

A Reference Resource for Copy Number Variations in Cancer

Implementing GA4GH Standards to Drive an Open Oncogenomics Resource

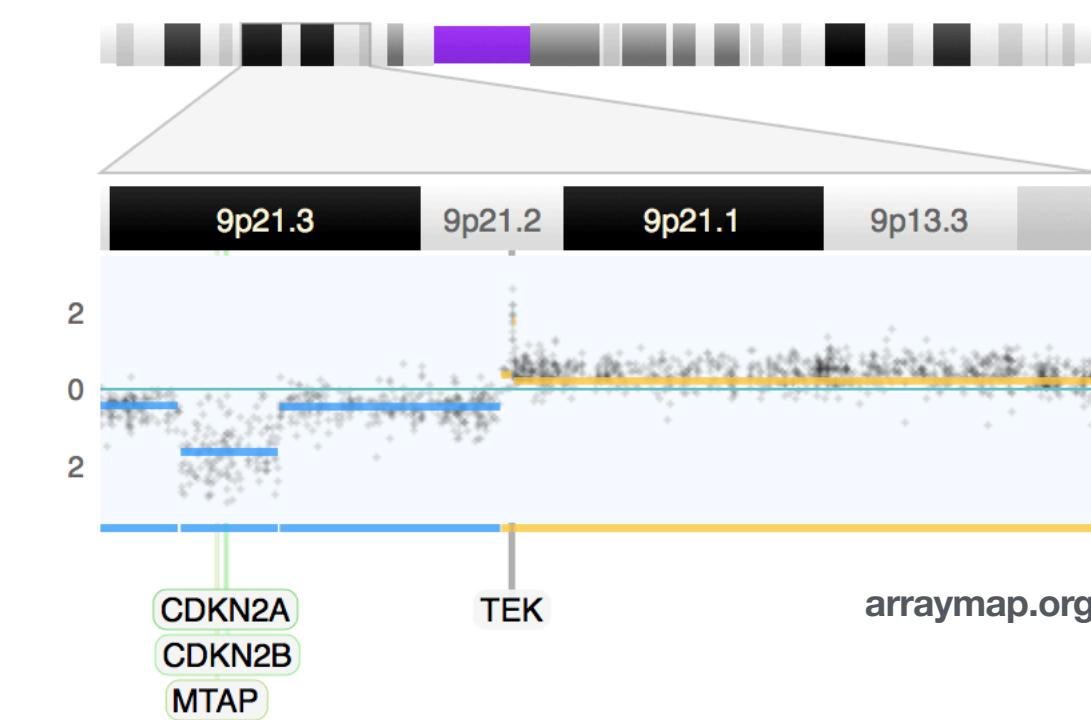
Qingyao Huang | CGC | 2021-08-03



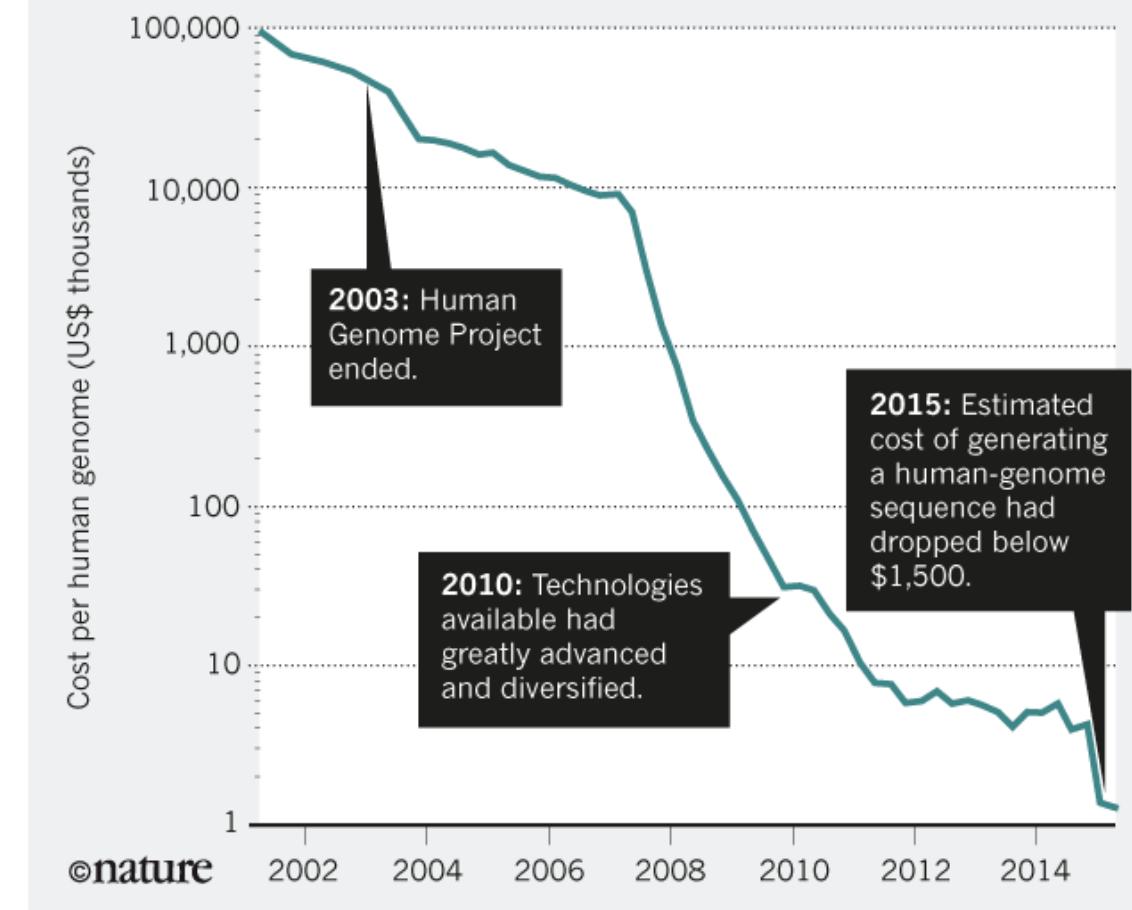


Genome screening at the core of “Personalized Health”

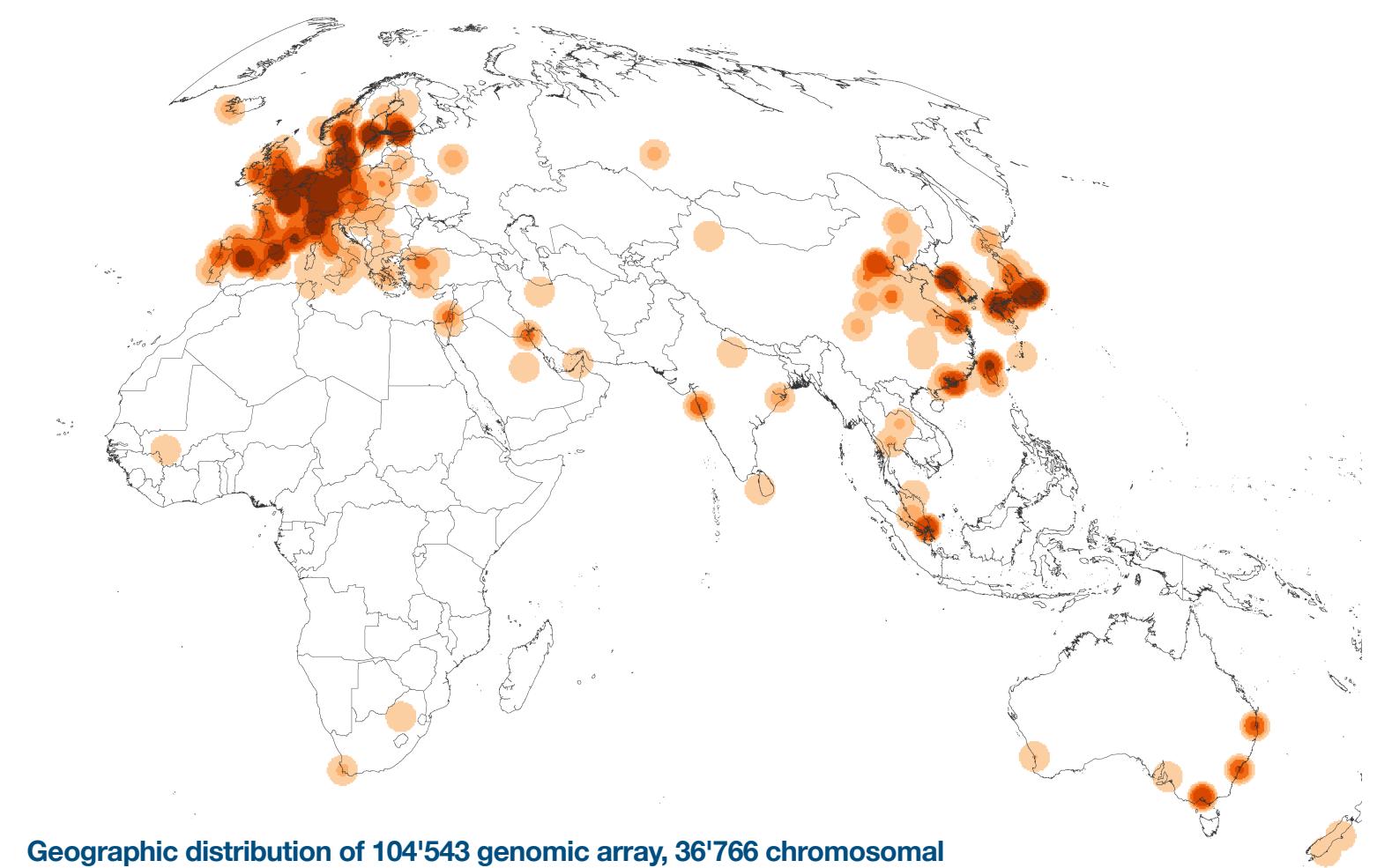
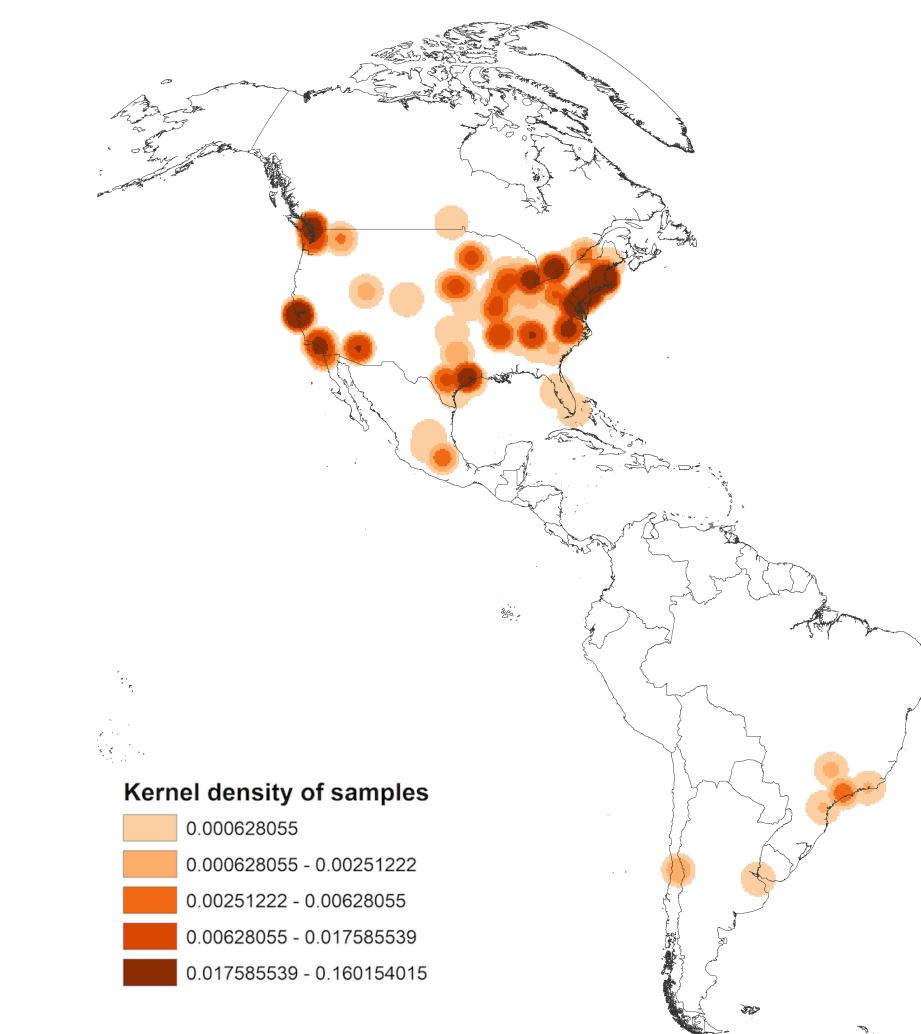
- ▶ **Genome analyses** (including transcriptome, metagenomics) are core technologies for Personalised Health™ applications
- ▶ The unexpectedly large amount of **sequence variants** in human genomes - germline and somatic/cancer - requires huge analysis efforts and creation of **reference repositories**
- ▶ **Standardized data formats** and **exchange protocols** are needed to connect these resources throughout the world, for reciprocal, international **data sharing** and **biocuration** efforts
- ▶ Our work @ UZH:
 - ▶ **cancer genome repositories**
 - ▶ **biocuration**
 - ▶ **protocols & formats**



BETTER, CHEAPER, FASTER
The cost of DNA sequencing has dropped dramatically over the past decade, enabling many more applications.



The future of DNA sequencing. Eric D. Green, Edward M. Rubin & Maynard V. Olson. Nature; 11 October 2017 (News & Views)



Geographic distribution of 104'543 genomic array, 36'766 chromosomal CGH and 15'409 whole genome/exome based cancer genome datasets

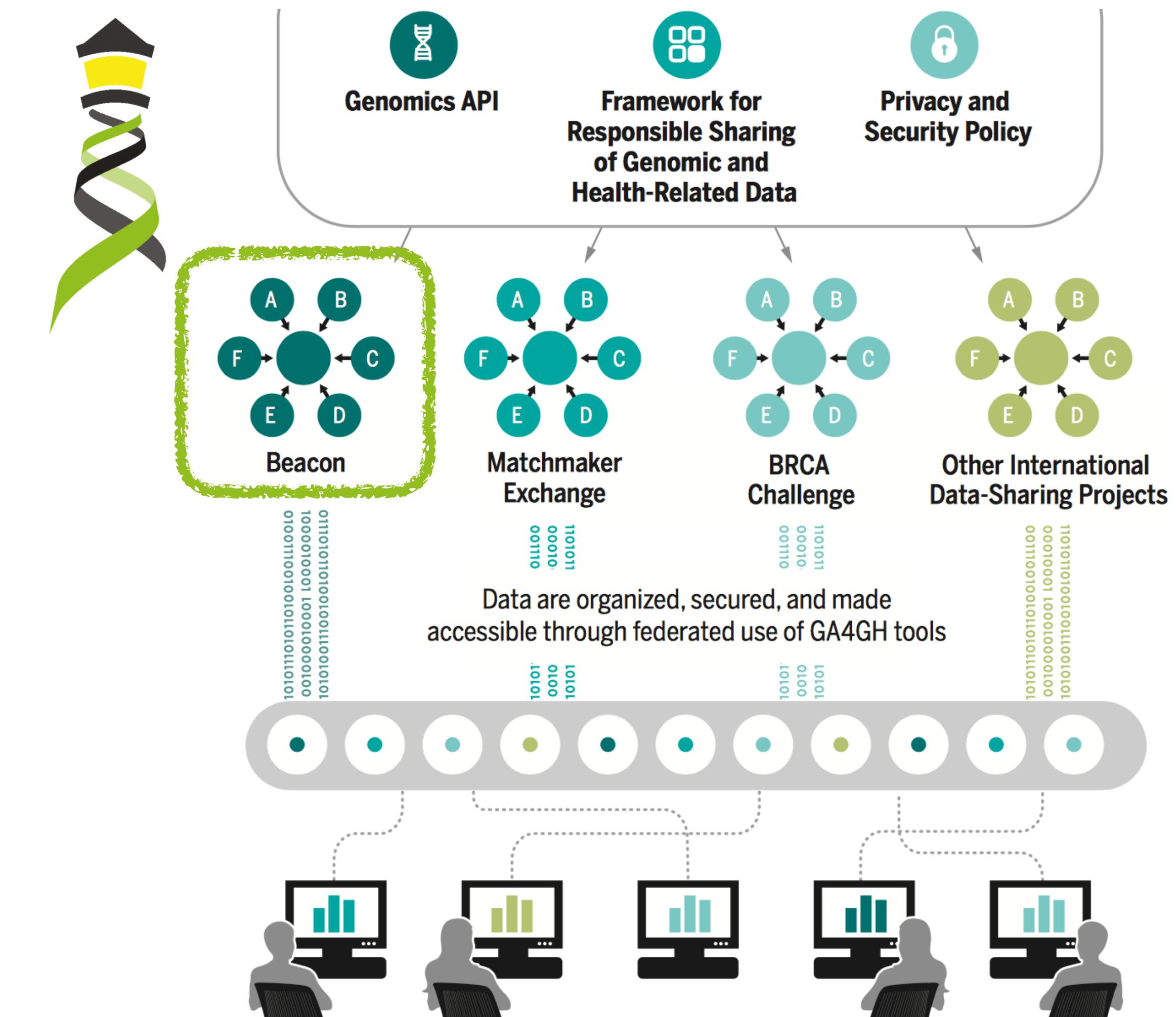


GENOMICS

A federated ecosystem for sharing genomic, clinical data

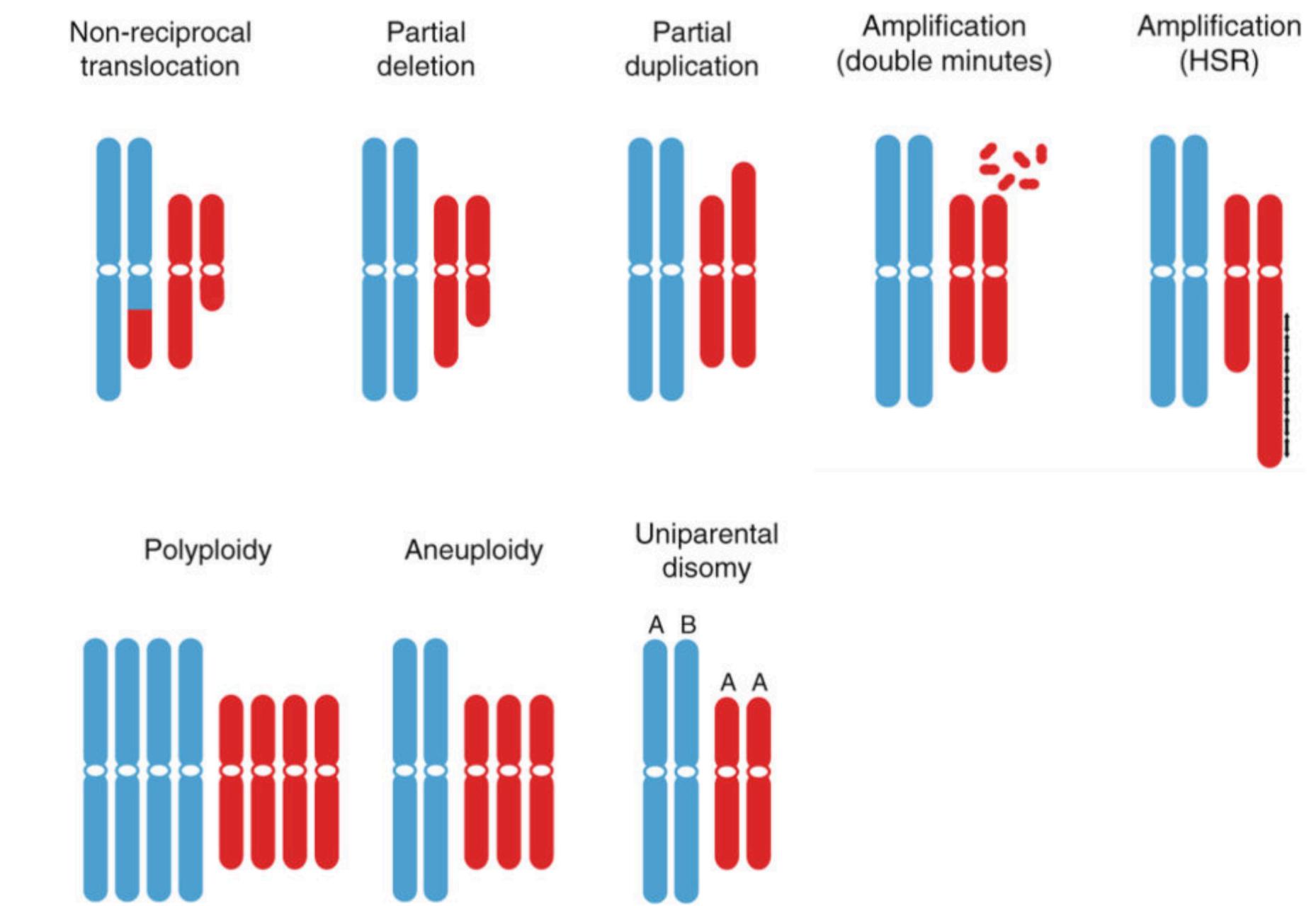
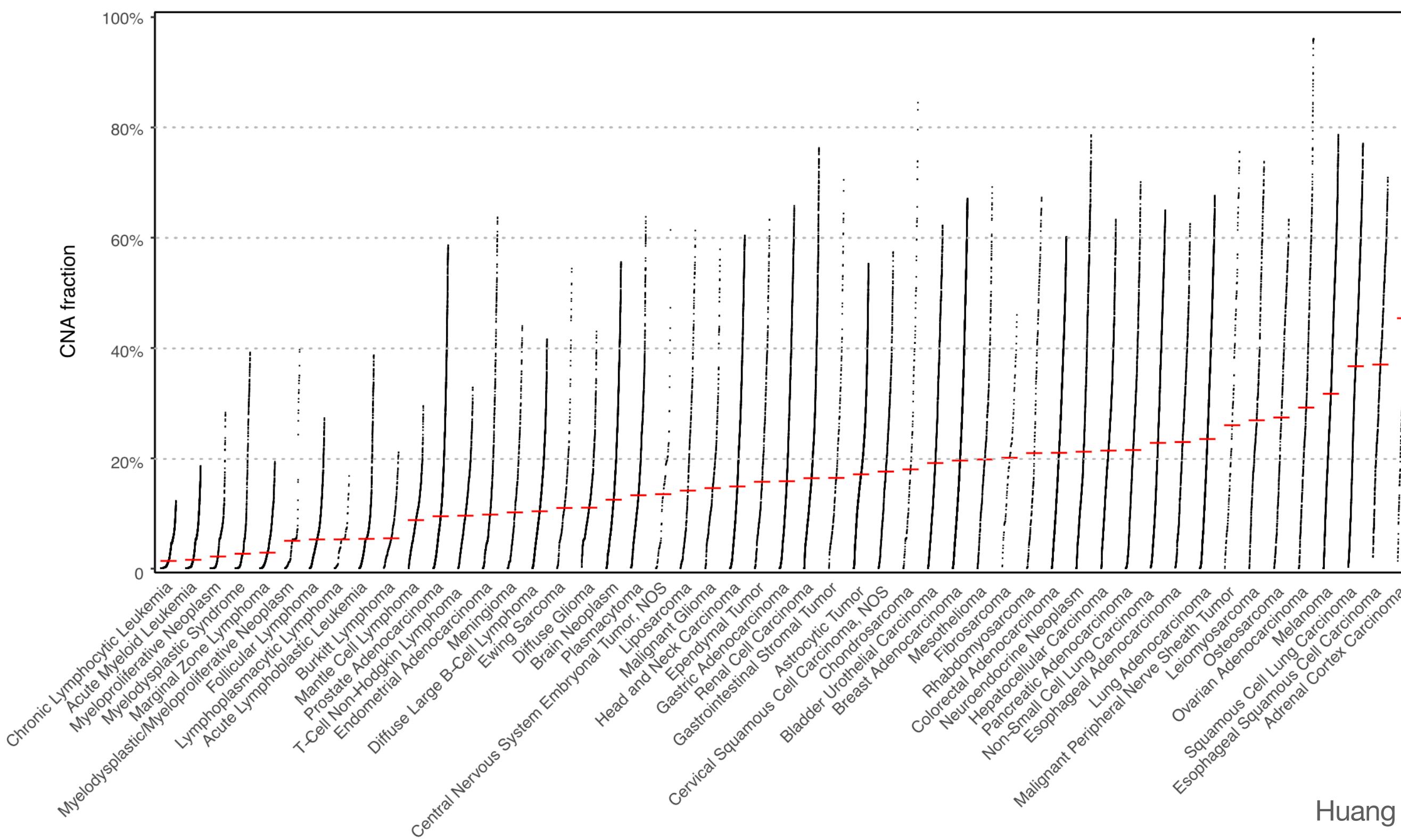
Silos of genome data collection are being transformed into seamlessly connected, independent systems

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



Genomic alterations in Cancer

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



Grade et al., 2015 Chromosomal Instability in Cancer Cells

History & Current State...

Origins & Trajectory of the Progenetix Resource



Progenetix, early 2000's

Text conversion for CNVs

- originally an internal, to store CGH profiles at Peter Lichter's group at the German Cancer Research Center (DKFZ), starting in 1998
- expansion to include **literature derived data**, with a focus on malignant non-Hodgkin's lymphomas
- based on articles and supplements with **cytoband-based** *revisish* CGH results
- sometimes rich, but **unstructured** associated information
- PDFs** readable, but **not well suited for data** extraction (character entities, text flow)

Domain Name: PROGENETIX.NET
 Registry Domain ID: 45678826_DOMAIN_NET-VRSN
 Registrar WHOIS Server: whois.enterprise.net
 Registrar URL: http://www.epag.de
 Updated Date: 2019-06-01T04:20:49Z
 Creation Date: 2000-11-29T18:17:38Z

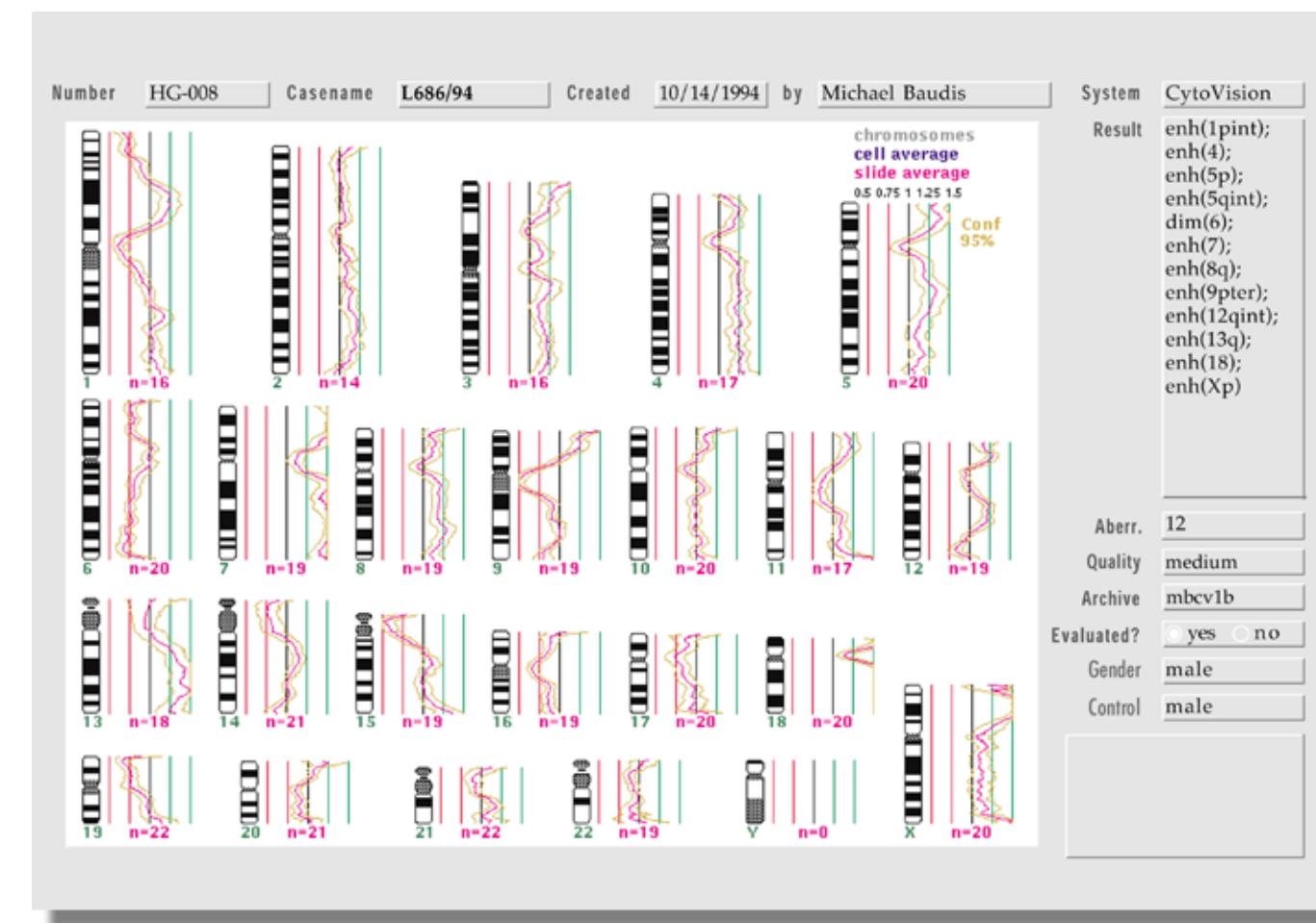


TABLE 3. Comparison of Primary Tumors and Metastases by CGH

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

TABLE 1. Clinical Data

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

^aAJCC/UICC staging system (Hutter and Sabin, 1986).^bGrade of primary tumor: 1–3, low, moderate, high grade; 9, grading unknown.^cSynchronous, diagnosis of metastatic disease within 12 months following diagnosis of primary tumor; metachronous, diagnosis of metastatic disease after 12 months or later.

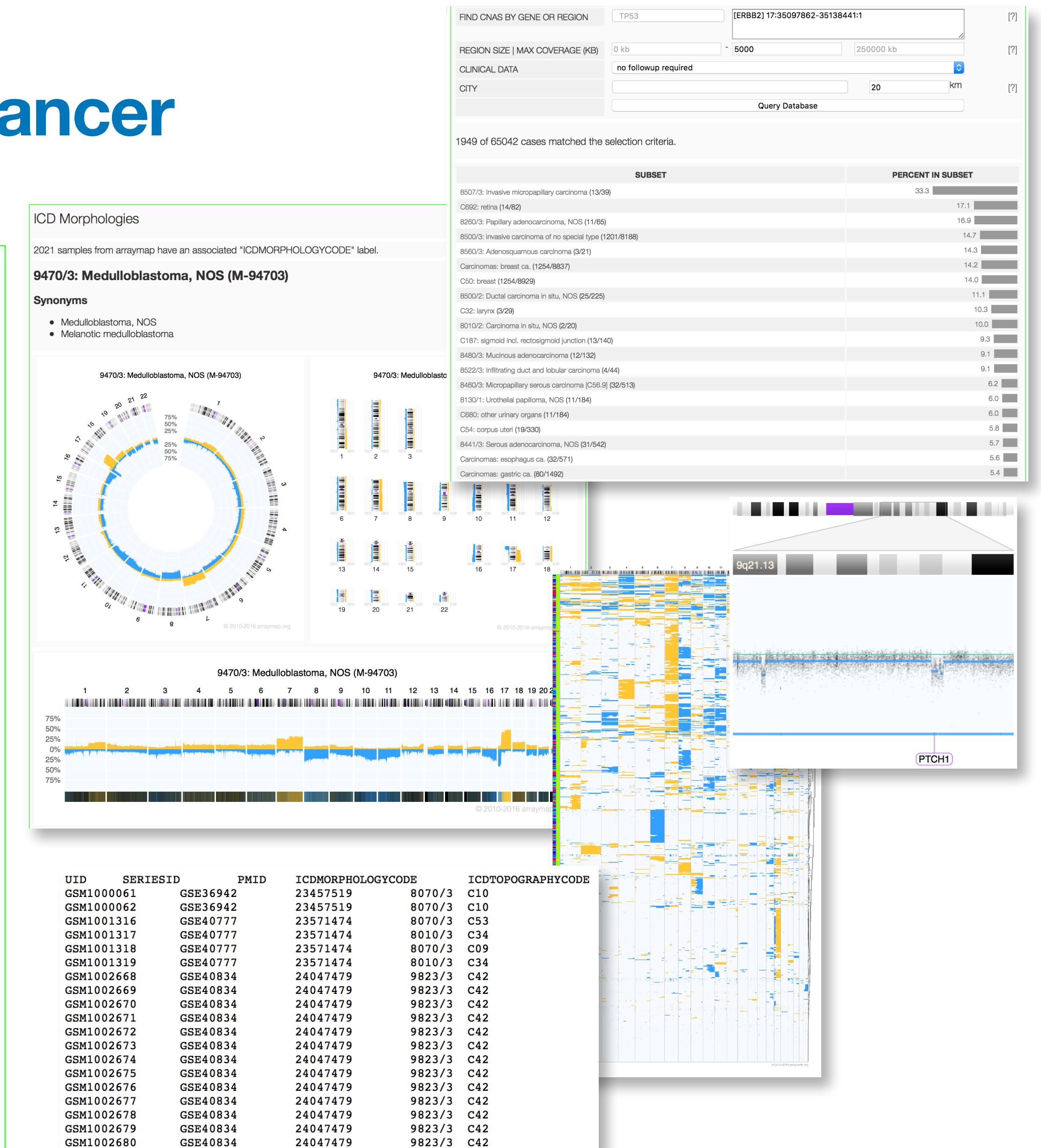
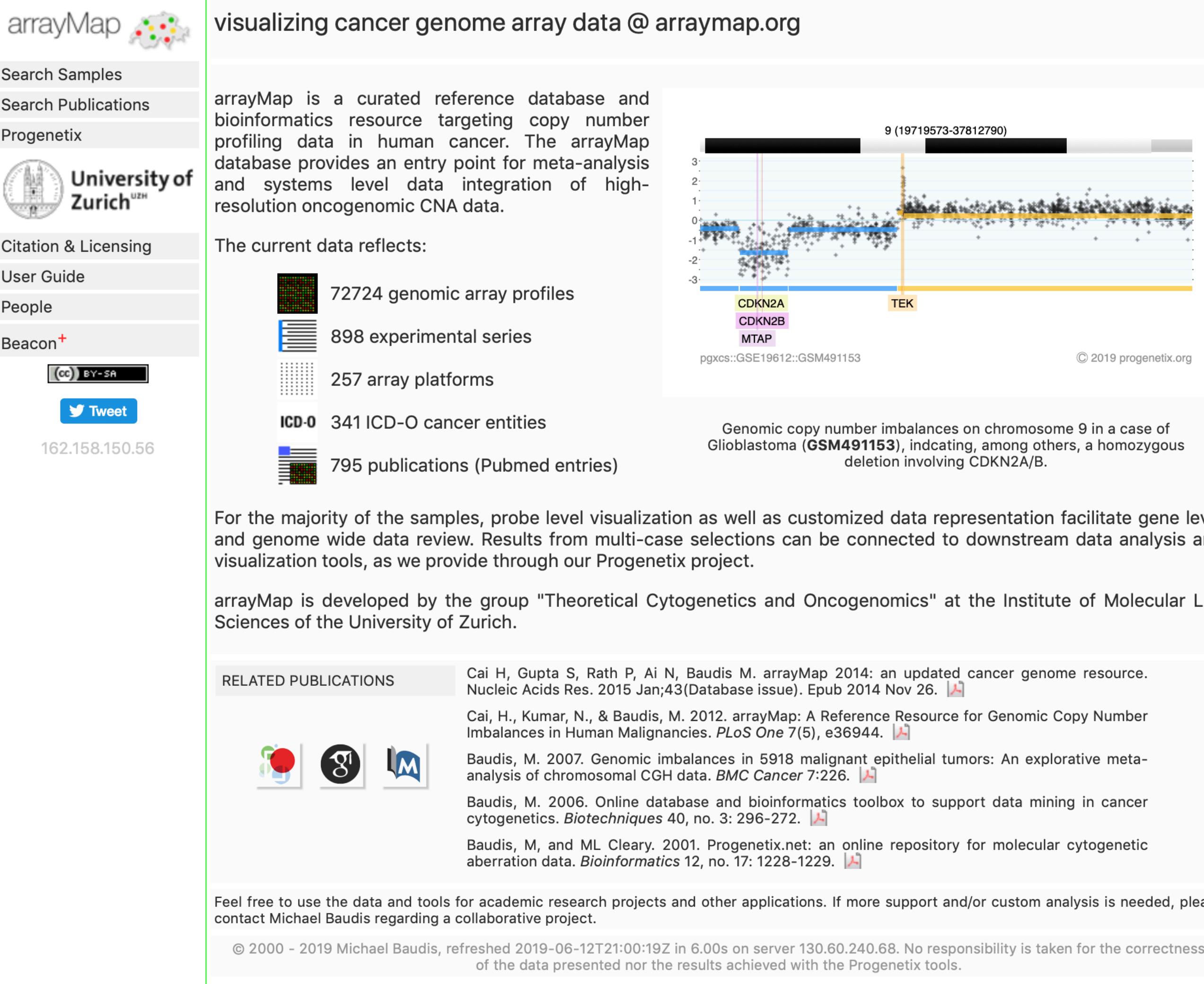
Chromosome Arm 20q Gains and Other Genomic Alterations in Colorectal Cancer Metastatic to Liver, as Analyzed by Comparative Genomic Hybridization and Fluorescence In Situ Hybridization

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

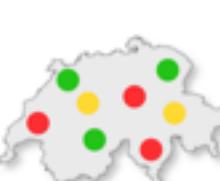
Example input Data: Korn et al., 1999 Genes, chromosomes and cancer

arrayMap (2012 - 2020)

Probe-Level Genomic Array Data in Cancer



Cai et al., 2012 PLoS One



Progenetix in 2021

Cross-platform Oncogenomics

- ▶ source data (i.e. array probe data access) and annotation derived (aCGH, WGS, WES, other arrays)
- ▶ >130'000 cancer and reference CNA profiles
- ▶ systematic metadata annotations following GA4GH standards
- ▶ unrestricted access w/o registration
- ▶ data access API
- ▶ online visualization
- ▶ CNA statistics



[Cancer CNV Profiles](#)

[Search Samples](#)

[Studies & Cohorts](#)

[arrayMap](#)

[TCGA Samples](#)

[DIPG Samples](#)

[Gao & Baudis, 2021](#)

[Cancer Cell Lines](#)

[Publication DB](#)

[Services](#)

[NCIt Mappings](#)

[UBERON Mappings](#)

[Upload & Plot](#)

[Beacon⁺](#)

[Progenetix Info](#)

[About Progenetix](#)

[Use Cases](#)

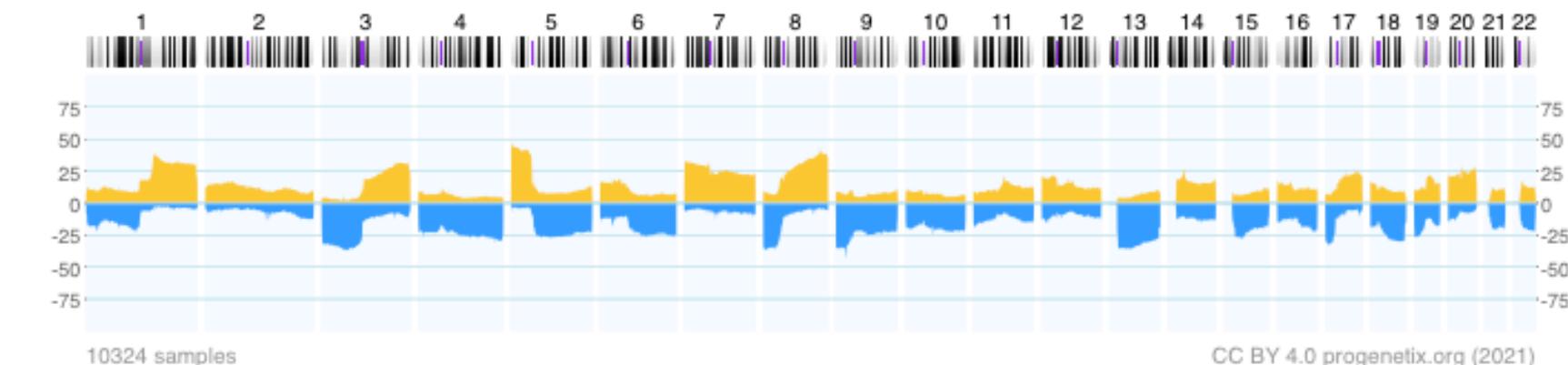
[Documentation](#)

[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* from currently **139448** samples.

Malignant Thoracic Neoplasm (NCIT:C3576)



[Download SVG](#) | [Go to NCIT:C3576](#) | [Download CNV Frequencies](#)

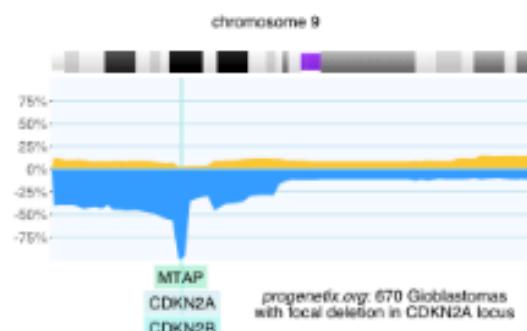
Example for aggregated CNV data in 10324 samples in Malignant Thoracic Neoplasm.

Here the frequency of regional **copy number gains** and **losses** are displayed for all 22 autosomes.

Progenetix Use Cases

[Local CNV Frequencies](#)

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



[Cancer CNV Profiles](#)

The progenetix resource contains data of **788** different cancer types (NCIt neoplasm classification), mapped to a variety of biological and technical categories. Frequency profiles of regional genomic gains and losses for all categories (diagnostic entity, publication, cohort ...) can be accessed through the [[Cancer Types](#)] page with direct visualization and options for sample retrieval and plotting options.

[Cancer Genomics Publications](#)

Through the [[Publications](#)] page Progenetix provides **4025** annotated references to research articles from cancer genome screening experiments (WGS, WES, aCGH, cCGH). The numbers of analyzed samples and possible availability in the Progenetix sample collection are indicated.

Progenetix Implements & Drives the GA4GH Beacon Protocol

CNV Queries, Ontology Based Filters, Data Delivery through Handovers...



Progenetix & Beacon

Demonstrator for GA4GH Beacon based genomic reference resource

- ▶ the CNV content of Progenetix has been a driver to develop the range and bracket variant query options
- ▶ extensive sample annotations using CURIEs with hierarchical ontologies for "biocharacteristics" (NCIT ...) and external references (cellosaurus, geo, PMID ...) serve implementation scenarios for Beacon testing and "production" environment
- ▶ custom implementation
 - ▶ "bycon" code on GH > /progenetix/bycon/
 - ▶ MongoDB backend
 - ▶ React front-end on GH > /progenetix/progenetix-web/



CDKN2A Deletion Example MYC Duplication TP53 Del. in Cell Lines K-562 Cell Line

Gene Spans Cytoband(s)

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 "highly focal" hits (here i.e. <= ~1Mbp in size). The query can be modified e.g. through changing the position parameters or diagnosis.

Gene Symbol i
Select...

Chromosome i
9

(Structural) Variant Type i
DEL (Deletion)

Start or Position i
21500001-21975098

End (Range or Structural Var.) i
21967753-22500000

Minimum Variant Length i

Maximal Variant Length i

Cancer Classification(s) i
NCIT:C3058: Glioblastoma (4375) x

Filter Precision i
exact

City i
Select...

Chromosome 9 i
21500001 21975098
21967753 22500000

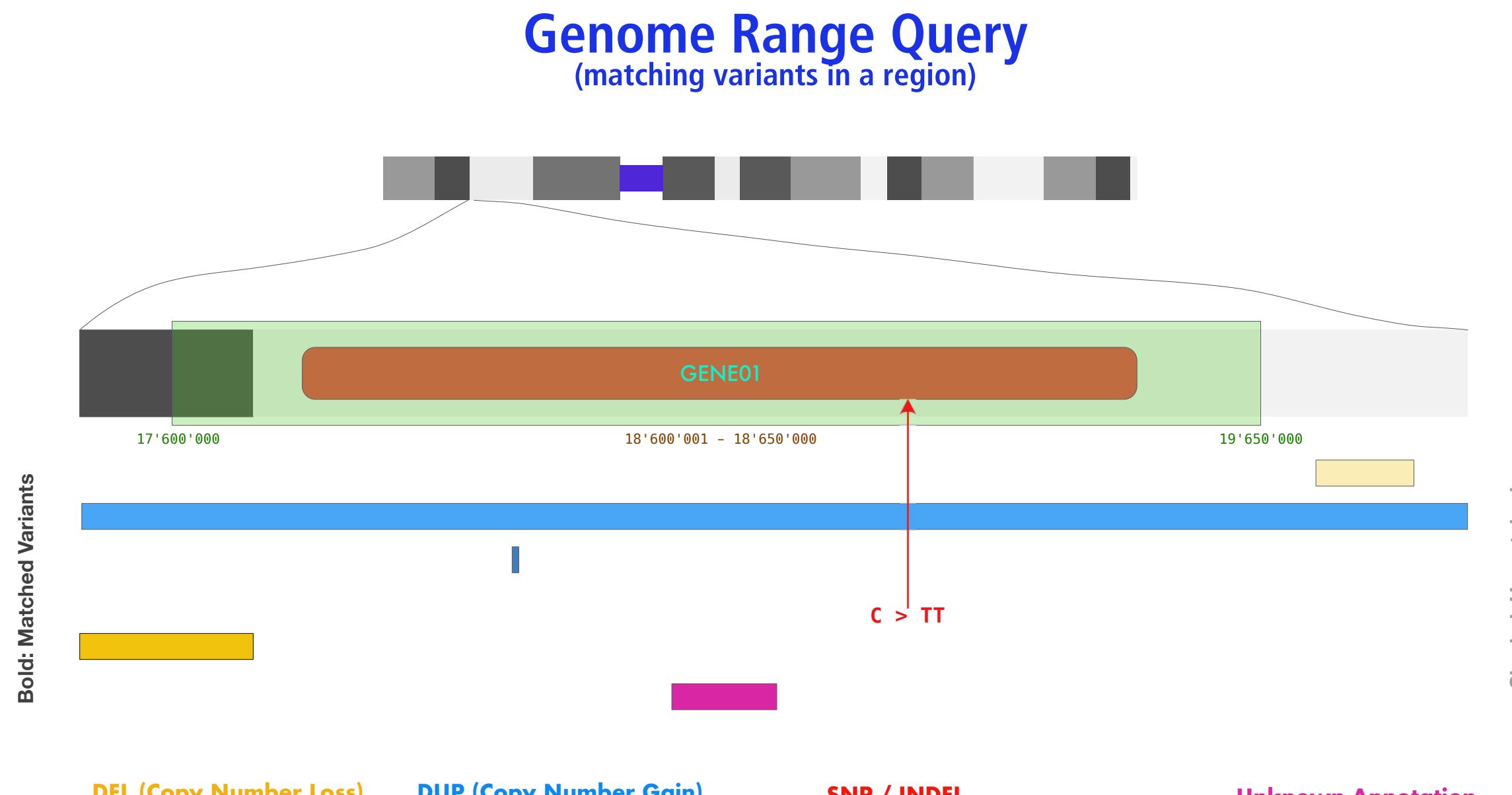
Query Database



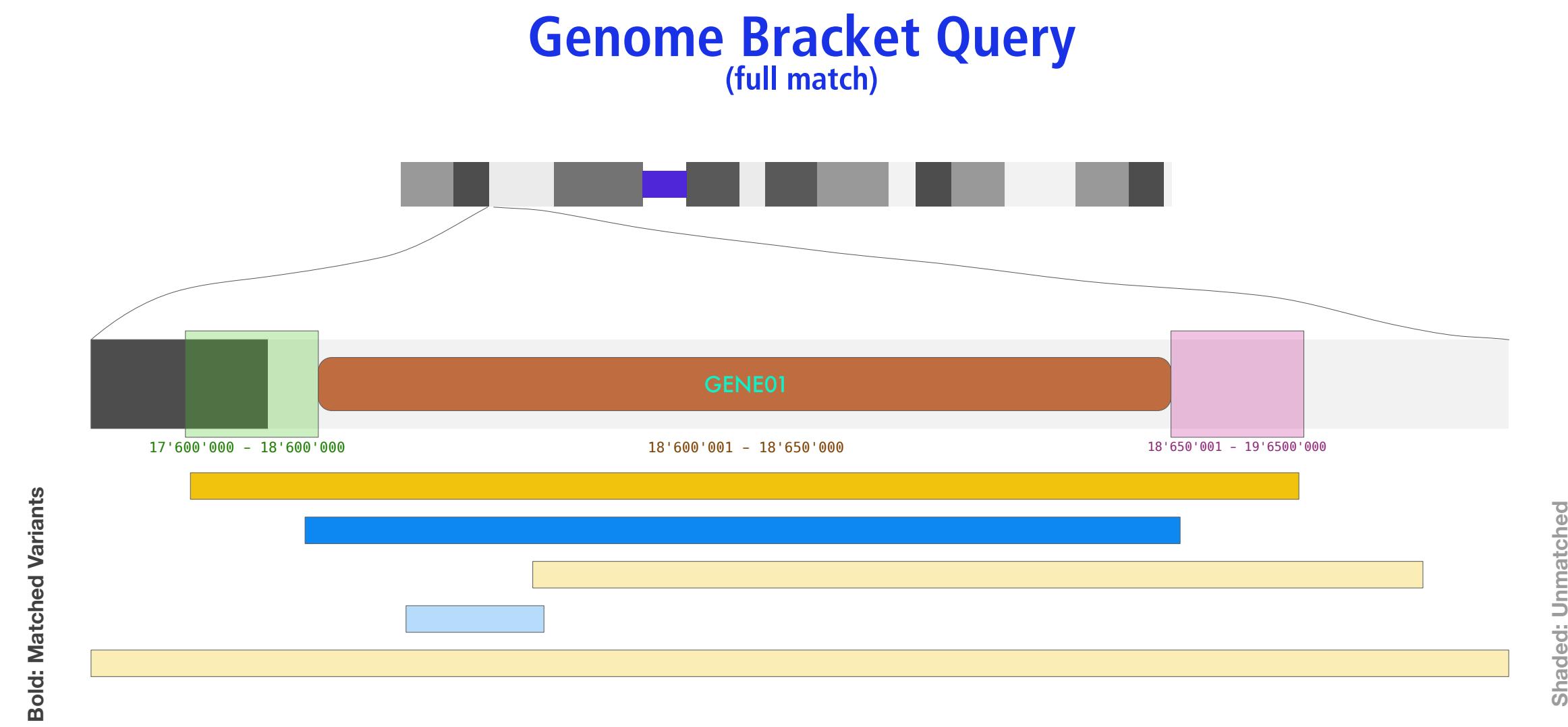
Beacon v2: Extended Variant Queries



Range and Bracket queries enable positional wildcards and fuzziness



- Genome Range Queries provide a way to "fish" for variants overlapping an indicated region, e.g. the CDR of a gene of interest
- Additional parameters (e.g. variant type, reference or alternate bases) limit the scope of the responses
- new Beacon v2 size parameters to limit structural variants (e.g. "focal" CNVs)

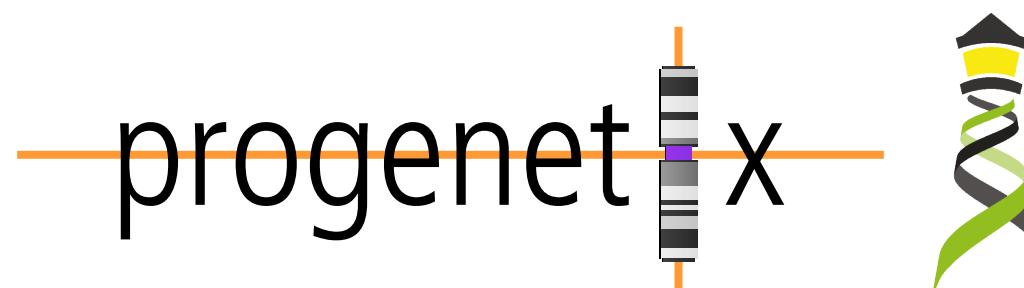


- Genome Bracket Queries allow to search for structural variants with start and end positions falling into defined sequence ranges
- allows to query any contiguous genomic variant (and in principle also can step in for range queries)
- typical use case is e.g the query for variants such as duplications covering the whole CDR of a gene, while limiting the allowed start or end regions

Beacon v2 Filters

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

- ▶ Beacon v2 "filters" assumes inclusion of child terms when using hierarchical classifications
 - ▶ implicit *OR* with otherwise assumed *AND*
 - ▶ implementation of hierarchical annotations overcomes some limitations of "fuzzy" disease annotations
 - ▶ data *handover* (Beacon v1.1+) enables further data exploration and export scenarios



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



progenetix

Variants: 0 falleles: 0 Callsets Variants ↗ UCSC region ↗ Calls: 0 Legacy Interface ↗ Samples: 523 [Show JSON Response](#)

Results Biosamples

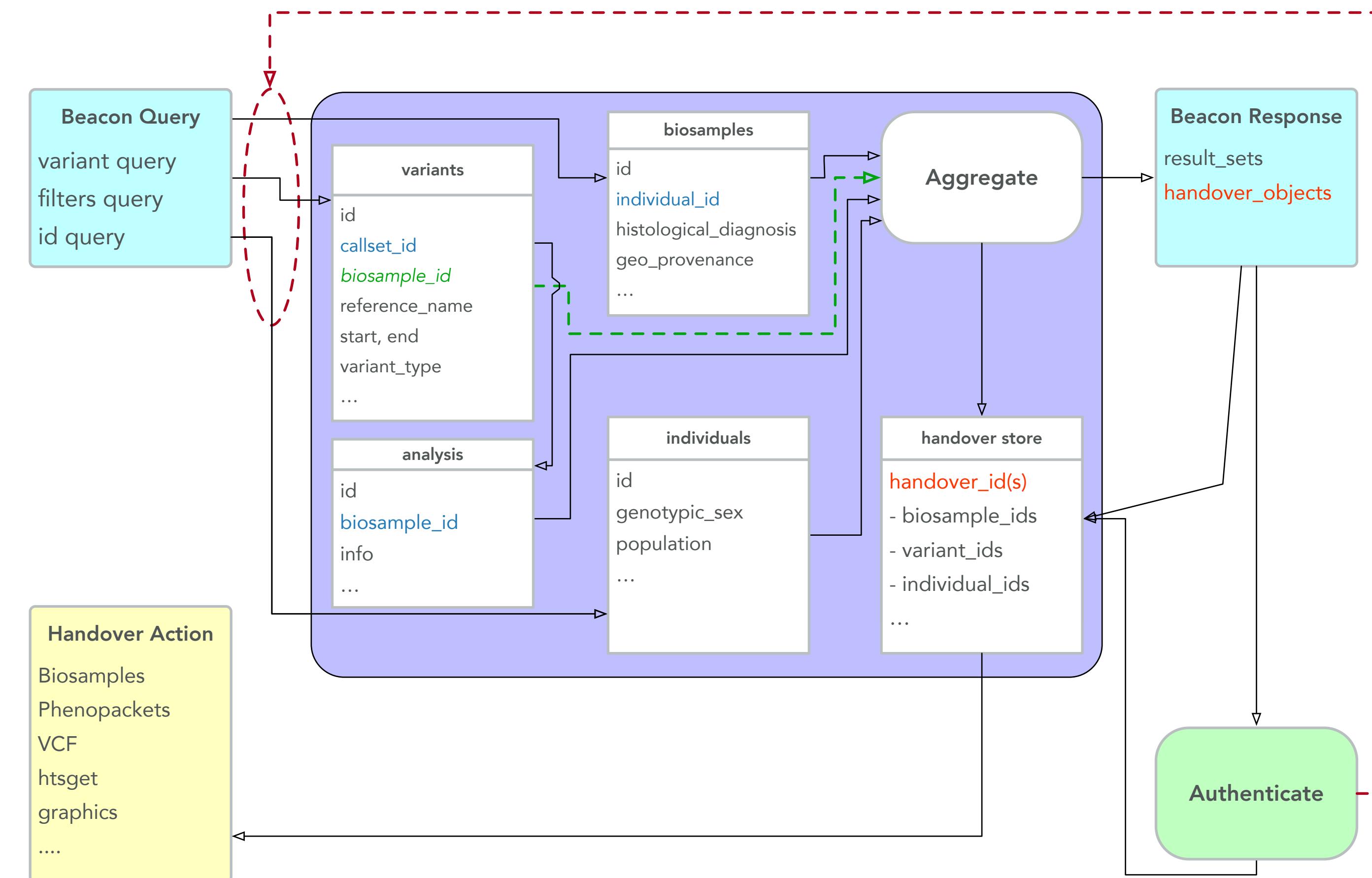
Id	Description	Classifications	Identifiers	DEL	DUP	CNV
PGX_AM_BS_MCC01	Merkel cell carcinoma	icdot-C44.9 Skin, NOS icdom-82473 Merkel cell carcinoma NCIT:C9231 Merkel Cell Carcinoma	PMID:9537255	0.116	0.104	0.22
PGX_AM_BS_MCC02	Merkel cell carcinoma	icdot-C44.9 Skin, NOS icdom-82473 Merkel cell carcinoma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

Progenetix & Beacon v1->2

Handover elements in Beacon responses

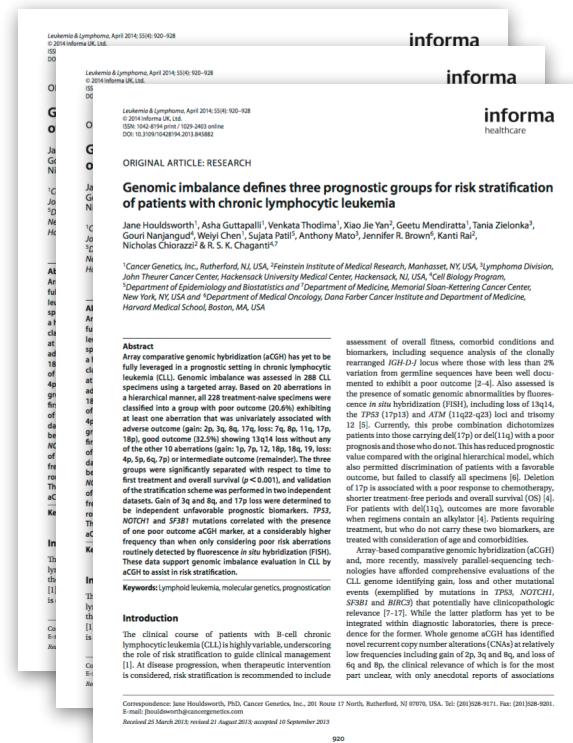
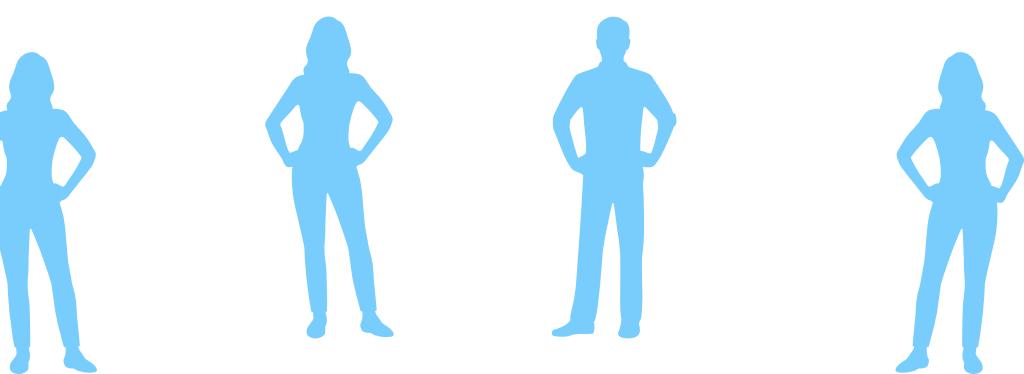
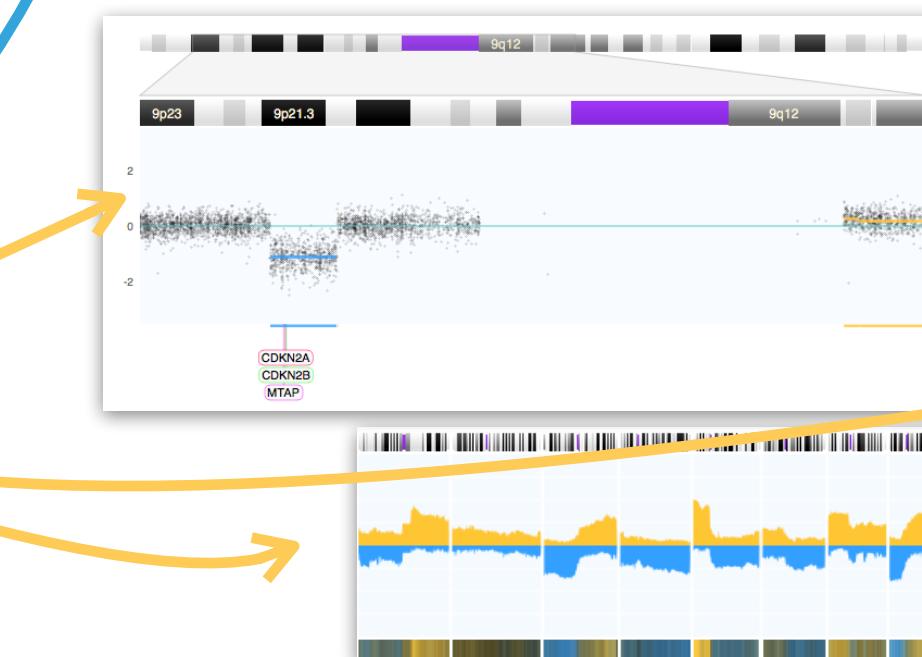
- ▶ Progenetix utilizes handovers to deliver data matched by the Beacon queries
- ▶ These handovers are interpreted by the front end to populate different parts of the UI, w/o the need of active selection
- ▶ Handovers are either standard Beacon v2 paths or dedicated custom functions

Handover Concept



DATA PIPELINE

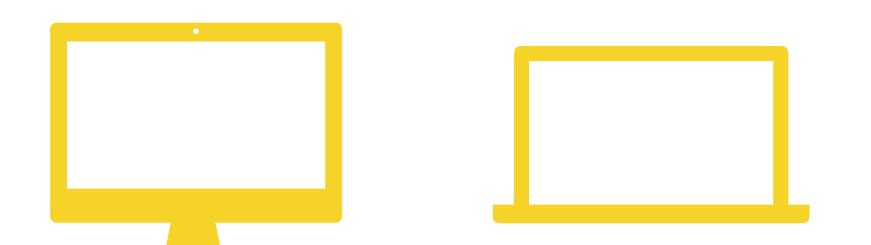
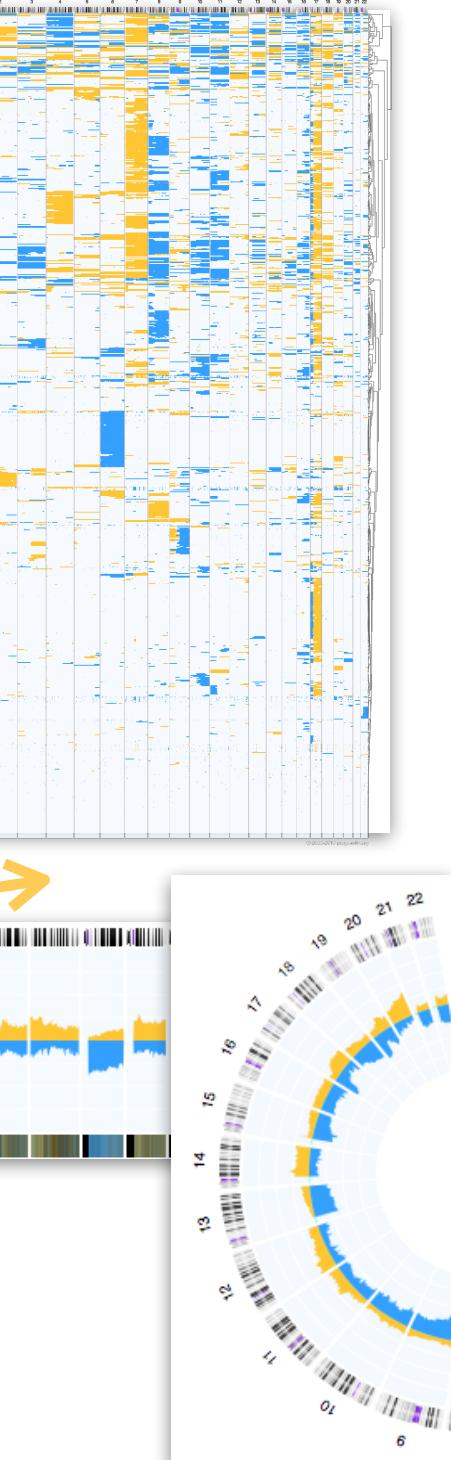
BIOCURATION
BIOINFORMATICS

arrayMap

I^{CD} Morphologies

Subsets	arrayMap	progenetix
00000: not classified in I ^{CD} (e.g. non-neoplastic or benign)	8814	370
00000: neoplasm, malignant	11	1
00100: epithelial tumor, benign	10	
00102: carcinoma in situ, nos	20	11
00102: carcinoma, in situ, nos	1430	258
00120: large cell carcinoma	46	54
00120: large cell carcinoma, nos	80	80
00120: large cell neuroendocrine carcinoma	3	2
00120: carcinoma, neoplastic type, nos	4	41
00220: mesothelioma	1	1
00310: giant cell carcinoma	4	3
00330: sarcomatoid carcinoma	1	1
00410: small cell carcinoma, nos	132	148
00460: non-small cell carcinoma	1195	164
00500: papillary carcinoma	16	
00701: preinvasive squamous epithelium, nos	46	162
00702: squamous cell carcinoma, in situ, nos	65	16
00703: squamous cell carcinoma, nos	2443	2097
00710: squamous cell carcinoma, acantholytic	11	12
00710: squamous cell carcinoma, acantholytic	11	22
00772: squamous metaplasia, neoplastic, grade II	136	22
00802: unifferentiated respiratory neoplasm, grade II	52	200
01200: transitional cell carcinoma, in situ, nos	28	6
01200: transitional cell carcinoma, nos	10	
01200: transitional cell carcinoma, nos	310	423
01300: papillary transitional cell carcinoma, non-invasive	184	39
01300: papillary transitional cell carcinoma	56	
01400: adenoma, nos	2	6
01401: atypical adenoma	385	361
01402: adenocarcinoma in situ	11	
01440: adenocarcinoma, intestinal type	9409	3240
01440: adenocarcinoma, intestinal type	167	206
01450: carcinoma, diffuse type	7	36
01501: islet cell adenoma	15	
01502: islet cell carcinoma	8	
01510: insulinoma, nos	1	18
01502: islet cell carcinoma	28	
01510: insulinoma, nos	1	18
01510: insulinoma, nos	1	18



arrayMap
progenetix

Metadata Curation

Standardized, machine-readable

- ▶ Developed by international initiatives (ELIXIR, GA4GH, MONARCH...) and resource providers (EBI, NCBI)
- ▶ Hierarchical NCI Neoplasm Core replaces heterogeneous primary annotations
- ▶ Representing varying granularity

"Colorectal Cancer"
or
"Rectal Mucinous Adenoca."

	Subsets	Samples
<input type="checkbox"/>	▼ NCIT:C3262: Neoplasm	88844
<input type="checkbox"/>	▼ NCIT:C3263: Neoplasm by Site	84747
<input type="checkbox"/>	▼ NCIT:C156482: Genitourinary System Neoplasm	11616
<input type="checkbox"/>	▼ NCIT:C156483: Benign Genitourinary System Neoplasm	219
<input type="checkbox"/>	▼ NCIT:C4893: Benign Urinary System Neoplasm	90
<input type="checkbox"/>	▼ NCIT:C4778: Benign Kidney Neoplasm	90
<input type="checkbox"/>	NCIT:C159209: Kidney Leiomyoma	1
<input type="checkbox"/>	NCIT:C4526: Kidney Oncocytoma	82
<input type="checkbox"/>	NCIT:C8383: Kidney Adenoma	7
<input type="checkbox"/>	▼ NCIT:C7617: Benign Reproductive System Neoplasm	129
<input type="checkbox"/>	▼ NCIT:C4934: Benign Female Reproductive System Neoplasm	129
<input type="checkbox"/>	▼ NCIT:C2895: Benign Ovarian Neoplasm	58
<input type="checkbox"/>	▼ NCIT:C4510: Benign Ovarian Epithelial Tumor	58
<input type="checkbox"/>	▼ NCIT:C40039: Benign Ovarian Mucinous Tumor	58
<input type="checkbox"/>	NCIT:C4512: Ovarian Mucinous Cystadenoma	58
<input type="checkbox"/>	▼ NCIT:C4060: Ovarian Cystadenoma	58
<input type="checkbox"/>	NCIT:C4512: Ovarian Mucinous Cystadenoma	58
<input type="checkbox"/>	▼ NCIT:C3609: Benign Uterine Neoplasm	71
<input type="checkbox"/>	▼ NCIT:C3608: Benign Uterine Corpus Neoplasm	71
<input type="checkbox"/>	NCIT:C3434: Uterine Corpus Leiomyoma	71
<input type="checkbox"/>	▼ NCIT:C156484: Malignant Genitourinary System Neoplasm	11171
<input type="checkbox"/>	▼ NCIT:C157774: Metastatic Malignant Genitourinary System Neoplasm	2
<input type="checkbox"/>	▼ NCIT:C146893: Metastatic Genitourinary System Carcinoma	2
<input type="checkbox"/>	NCIT:C8946: Metastatic Prostate Carcinoma	2
<input type="checkbox"/>	▼ NCIT:C164141: Genitourinary System Carcinoma	10561



Beacon & Phenopackets

Data *discovery* and *delivery* using standardized GA4GH formats and schemas

- ▶ modern standards and protocols are essential for federation and exchange of biomedical data
- ▶ emerging / established principles are the use of hierarchical coding systems and with widespread use of CURIEs
- ▶ Non-categorical annotations e.g. stage, grade, followup time, survival status, genomic sex, age at diagnosis based on international standards, e.g.
 - ▶ ISO (ISO 8601 time & period, ISO 3166 country codes ...)
 - ▶ IETF (GeoJSON ...)
 - ▶ W3C (CURIE ...)
- Beacon query **filters** correspond well to Phenopackets data
- Phenopackets as supported protocol for Beacon data delivery

```
"data_use_conditions" : {  
    "label" : "no restriction",  
    "id" : "DUO:0000004"  
},  
"material" : {  
    "id" : "EFO:0009656",  
    "label" : "neoplastic sample"  
},  
{  
    "ageAtDiagnosis": "P25Y3M2D"  
},  
"sampled_tissue" : {  
    "id" : "UBERON:0002037",  
    "label" : "cerebellum"  
},  
"histological_diagnosis" : {  
    "id" : "NCIT:C3222",  
    "label" : "Medulloblastoma"  
},
```

Progenetix is a Cancer CNV Reference Resource

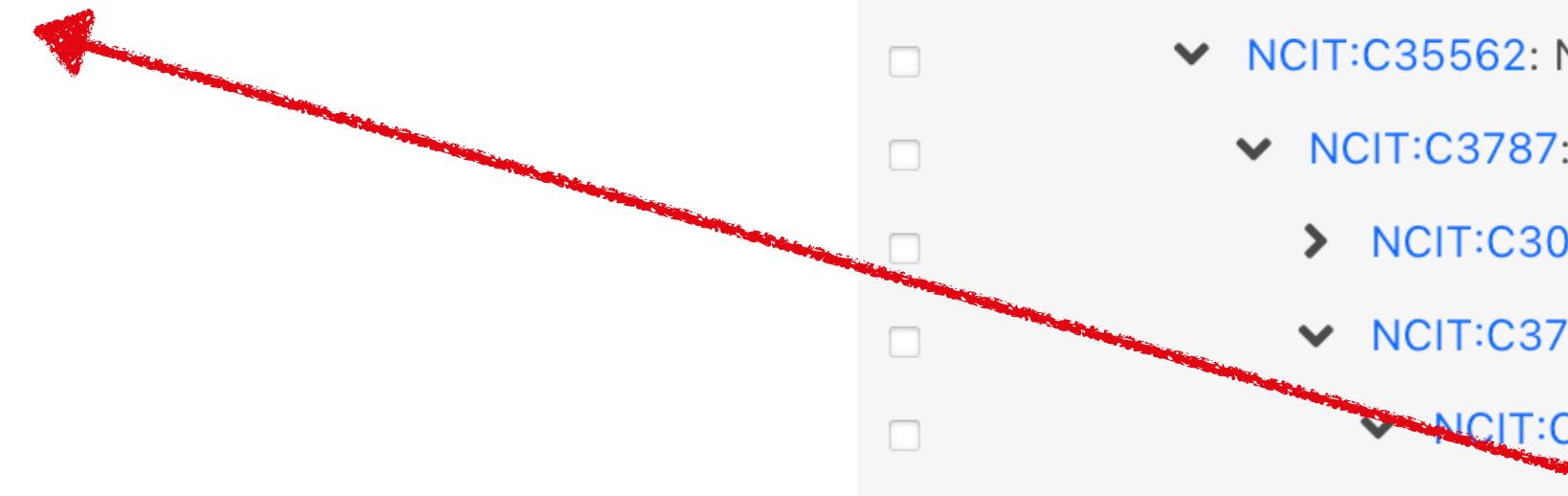
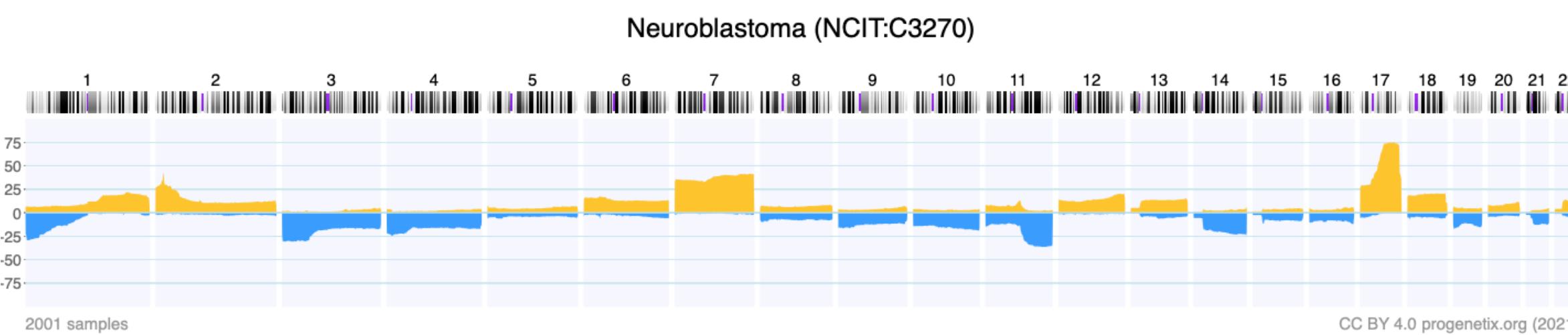
Cancer CNV Profiles, Gene CNV Frequencies, Cohorts & Publications



Progenetix in 2021

Cancer type representation

- ▶ Hierarchical aggregation
- ▶ CNV landscape overview
- ▶ Sample retrieval



Cancer Types

The cancer samples in Progenetix are mapped to several classification systems. For each of the classes, aggregated data is available by clicking the code. Additionally, a selection of the corresponding samples can be initiated by clicking the sample number or selecting one or more classes through the checkboxes.

Sample selection follows a hierarchical system in which samples matching the child terms of a selected class are included in the response.

Cancer Classification: NCIT Cancer Core

Filter subsets ...

Hierarchy Depth: collapsed

No Selection

- > NCIT:C3010: Endocrine Neoplasm (3319 samples)
- > NCIT:C3030: Eye Neoplasm (280 samples)
- > NCIT:C3052: Digestive System Neoplasm (15194 samples)
- > NCIT:C3077: Head and Neck Neoplasm (3769 samples)
- > NCIT:C3268: Nervous System Neoplasm (16270 samples)
 - > NCIT:C2963: Cranial Nerve Neoplasm (19 samples)
 - > NCIT:C3321: Peripheral Nervous System Neoplasm (901 samples)
- > NCIT:C35562: Neuroepithelial, Perineurial, and Schwann Cell N... (11690 samples)
 - > NCIT:C3787: Neuroepithelial Neoplasm (11214 samples)
 - > NCIT:C3059: Glioma (8688 samples)
 - > NCIT:C3716: Primitive Neuroectodermal Tumor (2213 samples)
 - > NCIT:C6963: Neu Click to retrieve samples for NCIT:C3270
- > NCIT:C3270: Neuroblastoma (2001 samples)



Cancer CNV Profiles

[Search Samples](#)

Studies & Cohorts

arrayMap

TCGA Samples

DIPG Samples

Gao & Baudis, 2021

Cancer Cell Lines

Publication DB

Services

NCIt Mappings

UBERON Mappings

Upload & Plot

Download Data

Beacon⁺

Progenetix Info

About Progenetix

Use Cases

Documentation

Baudisgroup @ UZH

Search
Samples

Modify Query

Glioblastoma

Assembly: GRCh38 Gene **CDKN2A** Filters **NCIT:C3058**

progenetix

Samples: 2426

Variants: 1733

Calls: 2566

Found Variants (.pgxseg)

All Sample Variants (.json)

All Sample Variants (.pgxseg)

Show Variants in UCSC

UCSC region

JSON Response

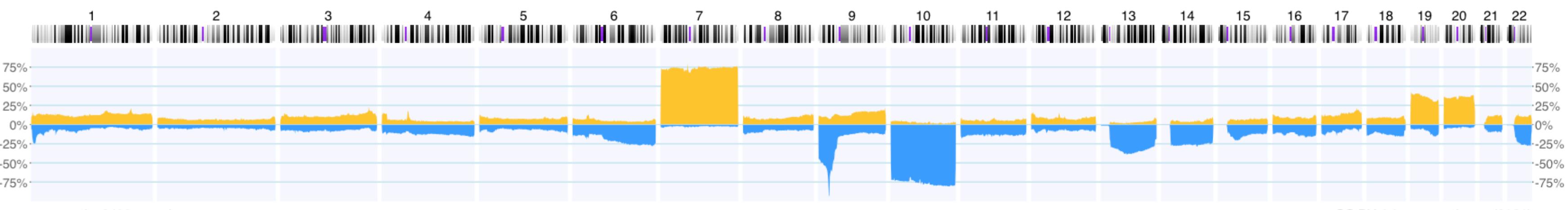
[Visualization options](#)

Results

Biosamples

Biosamples Map

Variants



progenetix: 2433 samples

CC BY 4.0 progenetix.org (2021)

Matched Subset Codes

Subset Samples

Matched Samples

Subset Match Frequencies

[icdot-C71.8](#)

1

1

1.000

[icdot-C71.4](#)

4

3

0.750

[UBERON:0002021](#)

4

3

0.750

[icdom-94403](#)

4291

2396

0.558

[NCIT:C3058](#)

4375

2396

0.548

[NCIT:C3796](#)

84

30

0.357



Search
Samples

Modify Query

Cancer CNV Profiles

[Search Samples](#)

Studies & Cohorts

arrayMap

TCGA Samples

DIPG Samples

Gao & Baudis, 2021

Cancer Cell Lines

Publication DB

Results Biosamples Biosamples Map

Variants

Services

NCIt Mappings

UBERON Mappings

Upload & Plot

Download Data

Beacon⁺

Progenetix Info

About Progenetix

Use Cases

Documentation

Baudisgroup @ UZH

progenetix

Samples: 2426

Variants: 1733

Calls: 2566

Found Variants (.pgxseg)

All Sample Variants (.json)

All Sample Variants (.pgxseg)

Show Variants in UCSC

UCSC region

JSON Response

[Visualization options](#)

JSON Download Variants

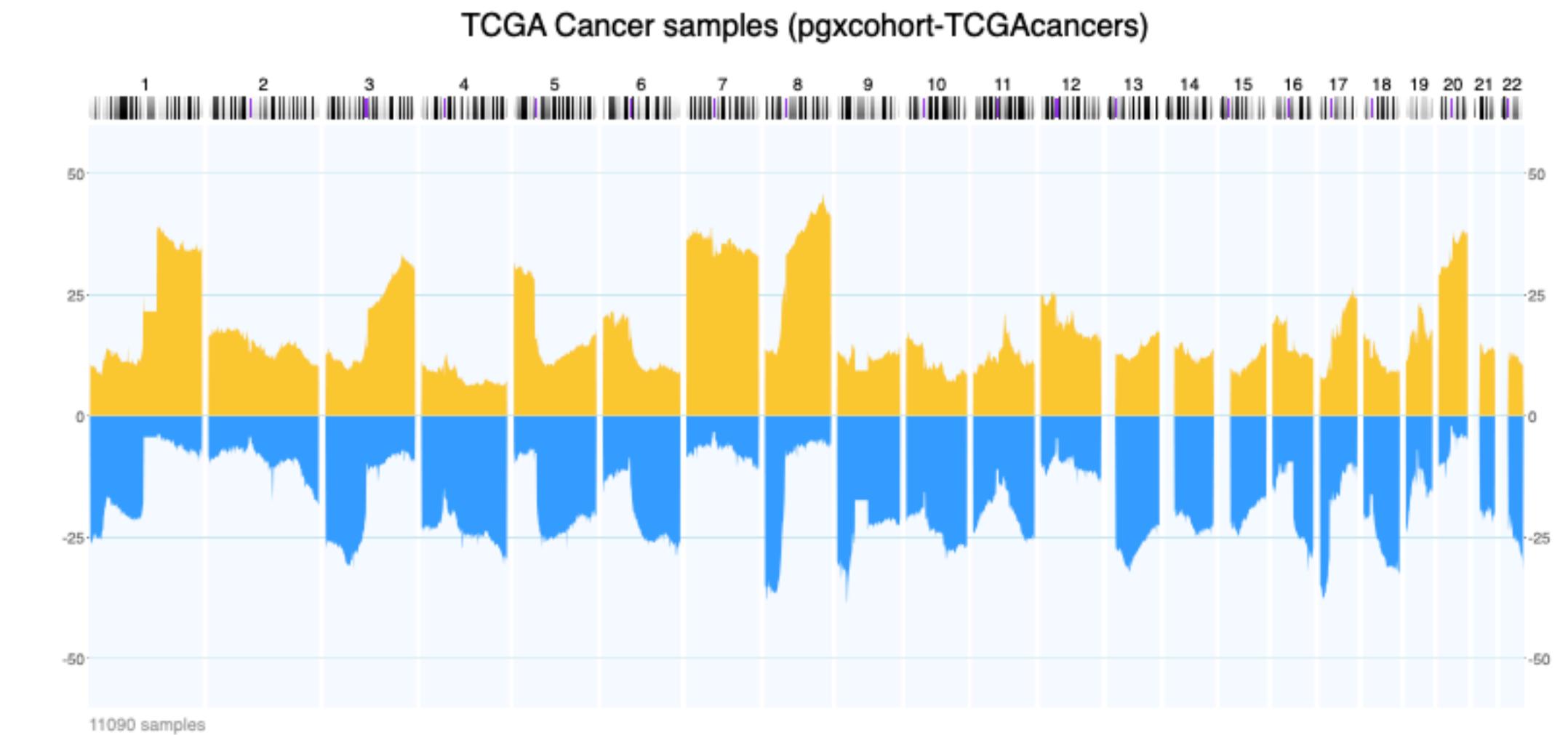
Int. ID	Digest	Biosample	Chr.	Ref. Base(s)	Alt. Base(s)	Type
pgxvar-5bab576a727983b2e00b9478	9:0-51799999:DEL	pgxbs-kftva5zv	9			DEL
pgxvar-5bab576a727983b2e00b9493	9:0-51799999:DEL	pgxbs-kftva601	9			DEL
pgxvar-5bab576a727983b2e00b94a4	9:0-51799999:DEL	pgxbs-kftva603	9			DEL
pgxvar-5bab576a727983b2e00b94ae	9:0-51799999:DEL	pgxbs-kftva606	9			DEL
pgxvar-5bab576a727983b2e00b94b3	9:0-140273252:DEL	pgxbs-kftva608	9			DEL
pgxvar-5bab576a727983b2e00b94be	9:0-51799999:DEL	pgxbs-kftva60b	9			DEL
pgxvar-5bab576d727983b2e00bae65	9:0-40199999:DEL	pgxbs-kftvgi7e	9			DEL
pgxvar-5bab576d727983b2e00bae7f	9:0-32799999:DEL	pgxbs-kftvgi7j	9			DEL
pgxvar-5bab576d727983b2e00bae97	9:0-51799999:DEL	pgxbs-kftvgi7t	9			DEL
pgxvar-5bab576d727983b2e00bae9b	9:0-51799999:DEL	pgxbs-kftvgi7u	9			DEL

Progenetix API

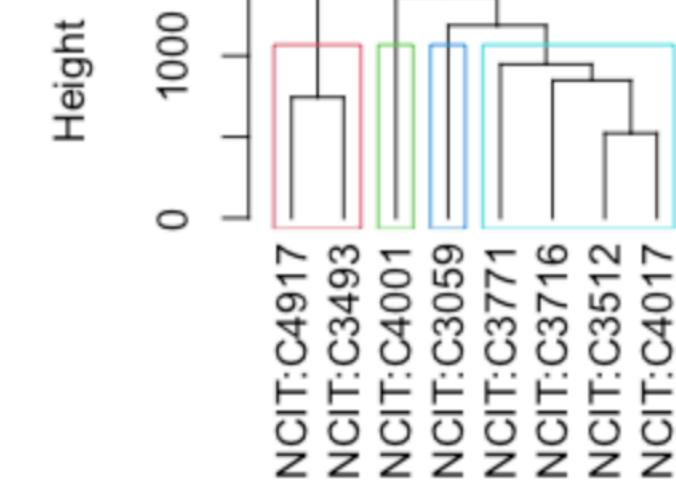
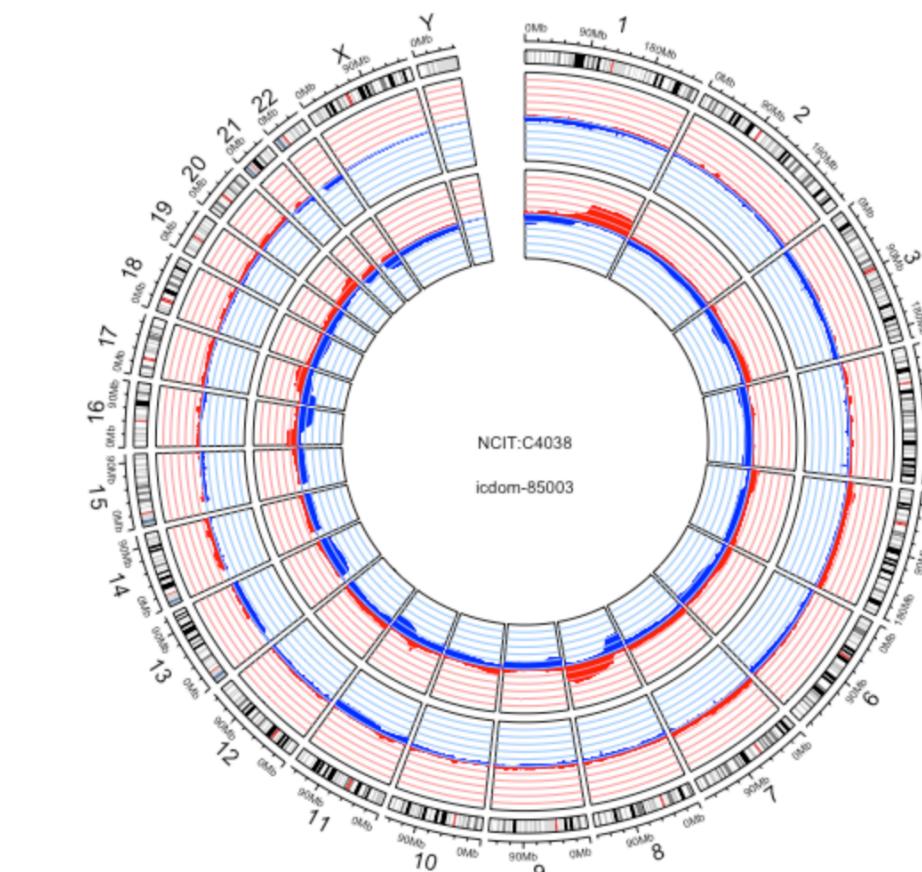
Data & Plots

- ▶ Through multiple API endpoints
- ▶ segmented CNV data in .pgxseg (tabular) and JSON format
- ▶ CNV frequency histograms for disease, study or cohort
- ▶ R package for circos, heatmap and more...

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



[https://progenetix.org/cgi/PGX/cgi/collationPlots.cgi?
datasetIds=progenetix&id=pgxcohort-TCGAcancers&-
size_plotimage_w_px=800&-size_plotarea_h_px=300&-value_plot_y_max=60](https://progenetix.org/cgi/PGX/cgi/collationPlots.cgi?datasetIds=progenetix&id=pgxcohort-TCGAcancers&-size_plotimage_w_px=800&-size_plotarea_h_px=300&-value_plot_y_max=60)



Progenetix

Services, Documentation...

- ▶ services e.g. for disease code translation (NCIt
<=> ICD-O; UBERON ...)
- ▶ documentation of data schema, API ...

progenetix

Services: Ontologymaps (NCIt)

The **ontologymaps** service provides equivalence mapping between ICD-O and other classification systems, notably NCIt and UBERON. It makes use of the sample-level mappings for NCIT and ICD-O 3 codes developed for the individual samples in the Progenetix collection.

NCIT and ICD-O 3

While NCIT treats diseases as **histologic** and **topographic** described entities (e.g. **NCIT:C7700: Ovarian adenocarcinoma**), these two components are represented separately in ICD-O, through the **Morphology** and **Topography** coding arms (e.g. here **8140/3 + C56.9**).

More documentation with focus on the API functionality can be found on the [documentation pages](#).

The data of all mappings can be retrieved through this API call: [{JSON}](#)

Code Selection

gli|

NCIT:C3058: Glioblastoma
NCIT:C3288: Oligodendrogloma
NCIT:C4326: Anaplastic Oligodendrogloma
NCIT:C3903: Mixed Glioma
NCIT:C3059: Glioma
NCIT:C3796: Gliosarcoma
NCIT:C4822: Malignant Glioma
NCIT:C3308: Paraganglioma
NCIT:C4831A: Renal Paraganglioma

About Progenetix

News
Cancer CNV Profiles
Search Samples
arrayMap
Studies & Cohorts
Publication DB
Services
Upload & Plot
Documentation
Beacon+
Baudisgroup @ UZH

NCI thesaurus

Progenetix :: Info

Structural Cancer Genomics Resource
Documentation and Example Pages

[News](#)
[About...](#)
[API](#)
[Documentation](#)
[Publications](#)
[Progenetix Home](#)

Related Sites

[Progentix Data](#)
[Baudisgroup @ UZH](#)
[Beacon+](#)
[SchemaBlocks {S}\[B\]](#)
[Beacon Project](#)

Github Projects

[baudisgroup](#)
[progenetix](#)
[ELIXIR Beacon](#)

Tags

[API](#) [Beacon](#) [BeaconPlus](#)
[BeaconV2](#) [GA4GH](#) [Perl](#) [Python](#)
[TCGA](#) [article](#) [bycon](#) [code](#)
[documentation](#) [identifiers](#) [licensing](#)
[ontologies](#) [prefixes](#) [schemas](#)
[services](#) [statistics](#) [tools](#) [website](#)

Welcome to the Progenetix documentation pages

The **Progenetix Resource Documentation** provides information and links related to the **Progenetix** cancer genome resource and the related **Progenetix code repositories** contains projects, such as data conversion scripts, ontology mappings and code for the **Beacon+** project.

Progenetix Website Code Repositories

- [Progenetix Source Code](#)
- [Related Projects](#)

Latest News

Progenetix File Formats

Standard Progenetix Segment Files [.pgxseg](#)

Progenetix uses a variation of a standard tab-separated columnar text file such as produced by array or sequencing CNV software, with an optional metadata header for e.g. plot or grouping instructions.

@mbaudis 2021-02-22: [more ...](#)

Gene Symbol	MYC
MYC (8:127736593-127740957)	
MYCBP (1:38864669-38873304)	
MYCBPAP (17:56508545-50531427)	
MYCL (1:39897371-39901887)	
MYCN (2:15940586-15946096)	

Beacon+ and Progenetix Queries by Gene Symbol

We have introduced a simple option to search directly by Gene Symbol, which will match to *any* genomic variant with partial overlap to the specified gene. This works by expanding the Gene Symbol (e.g. **TP53**, **CDKN2A** ...) into a range query for its genomic coordinates (maximum CDR).

Such queries - which would e.g. return all whole-chromosome CNV events covering the gene of interest, too - should be narrowed by providing e.g. **Variant Type** and **Maximum Size** (e.g. 2000000) values.

@mbaudis 2021-02-22: [more ...](#)

The Progenetix oncogenomic resource in 2021

Qingyao Huang, Paula Carrio Cordero, Bo Gao, Rahel Paloots, Michael Baudis

bioRxiv. doi: <https://doi.org/10.1101/2021.02.15.428237>

This article provides an overview of recent changes and additions to the Progenetix database and the services provided through the resource.

2021-02-15: [more ...](#)



Diffuse Intrinsic Pontine Glioma (DIPG) cohort

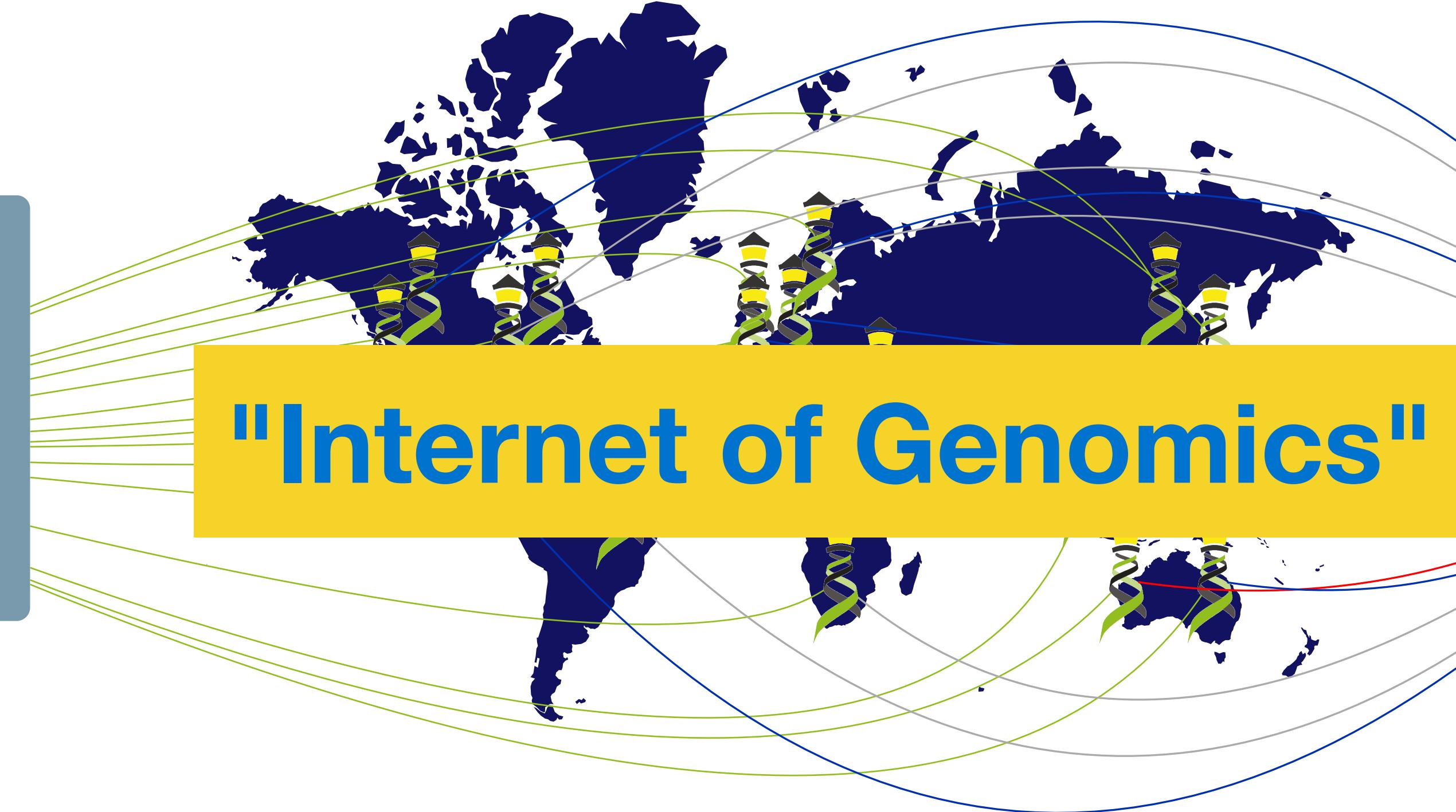
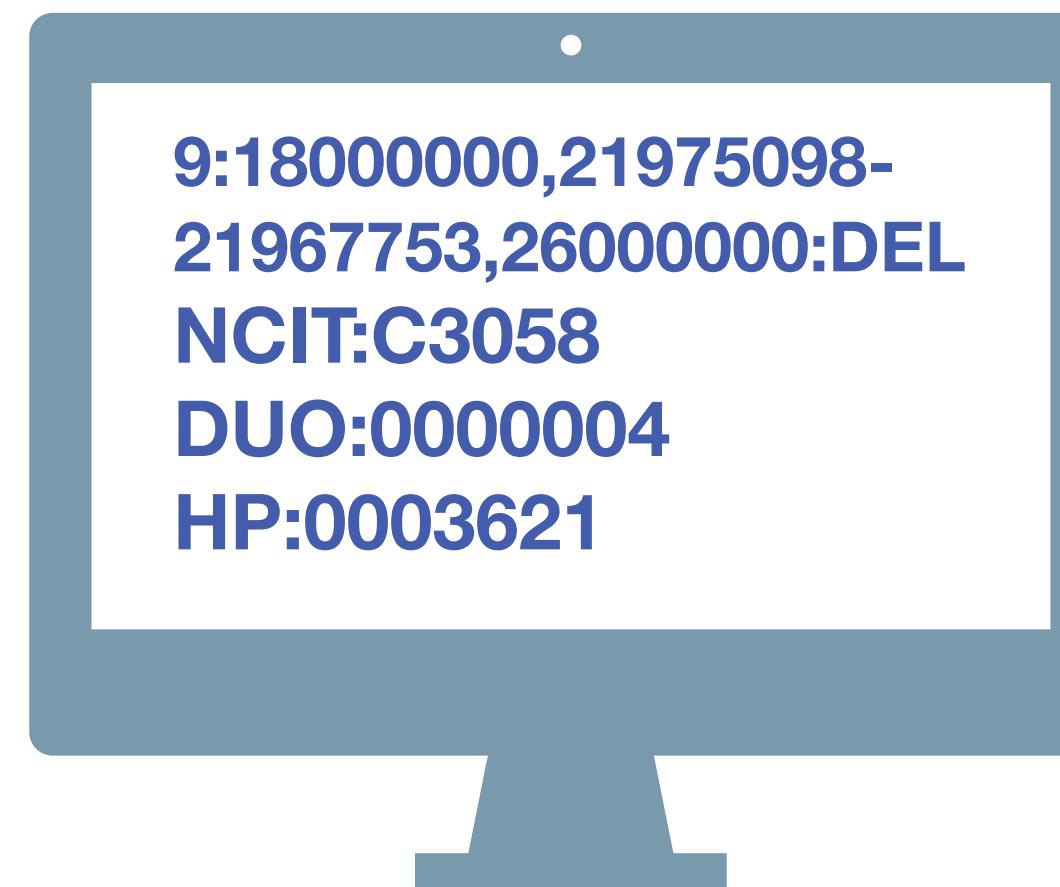
Diffuse Intrinsic Pontine Glioma (DIPG) is a highly aggressive tumor type that originates from glial cells in the pons area of the brainstem, which controls vital functions including breathing, blood pressure and heart rate. DIPG occurs frequently in the early childhood and has a 5-year survival rate below 1 percent. Progenetix has now incorporated the DIPG cohort, consisting of 1067 individuals from 18 publications. The measured data include copy number variation as well as (in part) point mutations on relevant genes, e.g. TP53, NF1, ATRX, TERT promoter.

@qingyao 2021-02-15: [more ...](#)

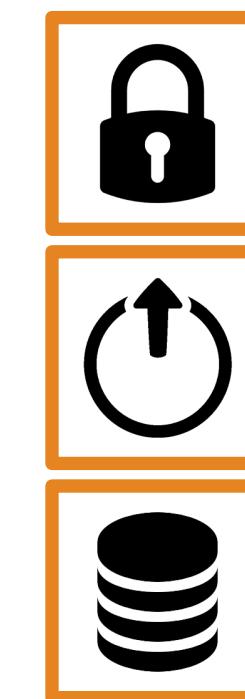
arrayMap is Back

After some months of dormancy, the **arrayMap** resource has been relaunched through integration with the new **Progenetix** site. All of the original arrayMap data has now been integrated into Progenetix, and of today the [arraymap.org](#) domain maps to a standard Progenetix search page, where only data samples with existing source data (e.g. probe specific array files) will be presented.

@mbaudis 2021-02-06: [more ...](#)



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



Beacon v2 API

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

GA4GH Genome Beacons A Driver Project of the Global Alliance for Genomics and Health GA4GH and supported through ELIXIR

Beacon Protocol for Genomic Data Sharing

Beacons provide discovery services of the Global Alliance for Genomics and Health GA4GH standard for genomics data discovery against genomic data collections in repositories.

Baudisgroup @ UZH

(Ni Ai)

Michael Baudis
 (Haoyang Cai)
 (Paula Carrio Cordo)
 Bo Gao
 Qingyao Huang
 (Saumya Gupta)
 (Nitin Kumar)
 Sofia Pfund
 Rahel Paloots
 Ziying Yang
 Hangjia Zhao

Pierre-Henri Toussaint

Have you seen deletions in this region on chromosome 10 in Glioblastomas from juvenile patient, in a dataset with unrestricted access?

beacon-project.io

{S}{B} and GA4GH

Melanie Courtot
 Helen Parkinson
 many more ...

Query Beacon

beacon.progenetix.org/ui/



Beacon API Leads

Jordi Rambla
 Anthony Brooks
 Juha Törnroos
Discovery WS

Michael Baudis (Beacon)
 Marc Fiume (Networks)
ELIXIR
 Gary Saunders
 David Lloyd
 Serena Scollen

The Beacon protocol defines an open standard for genomics data discovery, developed by members of the Global Alliance for Genomics & Health. It provides a framework for public web services responding to queries against genomic data collections, for instance from population based or disease specific genome repositories.

This repository contains the specification for the v2 major version upgrade of the Beacon API. It is now (2020) under active development and has *not* seen a stable code release.

For further information, please follow the work here and consult the [Beacon Project website](#).

github.com/ga4gh-beacon/

About
 GA4GH Beacon v2 specification.
 ga4gh beacon openapi

Readme
 Apache-2.0 License

Releases
 No releases published Create a new release

Sabela de la Torre Pernas

Contributors 3
 sdelatorrep sdelatorrep
 mbaudis mbaudis
 blankdots blankdots



University of
Zurich^{UZH}



Global Alliance
for Genomics & Health



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

[progenetix.org](#)
[info.baudisgroup.org](#)
[sib.swiss/baudis-michael](#)
[imls.uzh.ch/en/research/baudis](#)
[beacon-project.io](#)
[schemablocks.org](#)