Theoretical Cytogenetics and Oncogenomics

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

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

Universität Zürich[UZH]

SIB Swiss Institute of Bioinformatics

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

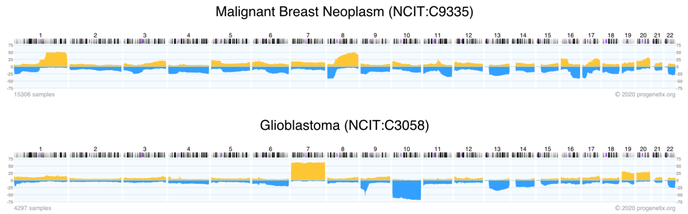
elixir

University of Zurich[UZH]
Department of Molecular Life Sciences

Theoretical Cytogenetics and Oncogenomics

... but what does this entail @baudisgroup?

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

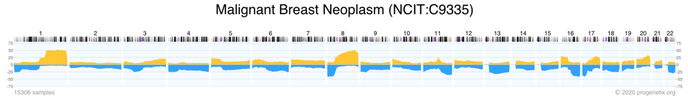


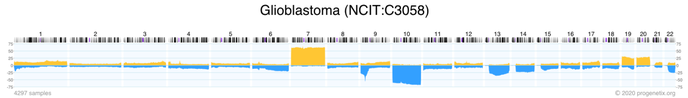
progenetix

Theoretical Cytogenetics and Oncogenomics

... but what does this entail @baudisgroup?

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



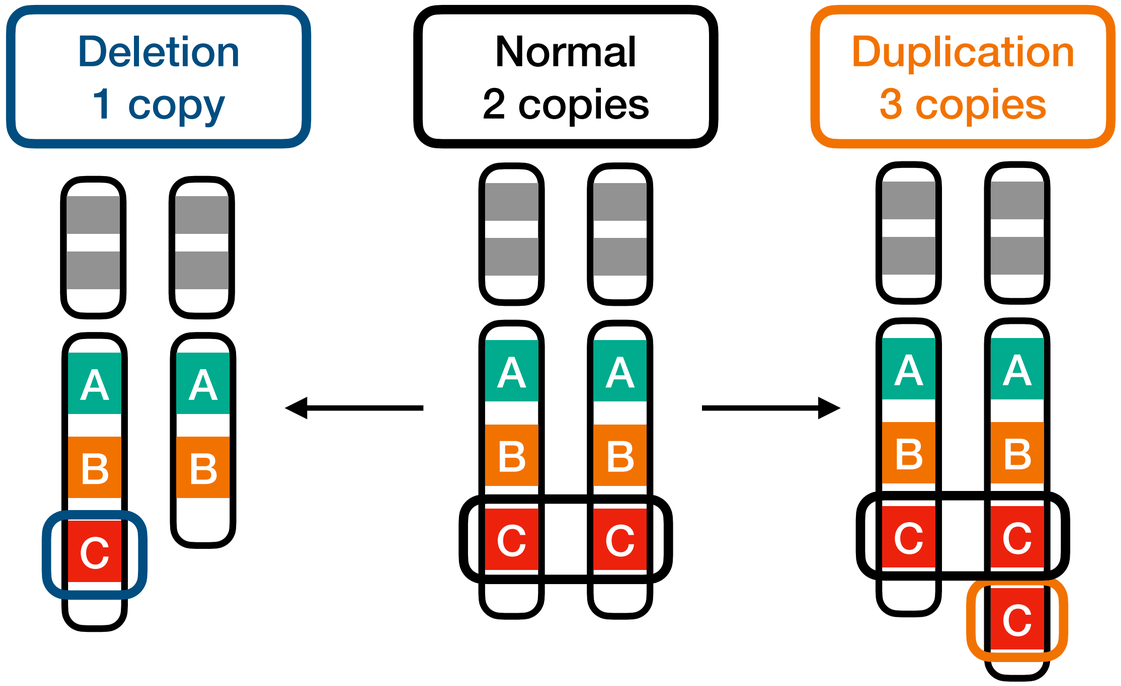


progenetix

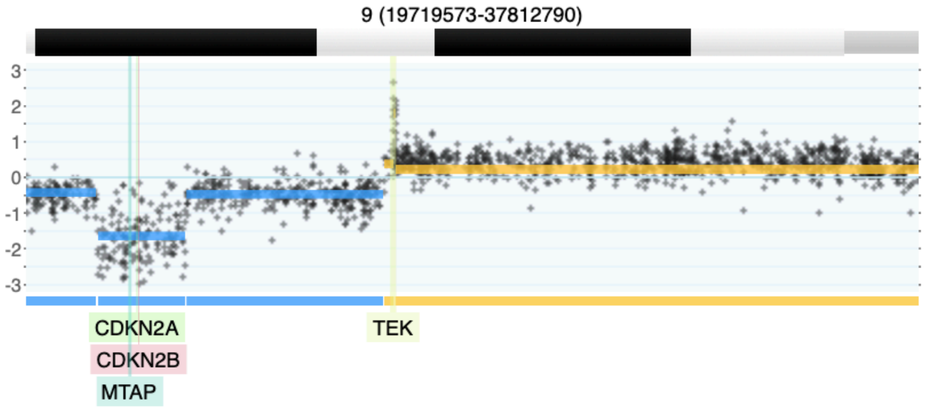
Curators

Data ~~Parasites~~

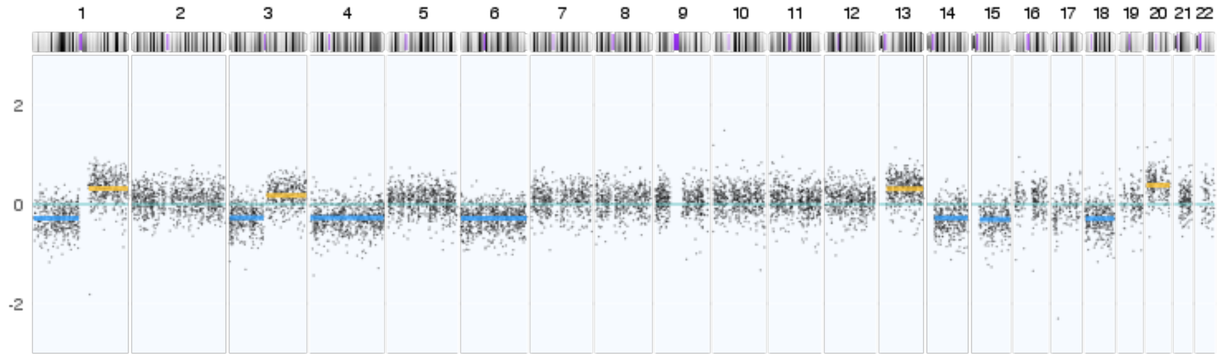
Copy Number Variant (CNV)



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

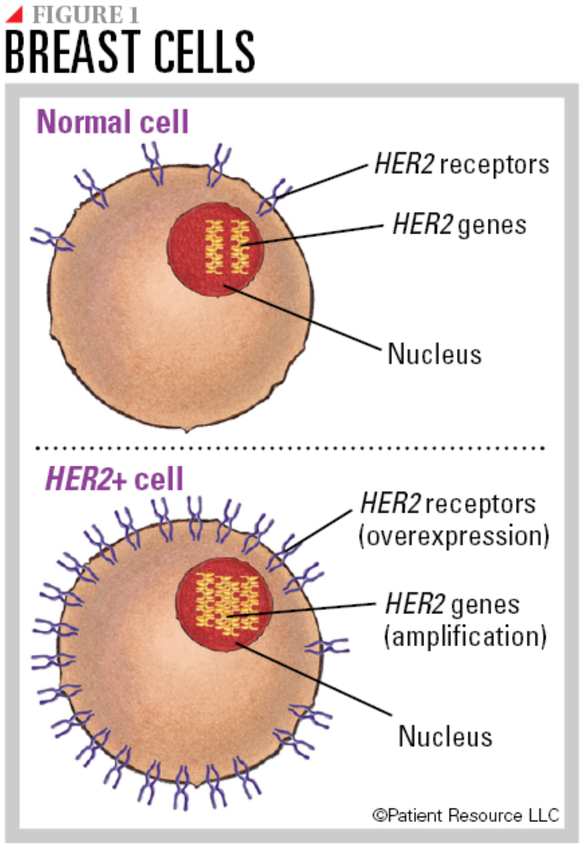


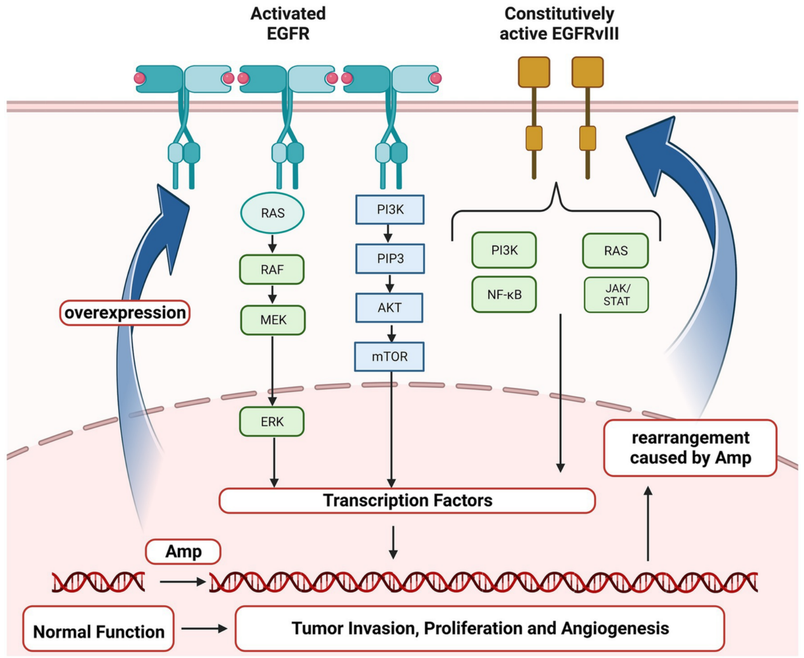
2-event, homozygous deletion in a Glioblastoma

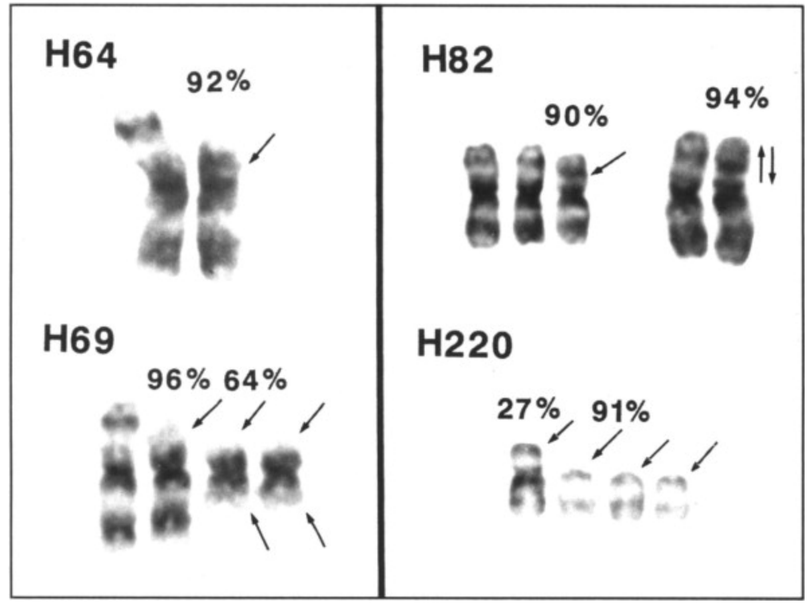


Gain of chromosome arm 13q in colorectal carcinoma

Somatic CNVs related to cancer



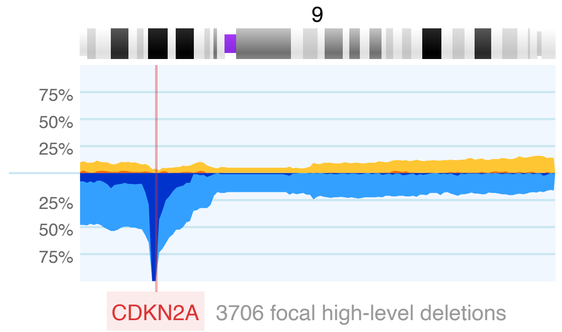




- Somatic CNVs in cancers:
  - HER2 Amplification in Breast Cancer
  - EGFR Amplification in Glioblastoma
  - Chromosome 3p Deletion in Lung Cancer

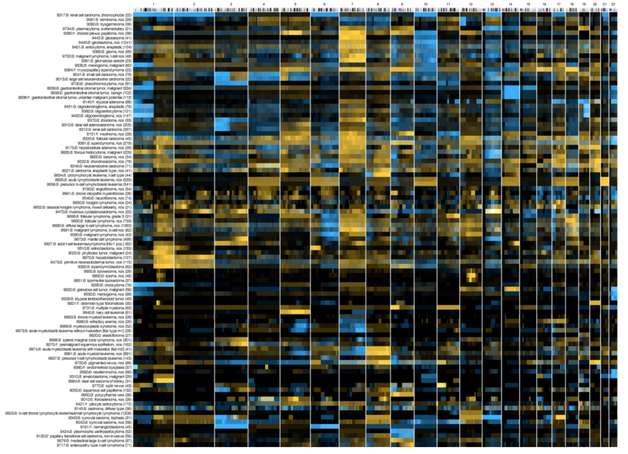
Wang H, Zhang X, Liu J, et al. Clinical roles of EGFR amplification in diffuse gliomas: a real-world study using the 2021 WHO classification of CNS tumors[J]. Frontiers in Neuroscience, 2024, 18: 1308627.

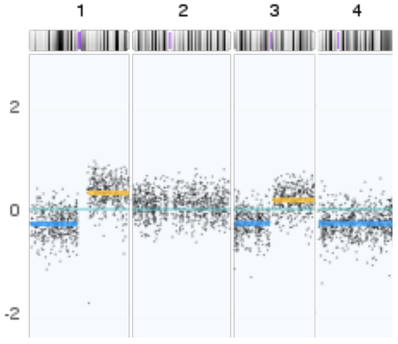
Theoretical Cytogenetics and Oncogenomics Research | Methods | Standards



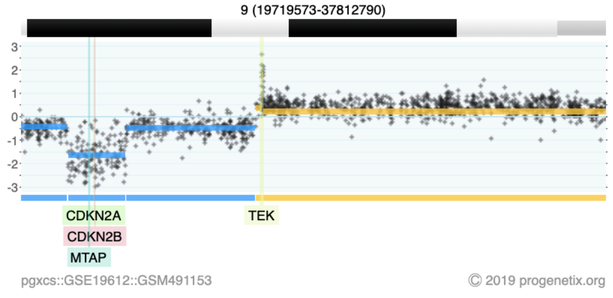
Genomic Imbalances in Cancer - Copy Number Variations (CNV)

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

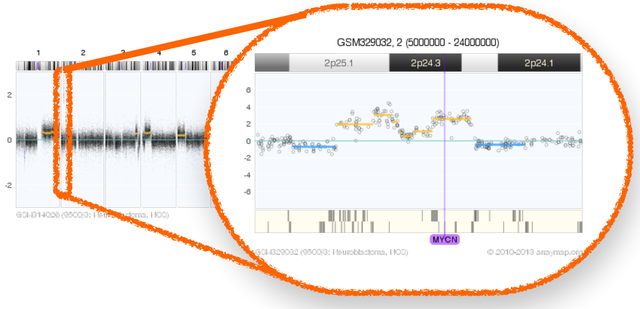




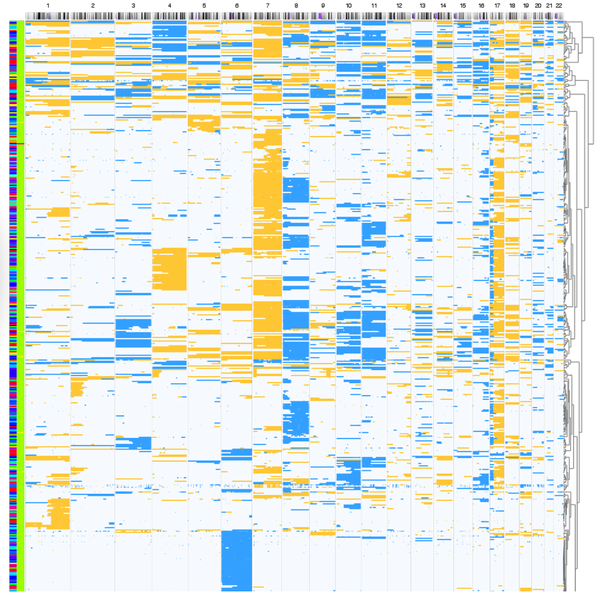
Large CNV in CRC



homozygous deletion in a Glioblastoma



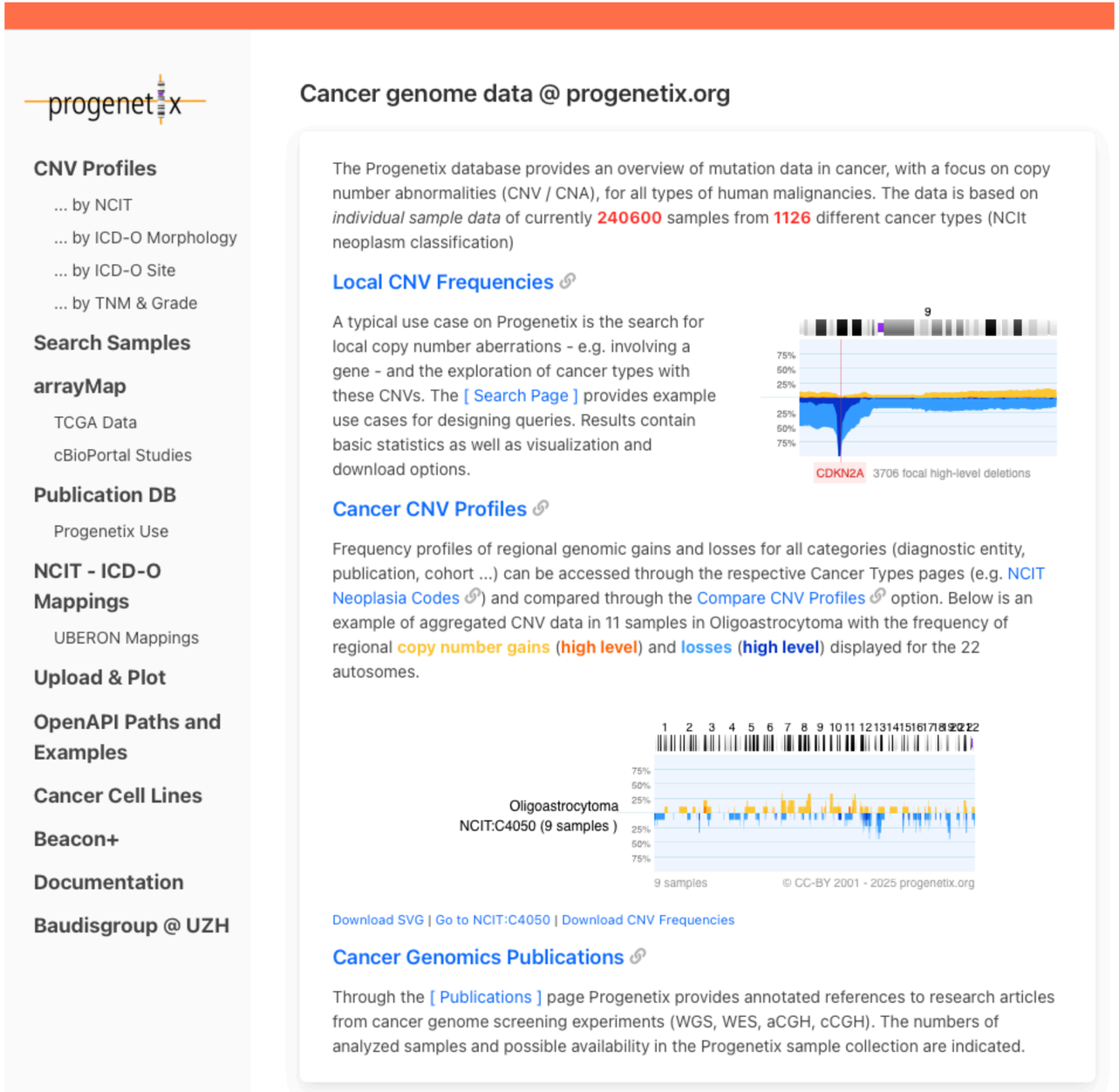
MYCN amplification in neuroblastoma
(GSM314026, SJNB8_N cell line)



progenetix.org

Cancer Genomics Reference Resource

- ***open*** resource for oncogenomic profiles
- over **240'000 cancer CNV profiles**
- more than 1'100 diagnostic types
- inclusion of reference sets (e.g. TCGA, GENIE...)
- standardized encodings (e.g. NCIt, ICD-O 3)
- identifier mapping for PMID, GEO, Cellosaurus, TCGA, cBioPortal where appropriate
- core clinical data (TNM, sex, survival ...)
- data mapping services
- recent addition of SNV data for some series



Universität Zürich[UZH]

progenetix

SIB Swiss Institute of Bioinformatics

Somatic CNV in Cancer
Example Frequency Profiles

progenetix.org

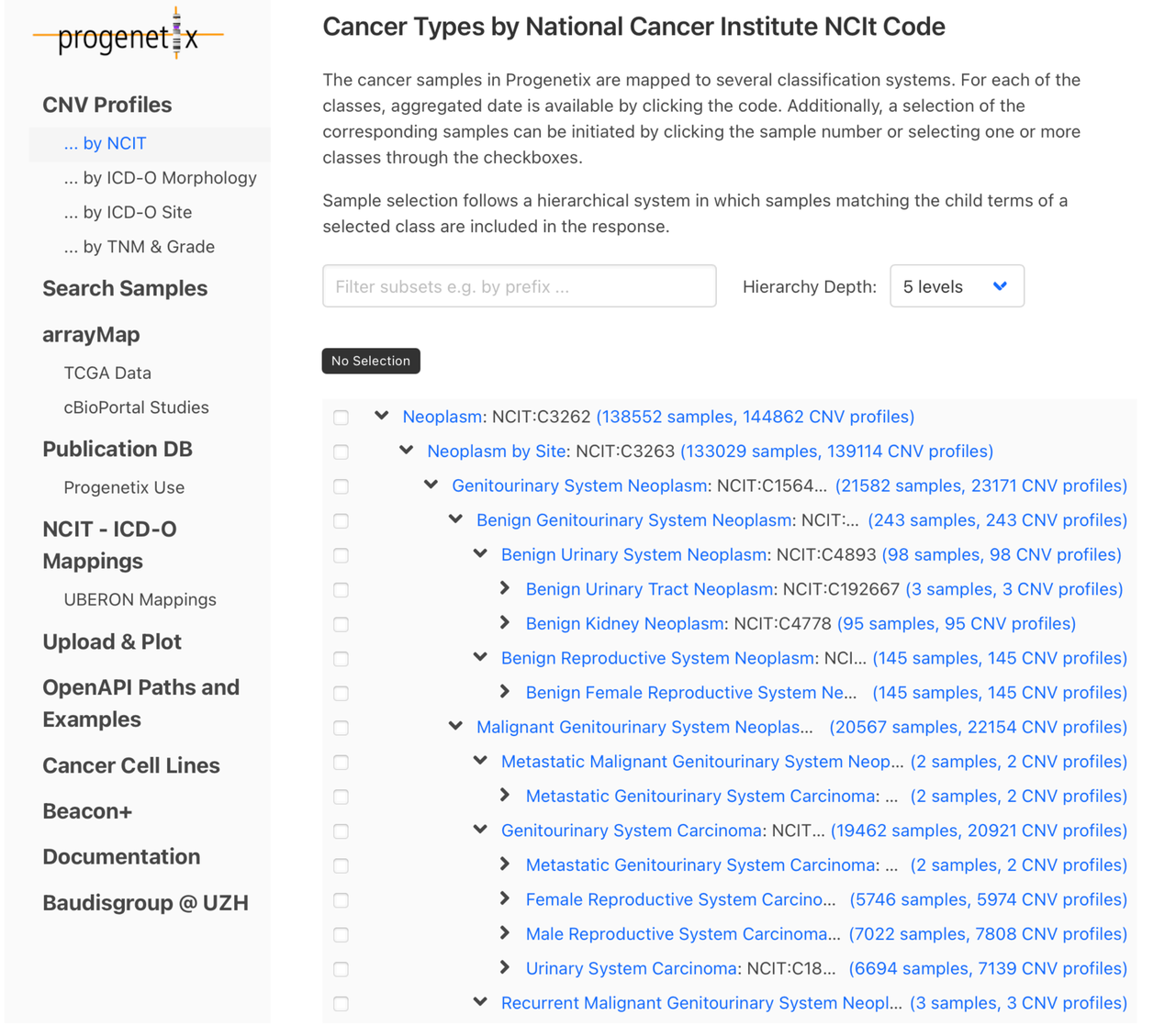
Cancer Genomics Reference Resource

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

Universität Zürich[UZH]

progenetix

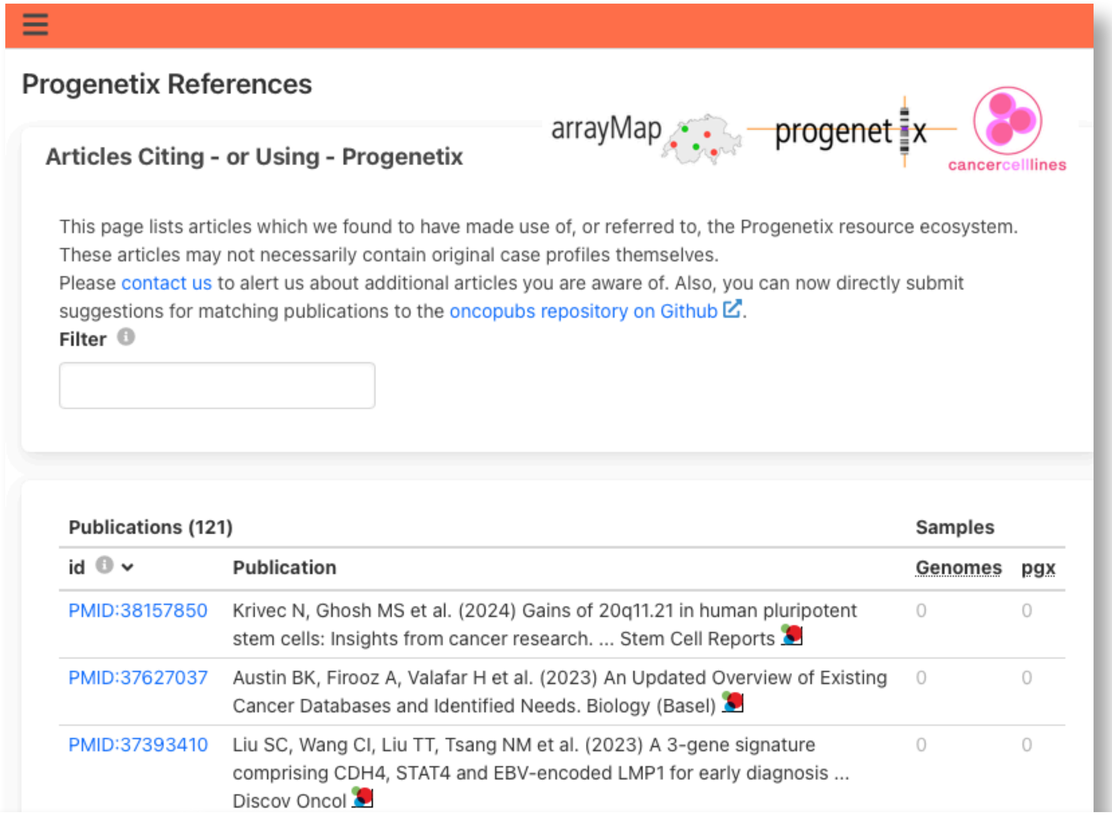
SIB Swiss Institute of Bioinformatics

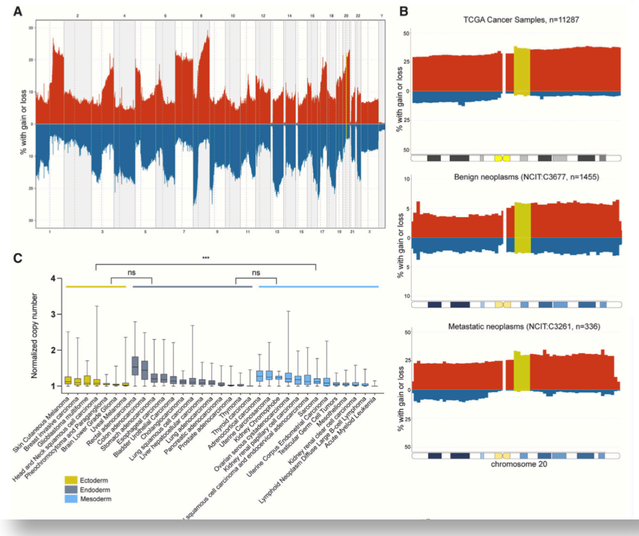


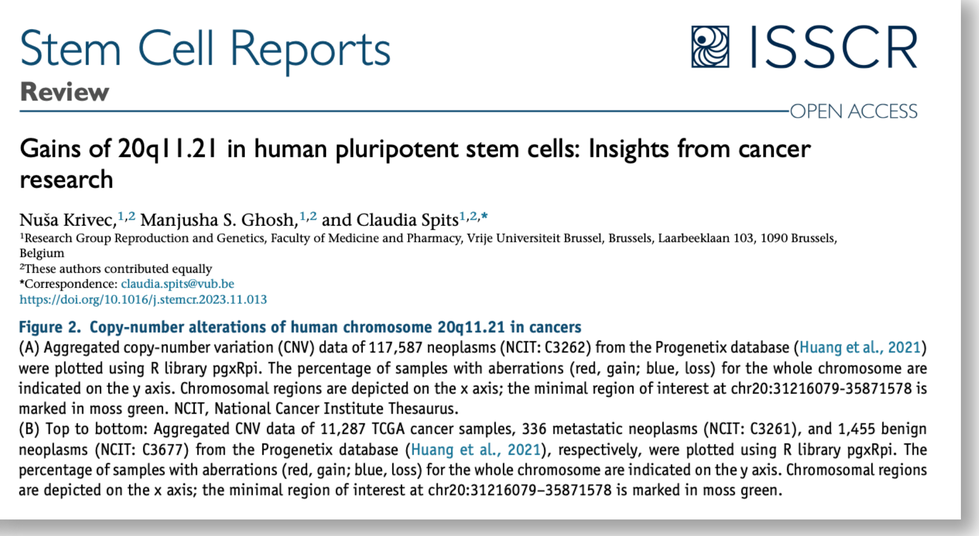
Progenetix Use

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

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






Stem Cell Reports
ISSCR
Review
OPEN ACCESS

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

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

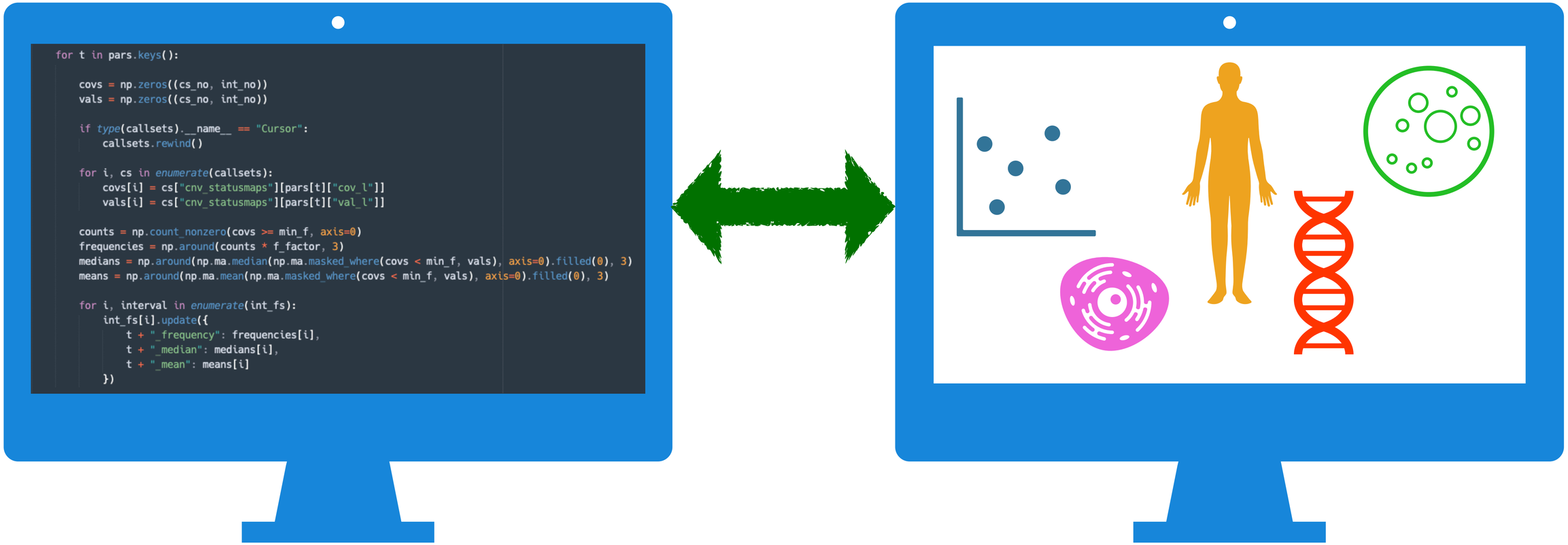
https://doi.org/10.1016/j.stemcr.2023.11.013

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

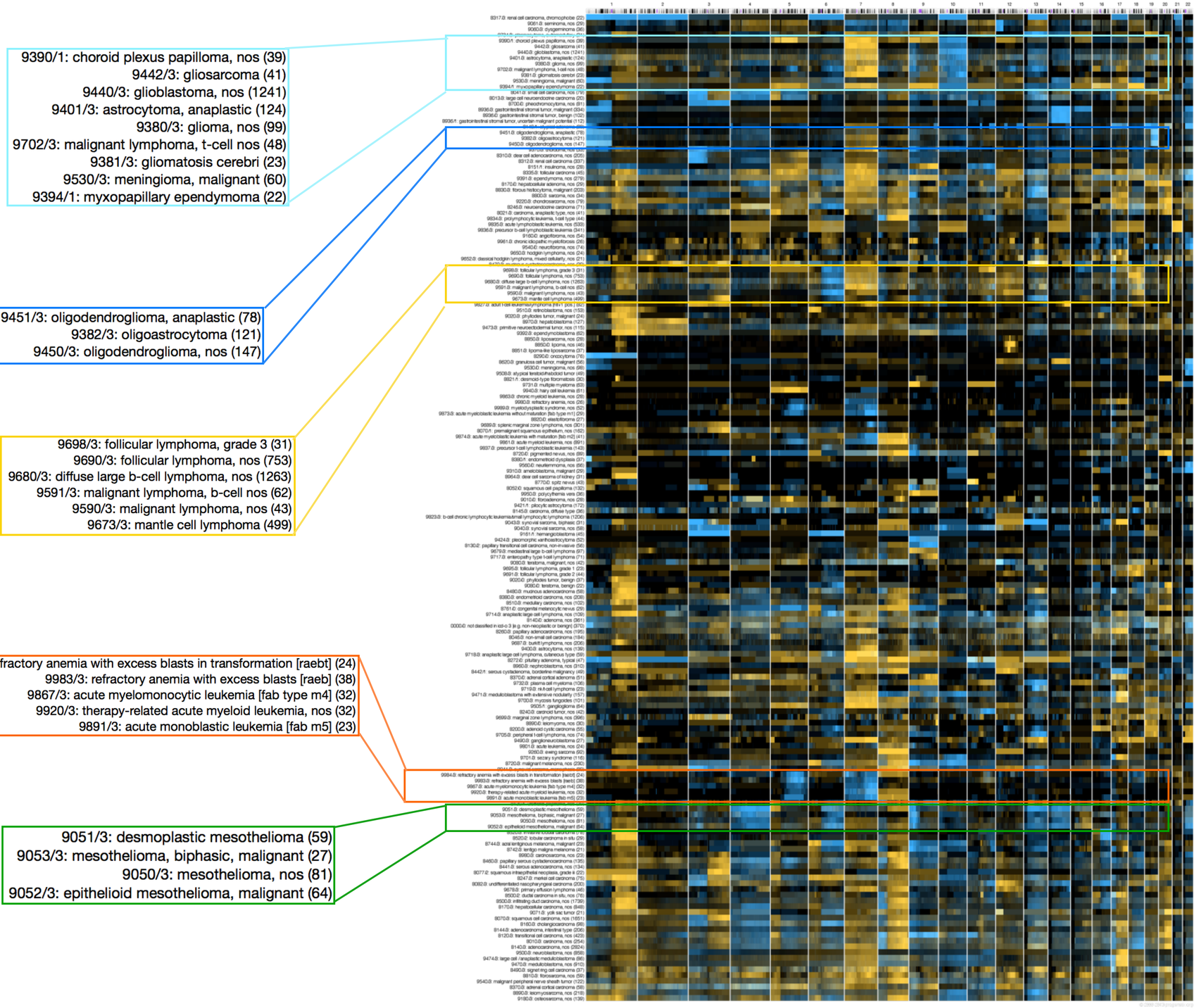
(A) Aggregated copy-number variation (CNV) data of 117,587 neoplasms (NCIT: C3262) from the Progenetix database (Huang et al., 2021) were plotted using R library pgxRpi. The percentage of samples with aberrations (red, gain; blue, loss) for the whole chromosome are indicated on the y axis. Chromosomal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

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

{BioInformaticsScience}



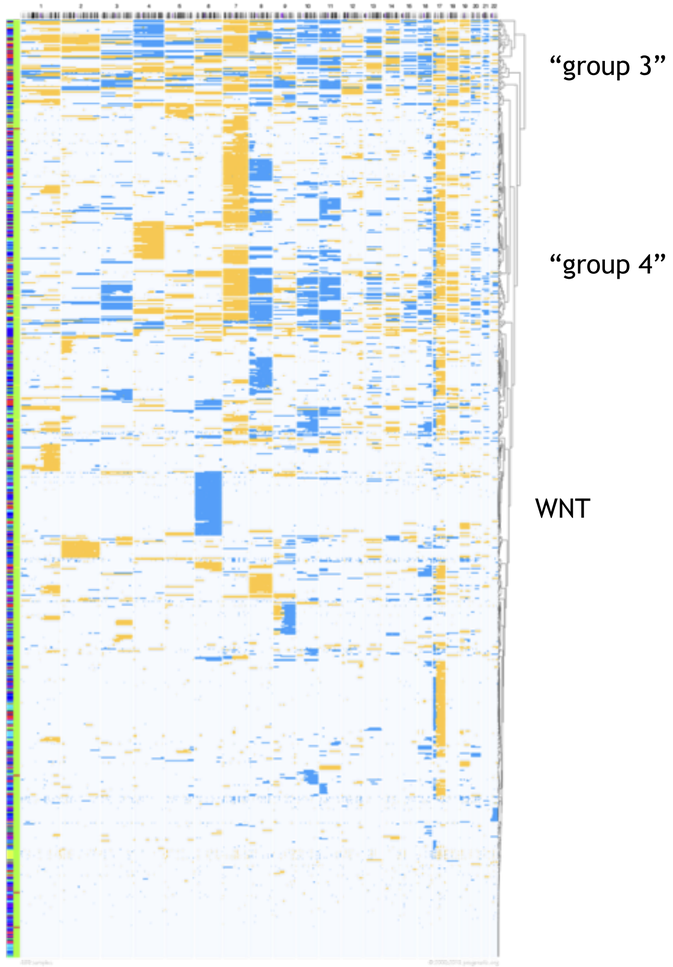
Somatic Mutations In Cancer: Patterns
Making the case for genomic classifications
Some related cancer entities show similar copy number profiles



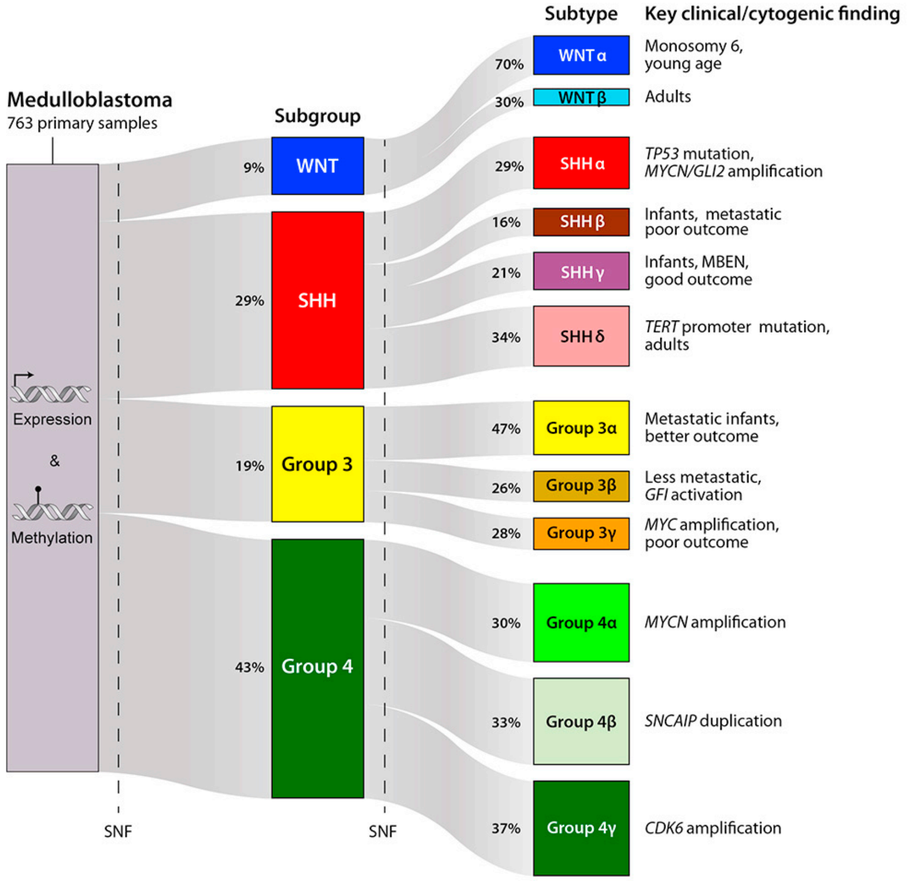
progenet x

CNA & Cancer heterogeneity

Cancer type definitions can be improved by the addition of molecular parameters as subtype markers or even complete re-evaluation of entity definitions from molecular subtypes with distinct functional mechanisms and clinical trajectories.



Copy number profiles from 889 primary medulloblastomas

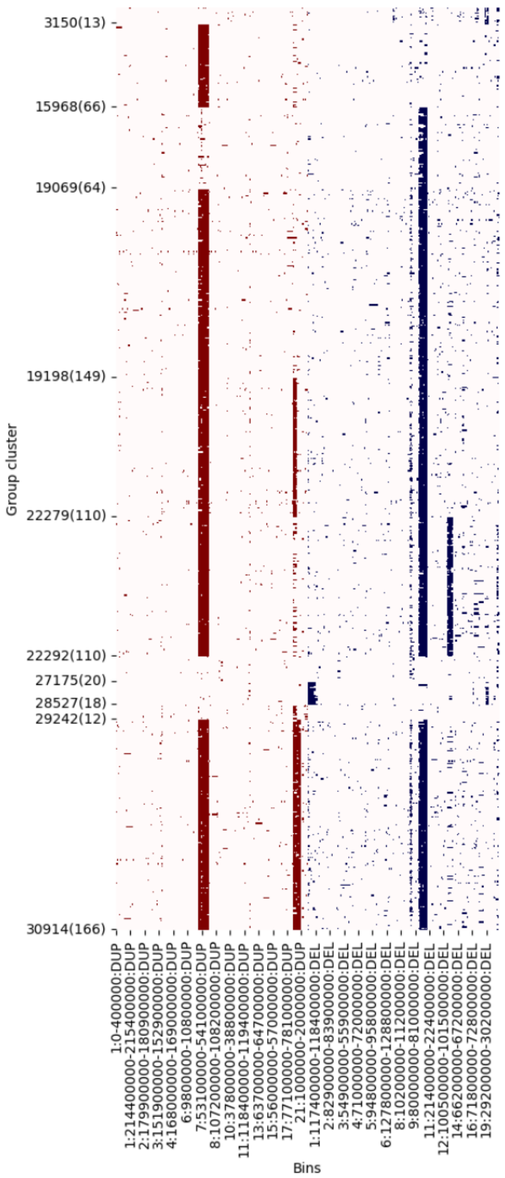


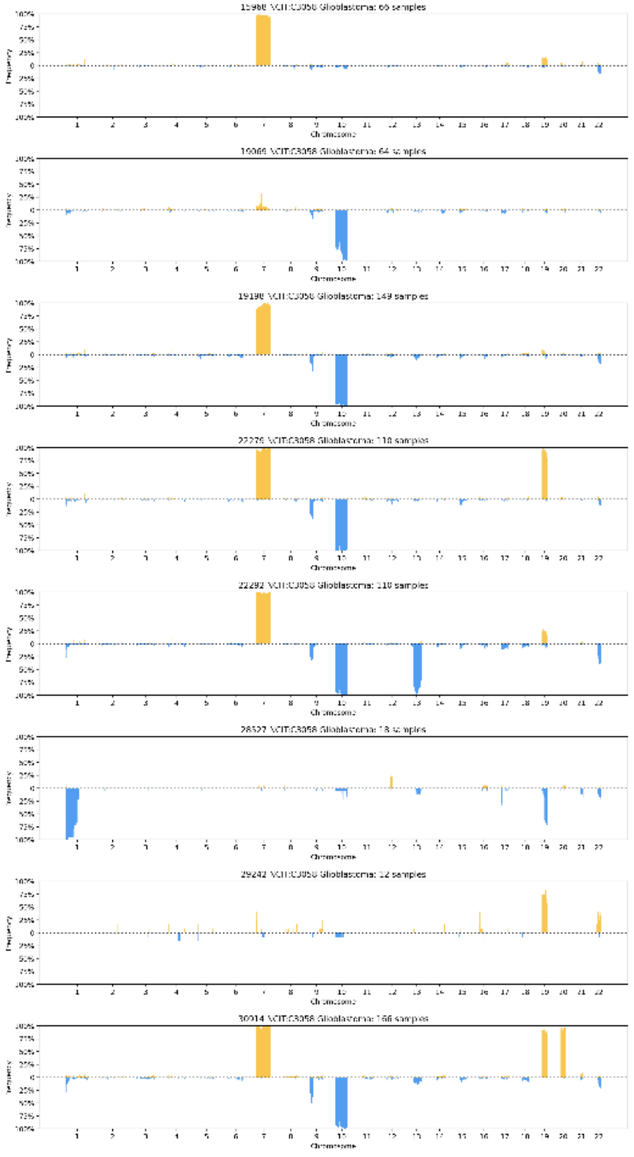
Intertumoral heterogeneity with medulloblastoma subgroups.
Cavalli, Florence MG, et al. "Intertumoral heterogeneity within medulloblastoma subgroups." *Cancer Cell* 31.6 (2017): 737-754.

Lead: Ziying Yang

Results

Entity CNV heterogeneity: Glioblastoma



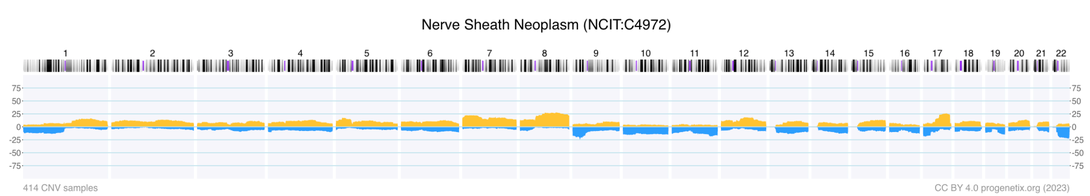


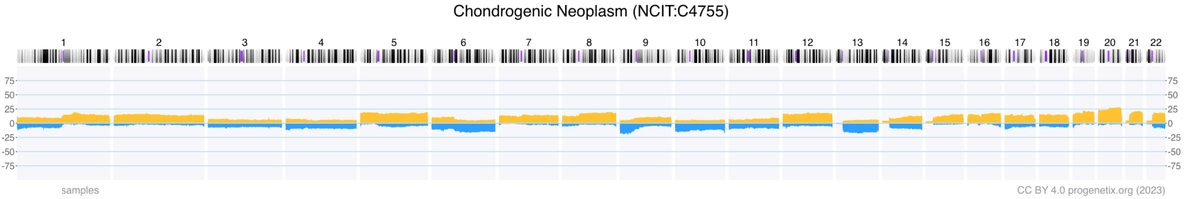
group cluster	CNV features
15968	Dup 7
19069	Del 10
19198	Dup 7, Del 10
22279	Dup 7, Del 10, Dup 19
22292	Dup 7, Del 10, Del 13
28527	Del 1p, Del 19q
29242	Dup 19
30914	Dup 7, Del 10, Dup 19, Dup 20

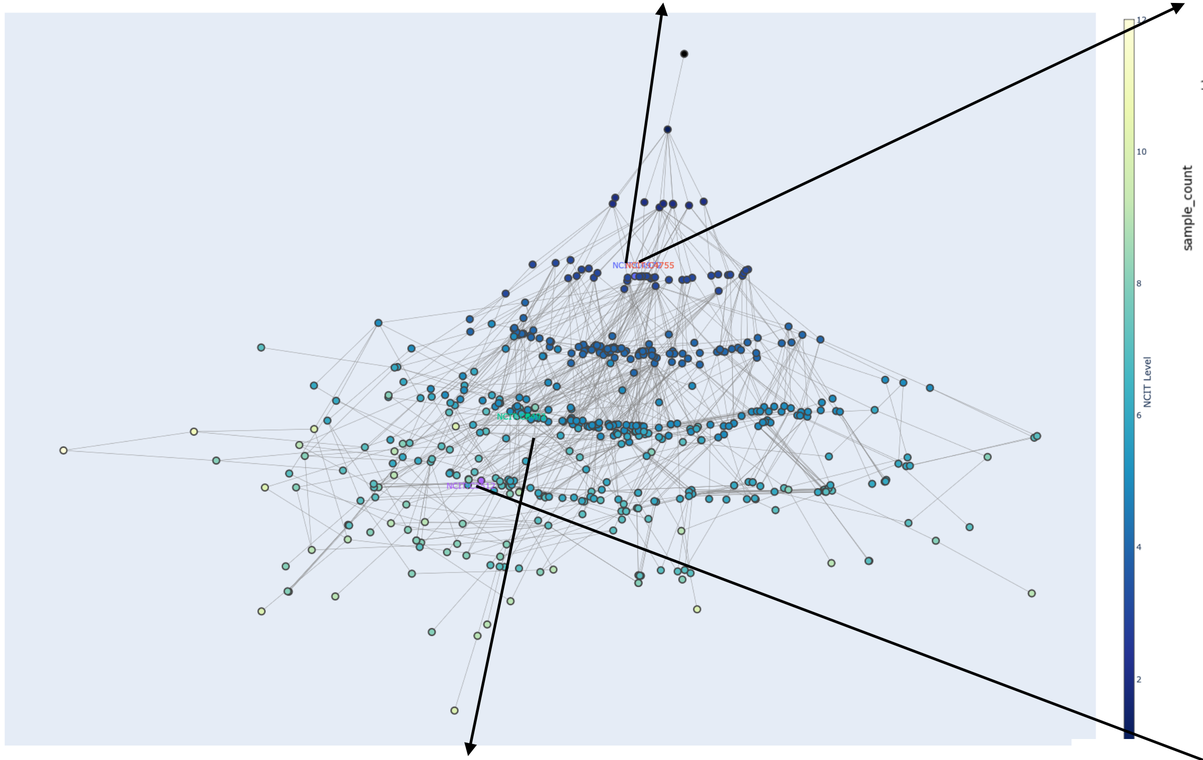
Lead: Ziying Yang

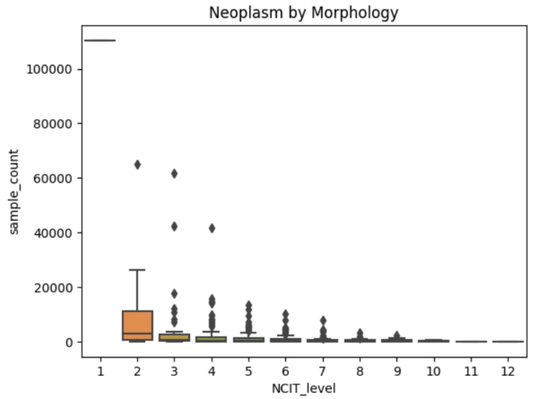
CNV profiles heterogeneity vs cancer classification

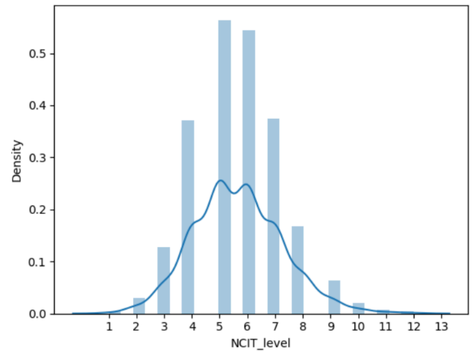
Correspondance of genomic profiles to NCIT cancer hierarchy

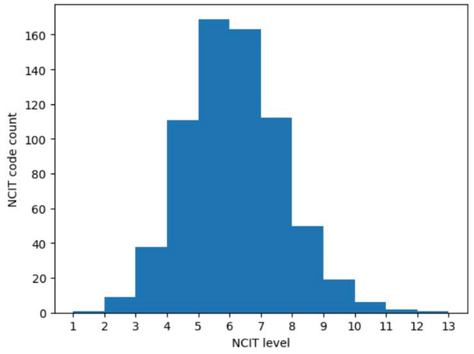


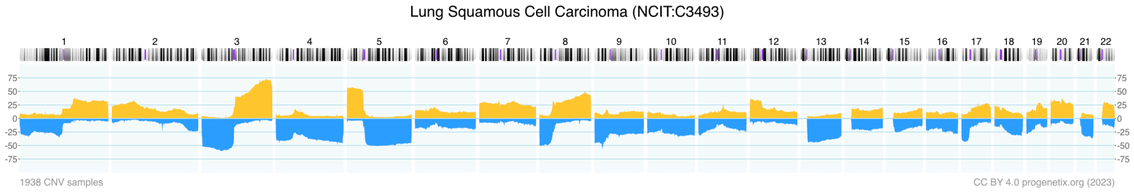


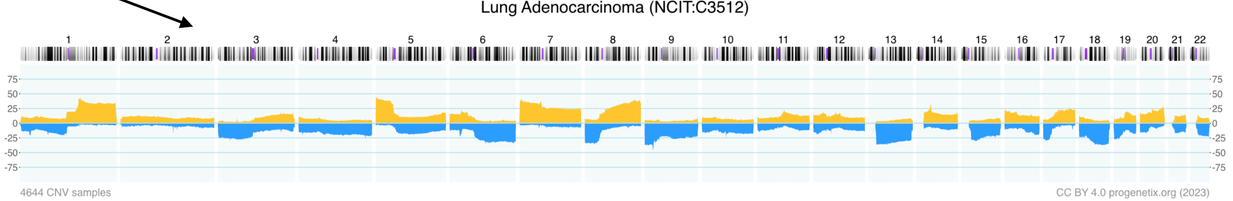








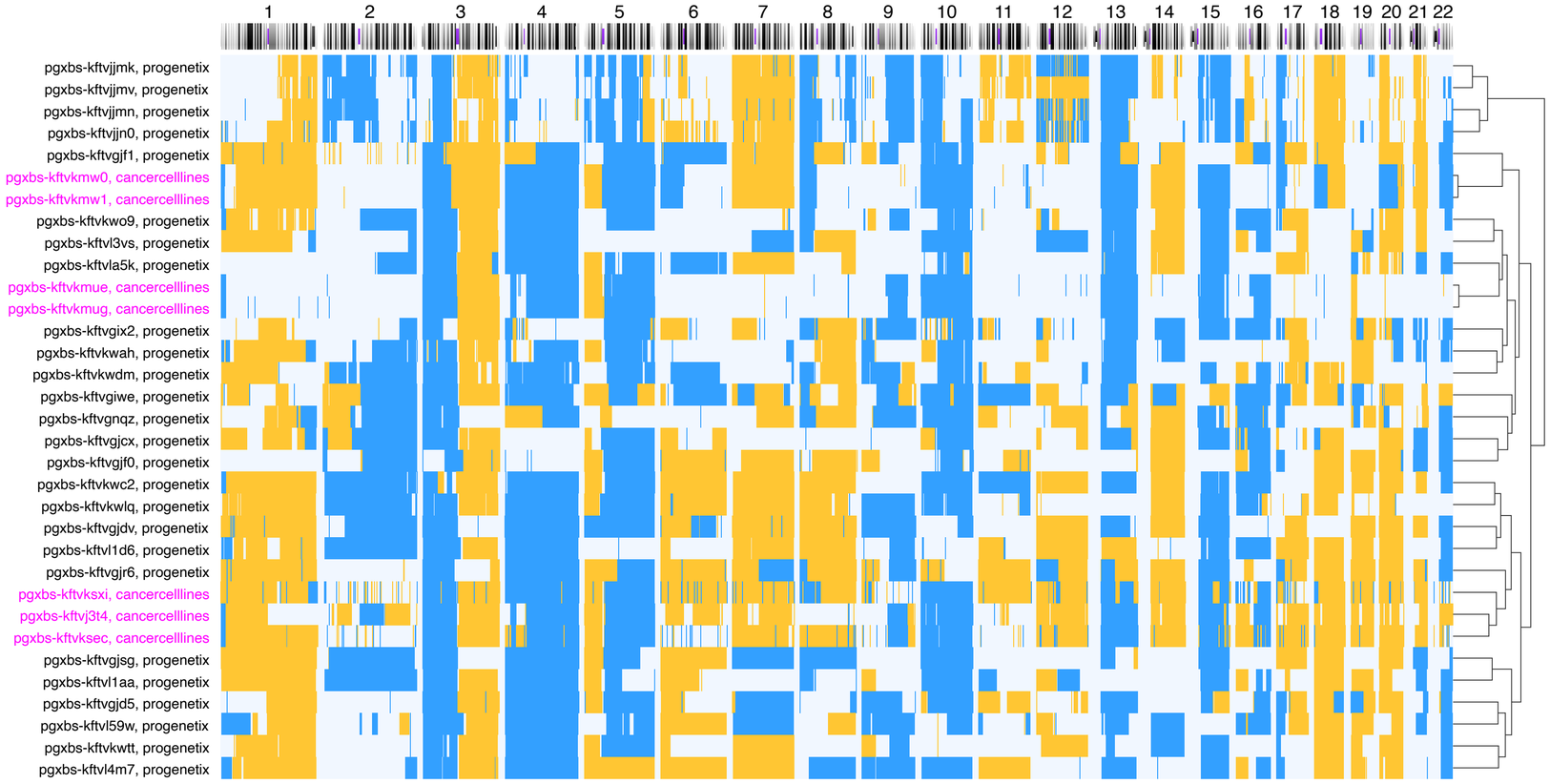




Lead: Ziying Yang

Tumor subpopulations can be matched with highly similar cell lines

- Lung Small Cell Carcinoma Subpopulation
- Cell Lines:
  - CVCL_1140: COR-L279
  - CVCL_1455: NCI-H1105
  - CVCL_1527: NCI-H2107



Lead: Rahel Paloots

Population stratification in cancer samples based on SNP array data

- Despite extensive somatic mutations of cancer profiling data, consistency between germline and cancer samples reached 97% and 92% for 5 and 26 populations
- Comparison of our benchmarked results with self-reported meta-data estimated a matching rate between 88 % to 92%.
- Ethnicity labels indicated in meta-data are vague compared to the standardized output from our tool

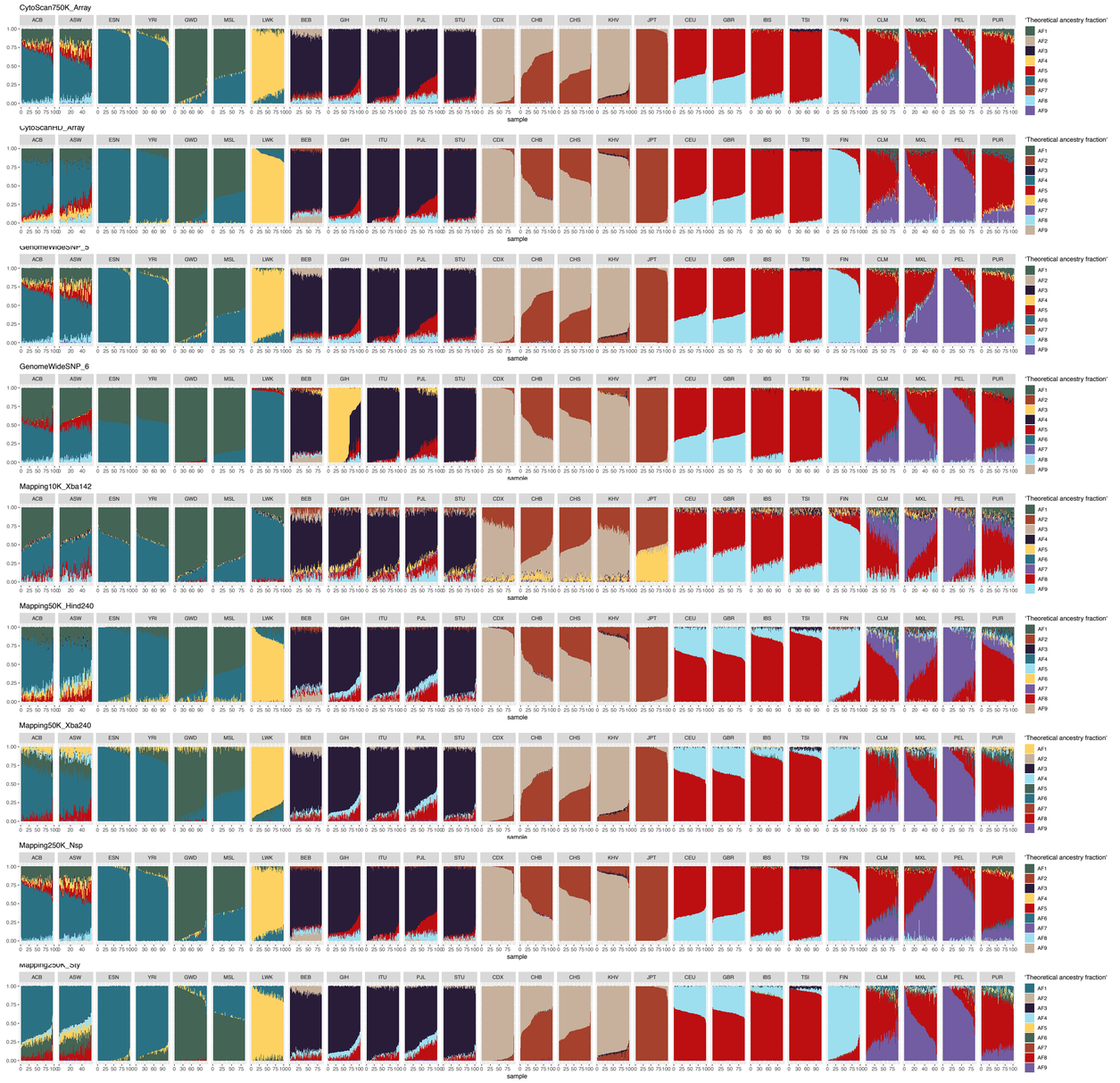


Figure S1 The fraction or contribution of theoretical ancestors (k=9) in reference individuals from 1000 Genomes Project with regard to nine SNP array platforms. The x-axis are individual samples, grouped by their respective population. Groups belonging to the same continent/superpopulation are placed neighboring to each other: AFR (1-7), SAS (8-12), EAS (13-17), EUR (18-22), AMR (23-26).

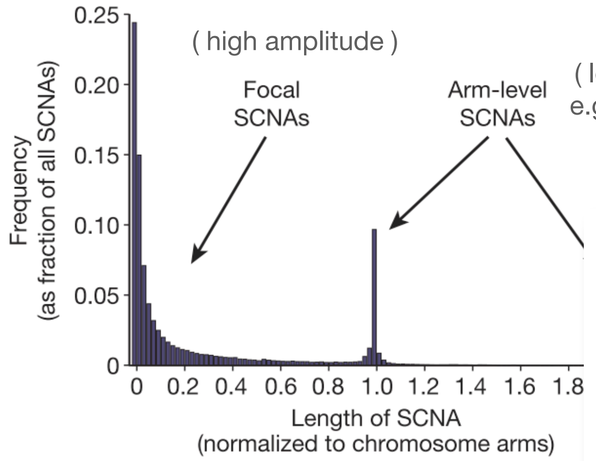
arrayMap

Lead: Qingyao Huang

CNV Categorization Method Supporting GA4GH Standards

GA4GH Variation Representation Specification

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



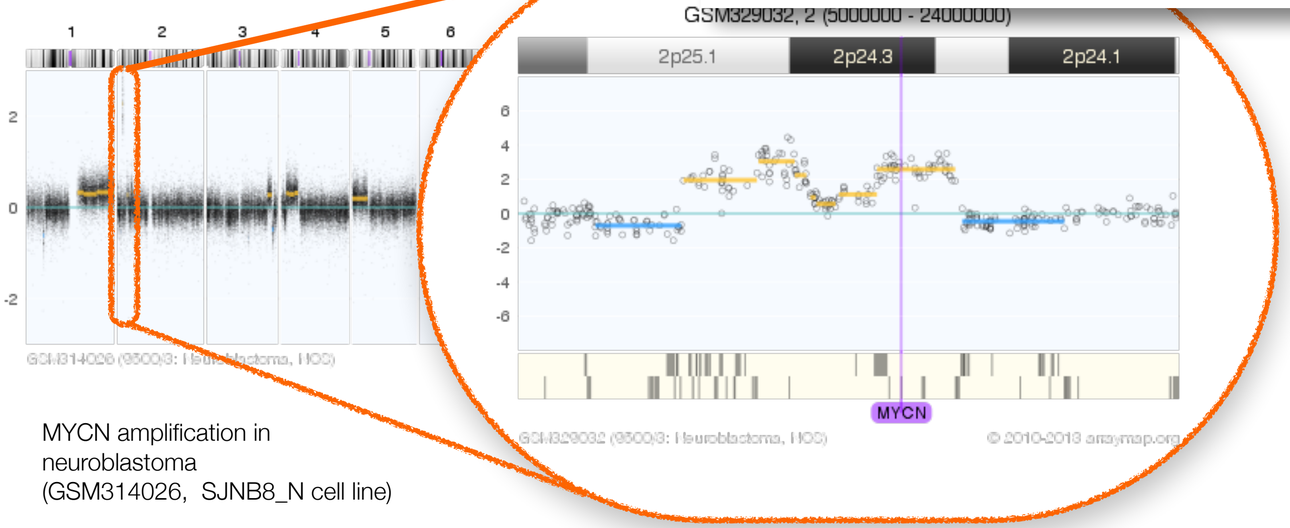
Rameen et al 20

OXFORD

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

labelSeg: segment annotation for tumor copy number alteration profiles

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



MYCN amplification in neuroblastoma (GSM314026, SJNB8_N cell line)

Lead: Hangjia Zhao

CopyNumberChange

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

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

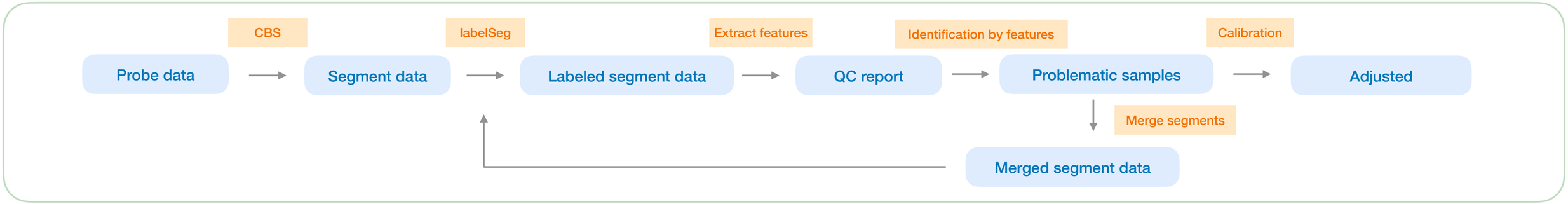
tes are inherited from Variation.

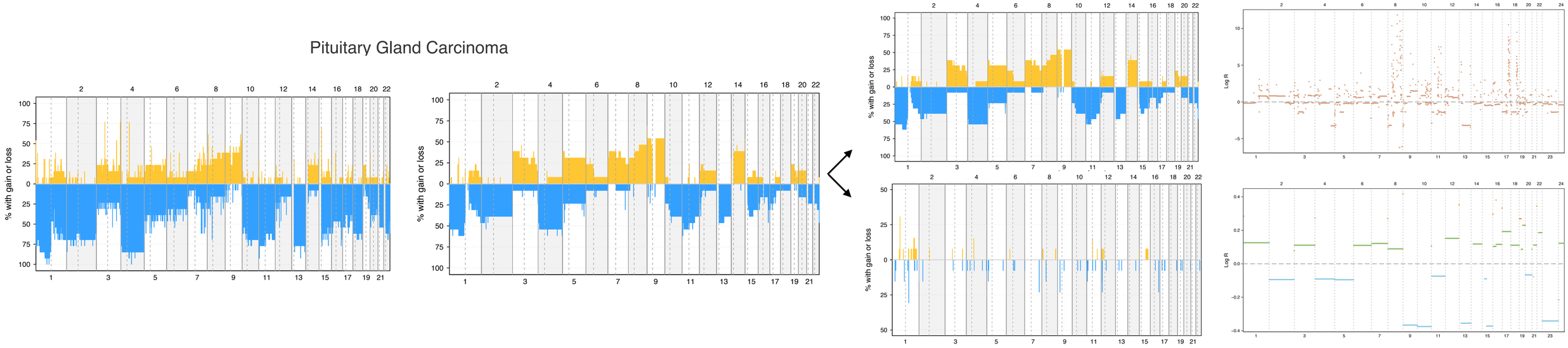
Field	Type	Limits	Description
_id	CURIE	0..1	Variation Id. MUST be unique within document.
type	string	1..1	MUST be “CopyNumberChange”
subject	Location | CURIE | Feature	1..1	A location for which the number of systemic copies is described.
copy_change	string	1..1	MUST be one of “efo:0030069” (complete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

Pipeline Development

improve CNV calling in large numbers of heterogeneous cancer samples

nextflow



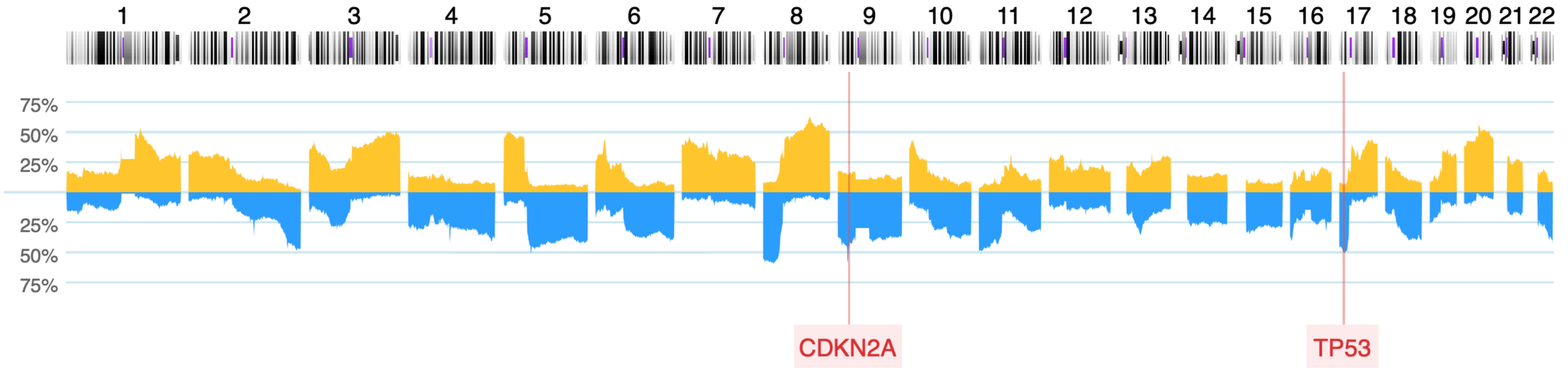


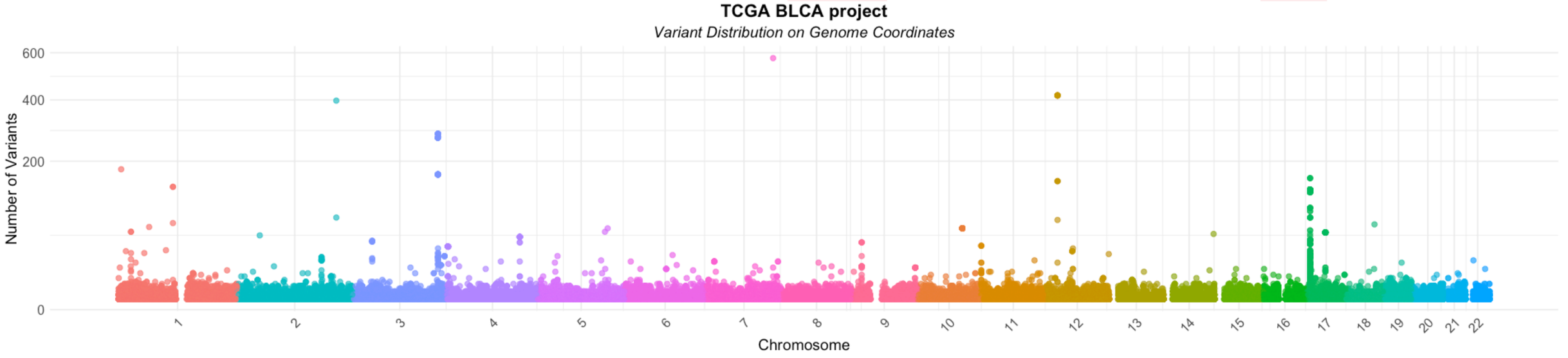
Lead: Hangjia Zhao

progenetix

TCGA

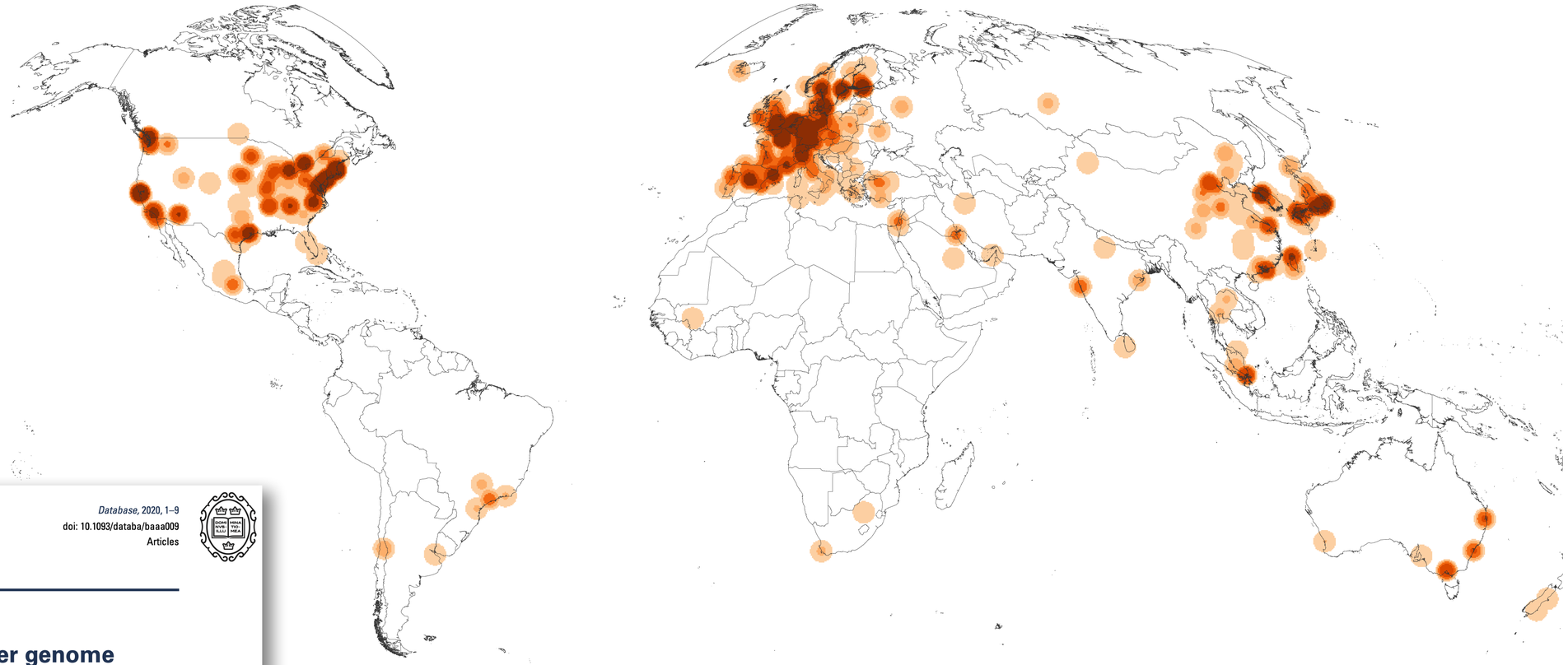
TCGA BLCA project (pgx:TCGA.BLCA)





Where does Genomic Data Come From?

Geographic bias in published cancer genome profiling studies



DATABASE
The Journal of Biological Databases and Curation

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

Articles

Geographic assessment of cancer genome profiling studies

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

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

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

Different Approaches to Data Sharing

progenetix

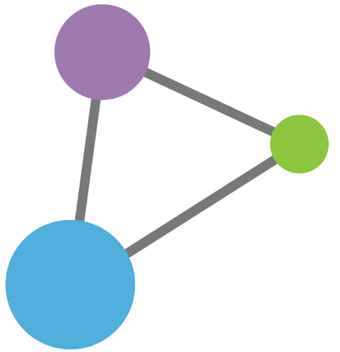
Centralized Genomic Knowledge Bases

European Genome-Phenome Archive

Data Commons
Trusted, controlled repository of multiple datasets

SPHN Swiss Personalized Health Network

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



Linkage of distributed and disparate datasets

Different Approaches to Data Sharing

progenetix

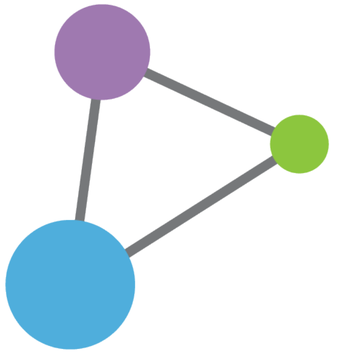
European Genome-Phenome Archive

SPHN Swiss Personalized Health Network

Centralized Genomic Knowledge Bases

Data Commons
Trusted, controlled repository of multiple datasets

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



Linkage of distributed and disparate datasets

Federation

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

INFORMATICS

Beacon v2 and Beacon networks: federated data discovery n biome

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

Cell Genomics

CellPress
OPEN ACCESS

Commentary

International federation of genomic medicine databases using GA4GH standards

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

Cell Genomics

OPEN ACCESS

Perspective

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

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

(Author list continued on next page)

Cell Genomics

OP

Technology

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

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

Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

GENOMICS

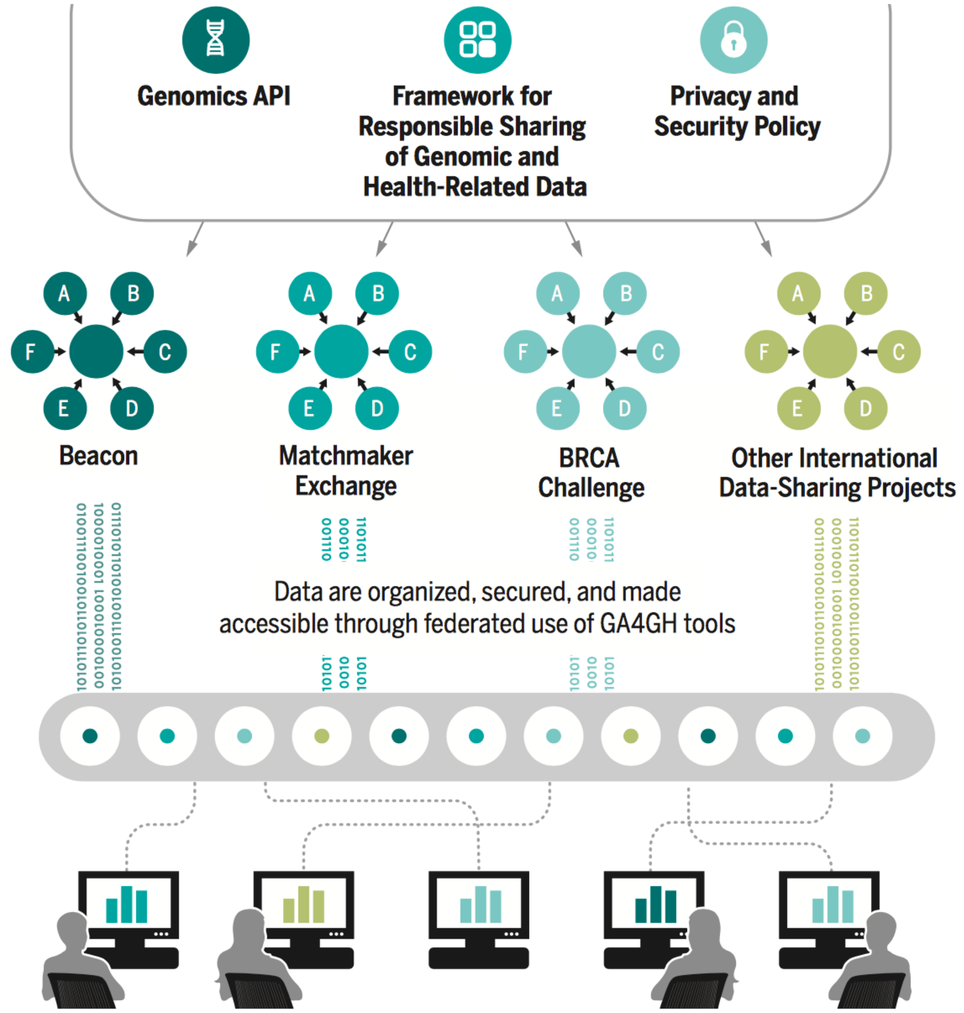
A federated ecosystem for sharing genomic, clinical data

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

The Global Alliance for Genomics and Health*

SCIENCE 10 JUNE 2016 • VOL 352 ISSUE 6291

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



A New Paradigm for Data Sharing

FROM

TO

Data Copying

Data Visiting

A New Paradigm for Data Sharing

FROM

STANDARDS

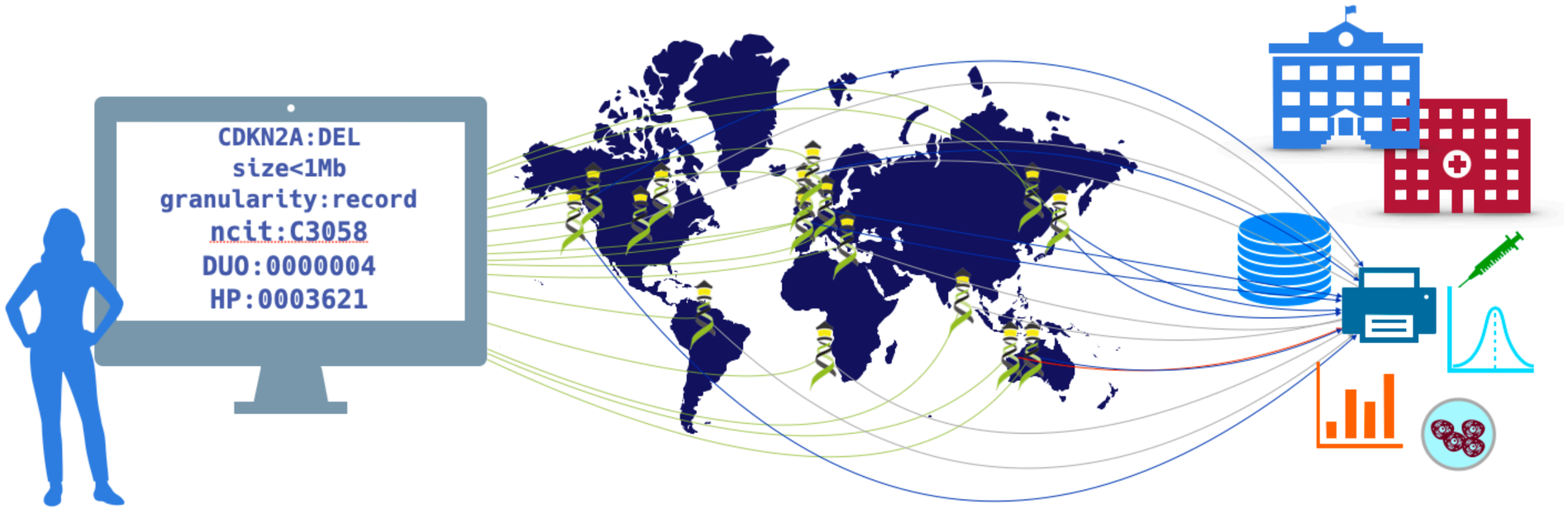
TO

Data Copying

Data Visiting

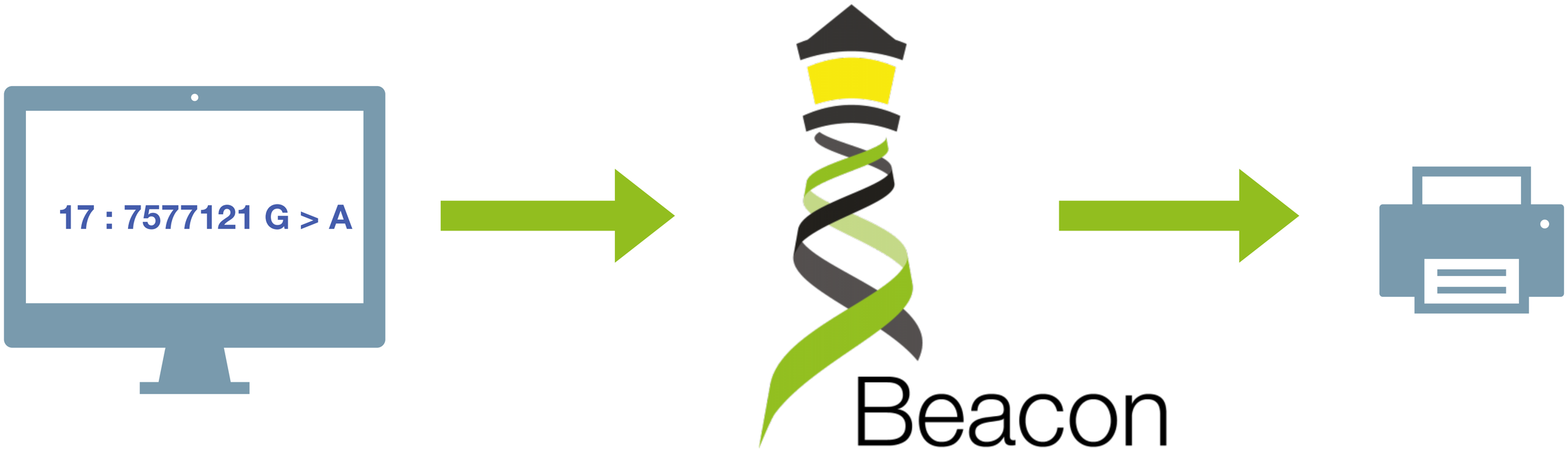
Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.



The GA4GH Beacon Protocol

Federating Genomic Discoveries



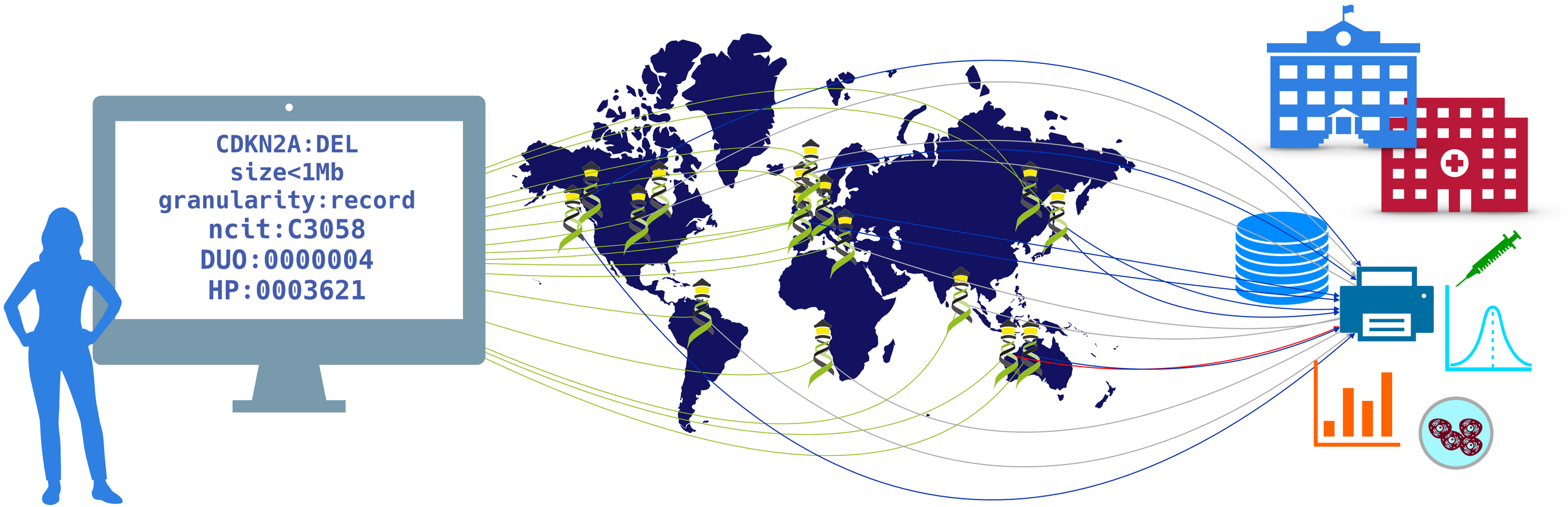
A ***Beacon*** answers a query for a specific genome variant against individual or aggregate genome collections

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

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

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

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



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

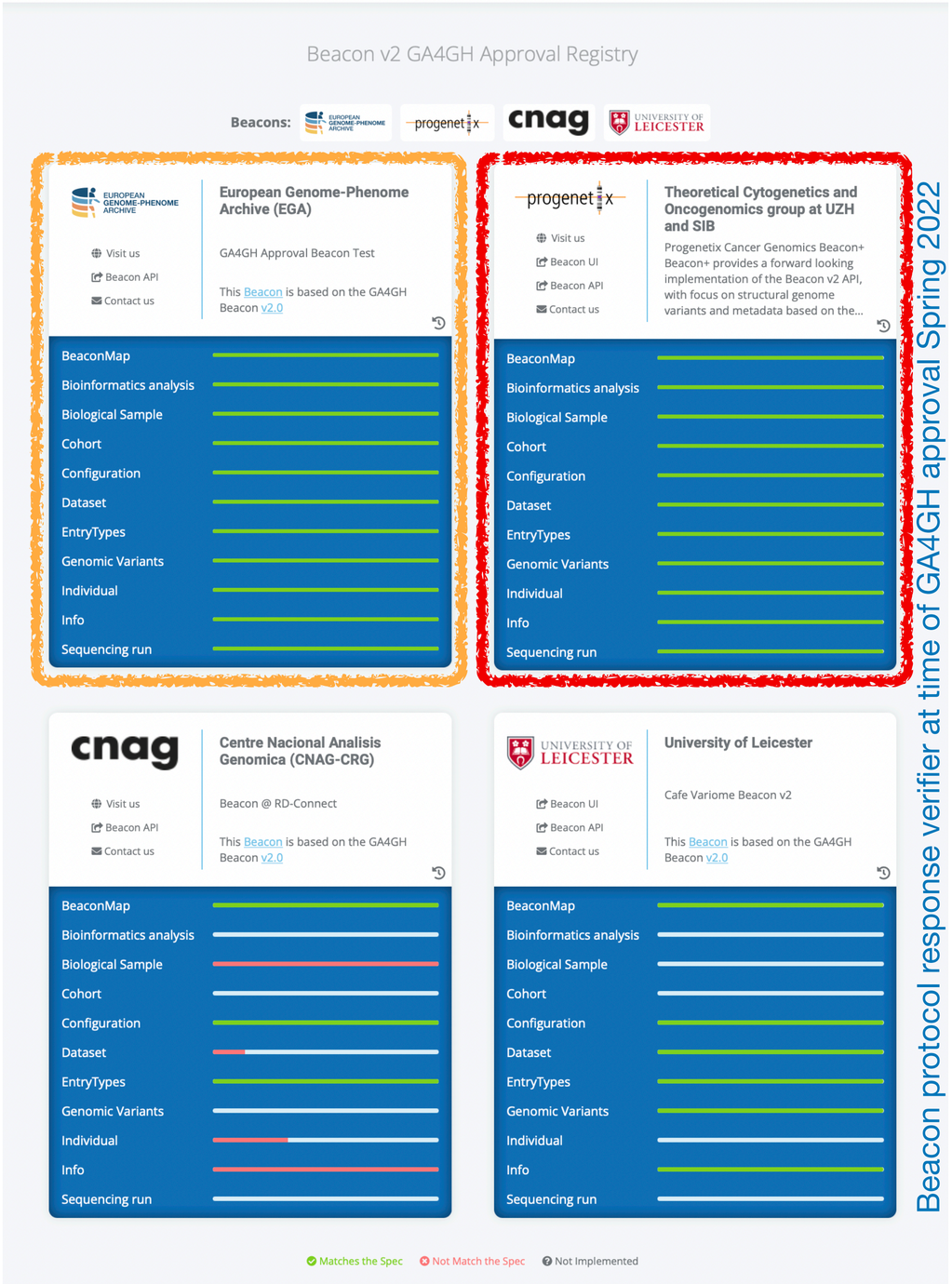
Beacon API

The Beacon API v2 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
  - ➡ ...



Global Alliance for Genomics & Health

progenetix

elixir

Website populated by asynchronous retrieval of Beacon query results using handovers

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**

Edit Query

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

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)

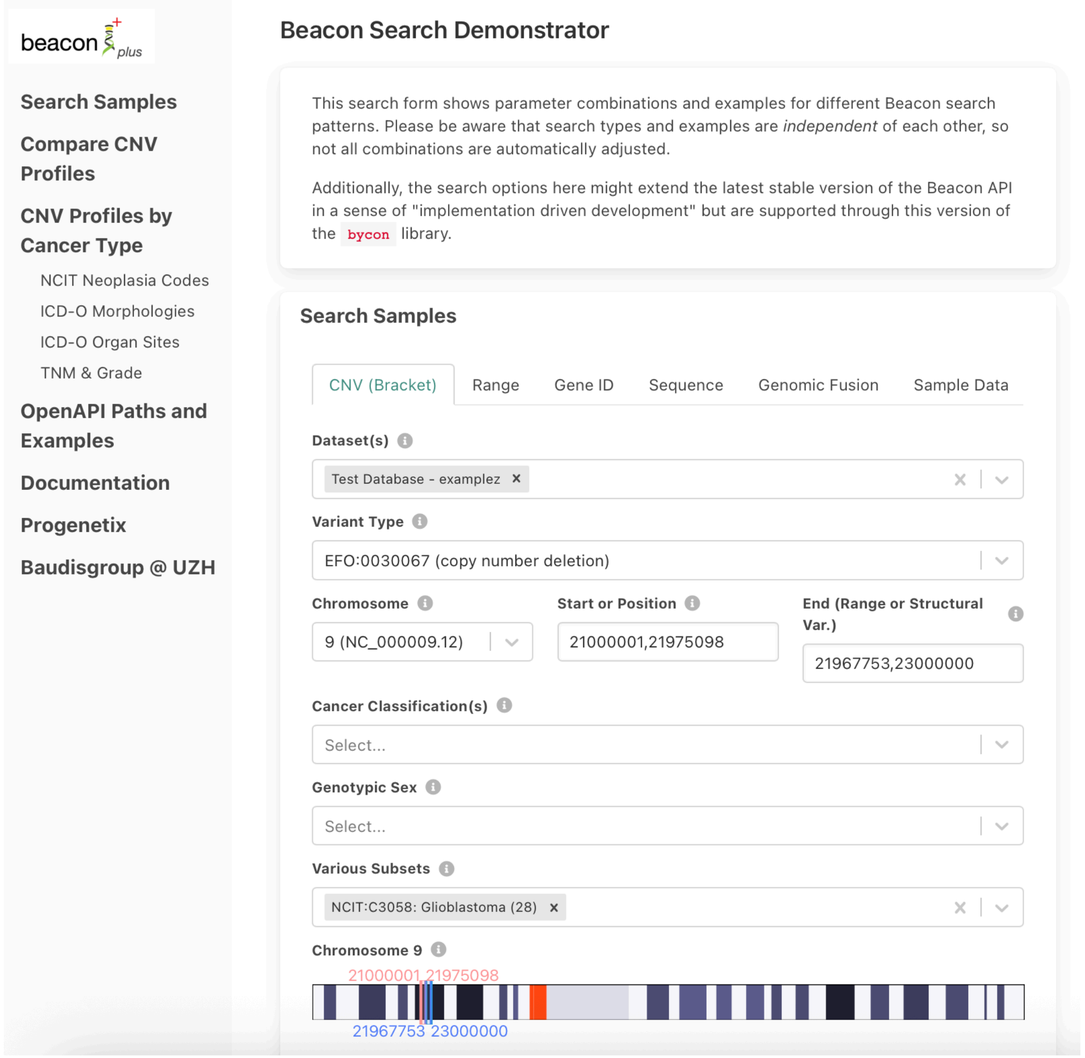
Results | Biosamples | Variants

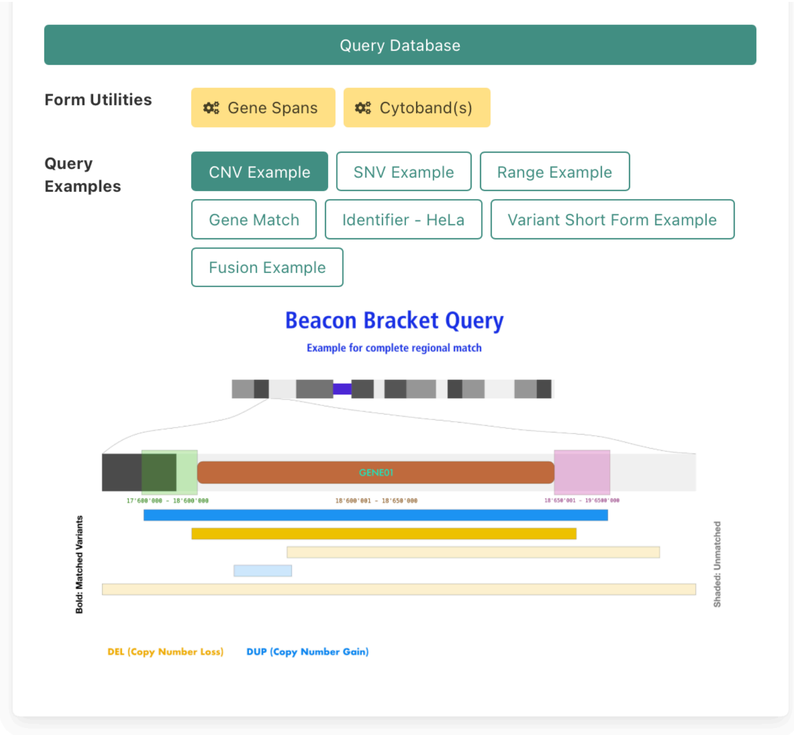
Biosample Id	Dx Classifications	Identifiers
pgxbs-kftvl1hz	pgx:icdom-94403 Glioblastoma, NOS pgx:icdot-C71.9 Brain, NOS NCIT:C3058 Glioblastoma	pubmed:28481359 Zehir A, Benayed R et al. (2017): Mutational landscape of metastatic cancer revealed... cbioportal:msk_impact_2017
pgxbs-kftvl7f4	pgx:icdom-94403 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... cbioportal: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...

Results | Biosamples | Variants

Digest	Gene	Pathogenicity	Variant type	Variant Instances
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-6657226e8f6b961582615aa6 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

beacon plus front-end showcases query strategies



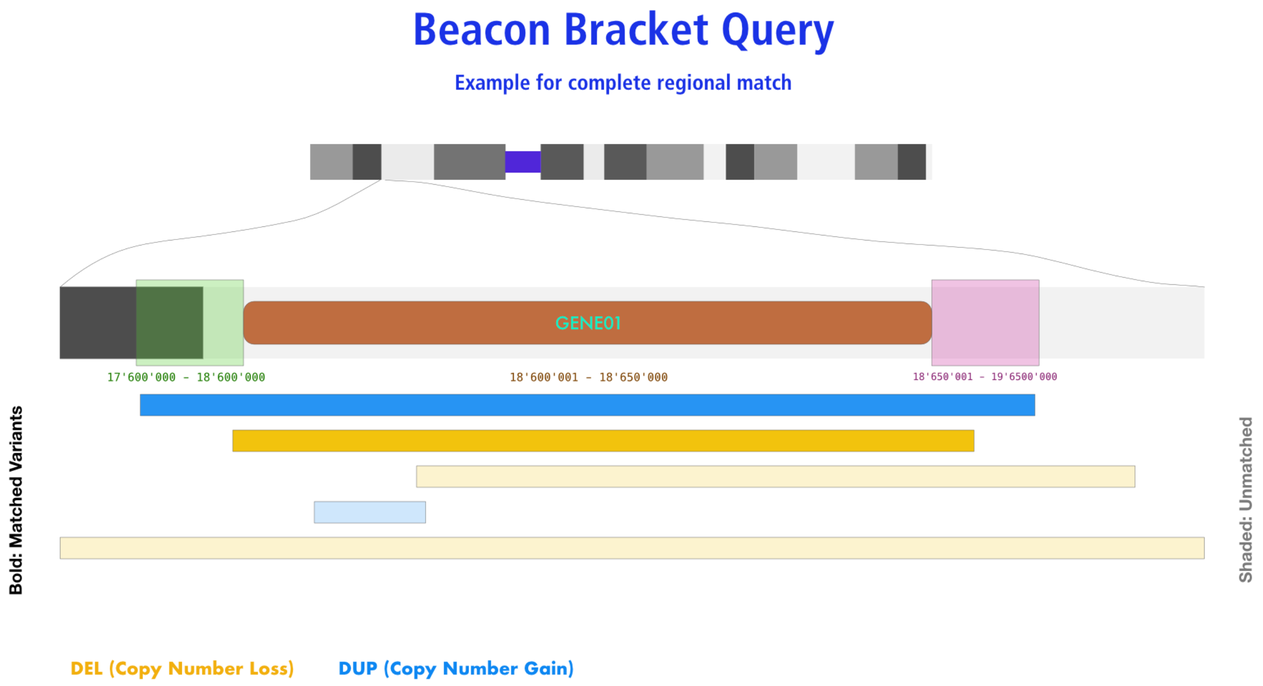


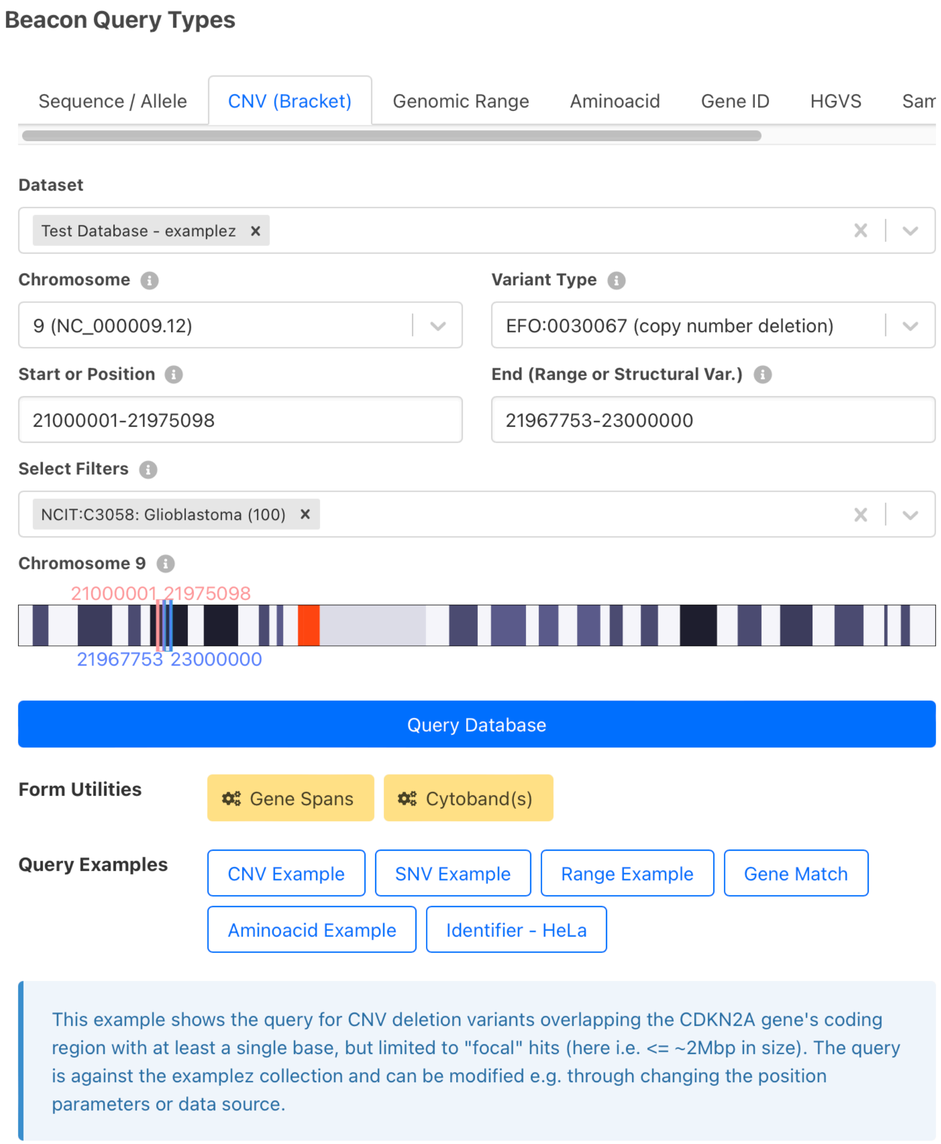
- React based website included in *bycon*
- exposes Beacon query options
- package provided examples can be extended by local ones

Beacon Queries

Bracket ("CNV") Query

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





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

Website populated by asynchronous retrieval of Beacon query results using handovers

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**

Edit Query

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

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

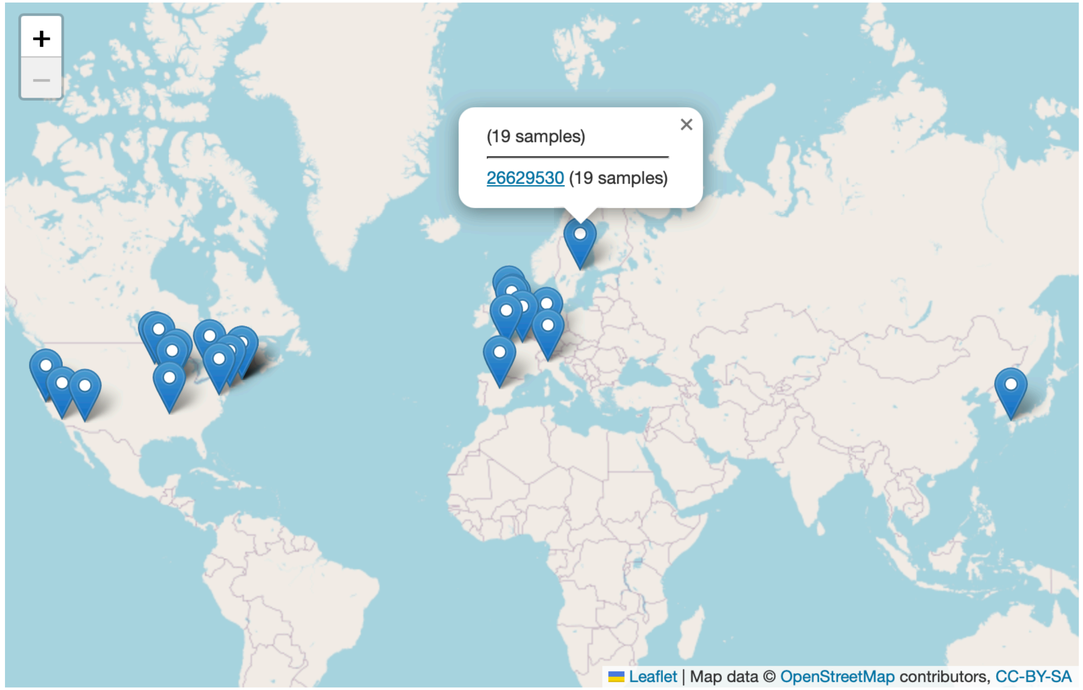
progenetix Data Downloads

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)



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

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

Select Summary Histograms

On this page you can combine one or multiple disease or other codes from one or more datasets frequencies). These based on pre-computed CNV frequencies for the given code; for a combina Page.

Dataset(s)

Progenetix cancer genome variants ×

Cancer Classification(s)

Select...

Various Subsets

pgx:TCGA-ACC: pgx:TCGA-ACC (180) × | pgx:TCGA-BLCA: pgx:TCGA-BLCA (810) × | pgx:TCGA-BRCA: pgx
pgx:TCGA-CESC: pgx:TCGA-CESC (586) × | pgx:TCGA-CHOL: pgx:TCGA-CHOL (85) × | pgx:TCGA-COAD: p
pgx:TCGA-DLBC: pgx:TCGA-DLBC (94) × | pgx:TCGA-ESCA: pgx:TCGA-ESCA (373) × | pgx:TCGA-GBM: pg
pgx:TCGA-HNSC: pgx:TCGA-HNSC (1081) × | pgx:TCGA-KICH: pgx:TCGA-KICH (132) × | pgx:TCGA-KIRC: p
pgx:TCGA-KIRP: pgx:TCGA-KIRP (594) × | pgx:TCGA-LAML: pgx:TCGA-LAML (285) × | pgx:TCGA-LGG: pgx
pgx:TCGA-LIHC: pgx:TCGA-LIHC (767) × | pgx:TCGA-LUAD: pgx:TCGA-LUAD (1110) × | pgx:TCGA-LUSC: p
pgx:TCGA-MESO: pgx:TCGA-MESO (173) × | pgx:TCGA-OV: pgx:TCGA-OV (1133) × | pgx:TCGA-PAAD: pgx:
pgx:TCGA-PCPG: pgx:TCGA-PCPG (363) × | pgx:TCGA-PRAD: pgx:TCGA-PRAD (1038) × | pgx:TCGA-READ
pgx:TCGA-SARC: pgx:TCGA-SARC (519) × | pgx:TCGA-SKCM: pgx:TCGA-SKCM (939) × | pgx:TCGA-STAD:
pgx:TCGA-TGCT: pgx:TCGA-TGCT (271) × | pgx:TCGA-THCA: pgx:TCGA-THCA (1025) × | pgx:TCGA-THYM:
pgx:TCGA-UCEC: pgx:TCGA-UCEC (1087) × | pgx:TCGA-UCS: pgx:TCGA-UCS (110) × | pgx:TCGA-UVM: pg

Plot Type

Select...

Gene Symbol(s) for Labeling

MYC (chr8:127735434-127742951) × | TP53 (chr17:7668421-7687490) × | CCND1 (chr11:69641156-69654
CDKN2A (chr9:21967752-21995324) × | EGFR (chr7:55019017-55211628) ×

Chromosomes to Plot

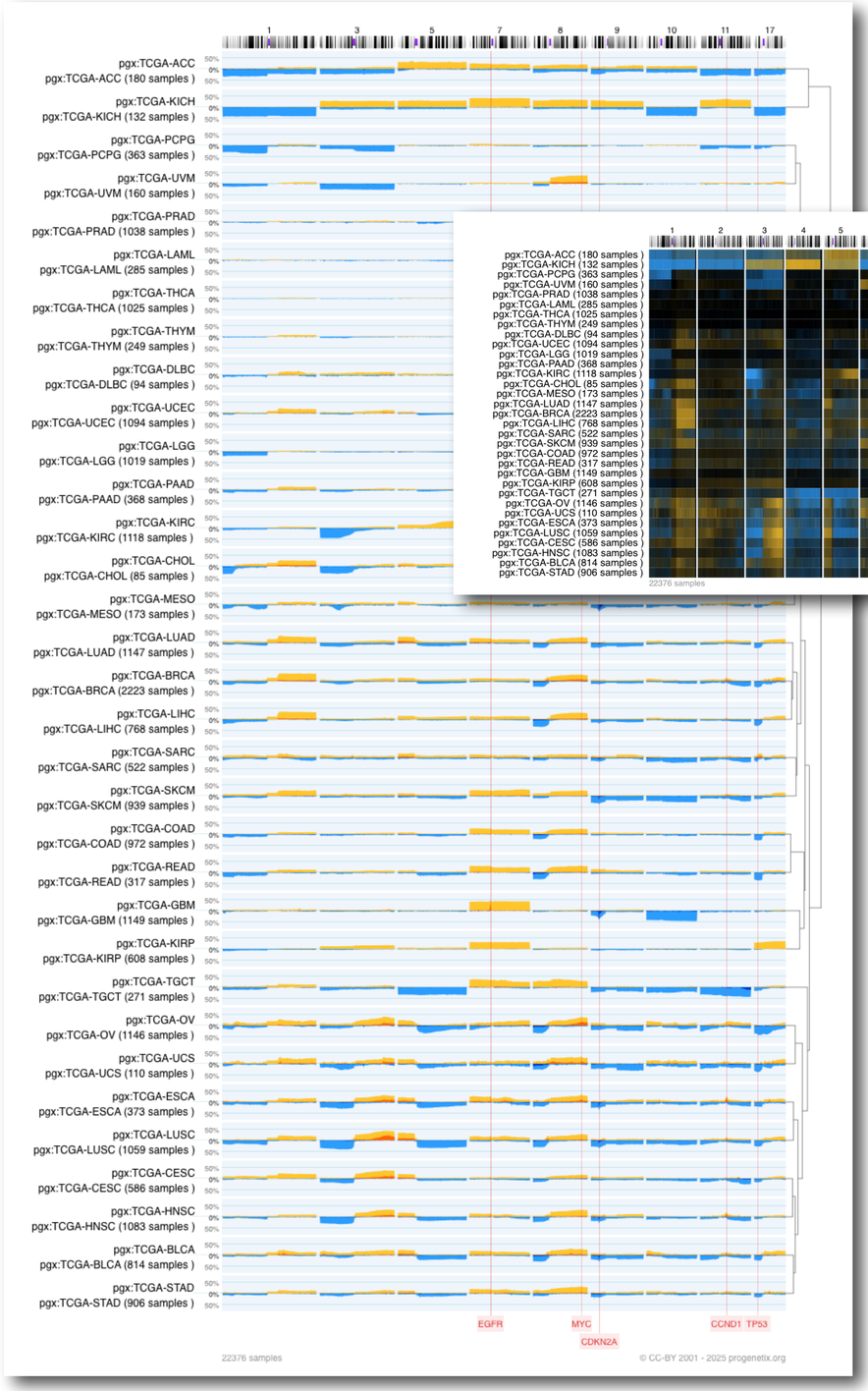
1,3,5,7,8,9,10,11,17

Plot Parameters

plot_axis_y_max:80;plot_area_height:40;plot_label_y_values:0,50

Retrieve CNV Profiles

B-Cell Lymphomas | TCGA Projects | Lung Cancers | Breast Cancers and Cell Lines



beacon plus

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

bycon based Beacon+ Stack

progenetix

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

React

APACHE HTTP SERVER PROJECT

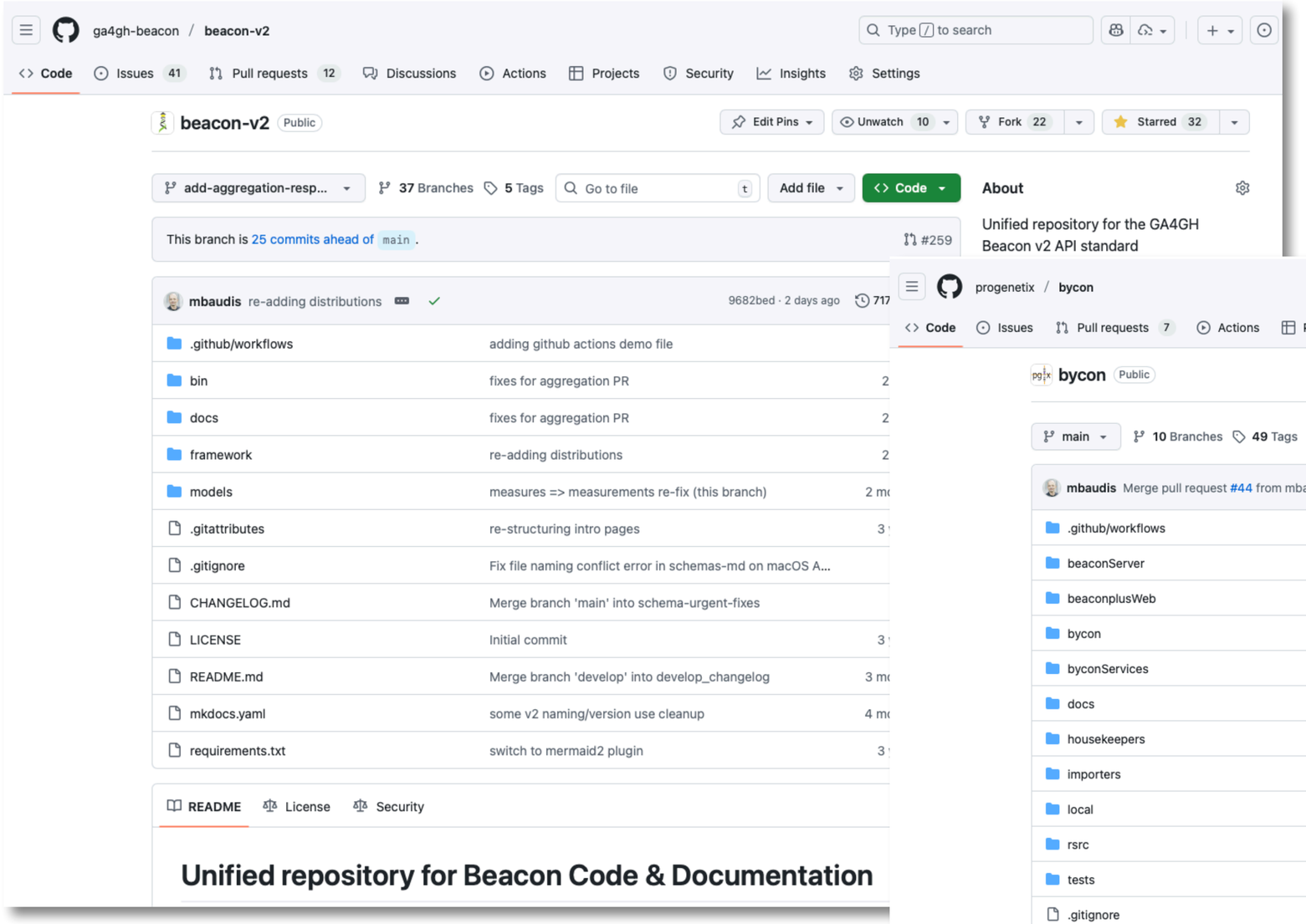
python™

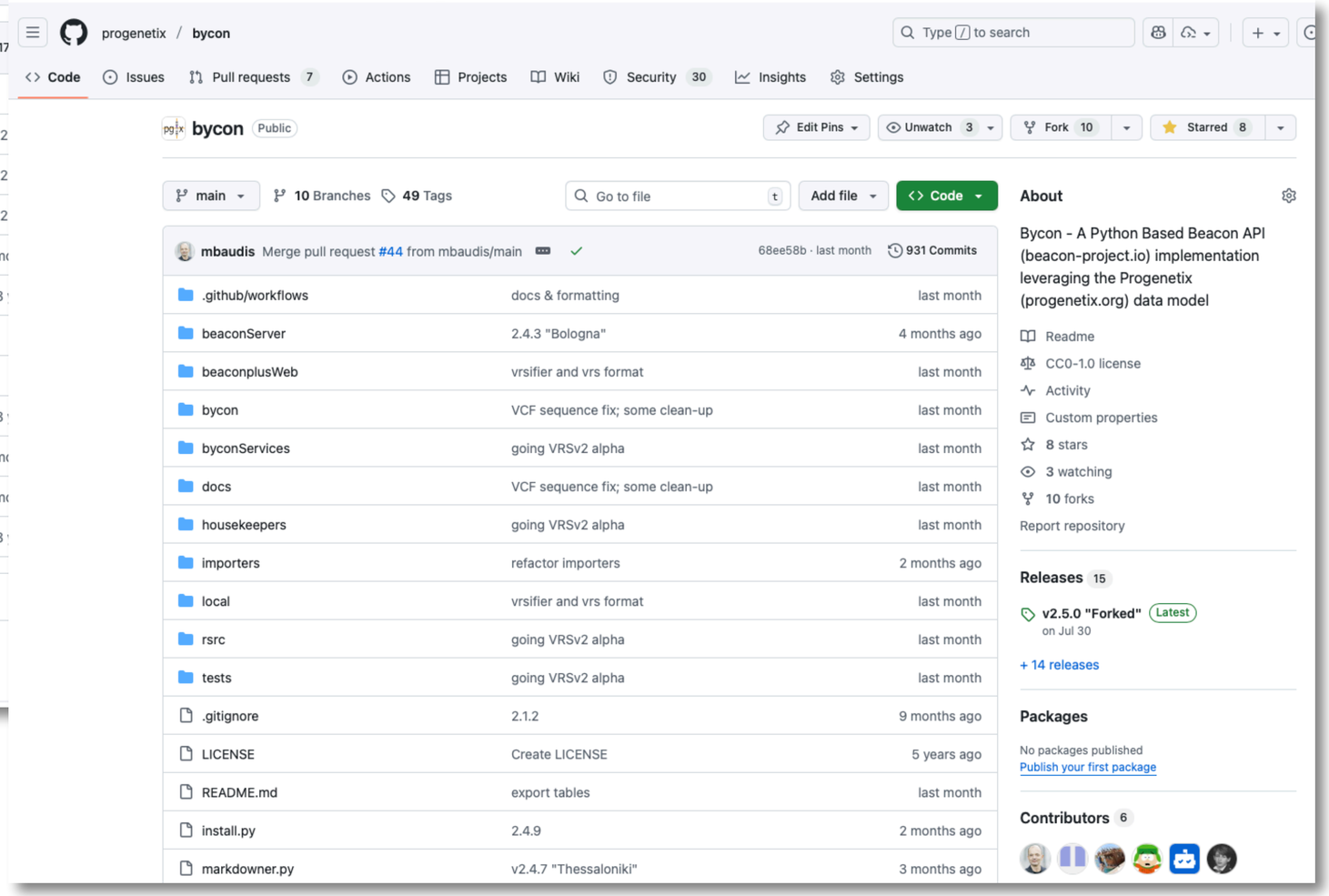
mongoDB

Entity collections

Utility collections

github.com/progenetix/bycon GitHub


ga4gh-beacon / beacon-v2
Type / to search
Code
Issues 41
Pull requests 12
Discussions
Actions
Projects
Security
Insights
Settings
beacon-v2
Public
Edit Pins
Unwatch 10
Fork 22
Starred 32
add-aggregation-resp...
37 Branches
5 Tags
Go to file
t
Add file
Code
About
Unified repository for the GA4GH Beacon v2 API standard
This branch is 25 commits ahead of main .
#259
mbaudis re-adding distributions
9682bed · 2 days ago
.github/workflows
adding github actions demo file
bin
fixes for aggregation PR
docs
fixes for aggregation PR
framework
re-adding distributions
models
measures => measurements re-fix (this branch)
.gitattributes
re-structuring intro pages
.gitignore
Fix file naming conflict error in schemas-md on macOS A...
CHANGELOG.md
Merge branch 'main' into schema-urgent-fixes
LICENSE
Initial commit
README.md
Merge branch 'develop' into develop_changelog
mkdocs.yaml
some v2 naming/version use cleanup
requirements.txt
switch to mermaid2 plugin
README
License
Security
Unified repository for Beacon Code & Documentation


progenetix / bycon
Type / to search
Code
Issues
Pull requests 7
Actions
Projects
Wiki
Security 30
Insights
Settings
bycon
Public
Edit Pins
Unwatch 3
Fork 10
Starred 8
main
10 Branches
49 Tags
Go to file
t
Add file
Code
About
Bycon - A Python Based Beacon API (beacon-project.io) implementation leveraging the Progenetix (progenetix.org) data model
Readme
CC0-1.0 license
Activity
Custom properties
8 stars
3 watching
10 forks
Report repository
mbaudis Merge pull request #44 from mbaudis/main
68ee58b · last month
931 Commits
.github/workflows
docs & formatting
last month
beaconServer
2.4.3 "Bologna"
4 months ago
beaconplusWeb
vrsifier and vrs format
last month
bycon
VCF sequence fix; some clean-up
last month
byconServices
going VRSv2 alpha
last month
docs
VCF sequence fix; some clean-up
last month
housekeepers
going VRSv2 alpha
last month
importers
refactor importers
2 months ago
local
vrsifier and vrs format
last month
rsrc
going VRSv2 alpha
last month
tests
going VRSv2 alpha
last month
.gitignore
2.1.2
9 months ago
LICENSE
Create LICENSE
5 years ago
README.md
export tables
last month
install.py
2.4.9
2 months ago
markdowner.py
v2.4.7 "Thessaloniki"
3 months ago
Releases 15
v2.5.0 "Forked"
Latest
on Jul 30
+ 14 releases
Packages
No packages published
Publish your first package
Contributors 6

pgxRpi

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

pgxRpi Public

Edit Pins | Watch 2

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

hangjiaz version bump 319d27c 4 days ago 87 commits

R	adapt to beacon variant export	4 days ago
data-raw	add data documentation; optimise get method which causes error in ...	3 months ago
data	change based on opinion from bioc reviewer	last month
inst	add data documentation; optimise get method which causes error in ...	3 months ago
man	documentation for parameters using match.arg	last month
tests		
vignettes		
.Rbuildignore		
.gitignore		
DESCRIPTION		
NAMESPACE		
NEWS.md		
README.md		

Bioinformatics Advances, 2025, **00**, vbaf172
https://doi.org/10.1093/bioadv/vbaf172
Advance Access Publication Date: 16 July 2025
Application Note

OXFORD

Data and text mining

pgxRpi: an R/bioconductor package for user-friendly access to the Beacon v2 API

Hangjia Zhao[1,2,*] and Michael Baudis[1,2,*]

[1]Department of Molecular Life Sciences, University of Zurich, Zurich 8057, Switzerland
[2]Computational Oncogenomics Group, Swiss Institute of Bioinformatics, Zurich 8057, Switzerland

*Corresponding authors. Hangjia Zhao, Department of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland. E-mail: hangjia.zhao@uzh.ch; Michael Baudis, Department of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, Zurich 8057, Switzerland. E-mail: michael.baudis@mls.uzh.ch.

2 Retrieve meatdata of samples

2.1 Relevant parameters

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

2.2 Search by filters

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

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

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

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

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

baudisgroup @ UZH

Project Opportunities

progenetix

cancercelllines.org

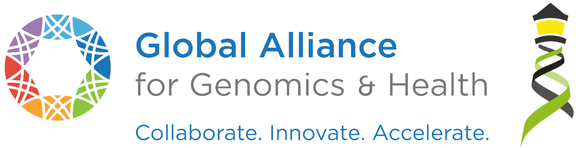
refCNV

arrayMap 

- data & metadata analysis projects
  - disease relations, trees, trajectories
  - geographies, disease statistics and clinical explorations
- data curation and resource generation
  - identification, retrieval and conversion of CNV and metadata from publications and repositories
  - ontology and classification mapping
  - clinical and metadata enrichment
  - beyond CNV mutation data & double hit events

baudisgroup @ UZH

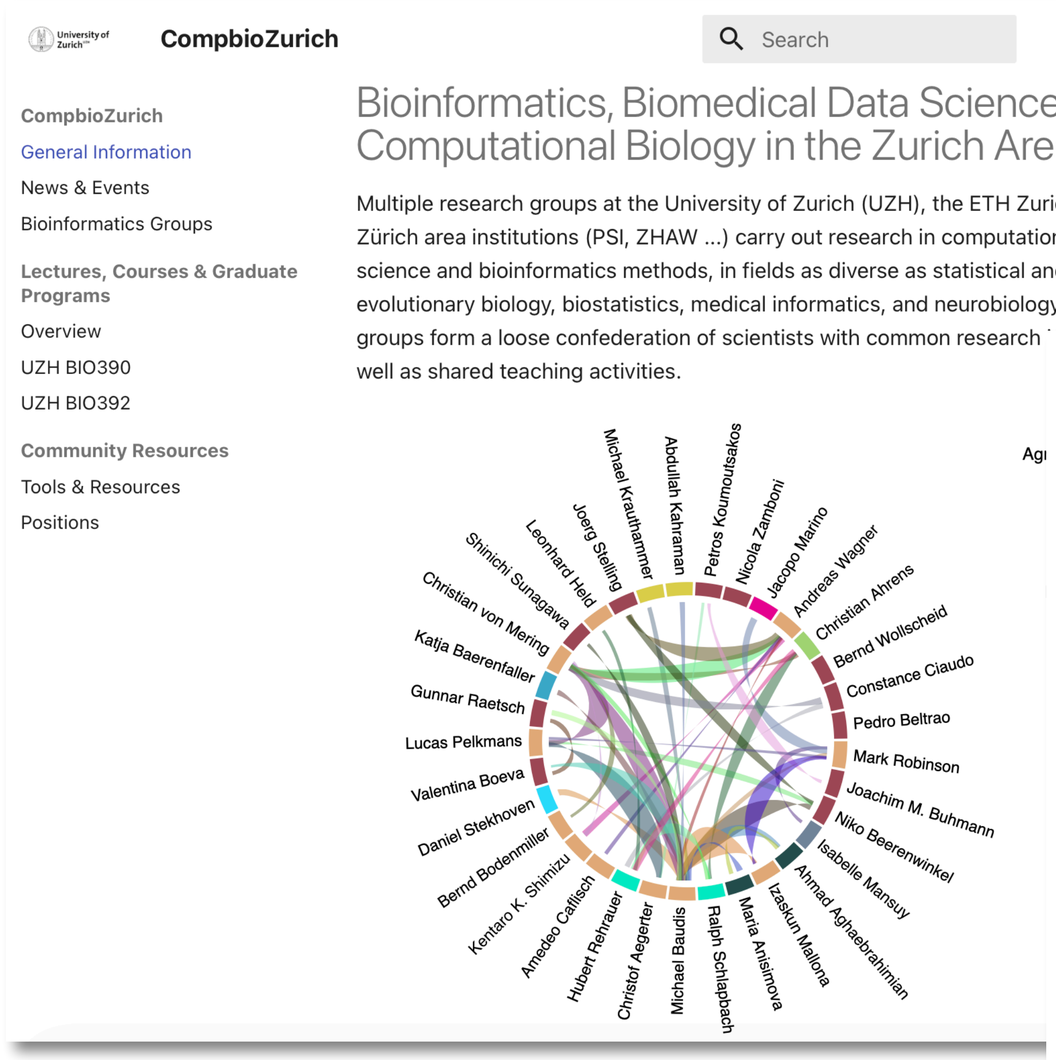
Project Opportunities: Beacon Ecosystem

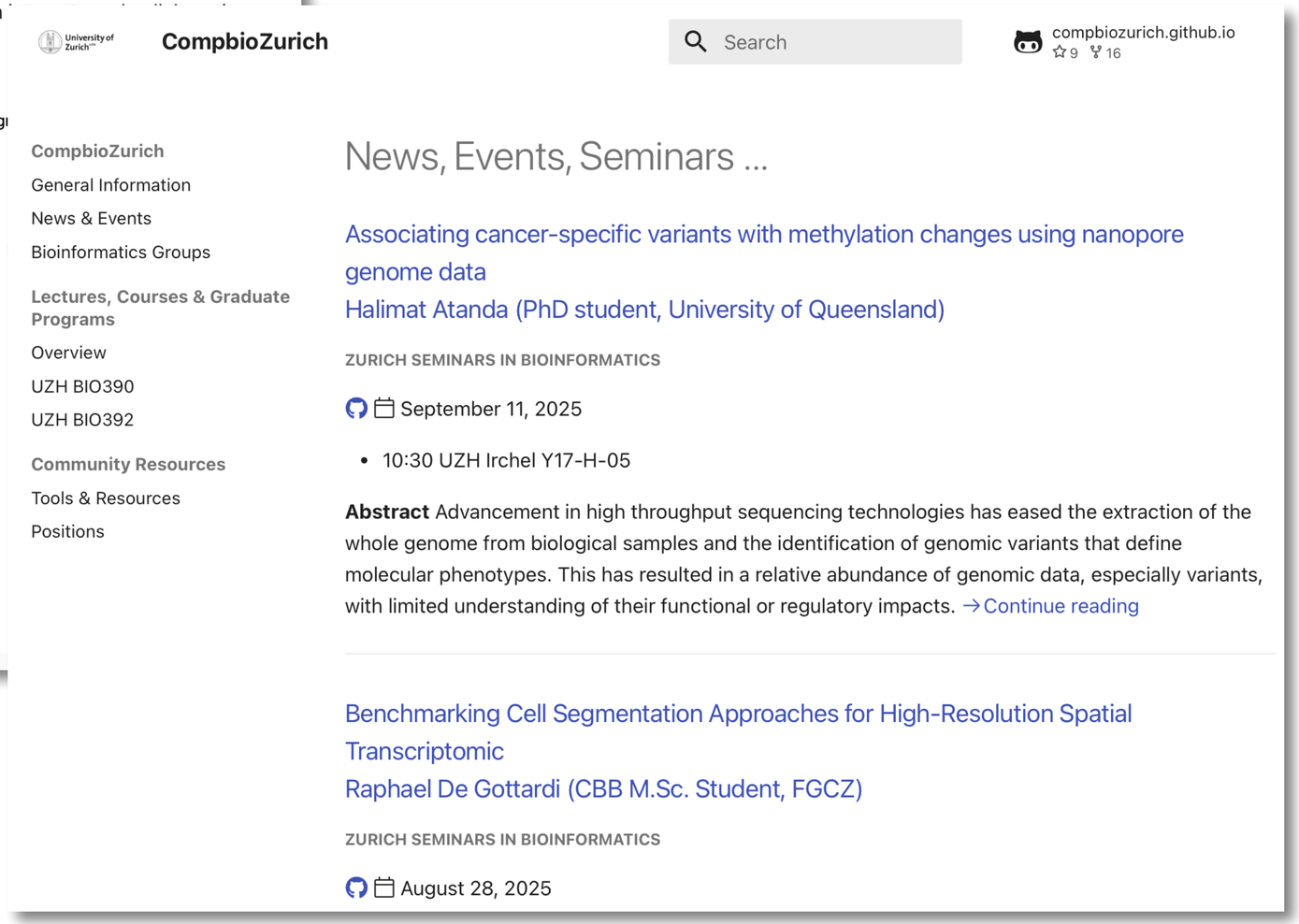


- schema development and implementation
  - adapting and contributing to **GA4GH** standards (Beacon but also VRS and core schemas)
- the **bycon** software stack
  - **Python** & **MongoDB** code development and optimization
  - packaging and distribution
  - security & authentication
  - aggregator...
- front-end, visualization and web deployment
  - **Java Script / React** development
  - queries, maps, genome plots ...

Zürich Seminars in Bioinformatics
Thursdays 12:15-13:00
UZH Irchel Campyus & Zoom

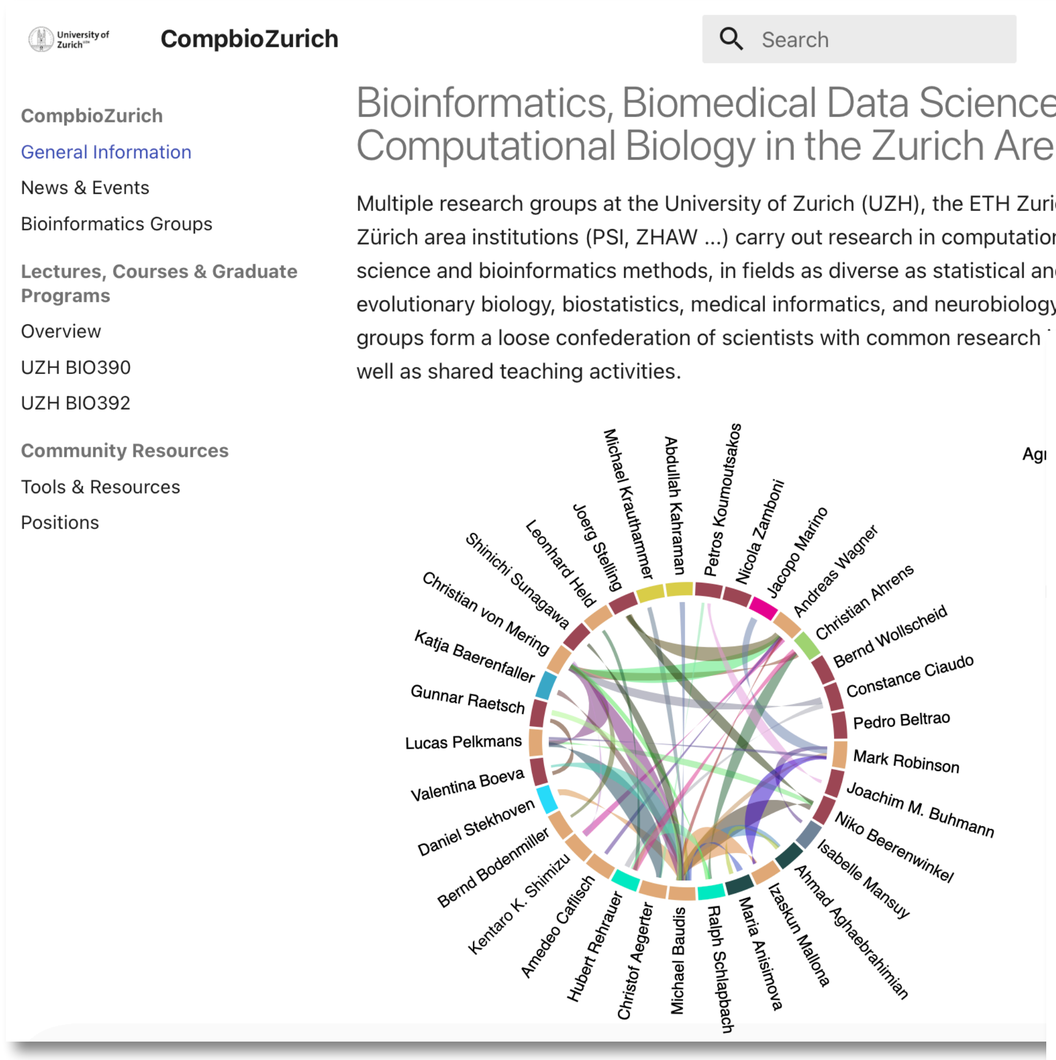
compbiozurich.org

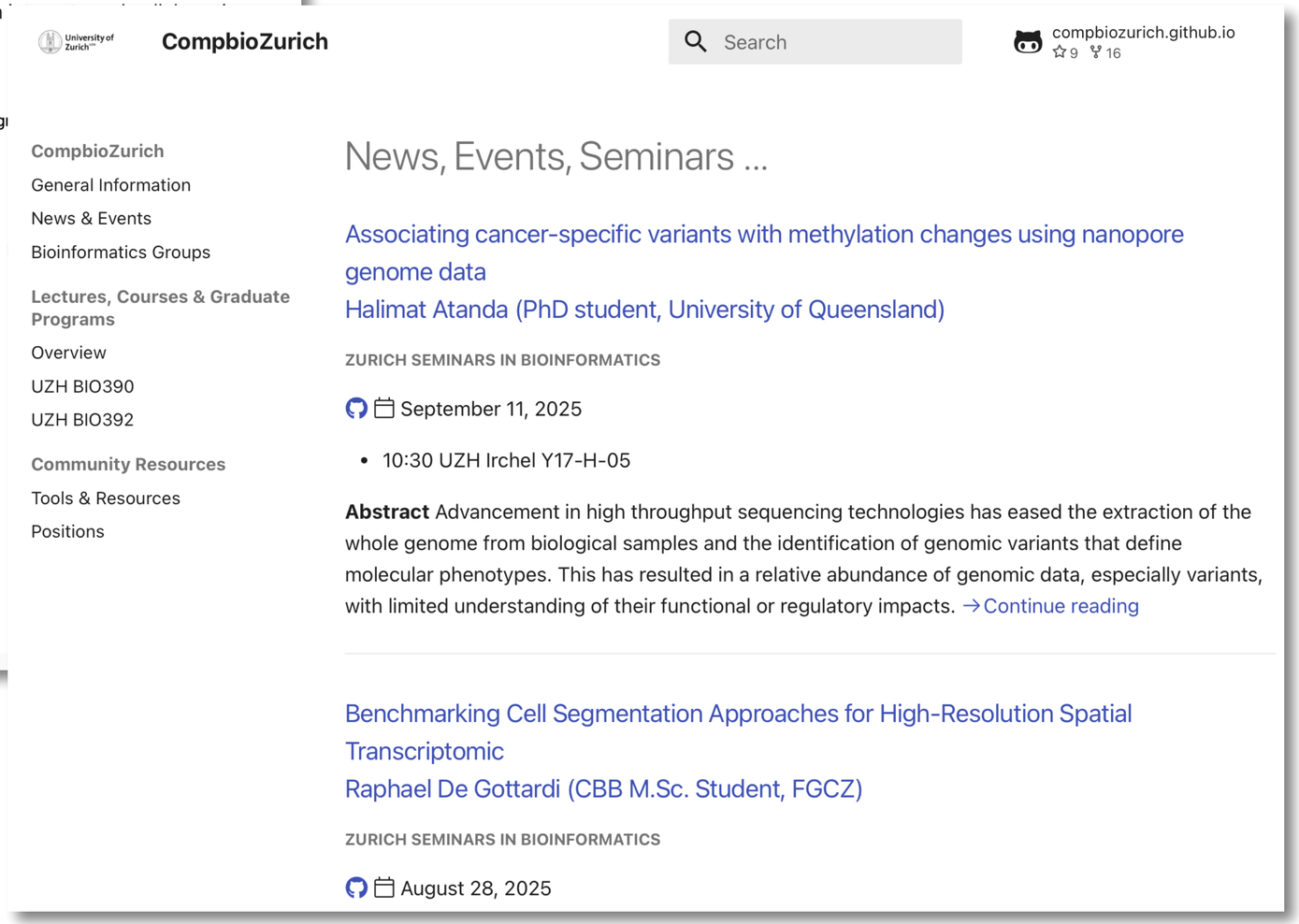




Zürich Seminars in Bioinformatics
Thursdays 12:15-13:00
UZH Irchel Campyus & Zoom

compbiozurich.org





University of Zurich[UZH]
Department of Molecular Life Sciences

Global Alliance for Genomics & Health

elixir

SIB Swiss Institute of Bioinformatics

Prof. Dr. Michael Baudis
Institute of Molecular Life Sciences
University of Zurich
SIB I Swiss Institute of Bioinformatics
Winterthurerstrasse 190
CH-8057 Zurich
Switzerland