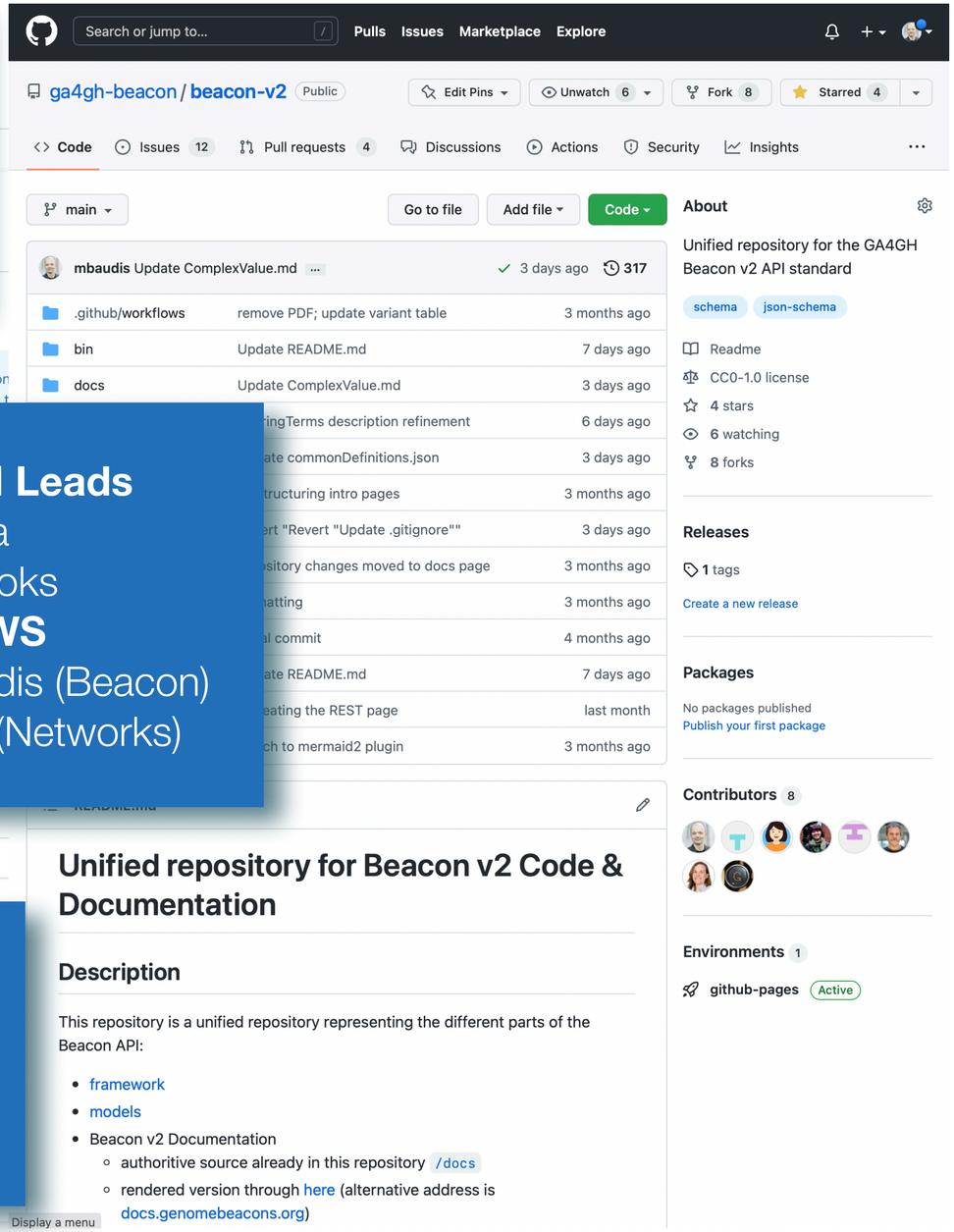
➡ genomic complexity requires ***data access beyond individual resources***

Beacon v2

➡ main GA4GH **data discovery** and sharing protocol, developed with support from the European bioinformatics organization **ELIXIR**

progenetix





Baudisgroup @ UZH
Ni Ai
Michael Baudis
Haoyang Cai
Paula Carrio Cordo
Bo Gao
Qingyao Huang
Saumya Gupta
Nitin Kumar
Rahel Paloots
Ziying Yang
Hangjia Zhao

Pierre-Henri Toussaint
Sofia Pfund

ELIXIR h-CNV
Christophe Béroux
David Salgado
many more ...

Beacon API Leads
Jordi Rambla
Anthony Brooks
Discovery WS
Michael Baudis (Beacon)
Marc Fiume (Networks)

{S}[B] and GA4GH
Melanie Courtot
Helen Parkinson
many more ...

beacon-project.io

beacon.progenetix.org/beaconPlus/

github.com/ga4gh-beacon/

Get Involved! Visit GA4GH.ORG

Global Alliance
for Genomics & Health

Join a Work Stream!

Contact secretariat@ga4gh.org

Become an Organizational Member
ga4gh.org/members

Subscribe to GA4GH Updates
ga4gh.org/subscribe