



Global Alliance
for Genomics & Health



Developing Beacons for Data Discovery

Advancing Beacons through *data-driven* implementations

Michael Baudis - #GA4GH2017



University of
Zurich^{UZH}



Beacon Project

An open web service that tests the willingness of international sites to share genetic data.



[Search Beacons](#)

A global search engine for genetic mutations.

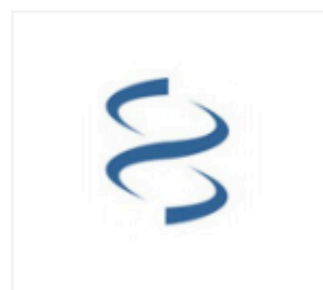
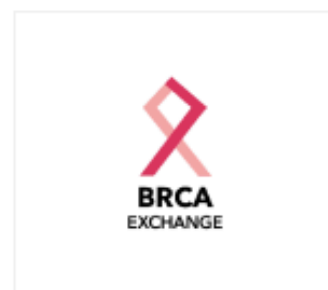
GRCh37 ▾

e.g. 1:100,000 A>C

Search

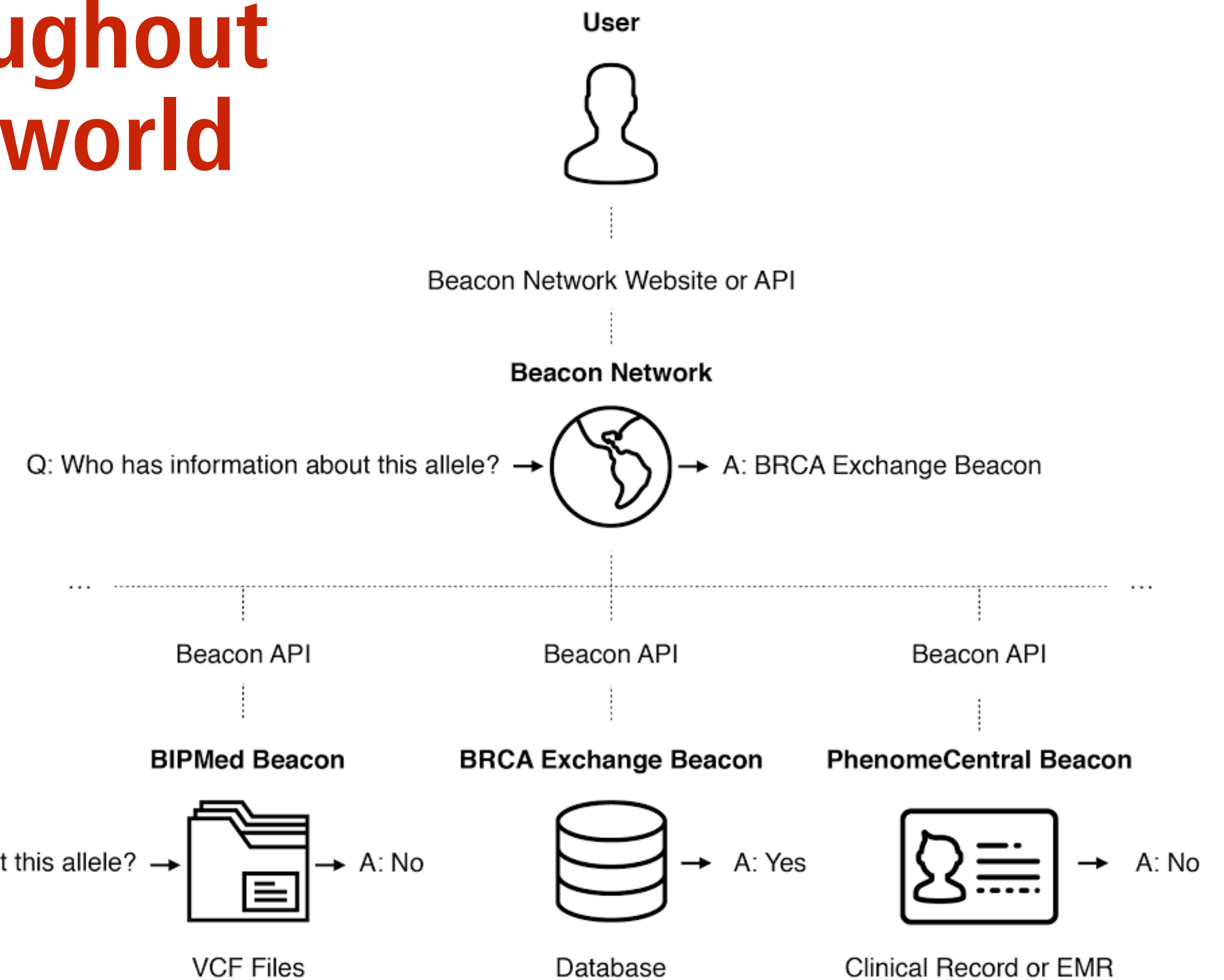
Quickstart: [Search for a BRCA2 variant](#)

Find genetic mutations shared by these organizations



[Browse Beacons](#)
»

> 50 Beacons throughout the world



Developing the GA4GH Metadata Schema

▶ arrayMap for GA4GH

- metadata schema development through implementation of arrayMap resource data
- OntologyTerm objects for biodata
- implementation w/ ontology services

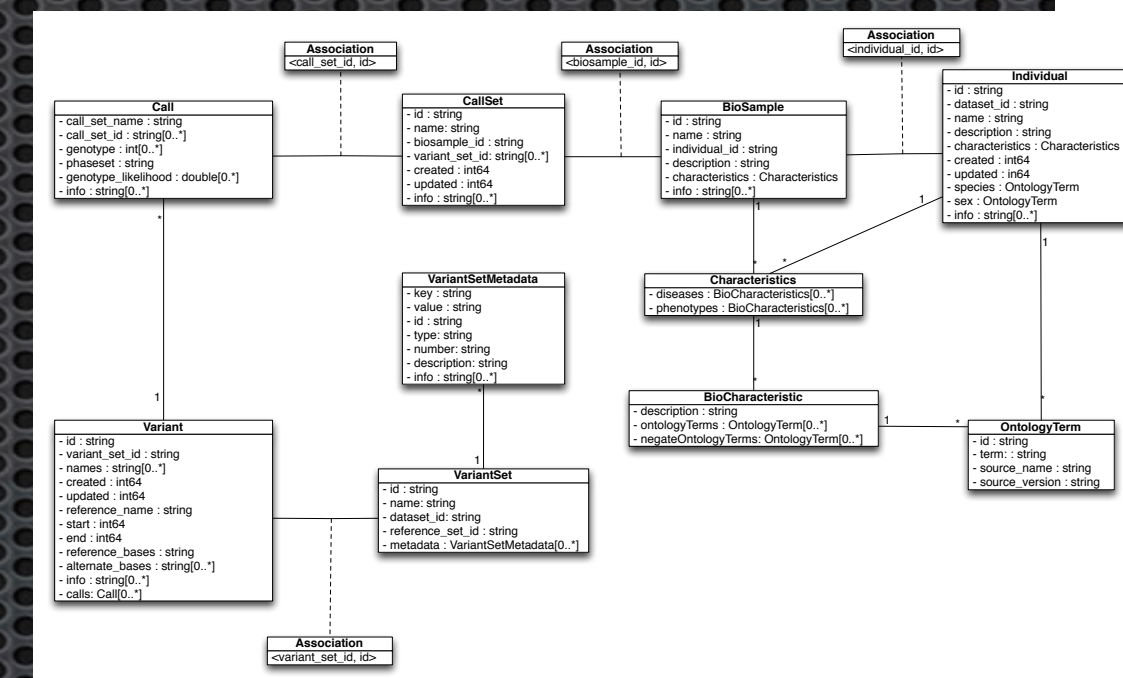


Swiss Institute of
Bioinformatics

Driving Beacon Development

▶ Beacon+

- CNV/CNA as first type of structural variants
- disease specific queries
- quantitative reporting



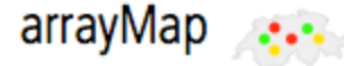
```

{
  "_id" : ObjectId("58297ca32ca4591e5a0df054"),
  "id" : "AM_V_1778741",
  "variant_set_id" : "AM_VS_HG18",
  "reference_name" : "10",
  "start" : 579049,
  "end" : 17236099,
  "alternate_bases" : "DUP",
  "reference_bases" : ".",
  "info" : {
    "svlen":16657050,
    "cipos":[
      -1000,
      1000
    ],
    "ciend":[
      -1000,
      1000
    ]
  },
  "calls" : [
    {
      "genotype" : [
        ".",
        "."
      ],
      "call_set_id" : "AM_CS_TCGA-61-1917-01A-01D-0648-01",
      "info" : {
        "segvalue" : 0.5491
      }
    }
  ],
  "created" : ISODate("2016-11-14T08:33:58.202Z"),
  "updated" : ISODate("2016-11-14T08:33:58.202Z"),
}


```


arrayMap

Resource for copy number variation data in cancer




- Search Samples
- Search Publications
- Gene CNA Frequencies
- User Data
- Array Visualization
- Progenetix



University of Zurich

- Citation
- User Guide
- Registration & Licensing
- People
- External Links






FOLLOW US ON 

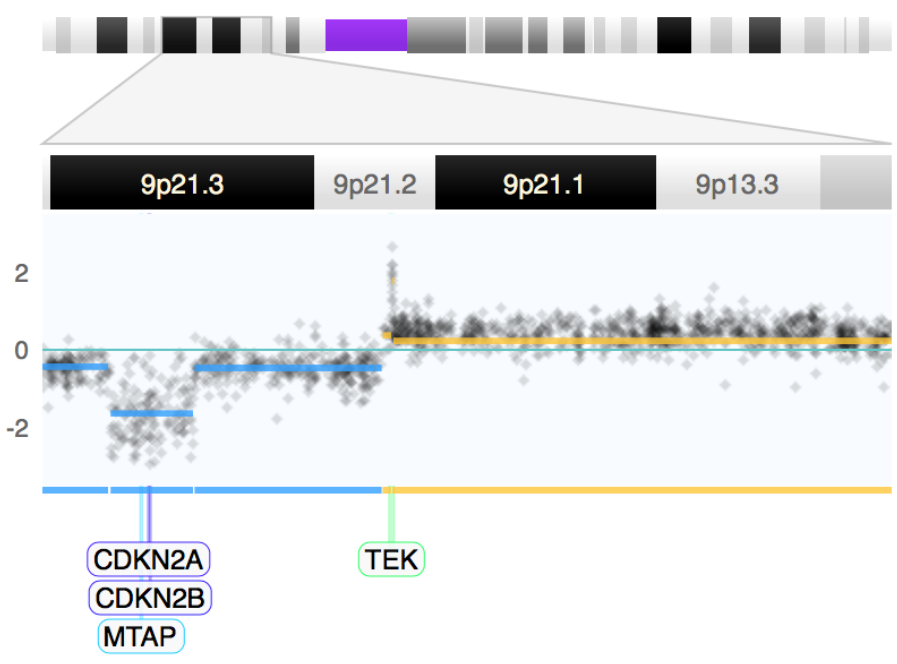
130.60.23.21

visualizing cancer genome array data @ arraymap.org

arrayMap is a curated reference database and bioinformatics resource targeting copy number profiling data in human cancer. The arrayMap database provides an entry point for meta-analysis and systems level data integration of high-resolution oncogenomic CNA data.

The current data reflects:

-  63060 genomic copy number arrays
-  763 experimental series
-  145 array platforms
-  141 ICD-O cancer entities
-  554 publications (Pubmed entries)



Genomic copy number imbalances on chromosome 9 in a case of Glioblastoma (GSM491153), indicating, among others, a homozygous deletion involving CDKN2A/B.

For the majority of the samples, probe level visualization as well as customized data representation facilitate gene level and genome wide data review. Results from multi-case selections can be connected to downstream data analysis and visualization tools, as we provide through our Progenetix project.

arrayMap is developed by the group "Theoretical Cytogenetics and Oncogenomics" at the Institute of Molecular Life Sciences of the University of Zurich.

BRAIN TUMOURS	5653 samples	[?]
BREAST CANCER	8329 samples	[?]
COLORECTAL CANCER	3238 samples	[?]
PROSTATE CANCER	991 samples	[?]
STOMACH CANCER	1062 samples	[?]

ARRAYMAP NEWS

2016-08-03: SVG graphics

2016-05-17: Transitioning to Europe PMC

[More news ...](#)

Feel free to use the data and tools for academic research projects and other applications. If more support and/or custom analysis is needed, please contact Michael Baudis regarding a collaborative project or a special license.

© 2000 - 2016 Michael Baudis, refreshed Mon, 19 Sep 2016 10:20:09 GMT in 6.87s on server 130.60.240.68. No responsibility is taken for the correctness of the data presented nor the results achieved with the Progenetix tools.

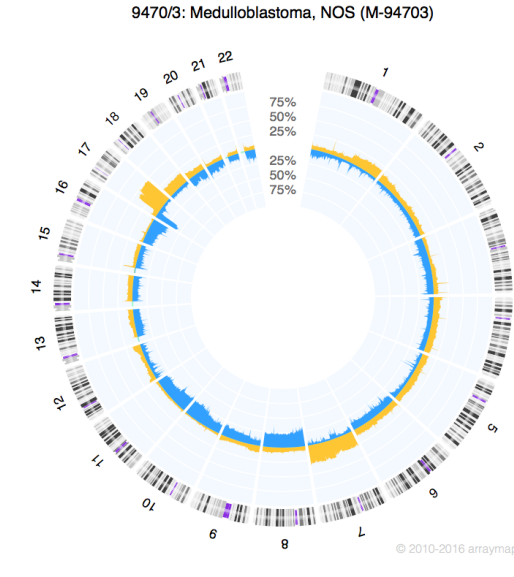
ICD Morphologies

2021 samples from arraymap have an associated "ICDMORPHOLOGYCODE" label.

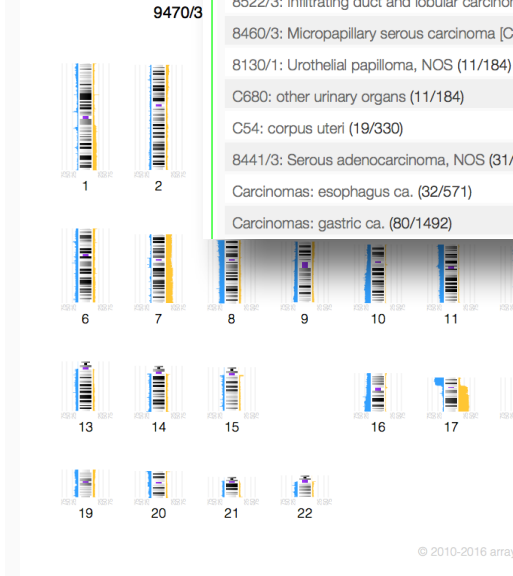
9470/3: Medulloblastoma, NOS (M-94703)

Synonyms

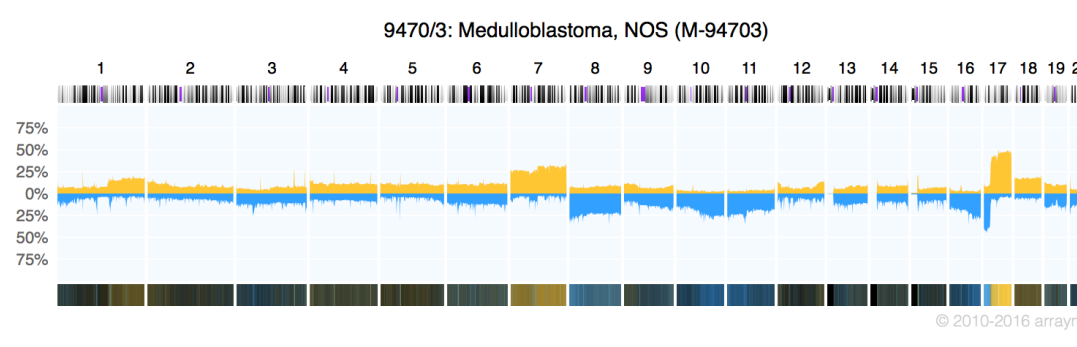
- Medulloblastoma, NOS
- Melanotic medulloblastoma



9470/3: Medulloblastoma, NOS (M-94703)



9470/3



9470/3: Medulloblastoma, NOS (M-94703)

FIND CNAS BY GENE OR REGION

REGION SIZE | MAX COVERAGE (KB) -

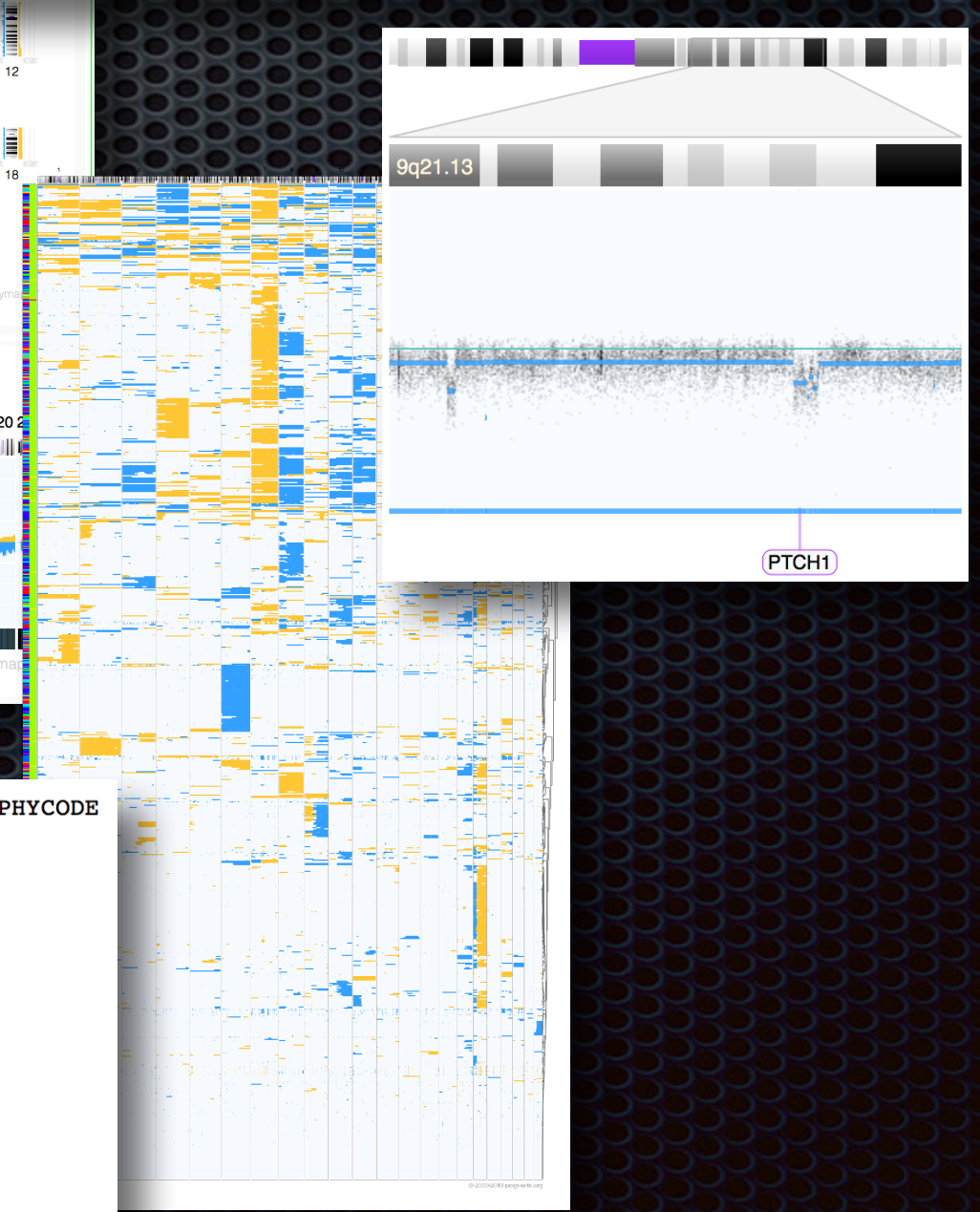
CLINICAL DATA

CITY km

[Query Database](#)

1949 of 65042 cases matched the selection criteria.

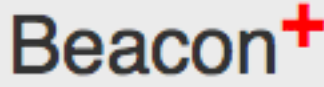

SUBSET	PERCENT IN SUBSET
8507/3: Invasive micropapillary carcinoma (13/39)	33.3
C692: retina (14/82)	17.1
8260/3: Papillary adenocarcinoma, NOS (11/65)	16.9
8500/3: Invasive carcinoma of no special type (1201/8188)	14.7
8560/3: Adenosquamous carcinoma (3/21)	14.3
Carcinomas: breast ca. (1254/8837)	14.2
C50: breast (1254/8929)	14.0
8500/2: Ductal carcinoma in situ, NOS (25/225)	11.1
C32: larynx (3/29)	10.3
8010/2: Carcinoma in situ, NOS (2/20)	10.0
C187: sigmoid incl. rectosigmoid junction (13/140)	9.3
8480/3: Mucinous adenocarcinoma (12/132)	9.1
8522/3: Infiltrating duct and lobular carcinoma (4/44)	9.1
8460/3: Micropapillary serous carcinoma [C56.9] (32/513)	6.2
8130/1: Urothelial papilloma, NOS (11/184)	6.0
C680: other urinary organs (11/184)	6.0
C54: corpus uteri (19/330)	5.8
8441/3: Serous adenocarcinoma, NOS (31/542)	5.7
Carcinomas: esophagus ca. (32/571)	5.6
Carcinomas: gastric ca. (80/1492)	5.4



UID	SERIESID	PMID	ICDMORPHOLOGYCODE	ICDPTOPOGRAPHYCODE
GSM1000061	GSE36942	23457519	8070/3	C10
GSM1000062	GSE36942	23457519	8070/3	C10
GSM1001316	GSE40777	23571474	8070/3	C53
GSM1001317	GSE40777	23571474	8010/3	C34
GSM1001318	GSE40777	23571474	8070/3	C09
GSM1001319	GSE40777	23571474	8010/3	C34
GSM1002668	GSE40834	24047479	9823/3	C42
GSM1002669	GSE40834	24047479	9823/3	C42
GSM1002670	GSE40834	24047479	9823/3	C42
GSM1002671	GSE40834	24047479	9823/3	C42
GSM1002672	GSE40834	24047479	9823/3	C42
GSM1002673	GSE40834	24047479	9823/3	C42
GSM1002674	GSE40834	24047479	9823/3	C42
GSM1002675	GSE40834	24047479	9823/3	C42
GSM1002676	GSE40834	24047479	9823/3	C42
GSM1002677	GSE40834	24047479	9823/3	C42
GSM1002678	GSE40834	24047479	9823/3	C42
GSM1002679	GSE40834	24047479	9823/3	C42
GSM1002680	GSE40834	24047479	9823/3	C42

Beacon+ Concept

- Implementation of cancer beacon prototype, backed by arrayMap and DIPG data set (MacKay *et al.*, Cancer Cell 2017, in print)
- structural variations (DUP, DEL) in addition to SNV
- diagnosis queries using ontology codes (NCIT, ICD-O)
- quantitative responses
- current version uses GA4GH schema compatible database

This forward looking Beacon interface implements additional, planned features beyond the current GA4GH specifications. [Info](#)

[SNV Example](#)
[CNV Example](#)

Query

Dataset: DIPG (CNV + selected SNV)

Reference name*: 17

Genome Assembly*: GRCh36 / hg18

Variant type*: SNV / indel

Position*: 7577121

Ref. Base(s)*: G



Alt. Base(s)*: A

Bio-ontology: pgx:icdom:9380_3




[Beacon Query](#)

Response

Dataset	Chro.	Assembly	Var. Type	Start Min	Start Max	End Min	End Max	Pos.	Ref.	Alt.	Bio Query	Call Count	Samples	f	Query
arraymap	9	GRCh36	DEL	19000000	21984490	21900000	25000000				pgx:icdom:8140_3	3781	403	0.0065	show JSON
dipg	17	GRCh36	SNV					7577121	G	A	pgx:icdom:9380_3	21	20	0.0187	show JSON

This Beacon implementation is developed by the [Computational Oncogenomics Group](#) at the [University of Zurich](#), with support from the [SIB Technology group](#) and [ELIXIR](#).

Beacon+ Concept

- Implementation of cancer beacon prototype, backed by arrayMap and DIPG data set (MacKay *et al.*, Cancer Cell 2017, in print)
- structural variations (DUP, DEL) in addition to SNV
- diagnosis queries using ontology codes (NCIT, ICD-O)
- quantitative responses
- current version uses GA4GH schema compatible database

Beacon+

This forward looking Beacon interface implements additional, planned features beyond the current GA4GH specifications. [Info](#)

Query [SNV Example](#) [CNV Example](#)

Dataset: arrayMap (CNV only)

Reference name*: 9

Genome Assembly*: GRCh36 / hg18

Variant type*: DEL (Deletion)

Start *min* Position*: 19000000

Start *max* Position: 21984490

End *min* Position: 21900000

End *max* Position: 25000000

Bio-ontology: ncit:C3059

[Beacon Query](#)

Response

Dataset	Chro.	Assembly	Var. Type	Start Min	Start Max	End Min	End Max	Pos.	Ref.	Alt.	Bio Query	Call Count	Samples	f	Query
dipg	17	GRCh36	SNV					7577121	G	A	pgx:icdom:9380_3	21	20	0.0187	show JSON
arraymap	9	GRCh36	DEL	19000000	21984490	21900000	25000000				ncit:C3059	3781	59	0.001	show JSON

arrayMap University of Zurich UZH This Beacon implementation is developed by the [Computational Oncogenomics Group](#) at the [University of Zurich](#), with support from the [SIB Technology group](#) and [ELIXIR](#).


```

    { "reference_name" : "0" },
    { "variant_type" : "DEL" },
    { "start" : { "$gte" : 19500000 } },
    { "start" : { "$lte" : 21984490 } },
    { "end" : { "$gte" : 21957751 } },
    { "end" : { "$lte" : 24500000 } }
  ],
  "api_version" : "0.4",
  "beacon_id" : "org.progenetix:progenetix-beacon",
  "exists" : true,
  "info" : {
    "query_string" :
    "dataset_id=arraymap&variants.reference_name=chr9&assembly_id=GRCh36&va
riants.variant_type=DEL&variants.start_max=19000000&variants.start_min=
21984490&variants.end_min=21900000&variants.end_max=25000000&biosamples
.bio_characteristics.ontology_terms.term_id=pgx:icdom:9440_3",
    "version" : "Beacon+ implementation based on a development branch
of the beacon-team project: https://github.com/ga4gh/beacon-team/pull/
94"
  },
  "url" : "http://progenetix.org/beacon/info/",
  "dataset_allele_responses" : [
    {
      "dataset_id" : "arraymap",
      "error" : null,
      "exists" : true,
      "external_url" : "http://arraymap.org",
      "sample_count" : 584,
      "call_count" : 3781,
      "variant_count" : 3244,
      "frequency" : 0.0094,
      "info" : {
        "description" : "The query was against database
\"arraymap_ga4gh\", variant collection \"variants_cnv_grch36\". 3781 /
59428 matched callsets for 3602919 variants. Out of 62105 biosamples in
the database, 2047 matched the biosample query; of those, 584 had the
variant.",
        "ontology_ids" : [
          "ncit:C3058",
          "pgx:icdom:9440_3",
          "pgx:icdot:C71.9",
          "pgx:icdot:C71.0"
        ]
      }
    }
  ]
}

```



Match using query ranges "at least one base in interval affected"

Region of Interest, e.g. CDR of Gene (here: CDKN2A)

Example "focal" matches (overlap w/ size limit)

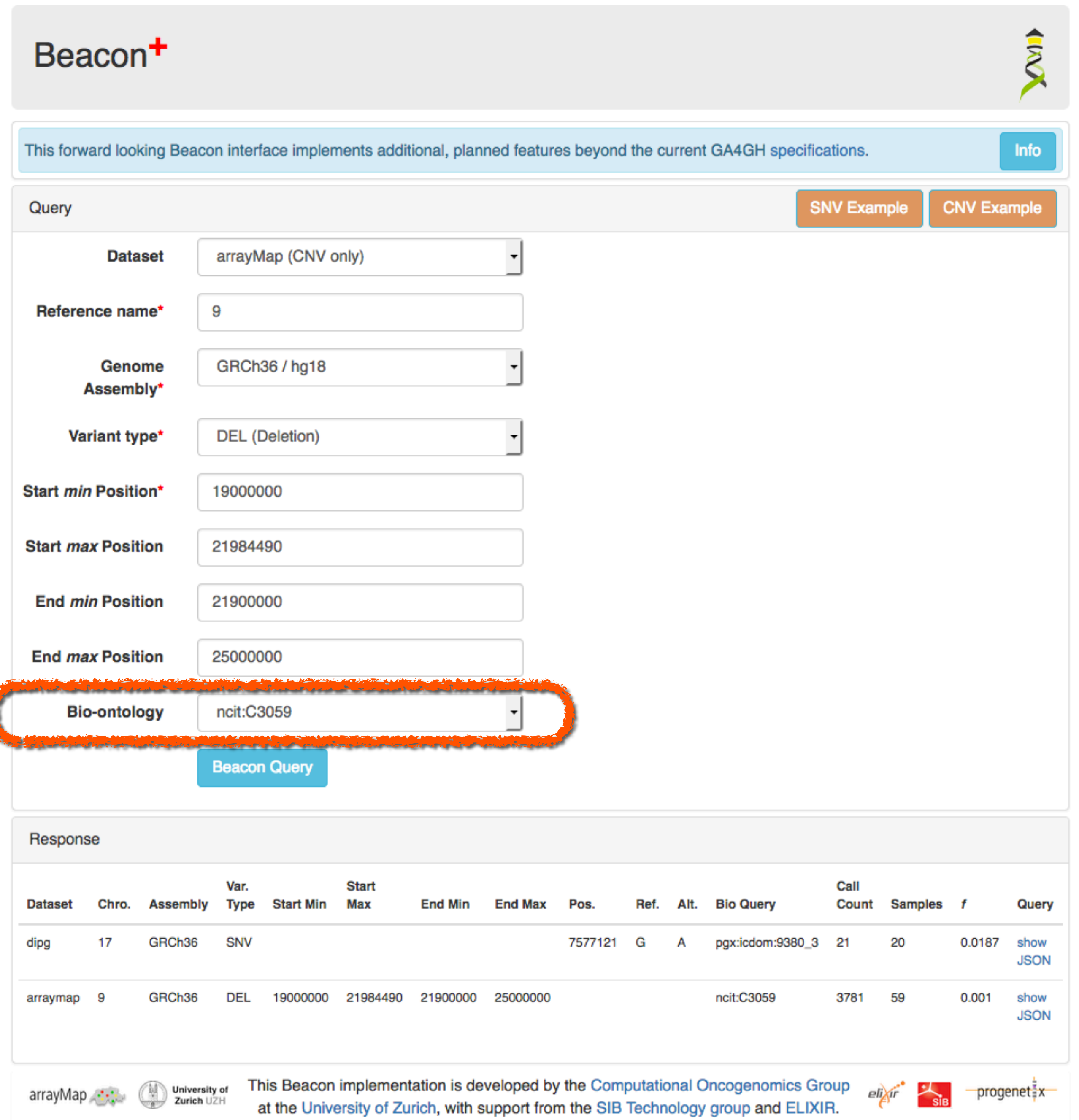
Mismatches
- too large
- end outside
- start outside


- Beacon+ **range queries** allow the definition of a genome region of interest, containing a specified variant or potentially other position related feature
- "fuzzy" matching of region ends essential for inexact features
- current Beacon implementation addresses CNV (<DUP>,), as are specified in VCF && GA4GH variant schema



Beacon+ Concept

- Implementation of cancer beacon prototype, backed by arrayMap and DIPG data set (MackKay *et al.*, Cancer Cell 2017, in print)
- structural variations (DUP, DEL) in addition to SNV
- diagnosis queries using ontology codes (NCIT, ICD-O)
- quantitative responses
- current version uses GA4GH schema compatible database



Beacon+ 

This forward looking Beacon interface implements additional, planned features beyond the current GA4GH specifications. [Info](#)

Query [SNV Example](#) [CNV Example](#)

Dataset: arrayMap (CNV only)

Reference name*: 9

Genome Assembly*: GRCh36 / hg18

Variant type*: DEL (Deletion)

Start *min* Position*: 19000000

Start *max* Position: 21984490

End *min* Position: 21900000






End *max* Position: 25000000

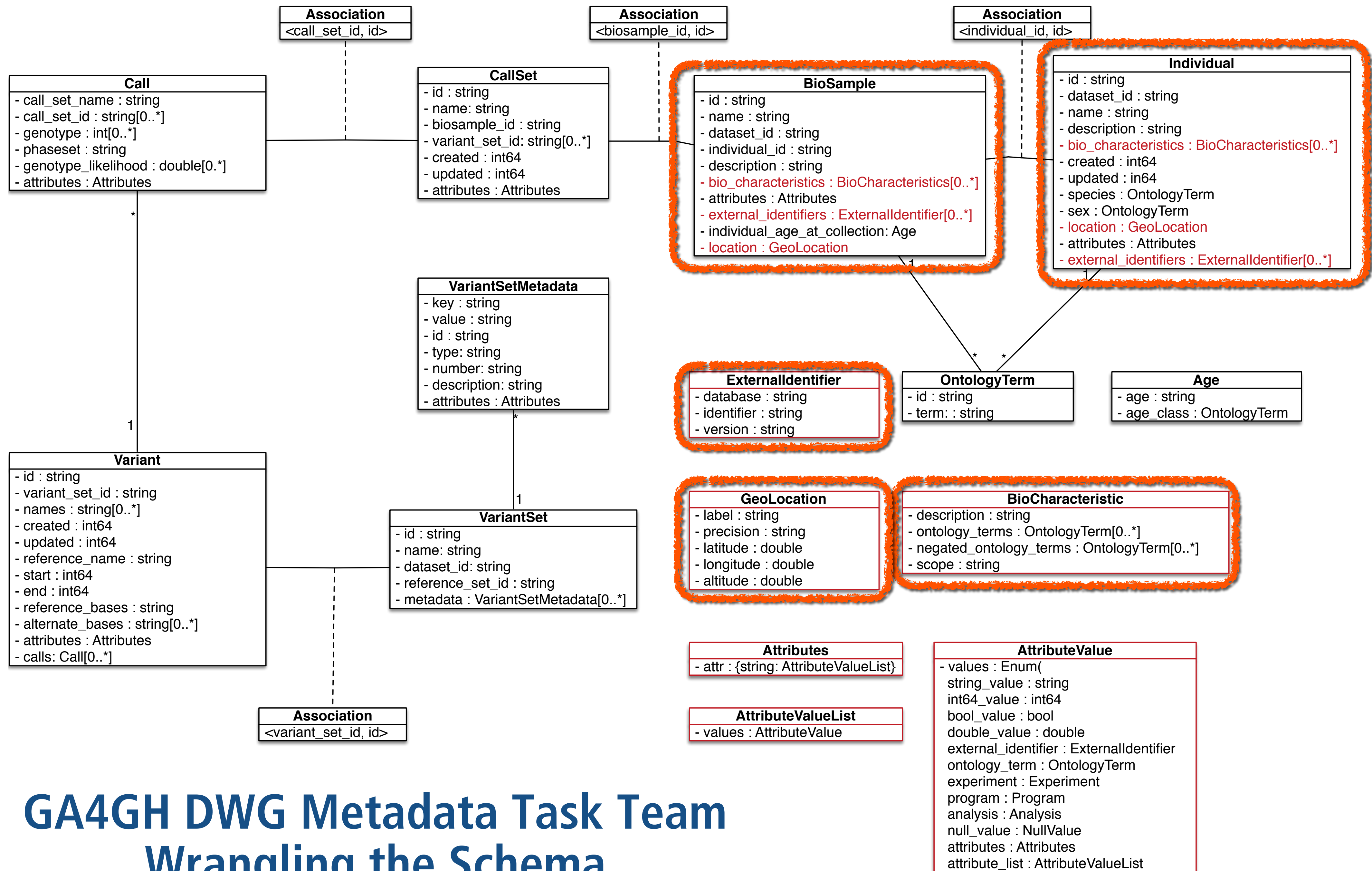
Bio-ontology: **ncit:C3059**

[Beacon Query](#)

Response

Dataset	Chro.	Assembly	Var. Type	Start Min	Start Max	End Min	End Max	Pos.	Ref.	Alt.	Bio Query	Call Count	Samples	f	Query
dipg	17	GRCh36	SNV					7577121	G	A	pgx:icdom:9380_3	21	20	0.0187	show JSON
arraymap	9	GRCh36	DEL	19000000	21984490	21900000	25000000				ncit:C3059	3781	59	0.001	show JSON

arrayMap   This Beacon implementation is developed by the [Computational Oncogenomics Group](#) at the [University of Zurich](#), with support from the [SIB Technology group](#) and [ELIXIR](#).   



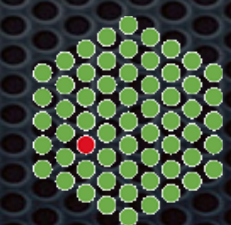
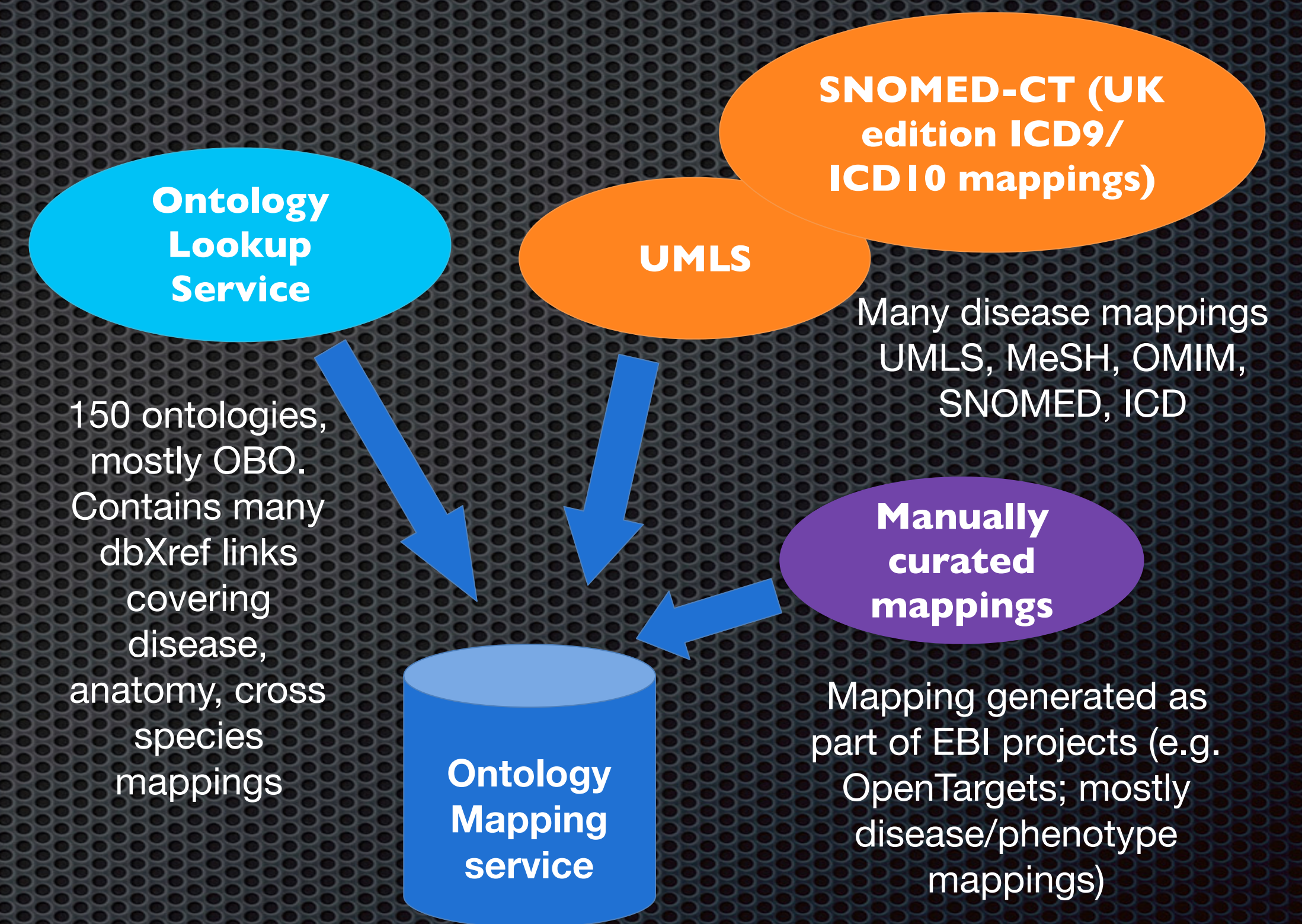
GA4GH DWG Metadata Task Team

Wrangling the Schema

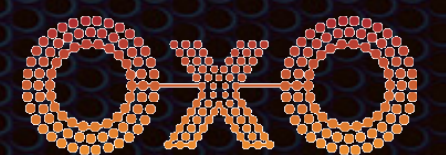
Making Ontologies Work for GA4GH Implementation Studies

- biomedical "metadata" in different resources frequently follows incompatible classification systems
- medical coding systems are driven by different paradigms compared to biological ontologies (e.g. for cross-species comparisons)
- frequently used classifications (ICD, Snomed...) are either not "ontologised" or cannot be referenced in open resources

Federated queries across resources need **curated mappings** of classifications/ontologies



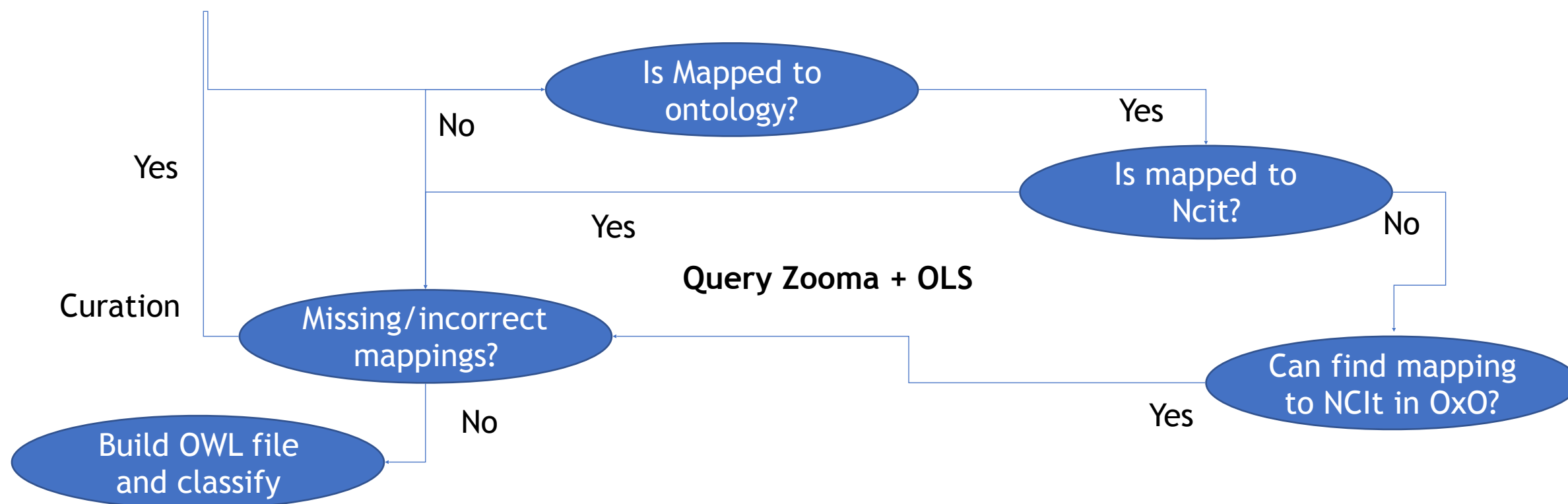
Ontology Mapping Service



Working towards ontologies w/ arrayMap: Mapping >55'000 samples from ICD-O to NCIt neoplasm core

ICDM	ICDMORPHOLOGY	NCItcode	NCItlabel	ICDT	ICDTOPOGRAPHY	NCItcode	NCItlabel	NCIt_mapper	NCIt_mapped_ICDM_T_label
8021/3	Carcinoma anaplastic type			C739	thyroid gland	C12400	thyroid gland	C3878	Thyroid Gland Undifferentiated (Anaplastic) Carcinoma
9451/3	Oligodendroglioma anaplastic	C4326	anaplastic oligodendroglioma	C719	Brain	C12439	brain	C4326	Anaplastic Oligodendroglioma
9051/3	Desmoplastic mesothelioma	C6747		C499	connective and soft tissue			C6747	Desmoplastic Mesothelioma
9732/3	Plasma cell myeloma	C3242	multiple myeloma	C42	hematopoietic and reticuloendothelial systems			C3242	Plasma Cell Myeloma
8070/3	Squamous cell carcinoma	C2926	non-small cell lung carcinoma	C140	pharynx			C102872	Pharyngeal Squamous Cell Carcinoma
8380/3	Endometrioid adenocarcinoma	C3769	endometrioid carcinoma	C54	corpus uteri	C12316		C6287	Endometrial Endometrioid Adenocarcinoma
8070/3	Squamous cell carcinoma	C2926	non-small cell lung carcinoma	C44	skin	C12470	zone of skin	C4819	Skin Squamous Cell Carcinoma
8430/3	Mucoepidermoid carcinoma	C45544	pulmonary mucoepidermoid carcinoma	C089	salivary gland	C12426	saliva-secreting gland	C5953	Minor Salivary Gland Mucoepidermoid Carcinoma
9680/3	Diffuse large B-cell lymphoma	C8851	diffuse large B-cell lymphoma	C42	hematopoietic and reticuloendothelial systems			C8851	Diffuse Large B-Cell Lymphoma
8800/3	Sarcoma	C9118	sarcoma	C559	uterus nos			C9306	Soft Tissue Sarcoma
8441/3	Serous adenocarcinoma	C7550	ovarian serous adenocarcinoma	C570	fallopian tube	C12403	fallopian tube	C40101	Serous Adenocarcinoma
9689/3	splenic marginal zone lymphoma nos			C422	spleen	C12432	spleen	C4663	Splenic Marginal Zone Lymphoma
8077/2	Squamous intraepithelial neoplasia grade III			C53	cervix uteri	C12311		C89476	Grade III Vaginal Intraepithelial Neoplasia
8140/0	Adenoma	C4196	adenoma	C189	large intestine excl. rectum and rectosigmoid junction			C4349	Colon Adenocarcinoma
8272/3	Pituitary carcinoma	C4536	Pituitary carcinoma	C751	pituitary gland	C12399	pituitary gland	C4536	Pituitary Gland Carcinoma
8500/2	Ductal carcinoma in situ	C3641	ductal carcinoma in situ	C50	breast	C12971	breast	C2924	Ductal Breast Carcinoma In Situ
8200/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx	C12420	larynx	C2970	Adenoid Cystic Carcinoma
9370/3	Chordoma	C2947	Chordoma	C419	bone	C13076	bone tissue	C2947	Chordoma
9717/3	Enteropathy type T-cell lymphoma			C17	small intestine	C12386	small intestine	C4737	Enteropathy-Associated T-Cell Lymphoma
9698/3	Follicular lymphoma grade 3			C42	hematopoietic and reticuloendothelial systems			C3460	Grade 3 Follicular Lymphoma
9863/3	Chronic myeloid leukemia	C3177	chronic myelogenous leukemia	C42	hematopoietic and reticuloendothelial systems			C3174	Chronic Myelogenous Leukemia BCR-ABL1 Positive
8852/3	Liposarcoma myxoid	C3735	myxoid liposarcoma	C499	connective and soft tissue			C27781	Myxoid Liposarcoma
9080/3	Teratoma malignant			C809	unknown	C35882	Hereditary elliptocytosis	C3403	Teratoma
8530/3	Inflammatory carcinoma	C4872	breast carcinoma	C50	breast	C12971	breast	C4001	Inflammatory Breast Carcinoma
8140/3	Adenocarcinoma	C27745	lung adenocarcinoma	C809	unknown	C35882	Hereditary elliptocytosis	C2852	Adenocarcinoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx	C12762	oropharynx	C2970	Adenoid Cystic Carcinoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx	C12415	kidney	C3158	Leiomyosarcoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx	C12499	internal ear	C2970	Adenoid Cystic Carcinoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx	C12683	bronchus	C2923	Bronchioloalveolar Carcinoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx	C12343	retina	C3224	Melanoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C8459	Hepatosplenic T-Cell Lymphoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx	C12393	pancreas	C8294	Pancreatic Adenocarcinoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C3996	Monoclonal Gammopathy of Undetermined Significance
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C4817	Ewing Sarcoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C3288	Oligodendroglioma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C4648	Tongue Squamous Cell Carcinoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C2862	Primary Myelofibrosis
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C4833	Oral Cavity Squamous Cell Carcinoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C9383	Rectal Adenocarcinoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C3158	Leiomyosarcoma
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx			C3898	Extranodal Marginal Zone Lymphoma of Mucosa-Associated Lymphoid Tissue
9060/3	Adenoid cystic carcinoma	C2970	adenoid cystic carcinoma	C32	larynx	C12404	female gonad	C4512	Ovarian Mucinous Cystadenoma

- From 456 pairs of ICD-O terms Morphology and Topography representative of cancer entities in arrayMap
- Develop Python script to take ICD-O Morphology and Topography labels separately QUERY ZOOMA, Oxo and OLS to find mapping to NCIt



From 456 pairs of ICD-O
 70% ICD-O Morphology - NCIt
 65% ICD-O Topography - NCIt

45% ICD-O-3 Pairs mapped to NCIt terms

=> MANUAL CURATION of >50%

Beacon+ Concept

Testing Beacons for Data Discovery

```
{ "reference_name" : "9" },
{ "variant_type" : "DEL" },
{ "start" : { "$gte" : 19500000 } },
{ "start" : { "$lte" : 21984490 } },
{ "end" : { "$gte" : 21957751 } },
{ "end" : { "$lte" : 24500000 } }
]
},
"api_version" : "0.4",
"beacon_id" : "org.progenetix:progenetix-beacon",
"exists" : true,
"info" : {
  "query_string" :
"dataset_id=arraymap&variants.reference_name=chr9&assembly_id=GRCh36&va
riants.variant_type=DEL&variants.start_max=19000000&variants.start_min=
21984490&variants.end_min=21957751&variants.end_max=24500000&biosamples
.bio_characteristics.ontology_terms.term_id=pgx:icdom:9440_3"
  "description" : "The query was against database
\"arraymap_ga4gh\", variant collection \"variants_cnv_grch36\". 3781 /
59428 matched callsets for 3602919 variants. Out of 62105 biosamples in
the database, 2047 matched the biosample query; of those, 584 had the
variant.",
  "ontology_ids" : [
    "ncit:C3058",
    "pgx:icdom:9440_3",
    "pgx:icdot:C71.9",
    "pgx:icdot:C71.0"
  ]
}
```

- standard Beacon payload (e.g. “exists”)
- testing GA4GH metadata “biocharacteristics” ontology term ids
- multiple datasets can be returned (only one shown here)
- quantitative reporting
- additional information about query & dataset(s)



Implementing real-world datasets for federated access using GA4GH schema specifications: pHGG

- Study in >1000 rare aggressive childhood brain tumors
- 157 of those not published previously
- copy number aberration data and selected gene panel represented in DIPG Beacon+
- interface with quantitative returns

DIPG Beacon+

This forward looking Beacon interface implements additional, planned features beyond the current GA4GH specifications, based on the pediatric high grade glioma dataset from MacKay et al. (Cancer Cell 2017). [Info](#)

Query SNV Example CNV Example

Reference name*

Genome Assembly*

Variant type*

Position*

Ref. Base(s)*

Alt. Base(s)*

Bio-ontology

[Beacon Query](#)

The SNV example test the Beacon+ UI and backend against the DIPG dataset, with a specific mutation.

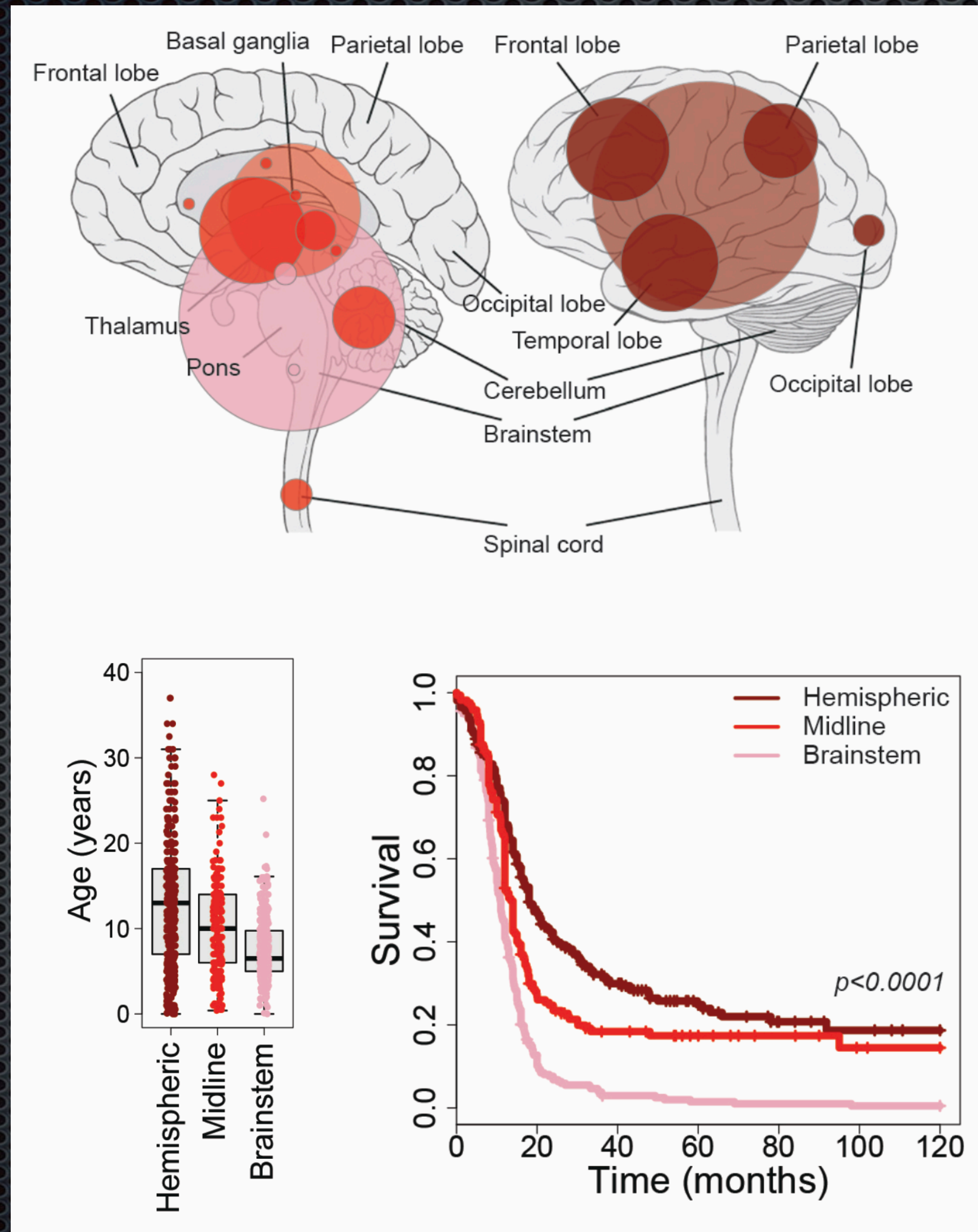
Response

Dataset	Assembly	Chro	Var Type	Start Range	End Range	Pos	Ref Alt	Bio Query	Variants Calls Callsets Samples	f_{callsets} f_{bio}	Response Context
dipg	hg18	17	SNV			7577121	G A	pgx:icdot:c71.7	1 21 21 12	0.0724 0.0367	JSON UCSC

arravMan This Beacon implementation is developed by the Computational Oncogenomics Group at the University of Zurich, with support from the SIB Technology group and ELIXIR.

Implementing real-world datasets for federated access using GA4GH schema specifications: pHGG

Mackay A, Jones C, Baudis M and many others:
 Integrated molecular meta-analysis of 1000 paediatric
 high grade and diffuse intrinsic pontine glioma (2017,
 Cancer Cell, in press)



```

    "id" : "DIPG_IND_0809",
    "individual_id" : "DIPG_IND_0809",
    "id" : "DIPG_BS_0809",
    "name" : "pHGG_META_0809",
    "description" : "glioma, paediatric, high grade",
    "individual_age_at_collection" : {
      "age_class" : {
        "term" : "Adult onset",
        "term_id" : "HP:0003581"
      },
      "age" : "P17Y0M"
    },
    "bio_characteristics" : [
      {
        "ontology_terms" : [
          {
            "term_label" : "Glioma",
            "term_id" : "ncit:C3059"
          },
          {
            "term_label" : "Brain NOS",
            "term_id" : "pgx:icdot:C71.9"
          }
        ]
      },
      {
        "description" : "Juvenile high grade glioma",
      }
    ],
    "external_identifiers" : [
      {
        "database" : "Pubmed",
        "identifier" : "25752754",
        "relation" : "reported_in"
      }
    ],
    "attributes" : {
      "grade" : { "values" : [ { "string_value" : "4" } ] },
      "histone" : { "values" : [ { "string_value" : "wt" } ] },
    }
  }

```


Check it Out!

- managed, participation driven projects living on Github: **ga4gh**
 - *beacon.arraymap.org*
 - *dipg.progenetix.org*
 - test datasets & code available through our **progenetix** repositories
- ➔ test
 - ➔ comment
 - ➔ suggest
 - ➔ propose
 - ➔ complain ...

The screenshot shows a GitHub repository page for 'progenetix / arraymap2ga4gh'. The repository has 5 stars, 2 forks, and 1 contributor. It contains 85 commits, 2 branches, 0 releases, and 3 contributors. The repository is currently on the 'master' branch. The file list includes:

File	Description	Last Commit
data	Multi genome editions	6 days ago
examples	update DIPG examples	2 months ago
tools	remove the per scripts => beaconplus-server	2 months ago
README.md	link	2 months ago
schema.pdf	updated schema diagram	6 months ago

The README.md file contains the following text:

Implementation of the **GA4GH** schema based on genome profiles and metadata from **arrayMap**

This repository will contain data and information regarding the **arrayMap** based implementation of a GA4GH schema structure. While it is not expected that GA4GH compliant resources mirror the schema in their internal structure, this project is aimed at showing the principle feasibility of such an approach, mainly to test & drive schema development.

Data & schemas represented here are not kept in a stable/versioned status, but are updated together with or anticipating GA4GH schema changes.

BAUDISGROUP @ UZH

NI AI
MICHAEL BAUDIS
(HAOYANG CAI)
PAULA CARRIO CORDO
BO GAO
(LINDA GROB)
SAUMYA GUPTA
(ROMAN HILLJE)
QINGYAO HUANG
(NITIN KUMAR)
(ALESSIO MILANESE)

SIB

HEINZ STOCKINGER
SÉVERINE DUVAUD
VASSILIOS IOANNIDIS
DANIEL TEIXEIRA

THOMAS EGGERMANN
ROSA NOGUERA
REINER SIEBERT
CAIUS SOLOVAN



GA4GH DWG + CWG

JACQUI BECKMANN
ANTHONY BROOKES
MARK DIEKHANS
MARC FIUME
MELISSA HAENDEL
DAVID HAUSSLER
SARAH HUNT
STEPHEN KEENAN
SUZY LEWIS
DAVID LLOYD
MICHAEL MILLER
HELEN PARKINSON
GUNNAR RÄTSCH
DAVID STEINBERG
JULIA WILSON

ELIXIR, CRG, EBI

JORDI RAMBLA DE ARGILA
MELANIE COURTOT
S. DE LA TORRE PERNAS
SUSANNA REPO
SERENA SCOLLEN
TRISH WHETZEL



University of
Zurich ^{UZH}



Global Alliance
for Genomics & Health