



Progenetix, *bycon* and More

A Platform for (Onco)genomic Resources



Michael Baudis

Professor of Bioinformatics

University of Zürich

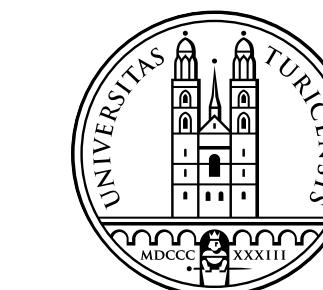
Swiss Institute of Bioinformatics **SIB**

Member GA4GH Strategic Leadership Committee

~~GA4GH Workstream Co-lead D/SCOVERY~~

Co-lead ELIXIR Beacon API Development

Co-lead ELIXIR hCNV Community



Universität
Zürich^{UZH}

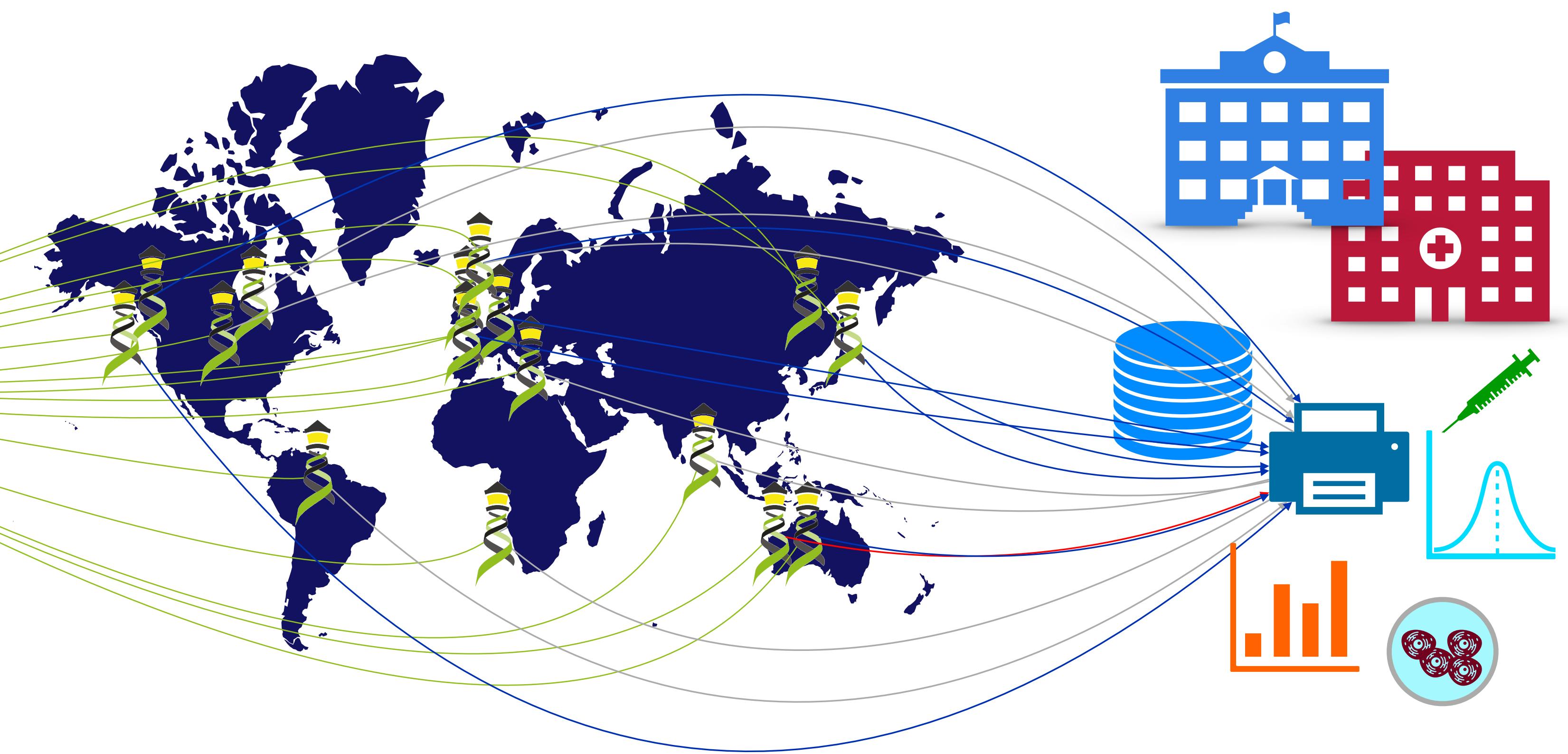
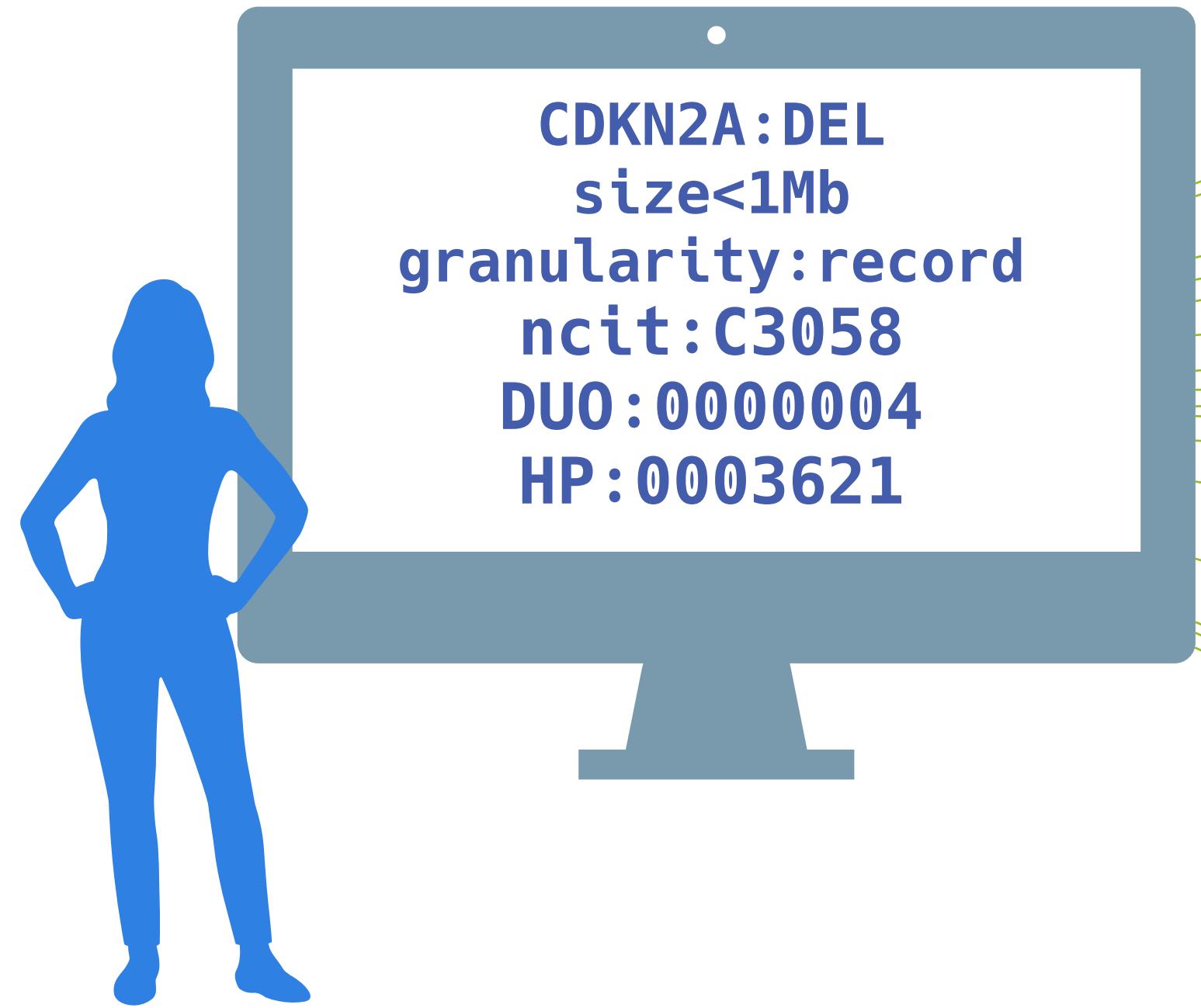


Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

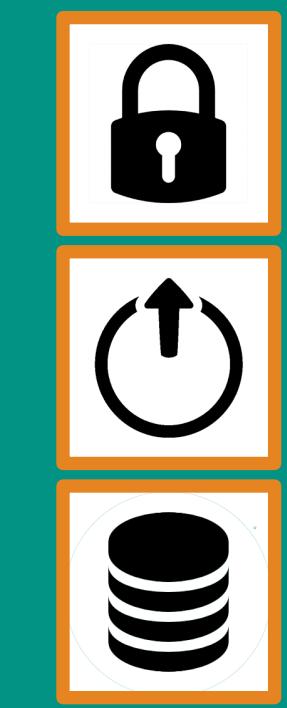


SIB
Swiss Institute of
Bioinformatics





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



Beacon API

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

bycon Beacon+

Implementation driven standards development

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

Beacon protocol response verifier at time of GA4GH approval Spring 2022

Beacon v2 GA4GH Approval Registry

Beacons: European Genome-Phenome Archive | progenetix | cnag | University of Leicester

European Genome-Phenome Archive (EGA)

GA4GH Approval Beacon Test

This [Beacon](#) is based on the GA4GH Beacon [v2.0](#)

BeaconMap	Green
Bioinformatics analysis	Green
Biological Sample	Green
Cohort	Green
Configuration	Green
Dataset	Green
EntryTypes	Green
Genomic Variants	Green
Individual	Green
Info	Green
Sequencing run	Green

progenetix

Theoretical Cytogenetics and Oncogenomics group at UZH and SIB

Progenetix Cancer Genomics Beacon+ provides a forward looking implementation of the Beacon v2 API, with focus on structural genome variants and metadata based on the...

Visit us	Green
Beacon UI	Green
Beacon API	Green
Contact us	Green
BeaconMap	Green
Bioinformatics analysis	Green
Biological Sample	Green
Cohort	Green
Configuration	Green
Dataset	Green
EntryTypes	Green
Genomic Variants	Green
Individual	Green
Info	Green
Sequencing run	Green

cnag

Centre Nacional Analisis Genomica (CNAG-CRG)

Beacon @ RD-Connect

This [Beacon](#) is based on the GA4GH Beacon [v2.0](#)

BeaconMap	Green
Bioinformatics analysis	White
Biological Sample	Red
Cohort	Green
Configuration	Green
Dataset	Red
EntryTypes	Green
Genomic Variants	White
Individual	Red
Info	Red
Sequencing run	White

University of Leicester

Cafe Variome Beacon v2

This [Beacon](#) is based on the GA4GH Beacon [v2.0](#)

BeaconMap	Green
Bioinformatics analysis	White
Biological Sample	White
Cohort	White
Configuration	Green
Dataset	Green
EntryTypes	Green
Genomic Variants	Green
Individual	White
Info	Green
Sequencing run	White

✓ Matches the Spec ✗ Not Match the Spec ● Not Implemented

Cancer Genomics Reference Resource

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



CNV Profiles

- ... by NCIT
- ... by ICD-O Morphology
- ... by ICD-O Site
- ... by TNM & Grade

Search Samples

arrayMap

- TCGA Data
- cBioPortal Studies

Publication DB

Progenetix Use

NCIT - ICD-O Mappings

UBERON Mappings

Upload & Plot

OpenAPI Paths and Examples

Cancer Cell Lines

Beacon+

Documentation

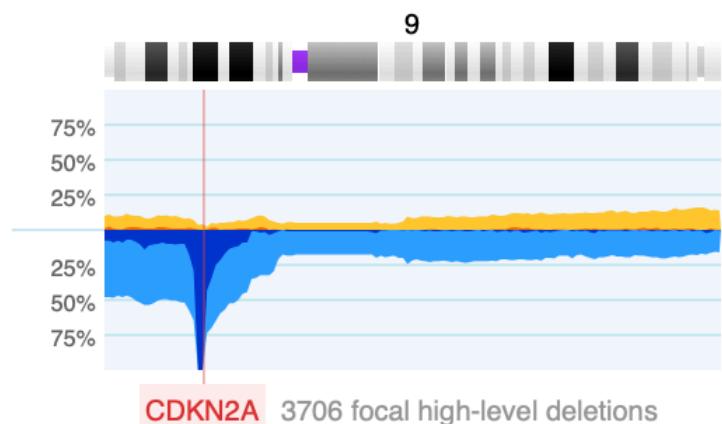
Baudisgroup @ UZH

Cancer genome data @ progenetix.org

The Progenetix database provides an overview of mutation data in cancer, with a focus on copy number abnormalities (CNV / CNA), for all types of human malignancies. The data is based on *individual sample data* of currently **240600** samples from **1126** different cancer types (NCIt neoplasm classification)

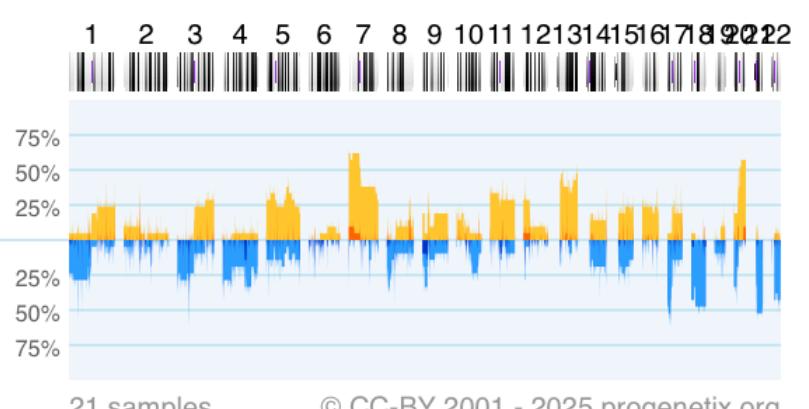
Local CNV Frequencies

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

Frequency profiles of regional genomic gains and losses for all categories (diagnostic entity, publication, cohort ...) can be accessed through the respective Cancer Types pages (e.g. [NCIT Neoplasia Codes](#)) and compared through the [Compare CNV Profiles](#) option. Below is an example of aggregated CNV data in 21 samples in Malignant Cecum Neoplasm with the frequency of regional **copy number gains (high level)** and **losses (high level)** displayed for the 22 autosomes.



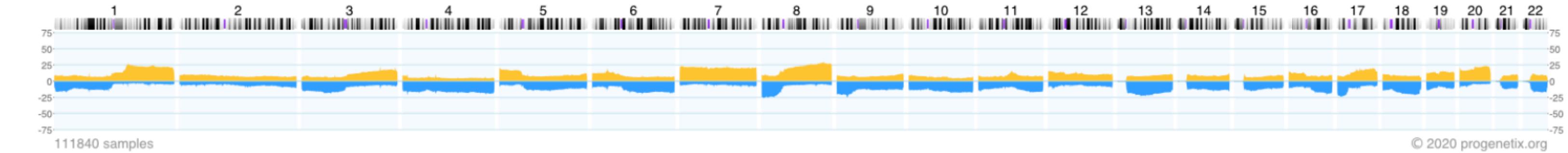
Malignant Cecum Neoplasm
NCIT:C9329 (21 samples)

[Download SVG](#) | [Go to NCIT:C9329](#) | [Download CNV Frequencies](#)

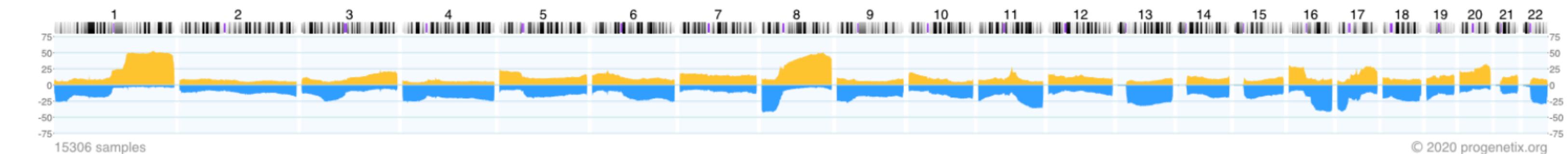
© CC-BY 2001 - 2025 progenetix.org

Somatic CNV in Cancer Example Frequency Profiles

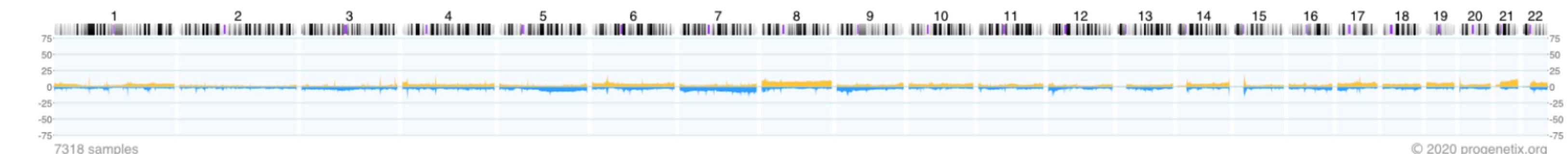
Progenetix: Regional CNV Frequencies in 111'840 Neoplasm (NCIT:C3262)



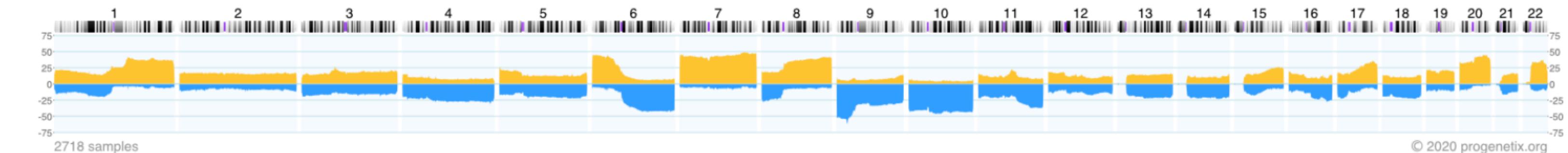
Malignant Breast Neoplasm (NCIT:C9335)



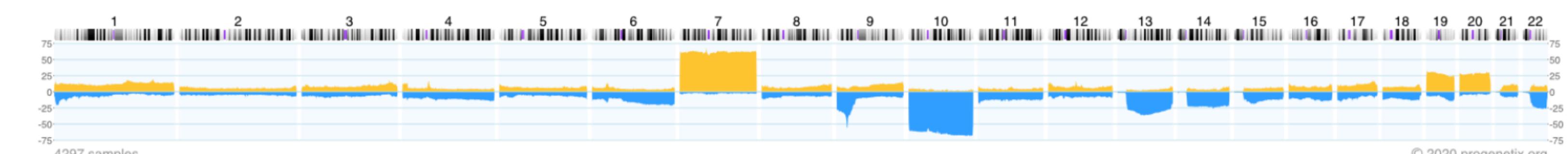
Acute Leukemia (NCIT:C9300)



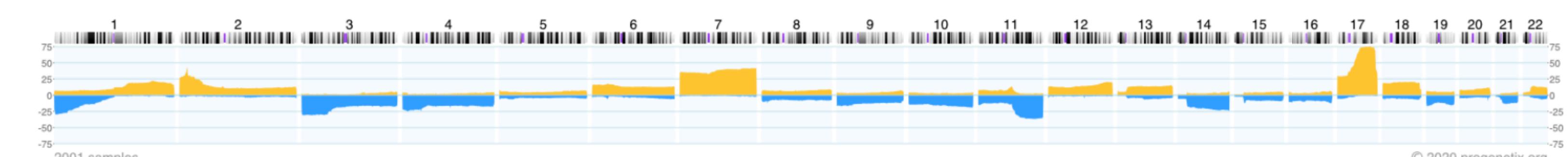
Melanoma (NCIT:C3224)



Glioblastoma (NCIT:C3058)



Neuroblastoma (NCIT:C3270)



2001 samples

© 2020 progenetix.org

Cancer Genomics Reference Resource

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

The screenshot shows the Progenetix interface with a sidebar on the left containing links like 'CNV Profiles', 'Search Samples', 'arrayMap', 'Publication DB', 'NCIT - ICD-O Mappings', 'Upload & Plot', 'OpenAPI Paths and Examples', 'Cancer Cell Lines', 'Beacon+', 'Documentation', and 'Baudisgroup @ UZH'. The main area is titled 'Cancer Types by National Cancer Institute NCI Code' and contains a detailed hierarchical tree of cancer types. A search bar at the top says 'Filter subsets e.g. by prefix ...' and a dropdown says 'Hierarchy Depth: 5 levels'. The tree starts with 'Neoplasm' (138552 samples, 144862 CNV profiles), which branches into 'Neoplasm by Site' (133029 samples, 139114 CNV profiles), then into 'Genitourinary System Neoplasm' (21582 samples, 23171 CNV profiles), 'Benign Genitourinary System Neoplasm' (243 samples, 243 CNV profiles), 'Benign Urinary System Neoplasm' (98 samples, 98 CNV profiles), 'Benign Urinary Tract Neoplasm' (3 samples, 3 CNV profiles), 'Benign Kidney Neoplasm' (95 samples, 95 CNV profiles), 'Benign Reproductive System Neoplasm' (145 samples, 145 CNV profiles), 'Benign Female Reproductive System Neoplasm' (145 samples, 145 CNV profiles), 'Malignant Genitourinary System Neoplasms' (20567 samples, 22154 CNV profiles), 'Metastatic Malignant Genitourinary System Neoplasms' (2 samples, 2 CNV profiles), 'Metastatic Genitourinary System Carcinoma' (2 samples, 2 CNV profiles), 'Genitourinary System Carcinoma' (19462 samples, 20921 CNV profiles), 'Metastatic Genitourinary System Carcinoma' (2 samples, 2 CNV profiles), 'Female Reproductive System Carcinoma' (5746 samples, 5974 CNV profiles), 'Male Reproductive System Carcinoma' (7022 samples, 7808 CNV profiles), 'Urinary System Carcinoma' (6694 samples, 7139 CNV profiles), and 'Recurrent Malignant Genitourinary System Neoplasms' (3 samples, 3 CNV profiles).

Cancer Genomics Reference Resource

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

progenetix

CNV Profiles

- ... by NCIT
- ... by ICD-O Morphology
- ... by ICD-O Site
- ... by TNM & Grade

Search Samples

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

Dataset(s) Progenetix cancer genome variants

Variant Type EFO:0030067 (copy number deletion)

Chromosome 9 (NC_000009.12) Start or Position 21000001,21975098 End 21967753,23000000

Cancer Classification(s) Select...

Genotypic Sex Select...

Various Subsets NCIT:C3058: Glioblastoma (4900)

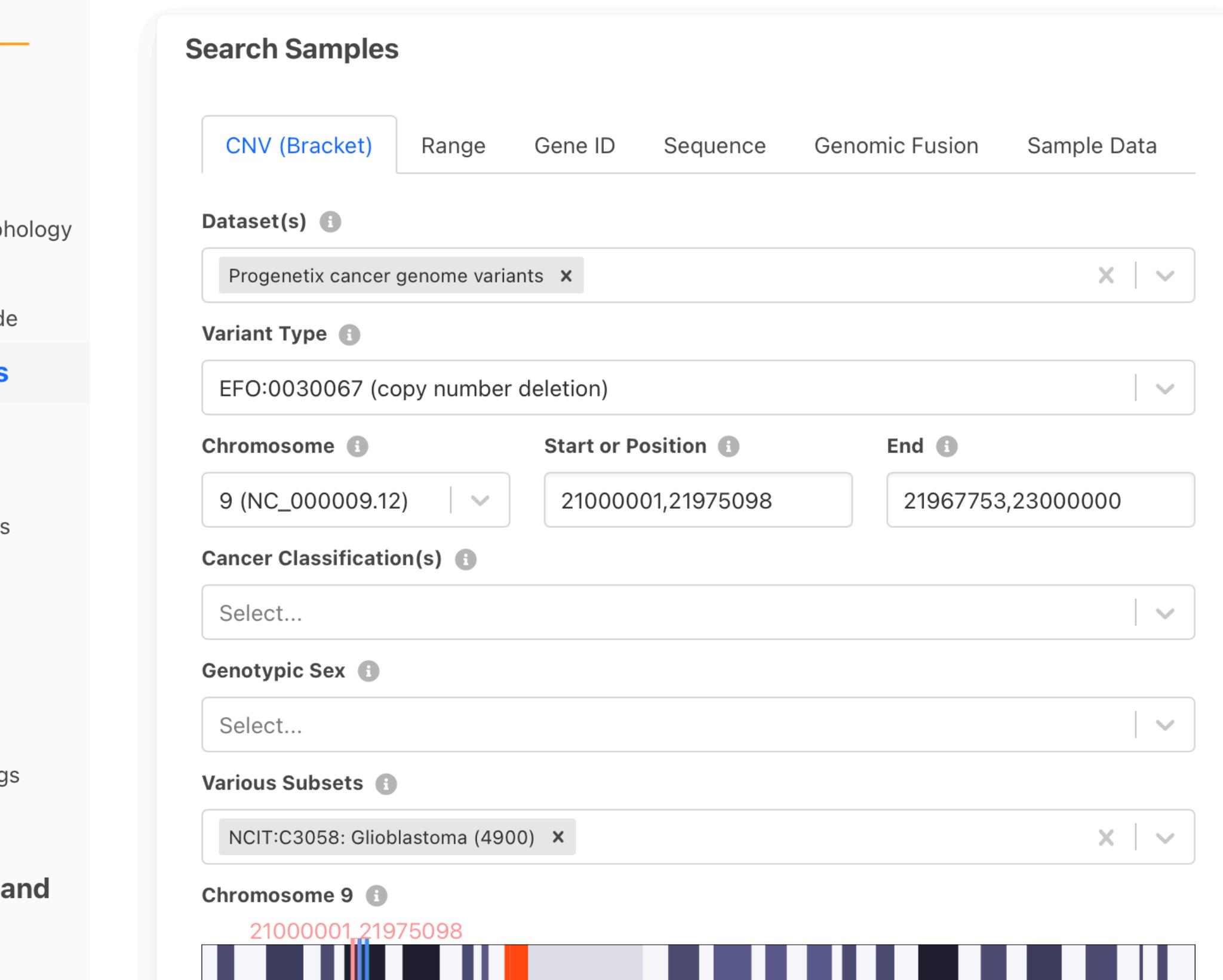
Chromosome 9 21000001,21975098

21967753 23000000

Query Database

Form Utilities Gene Spans Cytoband(s)

Query Examples CNV Example SNV Example Range Example



Cancer Genomics Reference Resource

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

progenetix

- CNV Profiles**
 - ... by NCIT
 - ... by ICD-O Morphology
 - ... by ICD-O Site
 - ... by TNM & Grade
- Search Samples**
- arrayMap**
 - TCGA Data
 - cBioPortal Studies
- Publication DB**
 - Progenetix Use
- NCIT - ICD-O Mappings**
 - UBERON Mappings
- Upload & Plot**
- OpenAPI Paths and Examples**
- Cancer Cell Lines**
- Beacon+**
- Documentation**
- Baudisgroup @ UZH**

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. Please submit information about additional articles to our [Github issue tracker](#) or email us at contact@progenetix.org. The inclusion criteria are described [in the documentation](#).

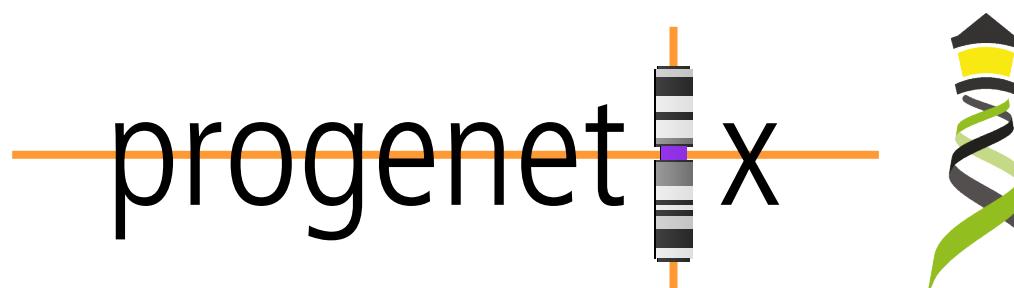
Filter 

Publications (3401)		Samples
id 	Publication	cCGH aCGH WES WGS pgx
pubmed:40379833	Guetter S, König C, Koerkel-Qu H et al. (2025) MCSP⁺ metastasis founder cells activate immunosuppression early in human melanoma metastatic colonization. ... Nat Cancer 	0 0 0 492 0
pubmed:37261085	Tang WF, Fan XJ, Bao H, Fu R, Liang Y et al. (2023) Acquired DNA damage repairs deficiency-driven immune evolution and involved immune factors of ... Oncoimmunology 	0 0 0 73 0
pubmed:37260182	Lee TH, Jang BS, Chang JH, Kim E et al. (2023) Genomic landscape of locally advanced rectal adenocarcinoma: Comparison between bEFOre and after ... Cancer Med 	0 0 0 54 0
pubmed:37255657	Gong C, Zhang W, Sun Y, Shou J, Jiang Z et al. (2023) Exploration of the immunogenetic landscape of hyperprogressive disease after combined immunotherapy in ... iScience 	0 0 0 12 0

Beacon v2 Filters

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

- Beacon v2 relies heavily on "filters"
 - ontology term / CURIE
 - alphanumeric
 - custom
 - Beacon v2 "filters" assumes inclusion of child terms when using hierarchical classifications
 - implicit *OR* with otherwise assumed *AND*
 - implementation of hierarchical annotations overcomes some limitations of "fuzzy" disease annotations



Beacon+ specific: Multiple term selection with OR logic

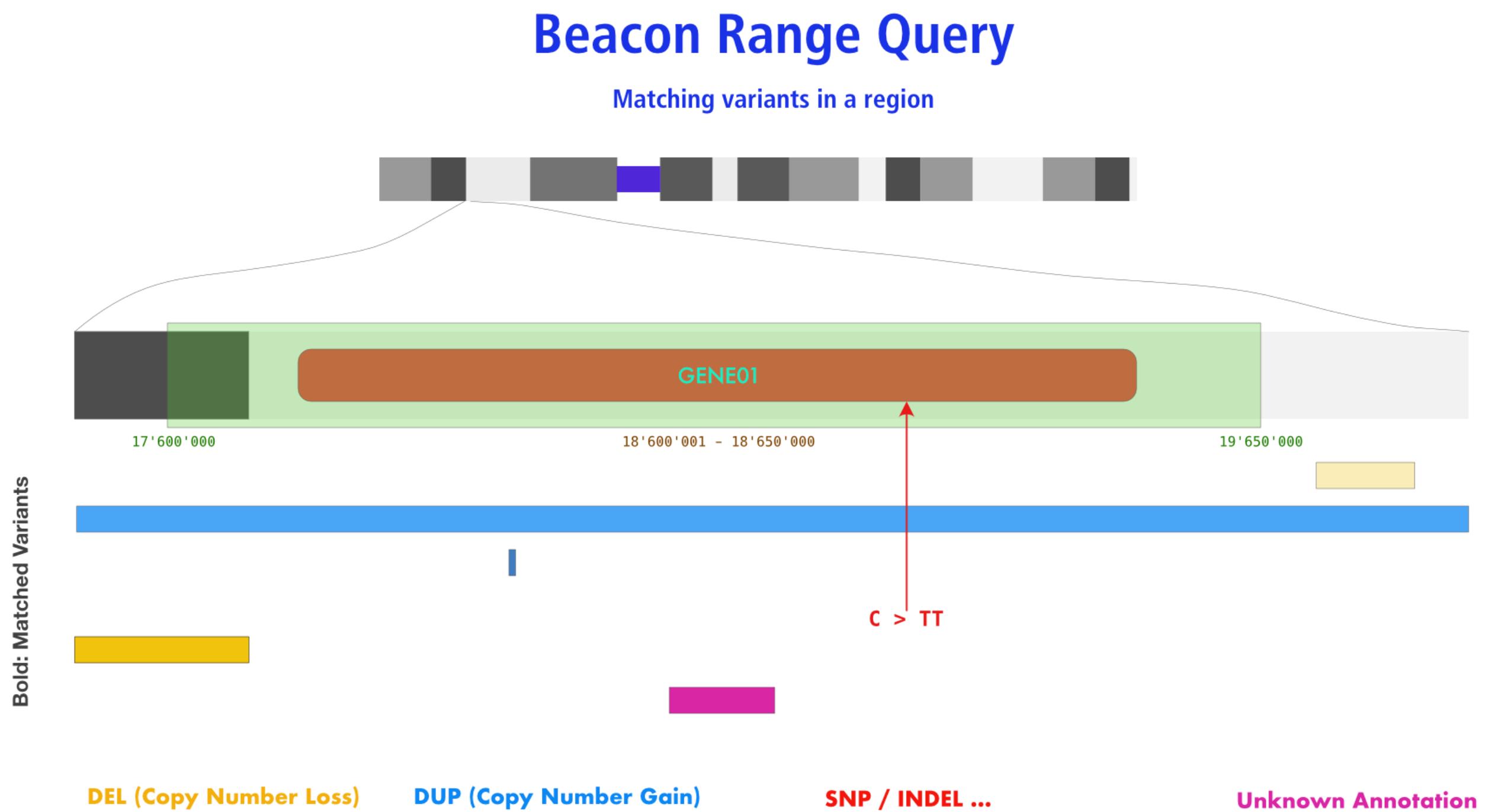
<input checked="" type="checkbox"/>	> NCIT:C4914: Skin Carcinoma	213
<input type="checkbox"/>	> NCIT:C4475: Dermal Neoplasm	109
<input checked="" type="checkbox"/>	▼ NCIT:C45240: Cutaneous Hematopoietic and Lymphoid Cell Neoplasm	310

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

Variation Queries

Range ("anything goes") Request

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



Beacon Query Types

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

Dataset: Test Database - examplez

Chromosome: 17 (NC_000017.11)

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

Start or Position: 7572826

End (Range or Structural Var.): 7579005

Reference Base(s): N

Alternate Base(s): A

Select Filters: Chromosome 17

Query Database

Form Utilities: Gene Spans, Cytoband(s)

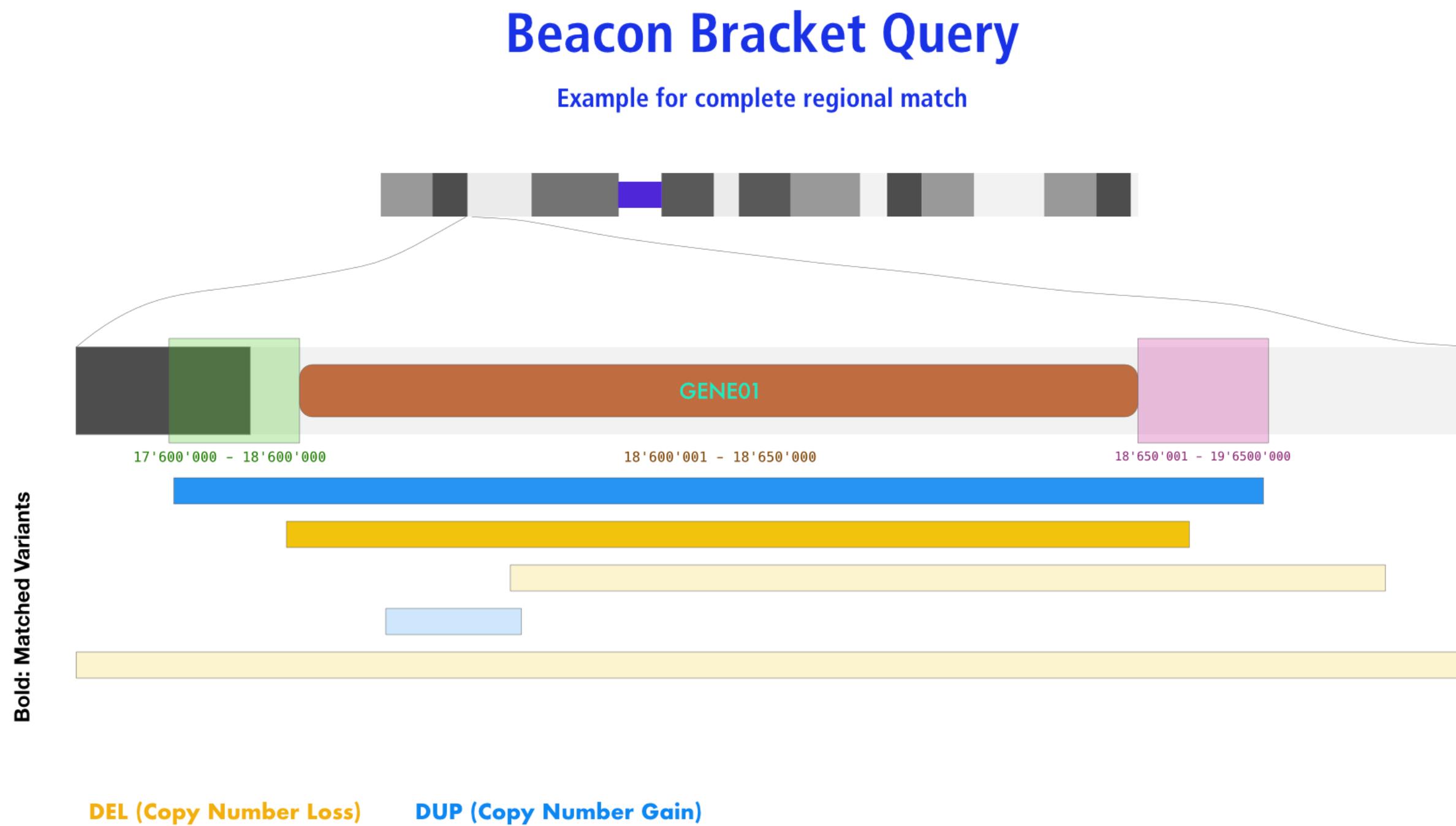
Query Examples: CNV Example, SNV Example, Range Example, Gene Match, Aminoacid Example, Identifier - HeLa

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

Variation Queries

Bracket ("CNV") Query

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



Beacon Query Types

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

Dataset

Test Database - examplez X | ▼

Chromosome

9 (NC_000009.12) | ▼

Variant Type

EFO:0030067 (copy number deletion) | ▼

Start or Position

21000001-21975098

End (Range or Structural Var.)

21967753-23000000

Select Filters

NCIT:C3058: Glioblastoma (100) X | ▼

Chromosome 9

21000001-21975098



Query Database

Form Utilities

Gene Spans

Cytoband(s)

Query Examples

[CNV Example](#)

[SNV Example](#)

[Range Example](#)

[Gene Match](#)

[Aminoacid Example](#)

[Identifier - HeLa](#)

This example shows the query for CNV deletion variants overlapping the CDKN2A gene's coding region with at least a single base, but limited to "focal" hits (here i.e. $\leq \sim 2\text{Mbp}$ in size). The query is against the examplez collection and can be modified e.g. through changing the position parameters or data source.

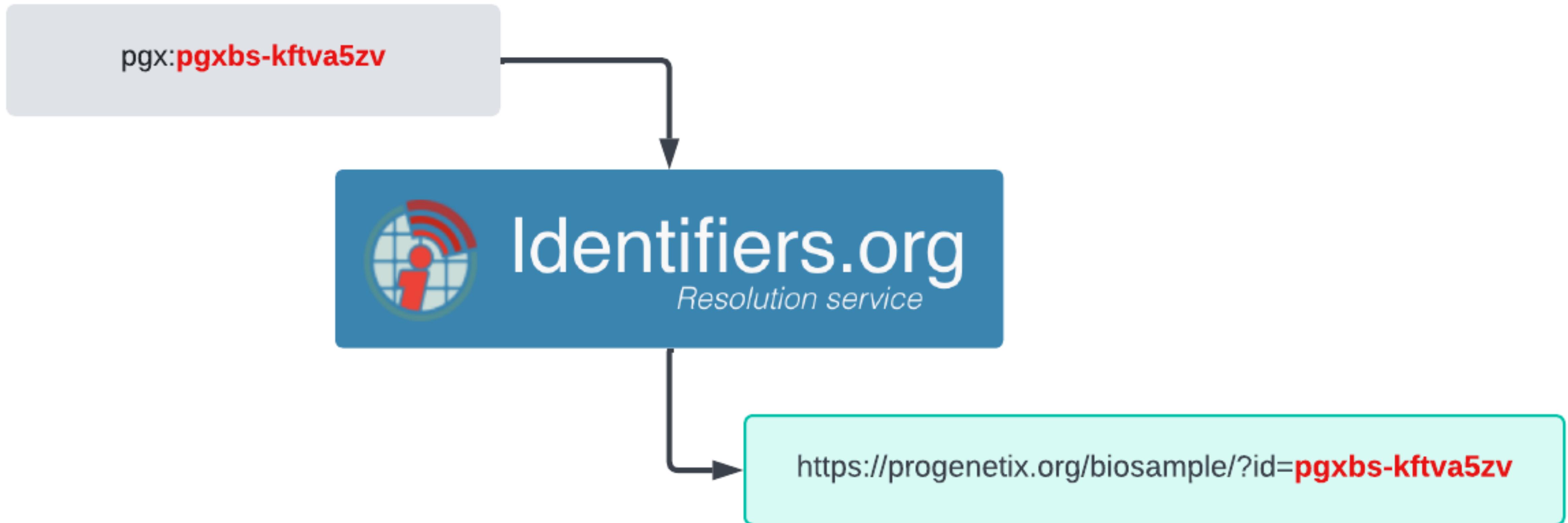
Standards Development & Implementation: CNV Terms

in computational (file/schema) formats

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

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

CURIE Resolution: Identifiers.org



Website populated by asynchronous retrieval of Beacon query results using handovers

 [Edit Query](#)

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

Search Samples

arrayMap
TCGA Data
cBioPortal Studies

Publication DB
Progenetix Use

NCIT - ICD-O Mappings
UBERON Mappings

Upload & Plot

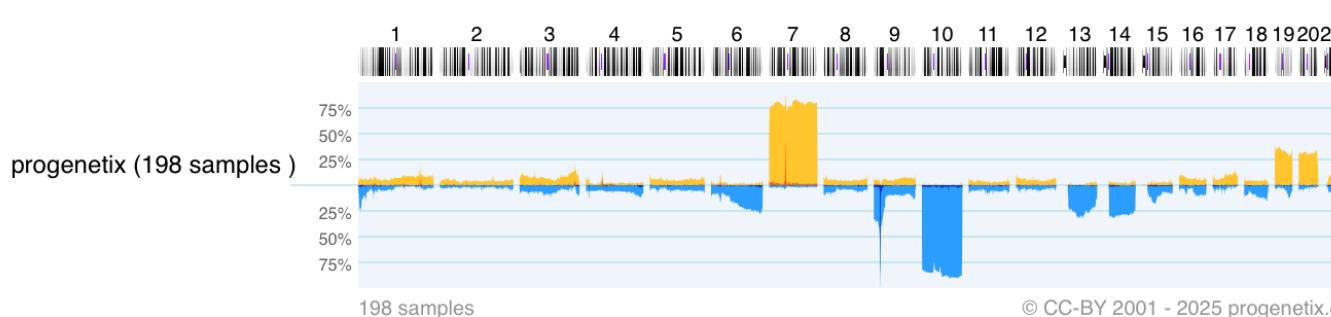
OpenAPI Paths and Examples

Cancer Cell Lines

progenetix

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

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



[Reload histogram in new window](#) 

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

progenetix Data Downloads

Download Sample Data (TSV)
Part1  Part2  Part3  Part4  Part5 

Download Sample Data (JSON)
Part1  Part2  Part3  Part4  Part5 

Download Variants (Beacon VRS)
Part1  Part2  Part3  Part4  Part5 

Download Variants (VCF)

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

Website populated by asynchronous retrieval of Beacon query results using handovers

progenetix

[Edit Query](#)

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

Search Samples

arrayMap
TCGA Data
cBioPortal Studies

Publication DB
Progenetix Use

NCIT - ICD-O Mappings
UBERON Mappings

Upload & Plot

OpenAPI Paths and Examples

Cancer Cell Lines

Chro: refseq:NC_000009.12 **Start:** 21000001,21975098 **End:** 21967753,23000000 **Type:** EFO:0030067
Filters: NCIT:C3058

progenetix

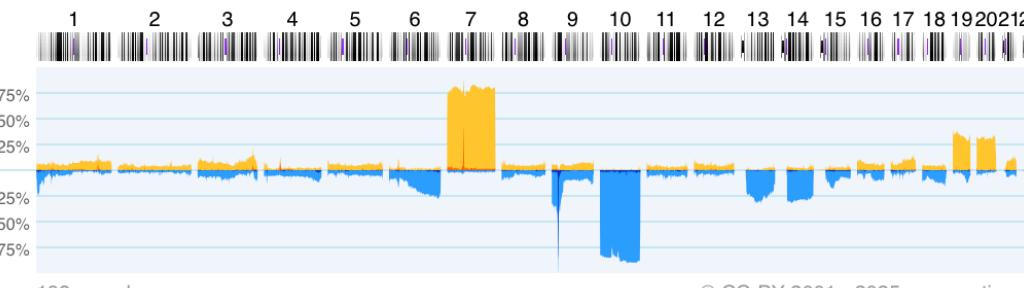
Matched Samples: 969 Retrieved Samples: 200 Variants: 984 Calls: 976

UCSC region Geographic Map Variants in UCSC Dataset Responses (JSON)

[Visualization options](#)

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

progenetix (198 samples)



© CC-BY 2001 - 2025 progenetix.org

[Reload histogram in new window](#)

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

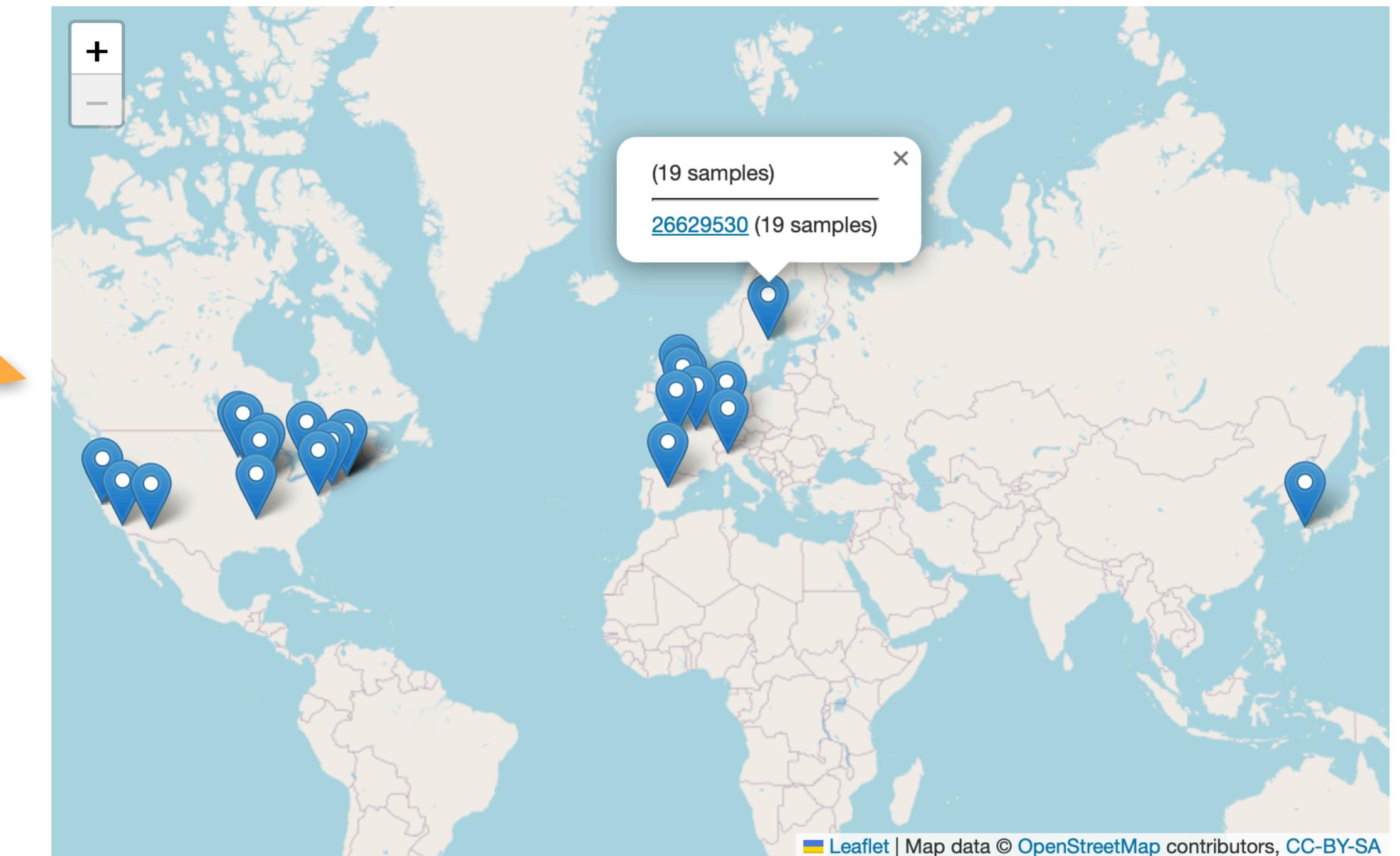
progenetix Data Downloads

[Download Sample Data \(TSV\)](#)
Part1 Part2 Part3 Part4 Part5

[Download Sample Data \(JSON\)](#)
Part1 Part2 Part3 Part4 Part5

[Download Variants \(Beacon VRS\)](#)
Part1 Part2 Part3 Part4 Part5

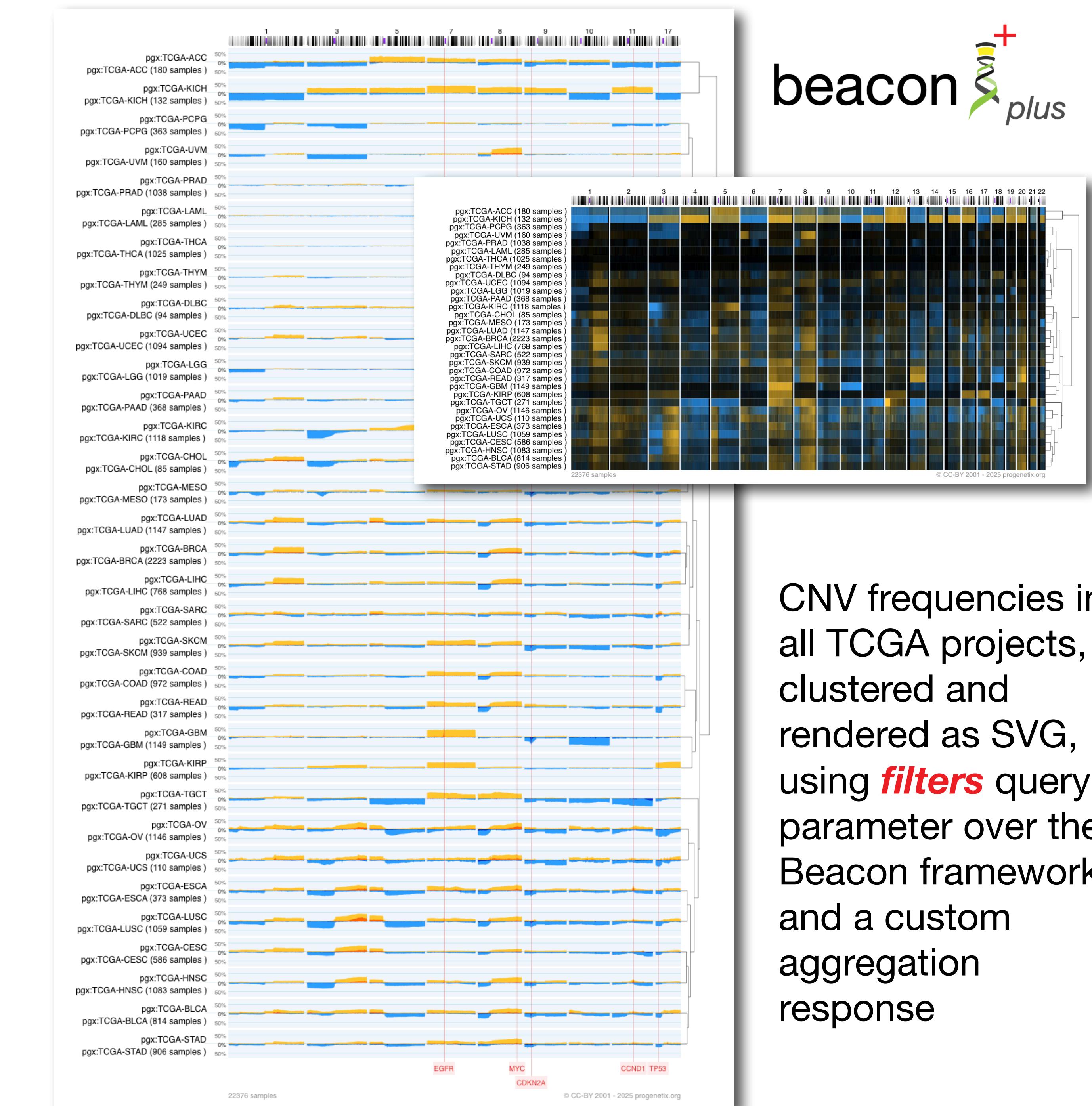
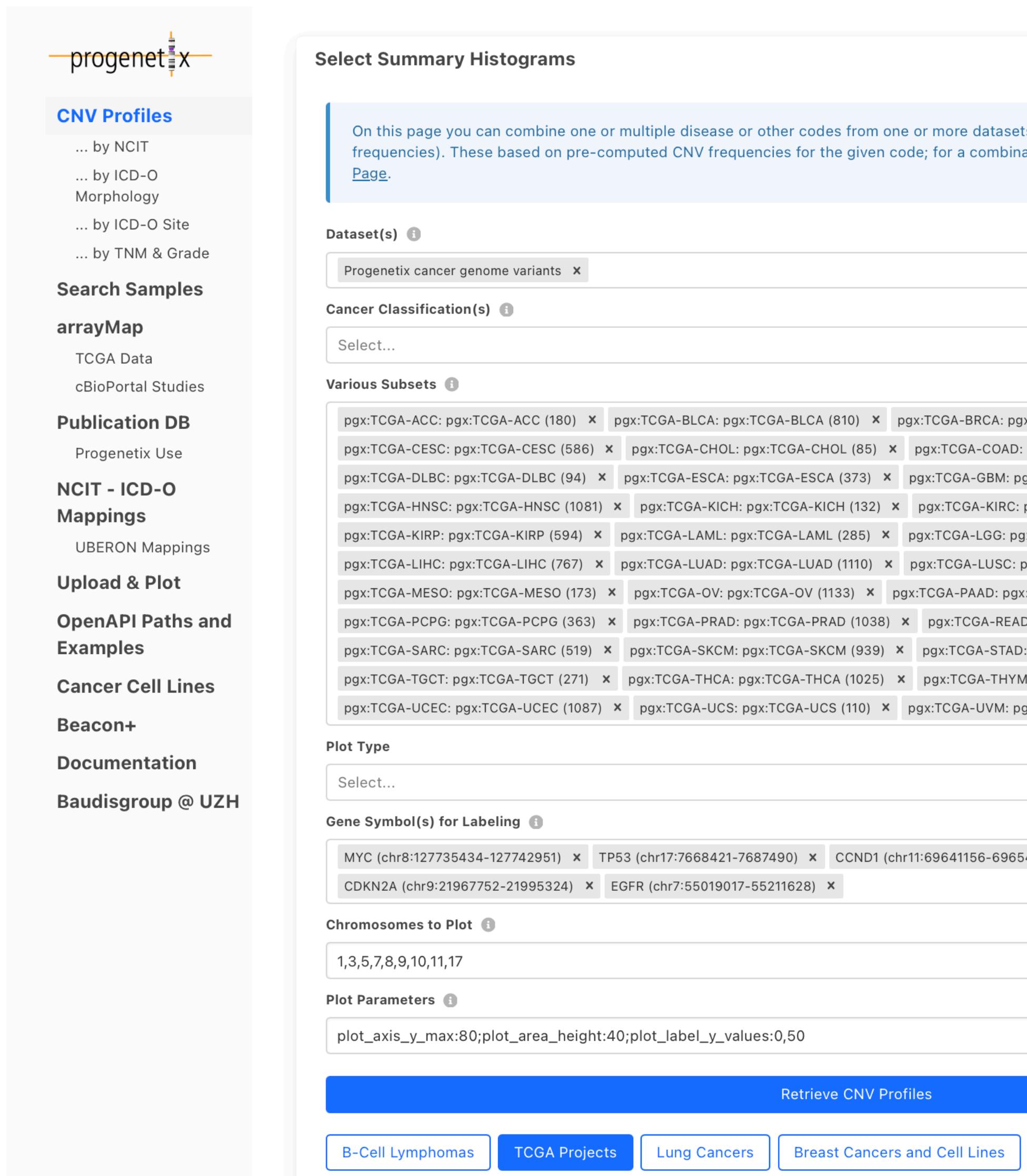
[Download Variants \(VCF\)](#)
Part1 Part2 Part3 Part4 Part5



Pushing the standard: Biosamples in Progenetix have geographic attribution in the form of GeoJSON objects, for query & display...

Pushing the envelope...

Custom Beacon aggregation response for displaying CNV frequencies



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

Beacon⁺: Phenopackets

Testing alternative response schemas...

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

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

```

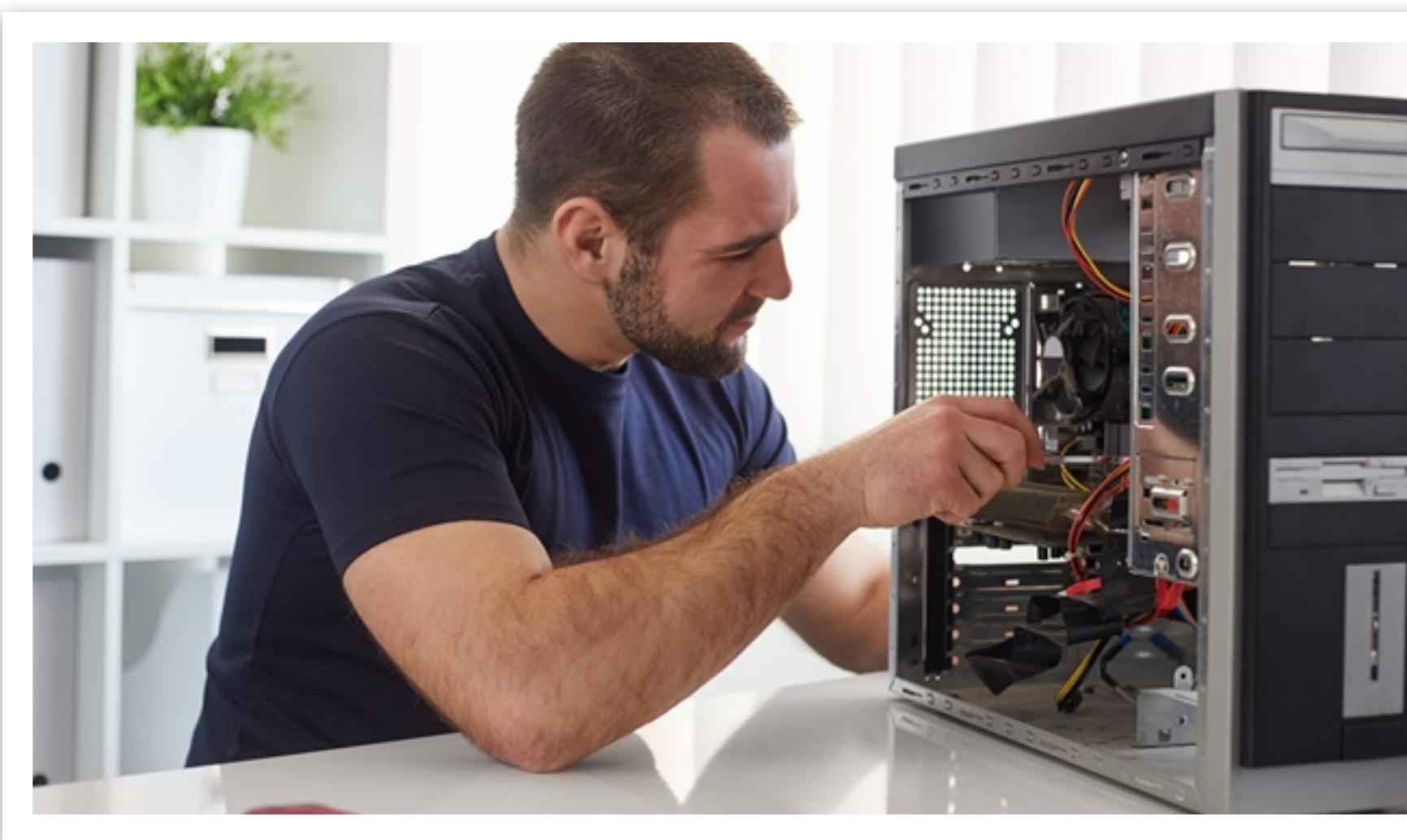
    "id": "pgpxpf-kftx3tl5",
    "metaData": {
      "phenopacketSchemaVersion": "v2",
      "resources": [
        {
          "id": "NCIT",
          "iriPrefix": "http://purl.obolibrary.org/obo/NCIT_",
          "name": "NCIt Plus Neoplasm Core",
          "namespacePrefix": "NCIT",
          "url": "http://purl.obolibrary.org/obo/ncit/neoplasm-core.c",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

```



Beacon v2 deployment

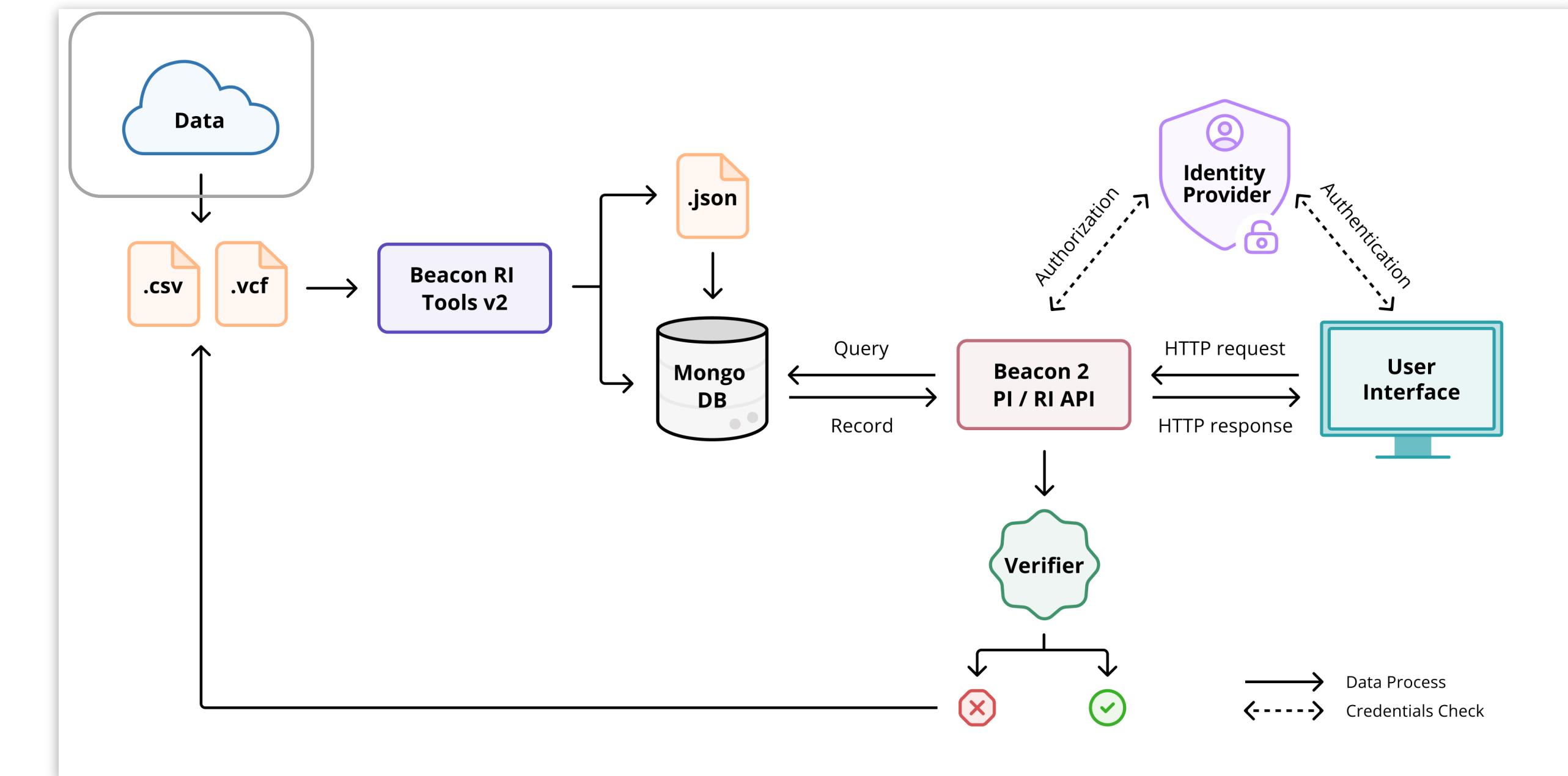
Build it yourself



Beacon v2 API

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

Toolkit for production environments



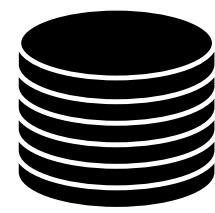
Beacon v2 Production Implementation (released Oct 2024)

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

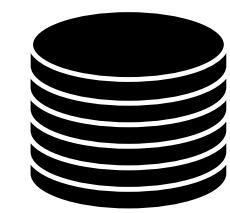
bycon based Beacon+ Stack

progenetix

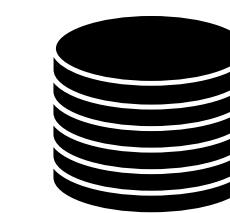
- *collations* contain pre-computed data (e.g. CNV frequencies, statistics) and information for all grouping entity instances and correspond to **filter values**
 - ▶ [pubmed:10027410](#), [NCIT:C3222](#), [pgx:cohort-TCGA](#), [pgx:icdom-94703...](#)
 - ▶ precomputed frequencies per collection informative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



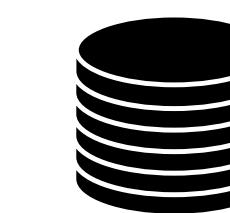
variants



analyses



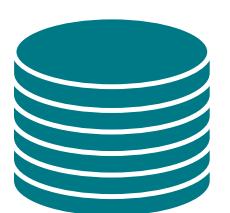
biosamples



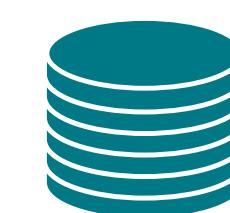
individuals



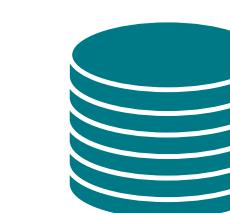
collations



geolocs



genespans



qBuffer

Entity collections

Utility collections

github.com/progenetix/bycon





front-end showcases query strategies

Beacon Search Demonstrator

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

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

Search Samples

Compare CNV Profiles

CNV Profiles by Cancer Type

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

OpenAPI Paths and Examples

Documentation

Progenetix

Baudisgroup @ UZH

Search Samples

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

Dataset(s) Test Database - examplez

Variant Type EFO:0030067 (copy number deletion)

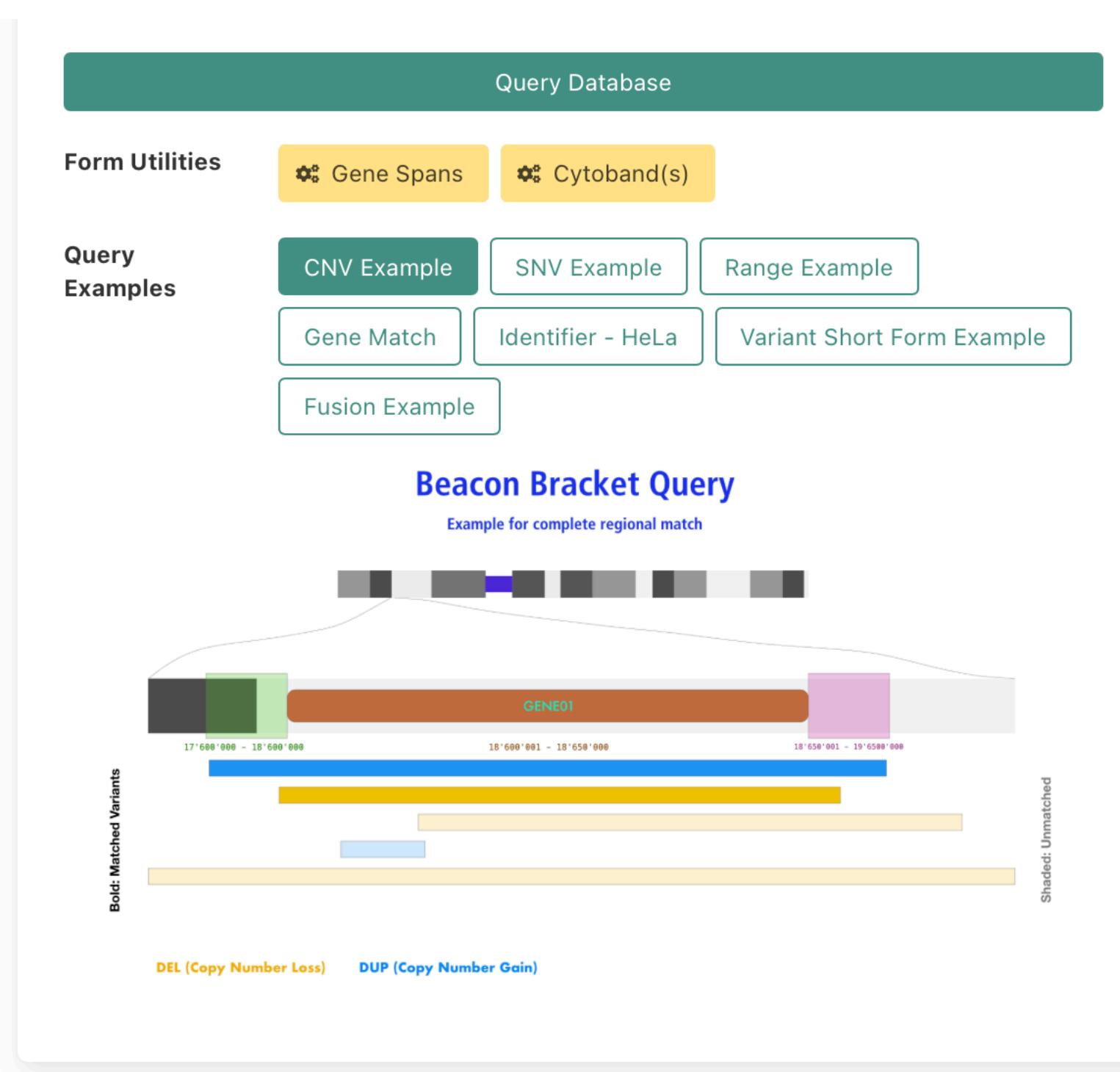
Chromosome 9 (NC_000009.12) **Start or Position** 21000001,21975098 **End (Range or Structural Var.)** 21967753,23000000

Cancer Classification(s) Select...

Genotypic Sex Select...

Various Subsets NCIT:C3058: Glioblastoma (28)

Chromosome 9 21000001,21975098 21967753,23000000



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

Progenetix Cancer Genomics Beacon+

/api

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

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

bycon and Data Aggregation

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

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

Contact the developer

Servers

<https://progenetix.org>

Beacon

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

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

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

[GET /beacon/g_variants](#) Get genomicVariant entries

[GET /beacon/g_variants/{id}](#) Get genomicVariant entries

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

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

[GET /beacon/g_variants/{id}/individuals](#) Get individual entries

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

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

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

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

Bv2minimalAlleleRequest

[GET /beacon/g_variants](#) Get genomicVariant entries

Parameters

Name Description

referenceName string (query) Examples: Chromosome 17

start array<integer> (query) Examples: Base position on chromosome 17

alternateBases string (query) Examples: An 'A' allele at the specified position

referenceBases string (query) Examples: A reference 'G' allele at the specified position

skip integer (query) Examples: G

limit integer (query) Examples: skip

requestedGranularity string (query) Examples: limit

Examples: The minimal boolean response

boolean

Bv2cnvbracketquery

[GET /beacon/g_variants](#) Get genomicVariant entries

Get genomicVariant entries

Parameters

Name Description

filters array<string> (query) Examples: Glioblastoma

NCIT:C3058

Add string item

referenceName string (query) Examples: Chromosome 9 (GRCh38)

refseq:NC_000009.12

start array<integer> (query) Examples: Range for start of CNV involving CDKN2A

21000001

21975098

Add integer item

end array<integer> (query) Examples: Range for end of CNV involving CDKN2A

21967753

23000000

Add integer item

variantType string (query) Examples: High-level copy number loss

EFO:0020073

VQSadjacencyRequest

[GET /beacon/g_variants](#) Get genomicVariant entries

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

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

Get biosample entries

Parameters

Name Description

filters array<string> (query) Examples: Malignant lymphoma, NOS (ICD-O 3 code 9680/3)

pgx:icd0-95903

Add string item

referenceAccession string (query) Examples: RefSeq ID for Chromosome 8 (GRCh38)

refseq:NC_000008.11

breakpointRange array<integer> (query) Examples: Range for band q24 on chromosome 8

11670000

145138636

Add integer item

adjacencyAccession string (query) Examples: RefSeq ID for Chromosome 14 (GRCh38)

refseq:NC_000014.9

adjacencyRange array<integer> (query) Examples: Range for band q32 on chromosome 14

89300000

107043718

Add integer item

vrsType string (query) Examples: Adjacency

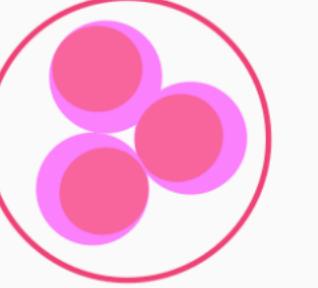
Adjacency

Cancer Cell Lines

&&

refCNV

cancercelllines.org
refcnv.org



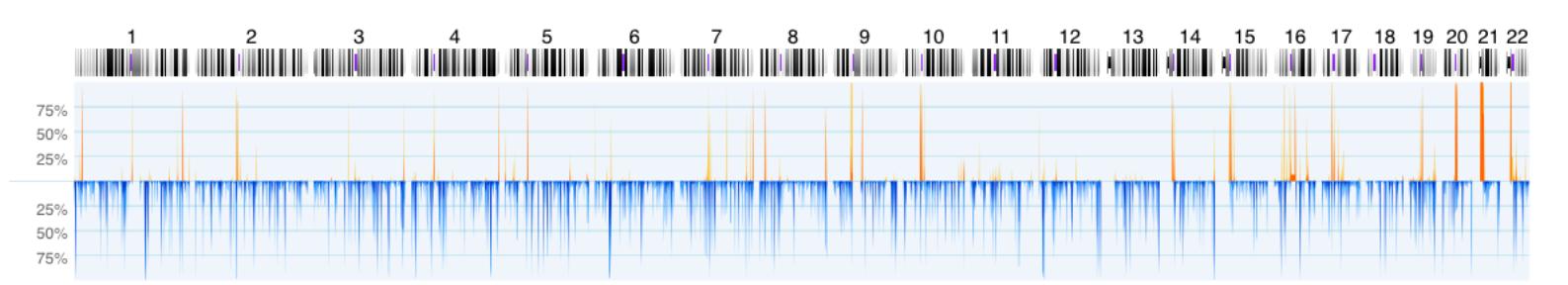
cancercelllines

Genomic Copy Number Variation (CNV) data from reference samples

Under Construction

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

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



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

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



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

Assembly: GRCh38 Chro: NC_000007.14 Start: 140713328 End: 140924929

Type: SNV

cellz

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

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

Visualization options

Results Biosamples Variants Annotated Variants

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

hierarchies

Cell Line Details

HOS (cellosaurus:CVCL_0312)

Subset Type

- Cellosaurus - a knowledge resource on cell lines cellosaurus:CVCL_0312 ↗

Sample Counts

- 204 samples
- 57 direct cellosaurus:CVCL_0312 code matches
- 21 CNV analyses

Search Samples

Select cellosaurus:CVCL_0312 samples in the Search Form

Raw Data (click to show/hide)

562/Ad...

umi-1 (9)

97 (2 sam...

(11 sam...

:U-1 (1 s...

M-3 (1 s...

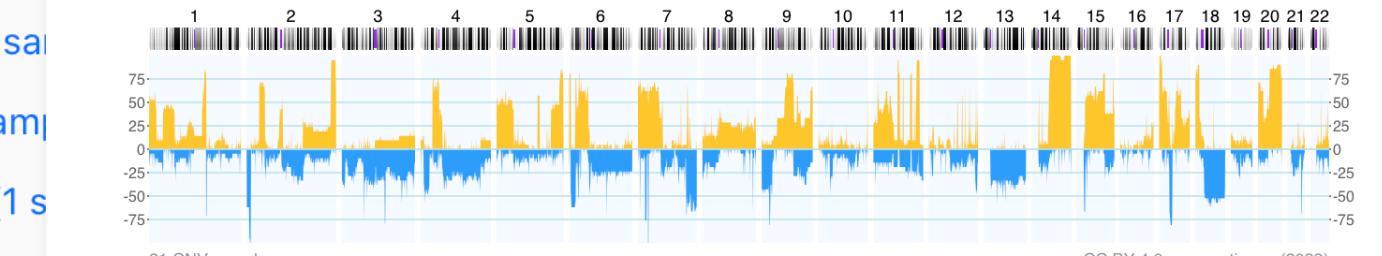
JOI-90 (1

Reh/Eph

NSU-CL

C827 (27)

HOS (cellosaurus:CVCL_0312)



Download SVG | Go to cellosaurus:CVCL_0312 | Download CNV Frequencies

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



bycon Documentation

Documentation Home

Recent Changes

Setup & Maintenance

Installation

Importing Data

Housekeeping

Beacon API

Beacon API

Services API

API Parameters

Front End

Code Repositories

bycon

Progenetix Front End

More Info

Progenetix Site

baudisgroup@UZH

Beacon Documentation

Changes & To Do

Changes Tracker

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

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

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

Table of contents

Changes Tracker

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

2025-05-02 (v2.4.2)

2025-04-25 (v2.4.1)

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

2025-04-15 (v2.3.1)

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

2025-03-10 (v2.2.6)

2025-03-06 (v2.2.5)

2025-03-03 (v2.2.4)

2025-02-26 (v2.2.3)

2025-02-21 (v2.2.2)

2025-02-21 (v2.2.1)

2025-02-14 (v2.2.0)

2025-02-08 (v2.1.5)

2025-01-29 (v2.1.4)

2025-01-16 (v2.1.3)

2024-12-20 (v2.1.2)

2024-12-19 (v2.1.1)

2024-12-09 (v2.1.0)

pgxRpi: an R/Bioconductor package

Client for Accessing Beaconized Data

- **Query and export variants**

https://progenetix.org/beacon/biosamples/pgxbs-kftvh94d/g_variants

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

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

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

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

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

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

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

- **Process local .pgxseg files**

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

pgxRpi

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

R wrapper for Progenetix

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

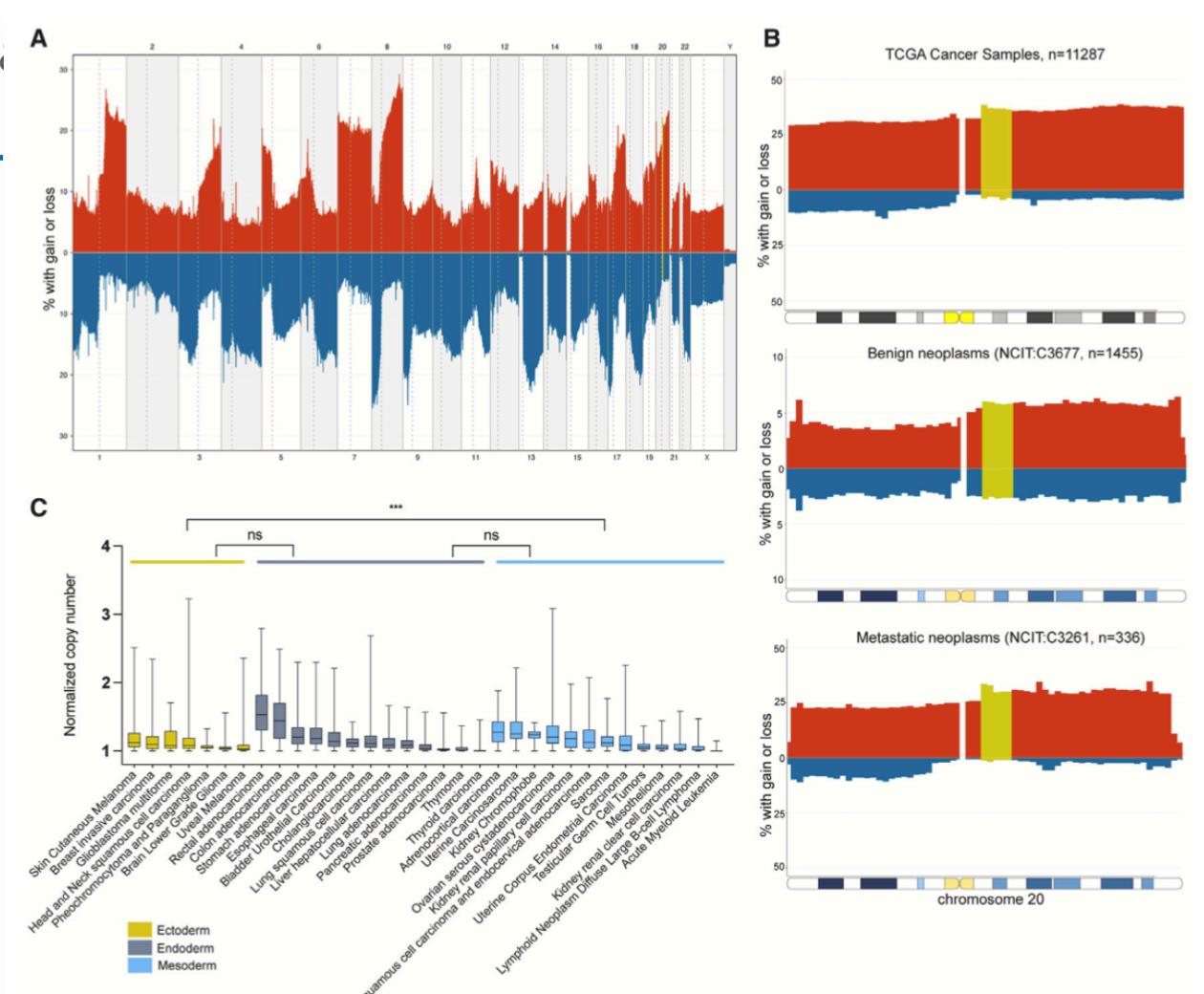
DOI: [10.18129/B9.bioc.pgxRpi](https://doi.org/10.18129/B9.bioc.pgxRpi)

Bioconductor version: Development (3.20)

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

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

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



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

Stem Cell Reports Review



OPEN ACCESS

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

Nuša Krivec,^{1,2} Manjusha S. Ghosh,^{1,2} and Claudia Spits^{1,2,*}

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

²These authors contributed equally

*Correspondence: claudia.spits@vub.be

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

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

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

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

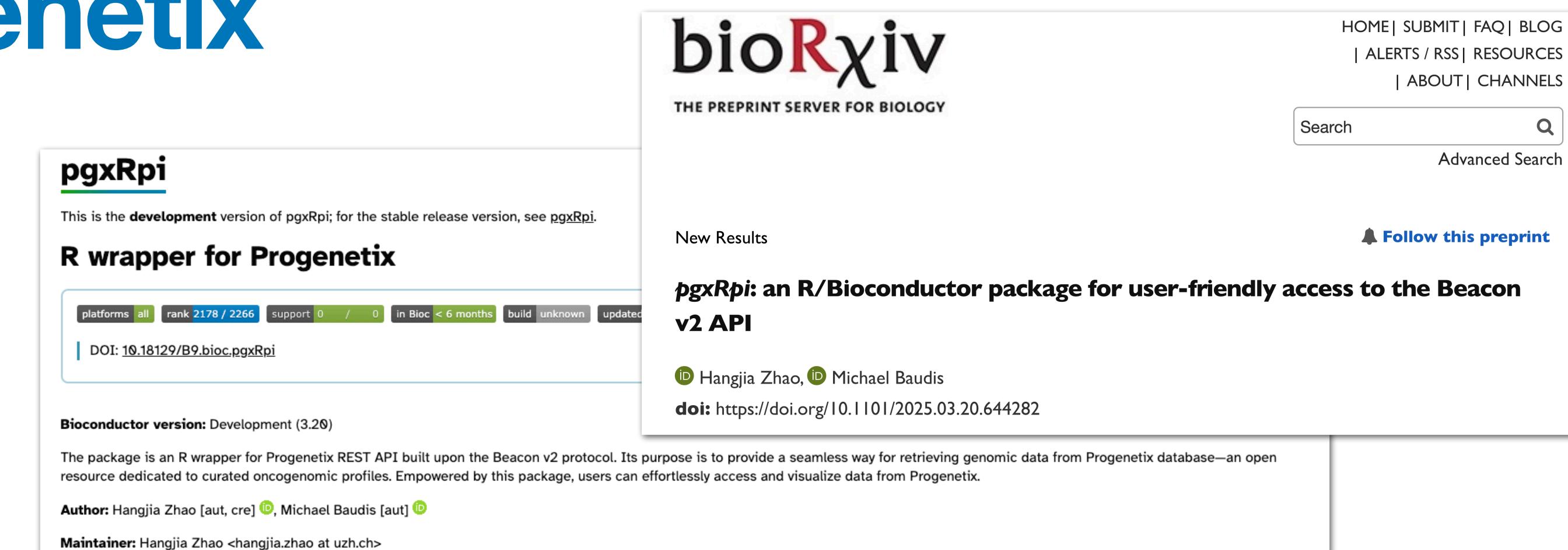
Client for Accessing Progenetix

pgxRpi: an R/Bioconductor package

- **Query and export variants**

https://progenetix.org/beacon/biosamples/pgxbs-kftvh94d/g_variants

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



The image shows a bioRxiv preprint page for the package pgxRpi. The title is "pgxRpi: an R/Bioconductor package for user-friendly access to the Beacon v2 API". It features a DOI: 10.1101/2025.03.20.644282. The page includes sections for "New Results", "Bioconductor version: Development (3.20)", and a brief description of the package's purpose: "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." It also lists authors: Hangjia Zhao and Michael Baudis.

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

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

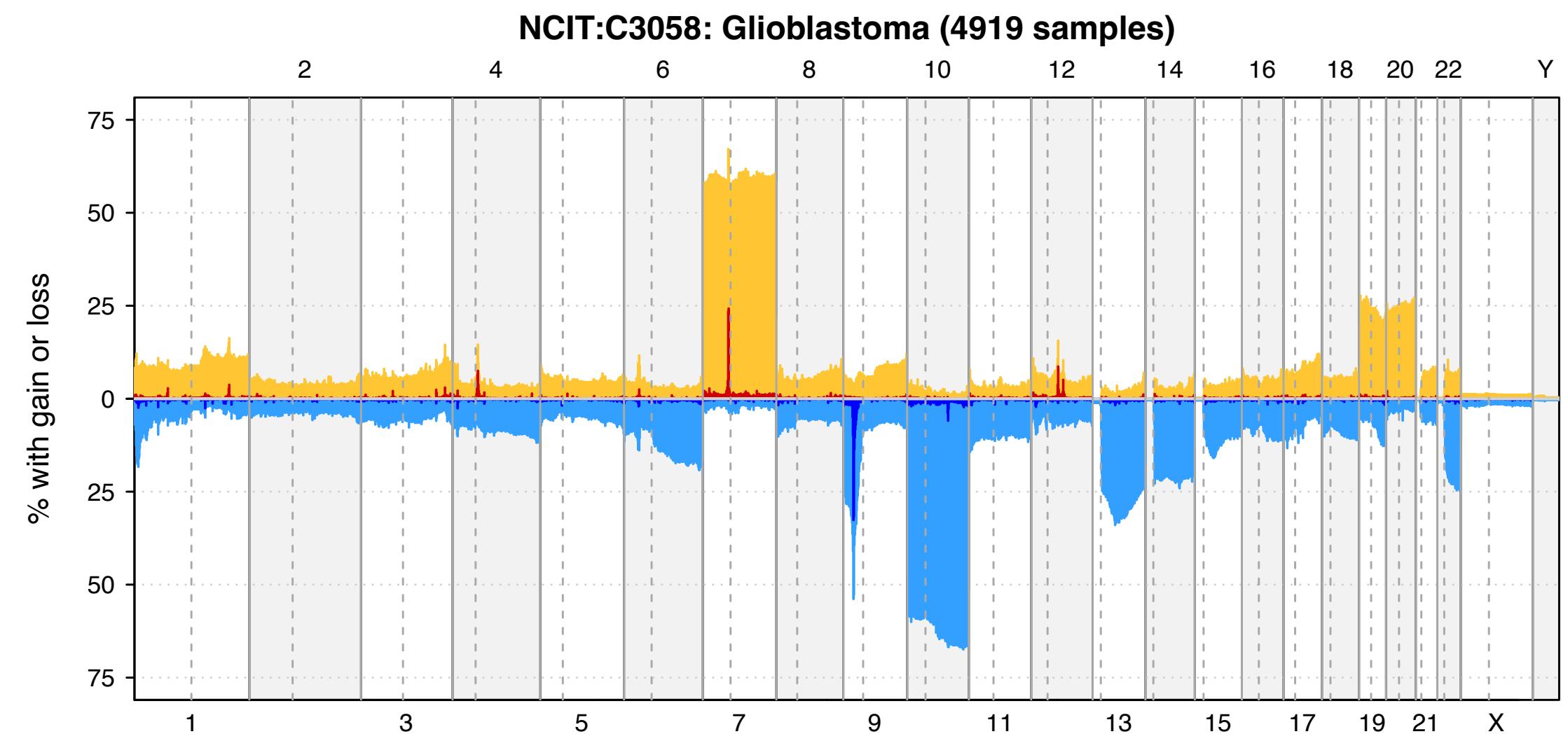
```
biosamples <- pgxLoader(  
  type = "biosamples",  
  filters = "NCIT:C3697"  
)
```

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

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

```
freq <- pgxLoader(  
  type = "cnv_frequency",  
  filters = "NCIT:C3058")  
  
pgxFreqplot(freq)
```

- **Process local .pgxseg files**



Client for Accessing Progenetix

pgxRpi: an R/Bioconductor package

- **Query and export variants**

https://progenetix.org/beacon/biosamples/pgxbs-kftvh94d/g_variants

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

- **Query metadata of biosamples**

<http://progenetix.org/services/>

```
biosamples <- pgxLoader(  
  type = "biosample",  
  filters = "NCIT:C3512",  
)
```

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

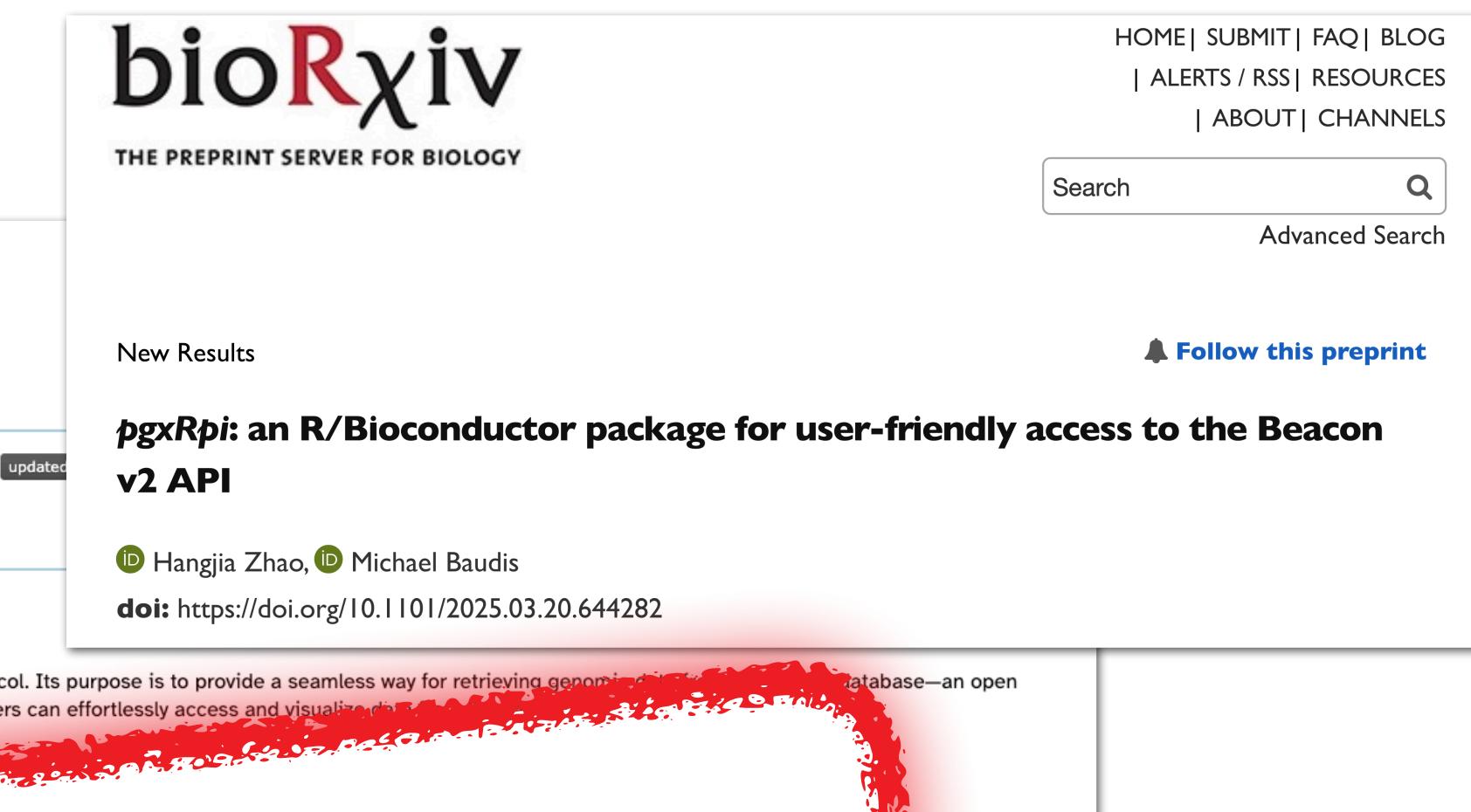
<http://www.progenetix.org/services/intervalFrequencies/?>

```
filters=NCIT:C3512&output=pgxfreq
```

```
freq <- pgxLoader(  
  type = "cnv_frequency",  
  filters = "NCIT:C3058")
```

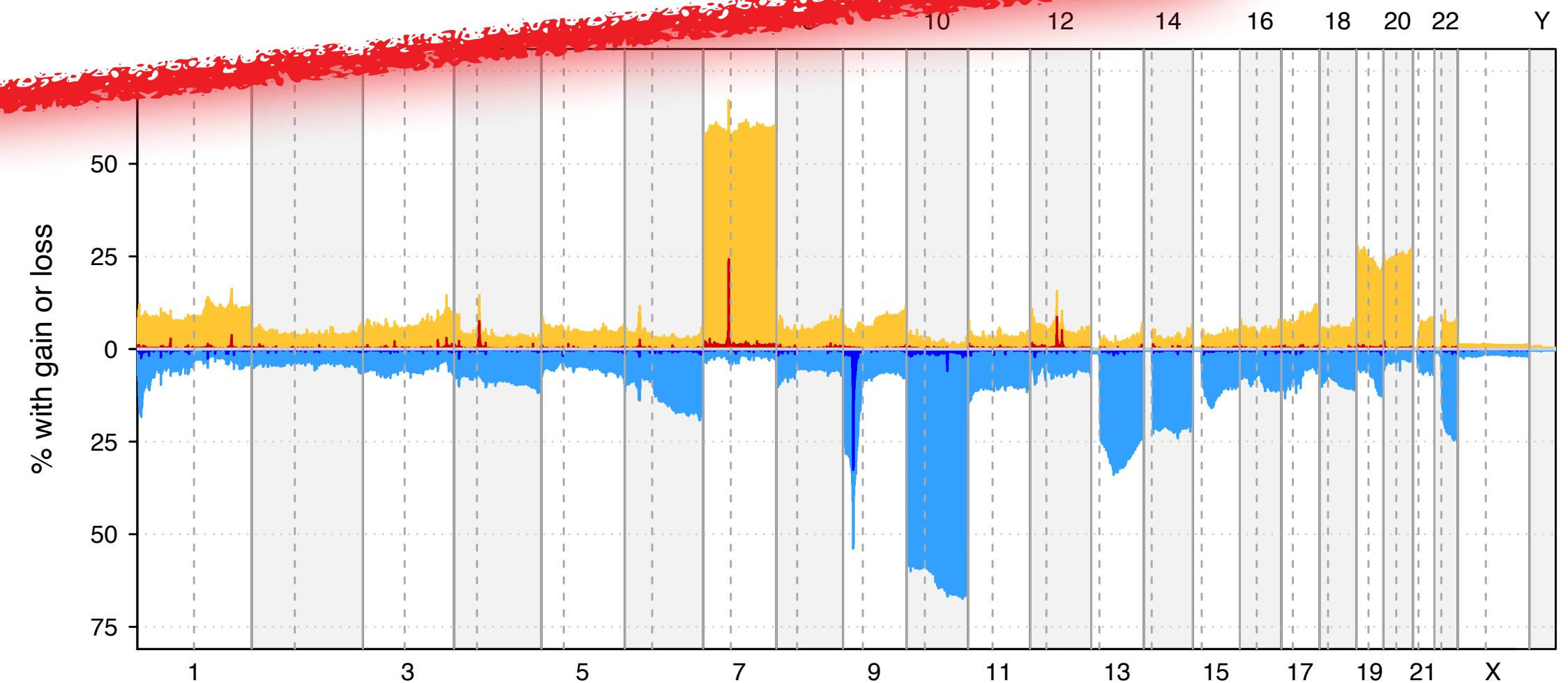
```
pgxFreqplot(freq)
```

- **Process local .pgxseg files**

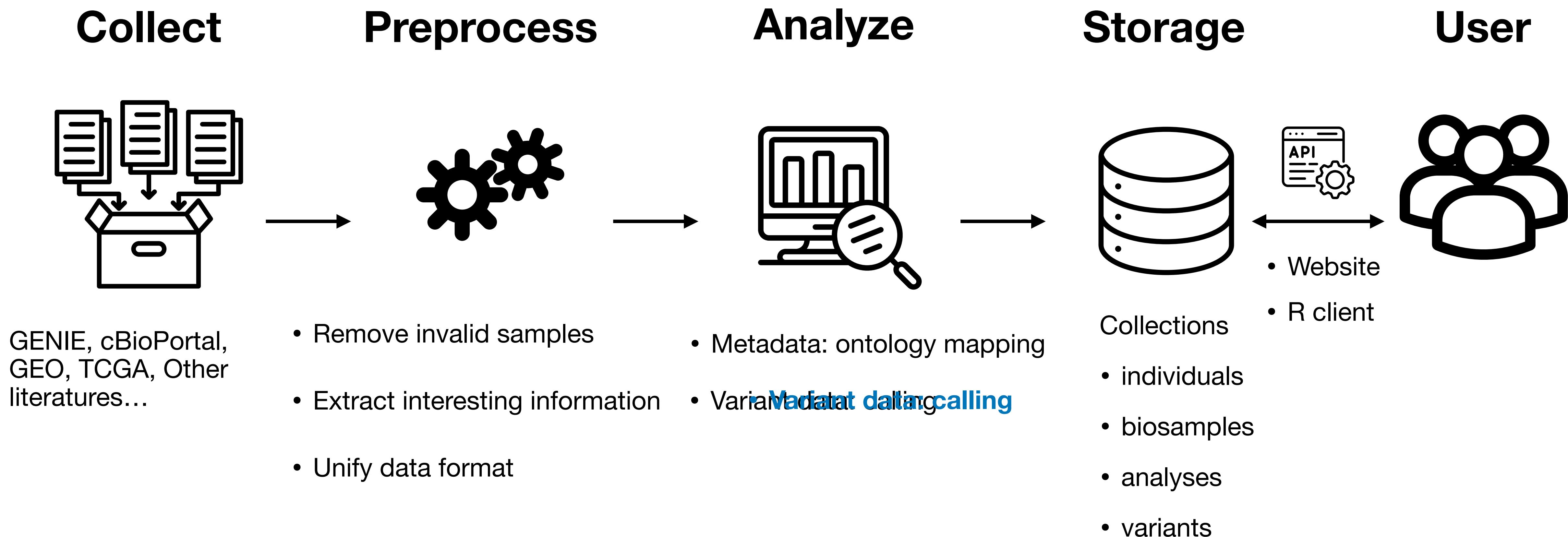


The image shows a bioRxiv preprint page for the package pgxRpi. The title is "R wrapper for Progenetix". It includes a DOI link (10.18129/B9.bioc.pgxRpi), Bioconductor version (Development 3.20), and author information (Hangjia Zhao, Michael Baudis). A red box highlights the title and some text below it.

ACCEPTED AT
BIOINFORMATICS ADVANCES

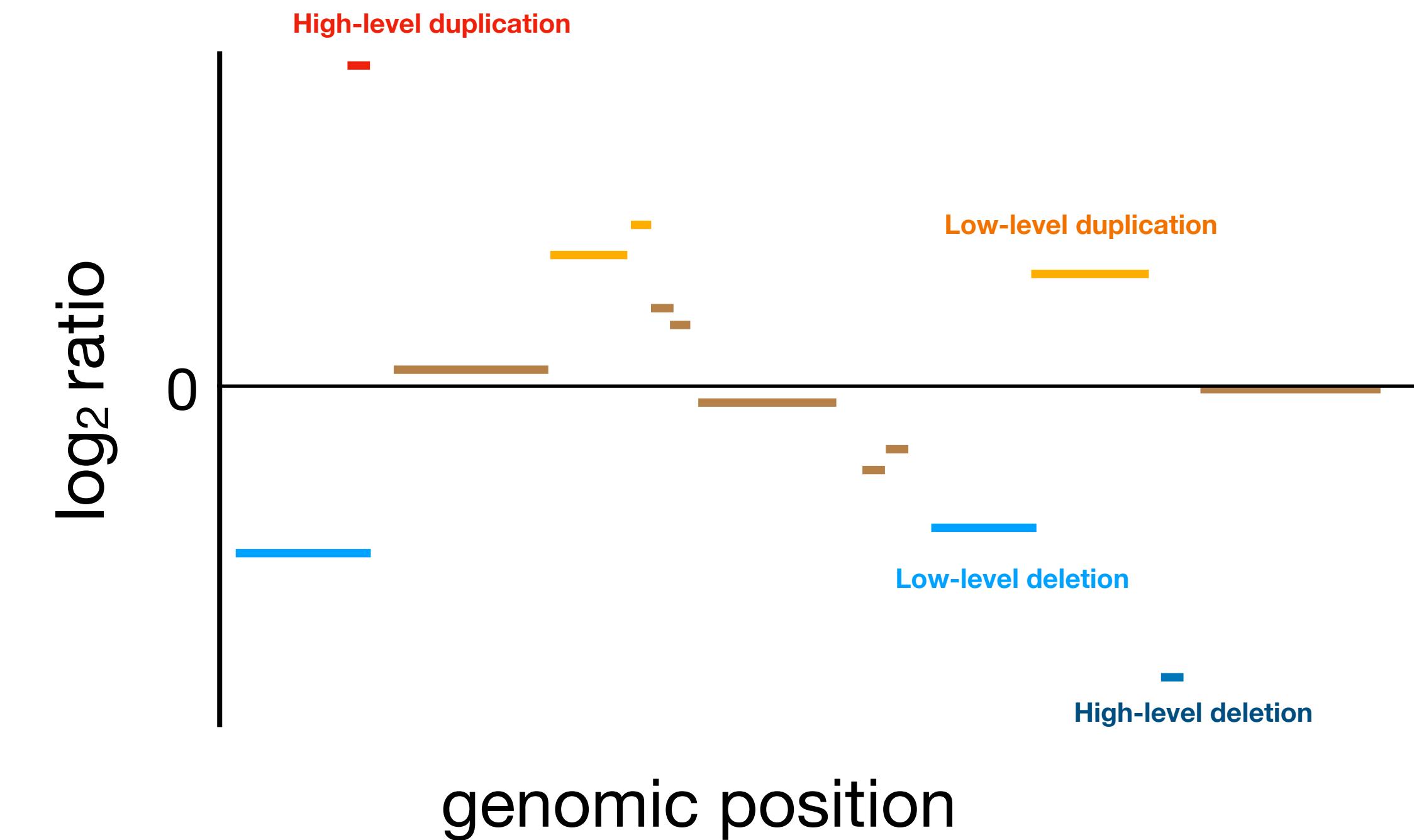
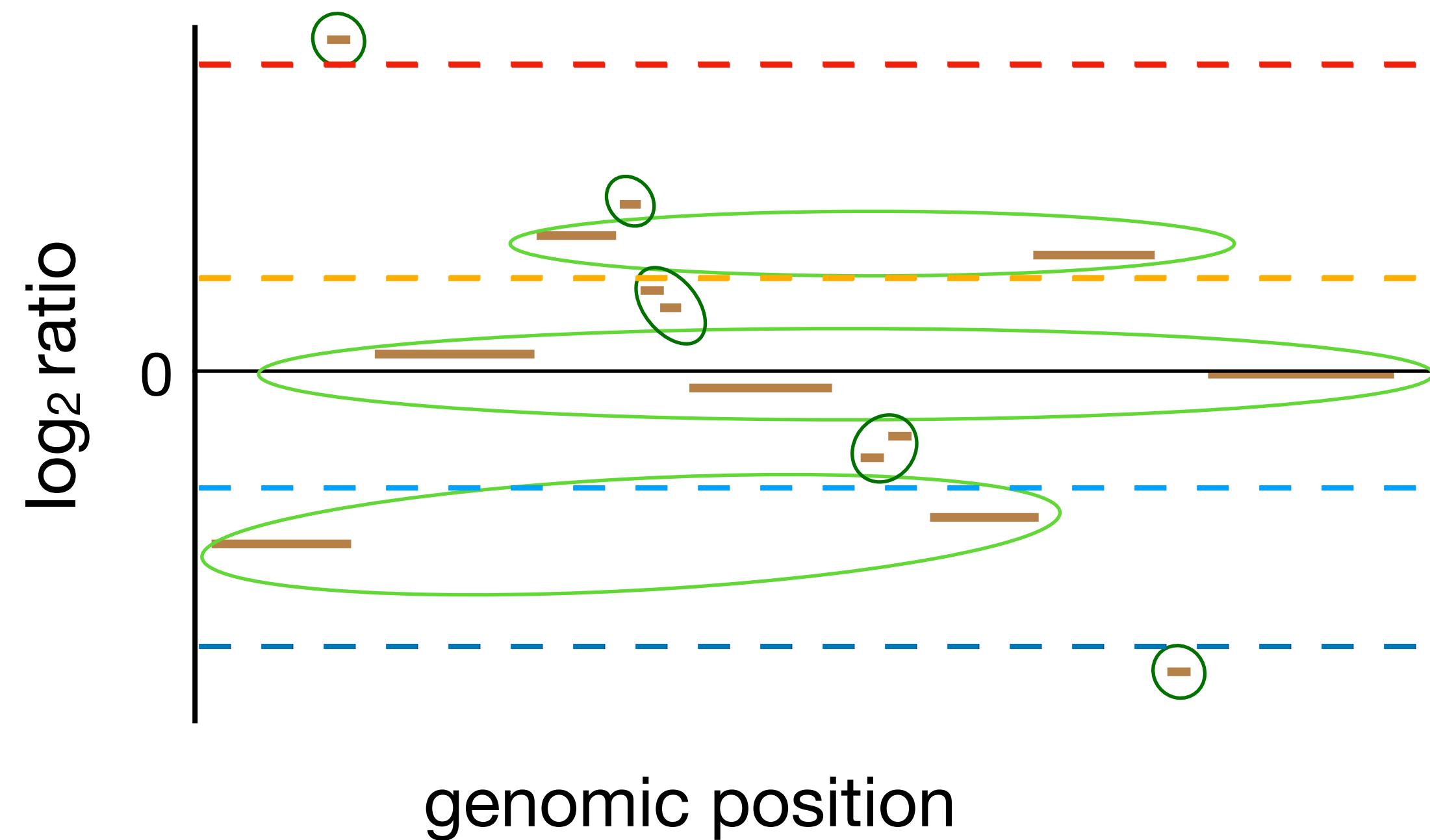


End-to-End Data Pipeline



labelSeg: segment annotation for tumor copy number alteration profiles

Method



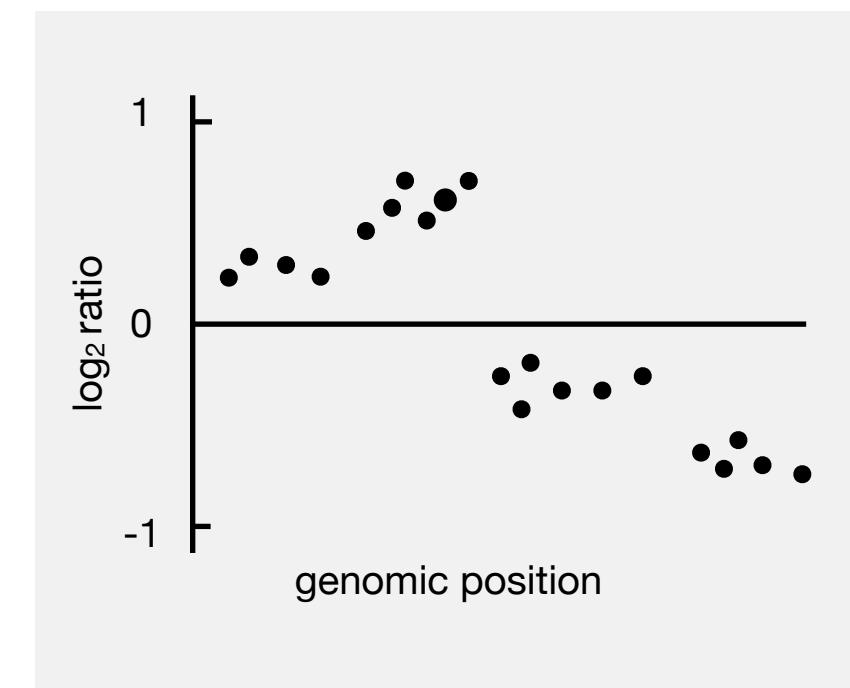
Leveraging **density-based clustering** and exploiting **the length–amplitude relationships of SCNA**, labelSeg proficiently identifies distinct relative copy number states from individual segment profiles.

Somatic CNV Calling Challenge

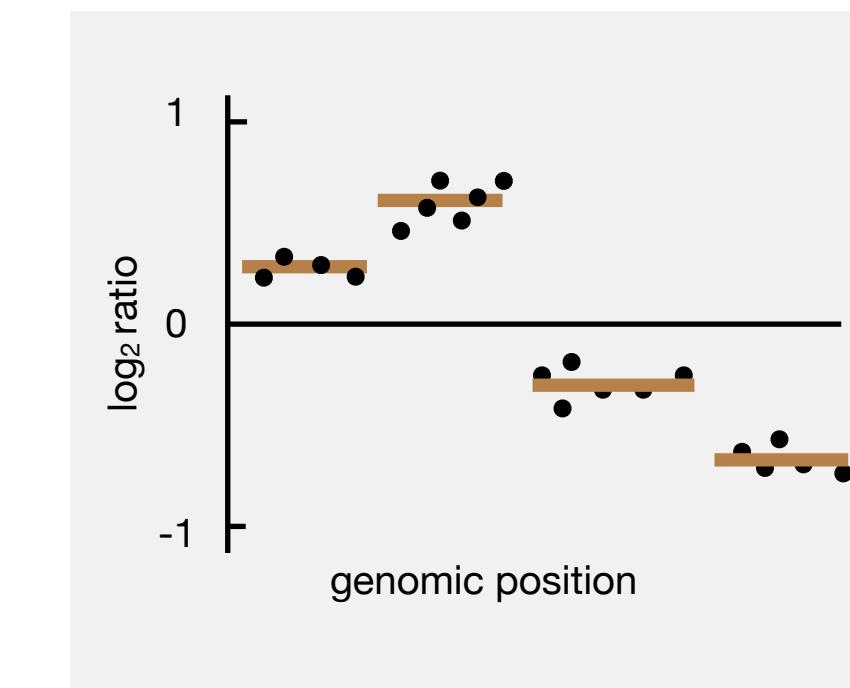
Baseline ambiguity



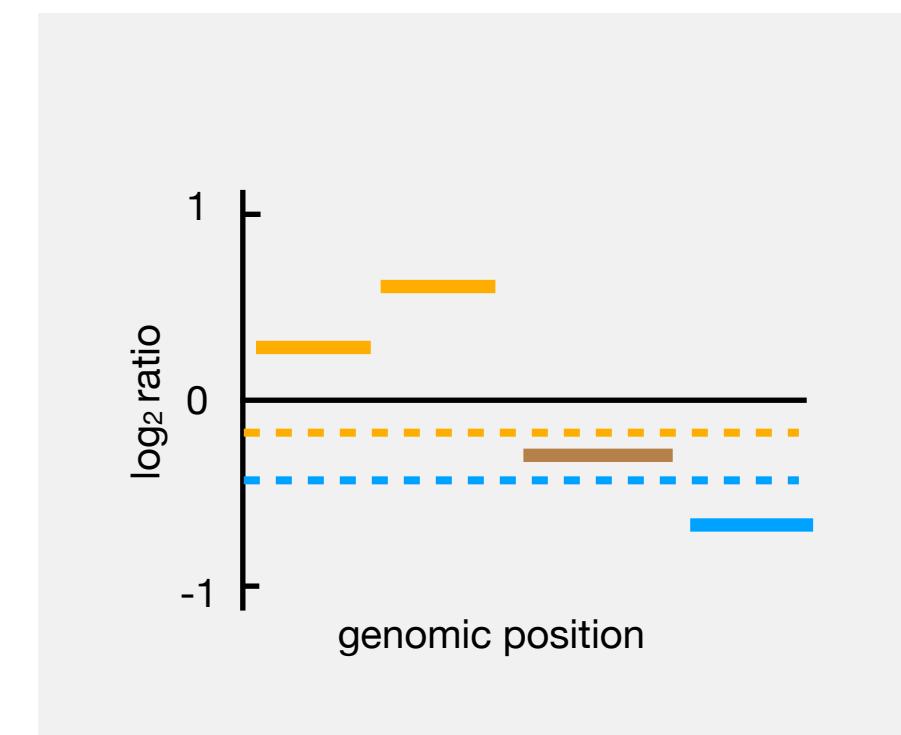
QC, Normalization



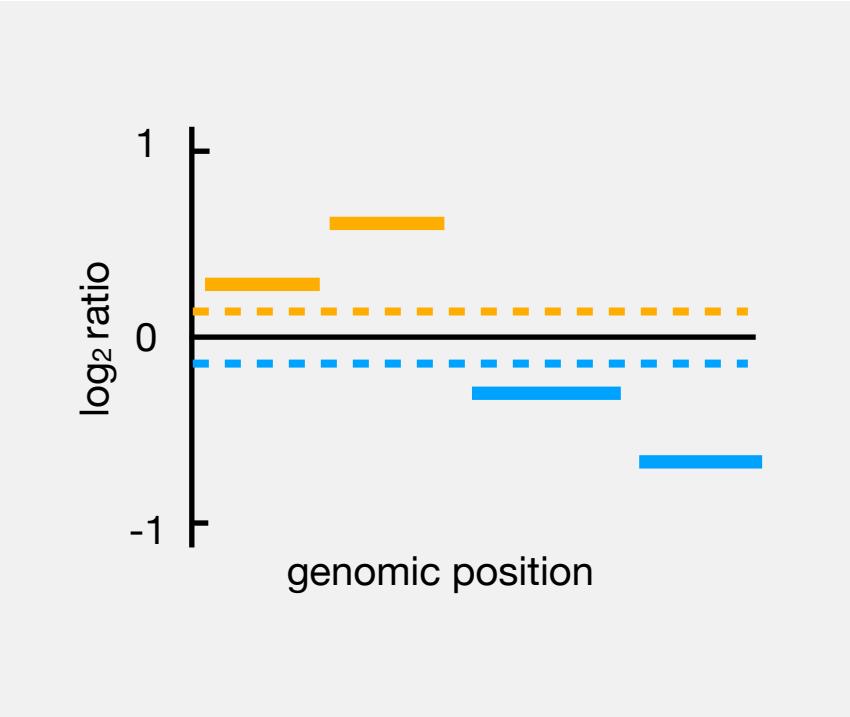
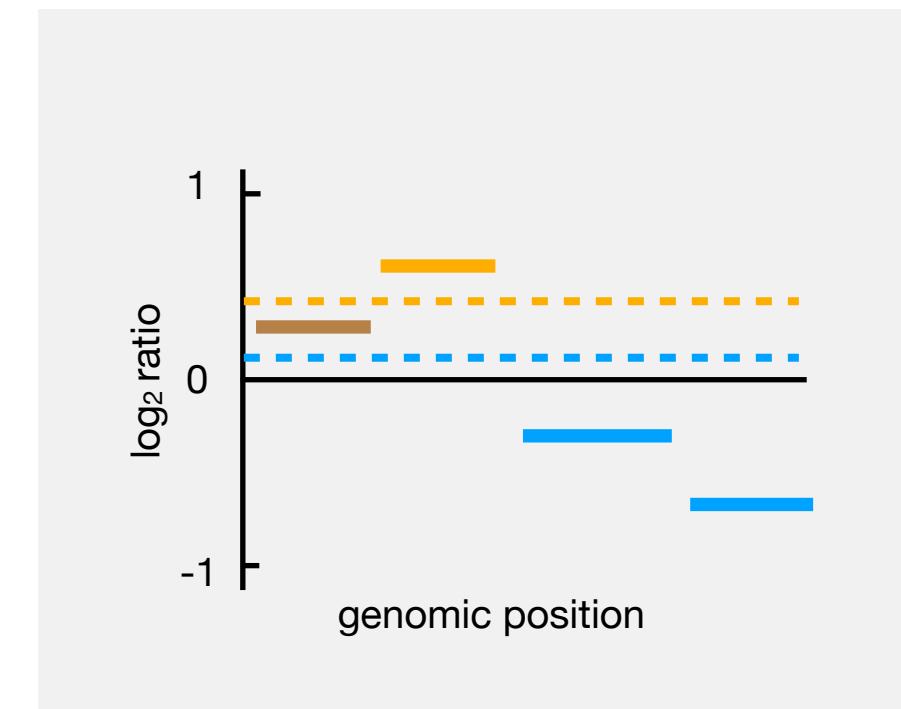
Segmentation



Calling

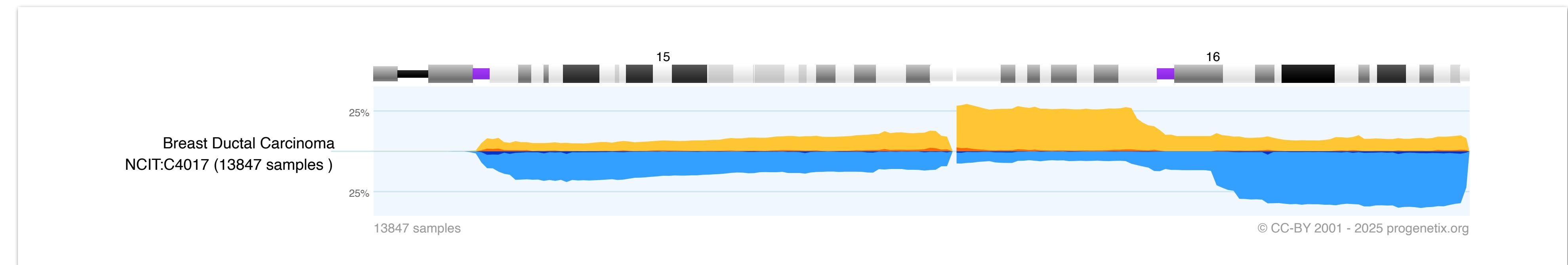


Which one is better?

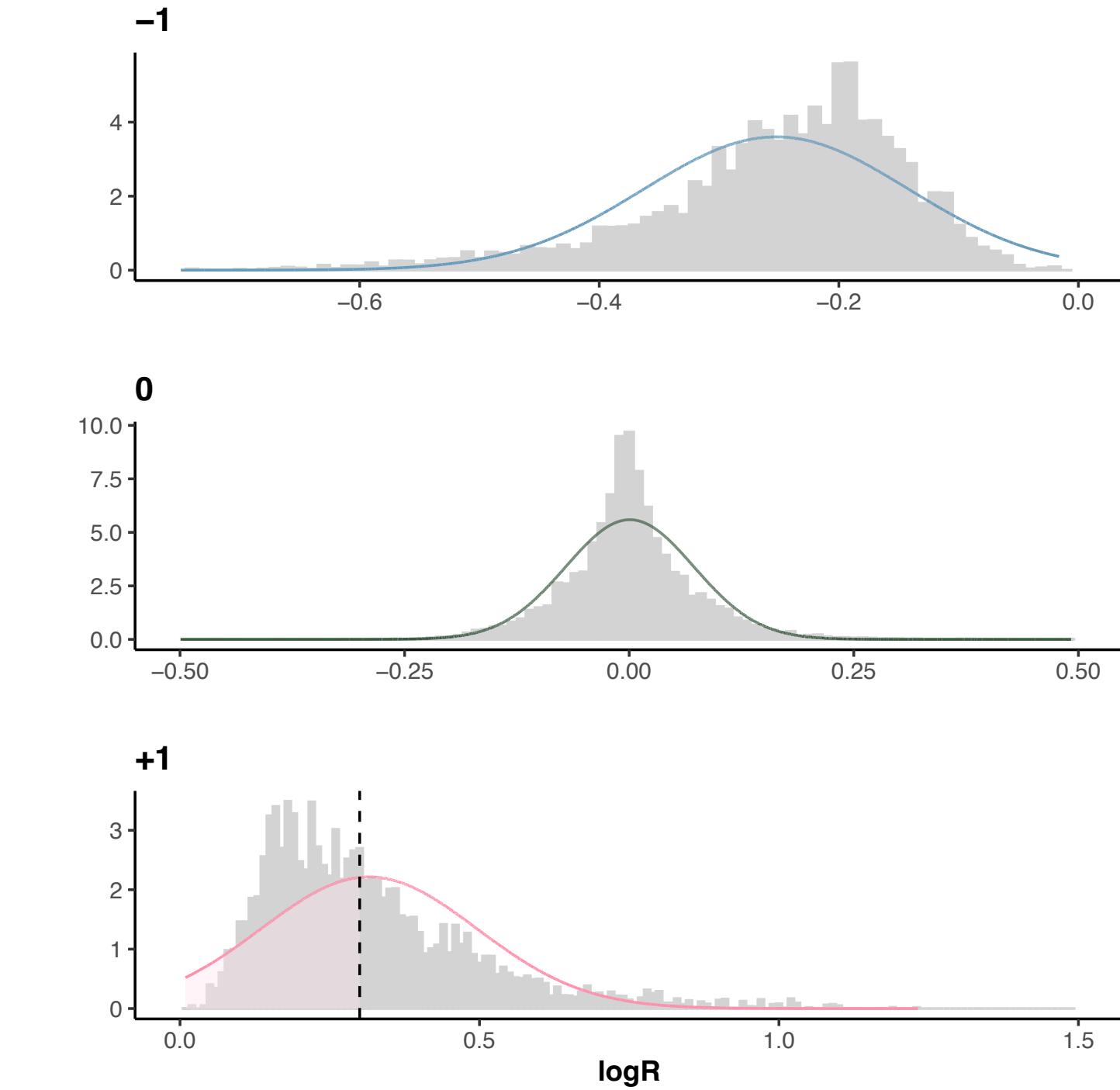
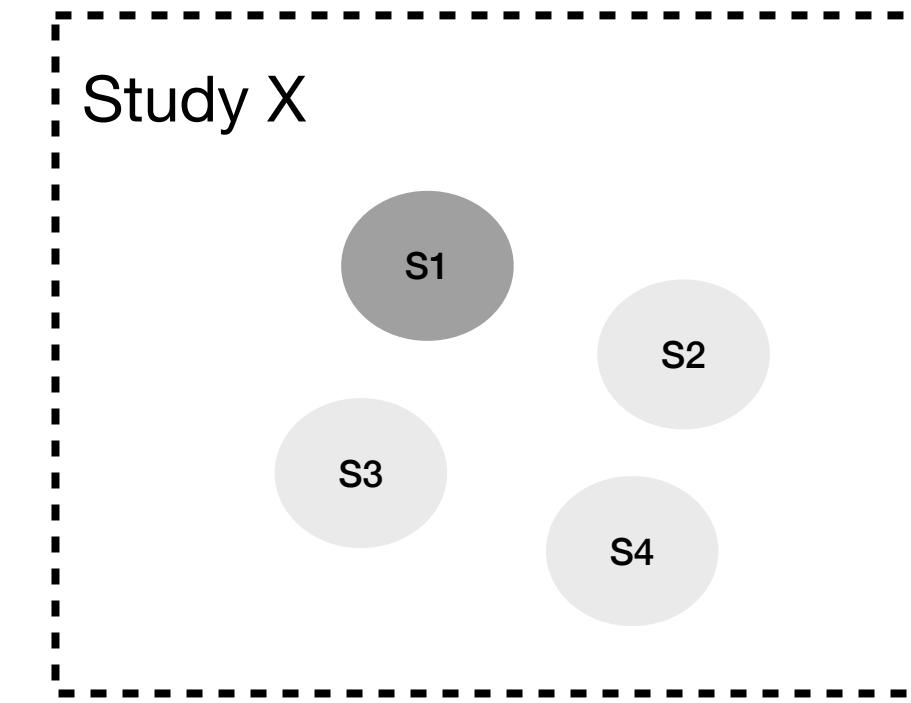
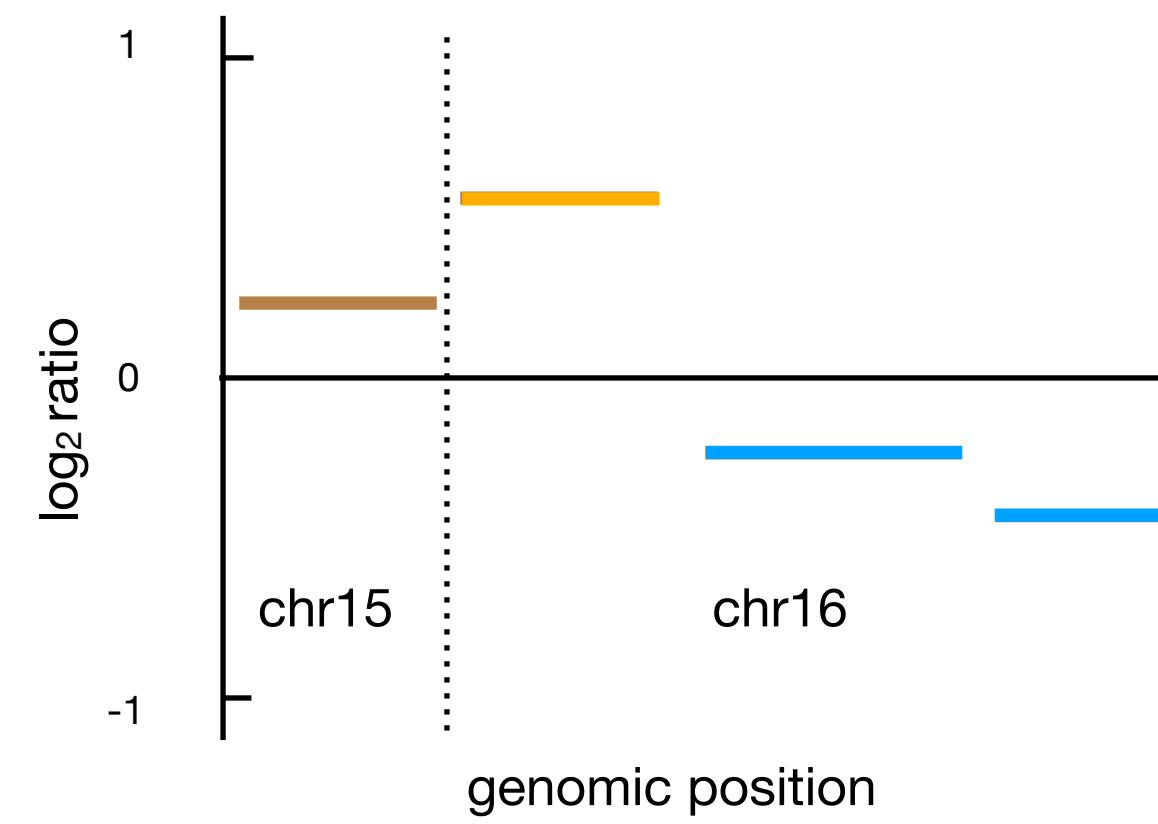


CNAdjust: enhancing CNA calling accuracy through systematic baseline adjustment

Method



from breast ductal carcinoma sample S1



- CNA patterns observed within the same population, known to share similar CNV characteristics (**Prior**)

- logR signals generated within the same batch (**Likelihood**)

Looking for implementers and contributors

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

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

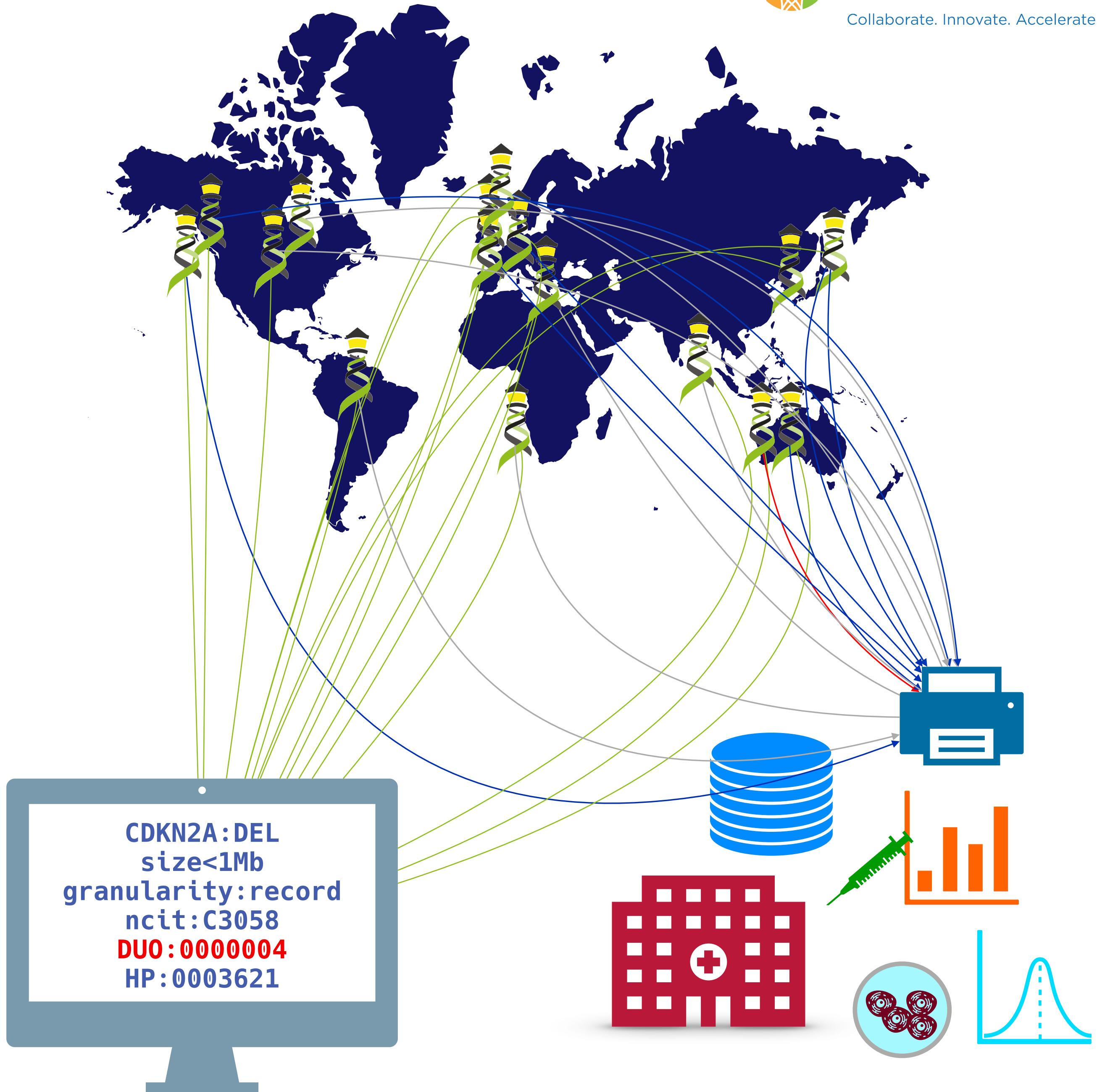
File / Commit	Description	Date
.github/workflows	Create mk-bycon-docs.yaml	8 months ago
bycon	1.3.6	3 days ago
docs	1.3.6	3 days ago
local	1.3.5 preparation	2 weeks ago
.gitignore	Update .gitignore	3 months ago
LICENSE	Create LICENSE	3 years ago
MANIFEST.in	major library & install disentanglement	9 months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

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

What Can You Do?

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

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





Universität
Zürich^{UZH}



Swiss Institute of
Bioinformatics

