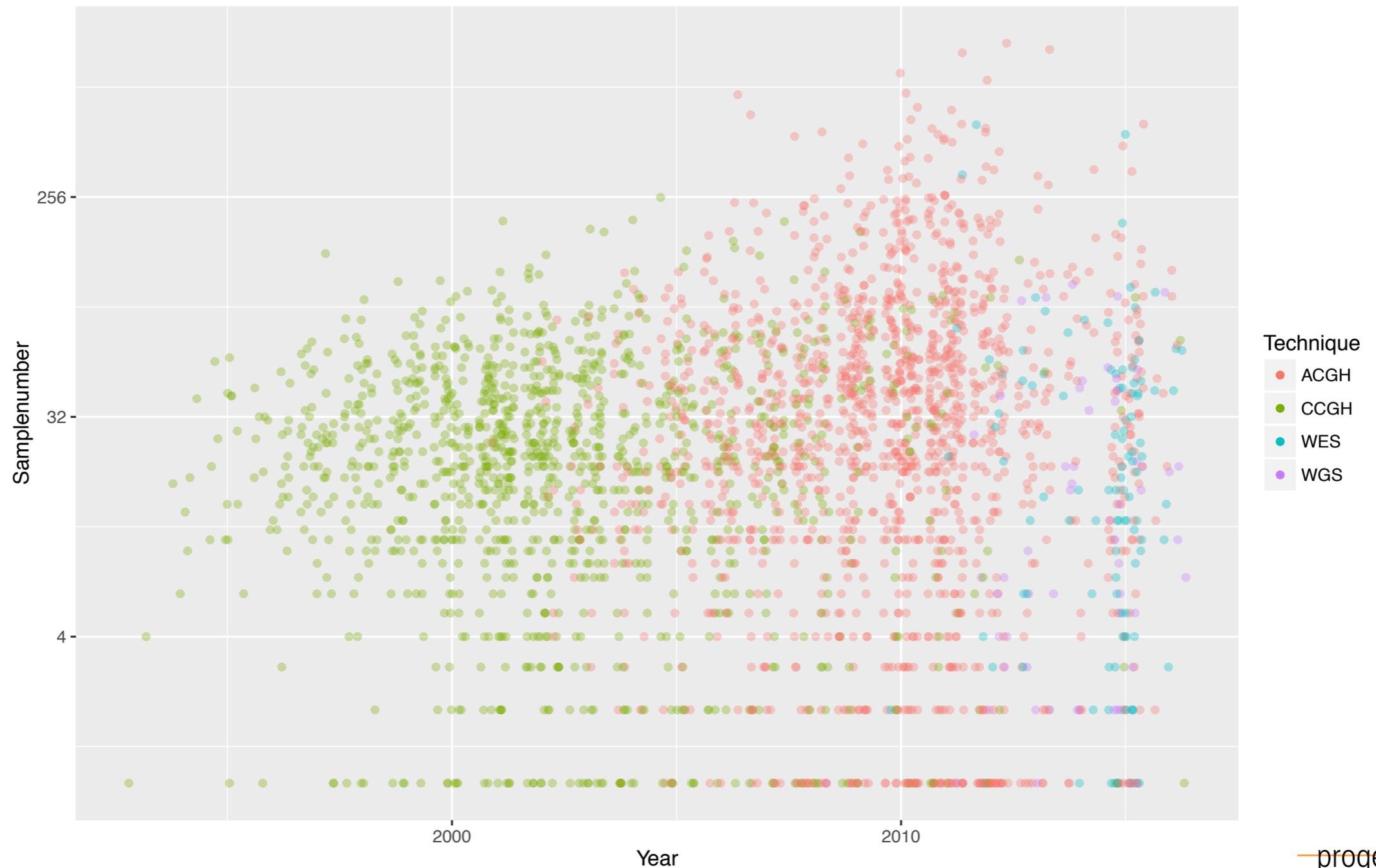


CURRENT OPPORTUNITIES AND FUTURE CHALLENGES

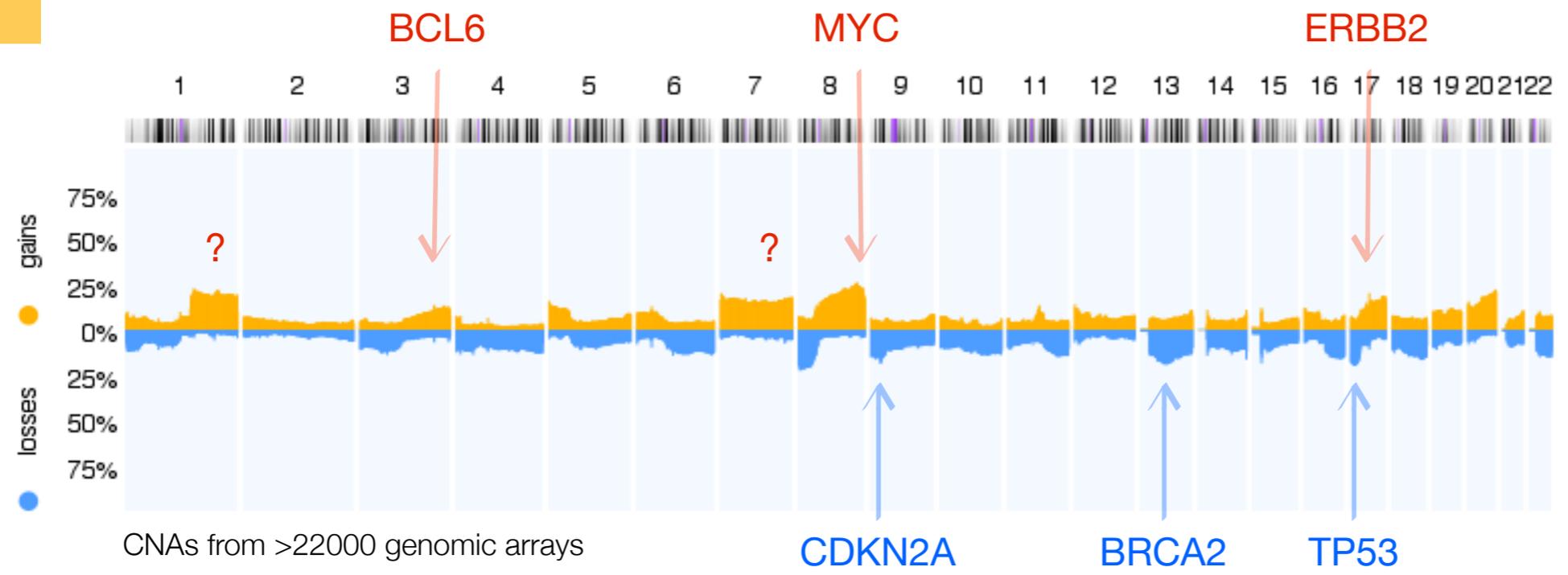
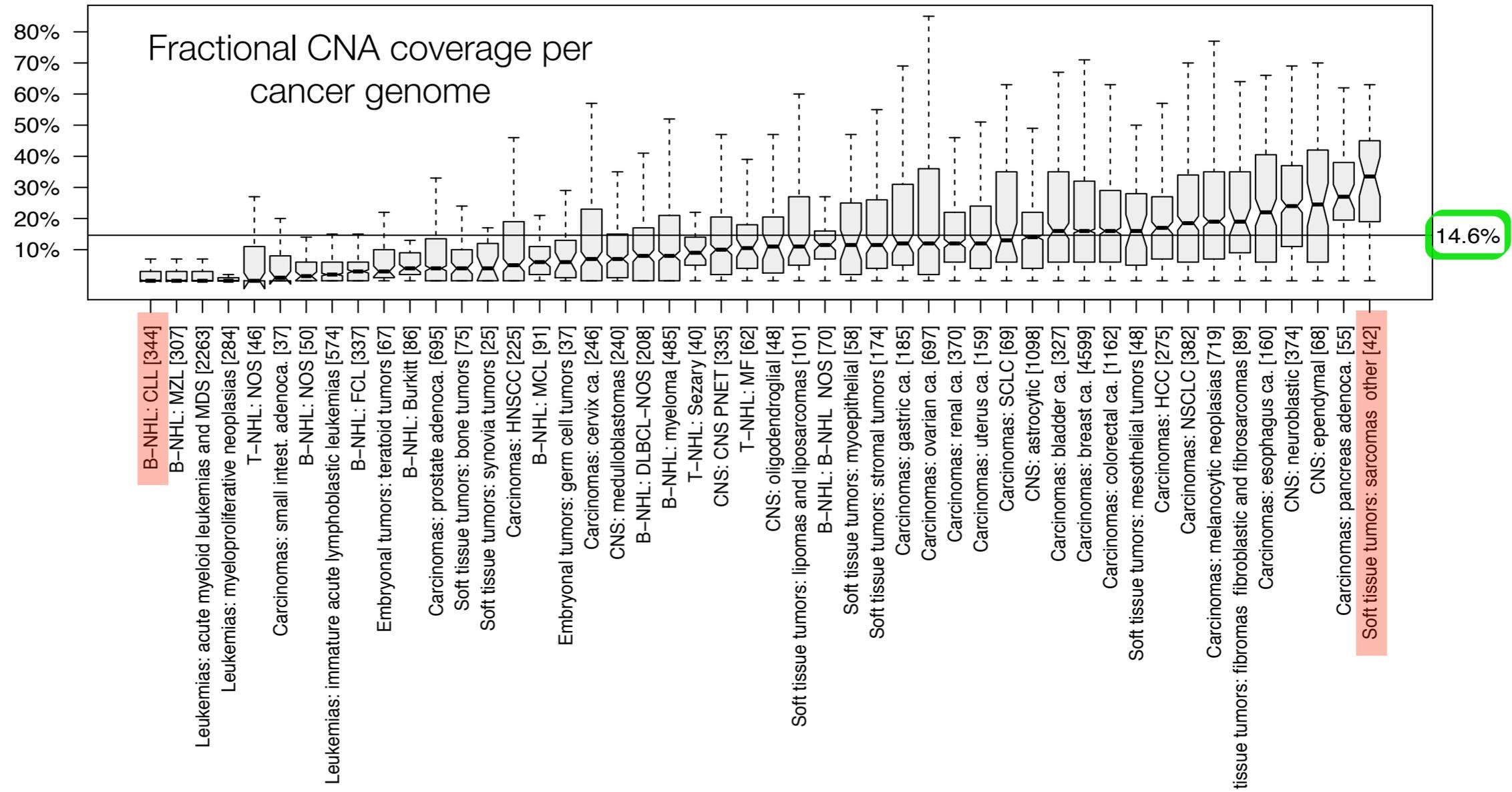
HARVESTING CANCER GENOME DATA

MOLECULAR CYTOGENETICS & SEQUENCING STUDIES FOR WHOLE GENOME PROFILING

Cancer Samples per Publication for Different Techniques
[129417 samples from 2747 publications]



GENOMIC COPY NUMBER IMBALANCED PROVIDE WIDESPREAD SOMATIC VARIANTS IN CANCER



Gain => Oncogenes

Deletion => Tumor
Suppressor Genes

Abstract: Through increasing sample size and more recently, high-throughput sequencing approaches, a wealth of genetic data has become available in cancer research. In addition, the use of the Progenetix.net repository has allowed for the collection and analysis of a large number of data sets, providing an international, open access platform for the storage and analysis of molecular cytogenetic aberration data. The availability of a central, open access platform for the storage and analysis of molecular cytogenetic aberration data is essential for the advancement of cancer research. Progenetix.net provides a central, open access platform for the storage and analysis of molecular cytogenetic aberration data, and is essential for the advancement of cancer research.

Background: Through increasing sample size and more recently, high-throughput sequencing approaches, a wealth of genetic data has become available in cancer research. In addition, the use of the Progenetix.net repository has allowed for the collection and analysis of a large number of data sets, providing an international, open access platform for the storage and analysis of molecular cytogenetic aberration data. The availability of a central, open access platform for the storage and analysis of molecular cytogenetic aberration data is essential for the advancement of cancer research. Progenetix.net provides a central, open access platform for the storage and analysis of molecular cytogenetic aberration data, and is essential for the advancement of cancer research.



techniques

scope

content

raw data presentation

per sample re-analysis

final data

main purposes

cCGH, aCGH, WES, WGS	aCGH (+?)
sample (e.g. combination of several experiments); literature tracking	experiment
>31000 samples >2700 publications	>60000 arrays
no (link to sources if available)	yes (raw, log2, segmentation if available)
no; supervised result (mostly as provided through publication)	yes (re-segmentation, thresholding, size filters ...)
annotated/interpreted CN status for GP and cytogenetic regions	unsupervised CN status for GP and cytogenetic regions
<ul style="list-style-type: none"> ● Distribution of CNA target regions in most tumor types (>350 ICD-O) ● Cancer classification 	<ul style="list-style-type: none"> ● Gene specific hits ● Genome feature correlation (fragile sites ...)



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



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

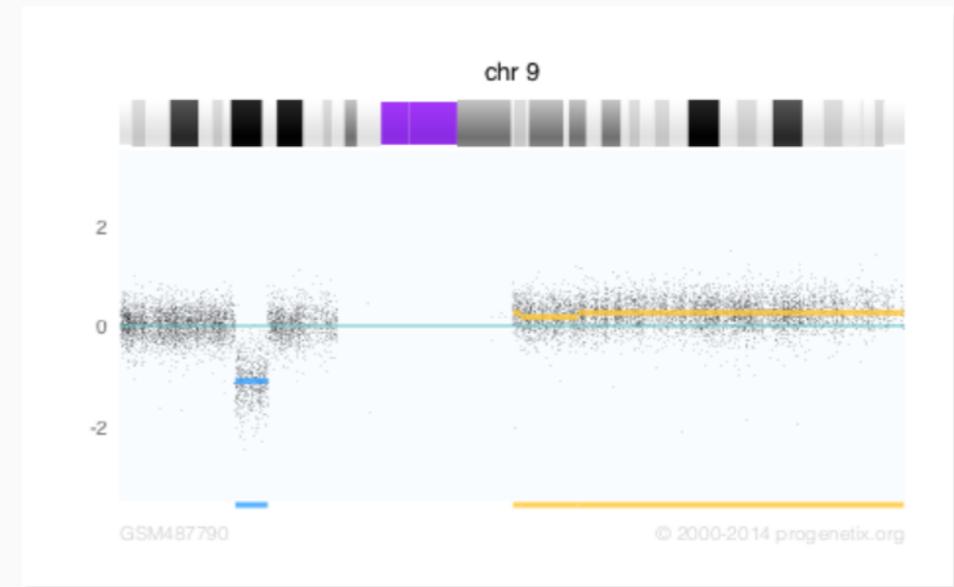
FOLLOW US ON 



130.60.23.21

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:

-  65042 genomic copy number arrays
-  986 experimental series
-  333 array platforms
-  253 ICD-O cancer entities
-  716 publications (Pubmed entries)



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	5791 samples ↗	[?]
BREAST CANCER	8594 samples ↗	[?]
COLORECTAL CANCER	3470 samples ↗	[?]
PROSTATE CANCER	1366 samples ↗	[?]
STOMACH CANCER	1457 samples ↗	[?]

ARRAYMAP NEWS

2016-04-11: Sorting cancer subset tables

2015-03-23: SIB Profile 2015

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.

THE ARRAYMAP CANCER GENOME RESOURCE

FIND CNAS BY GENE OR REGION [?]

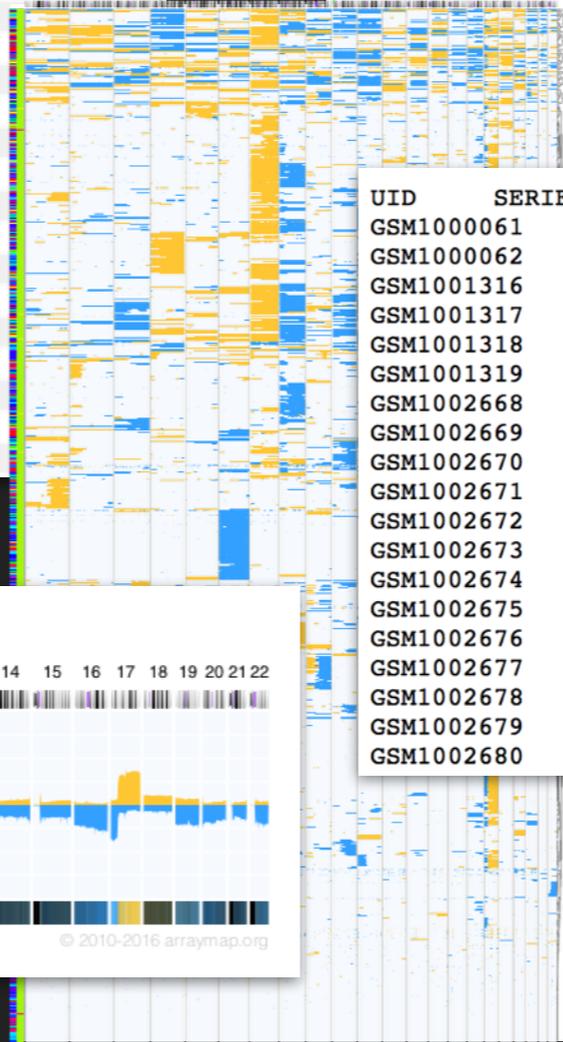
REGION SIZE | MAX COVERAGE (KB) - [?]

CLINICAL DATA [?]

CITY km [?]

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)	
C50: breast (1254/8929)	
8500/2: Ductal carcinoma in situ, NOS (25/225)	
C32: larynx (3/29)	
8010/2: Carcinoma in situ, NOS (2/20)	
C187: sigmoid incl. rectosigmoid junction (13/140)	
8480/3: Mucinous adenocarcinoma (12/132)	
8522/3: Infiltrating duct and lobular carcinoma (4/44)	
8460/3: Micropapillary serous carcinoma [C56.9] (32/513)	
8130/1: Urothelial papilloma, NOS (11/184)	
C680: other urinary organs (11/184)	
C54: corpus uteri (19/330)	
8441/3: Serous adenocarcinoma, NOS (31/542)	
Carcinomas: esophagus ca. (32/571)	
Carcinomas: gastric ca. (80/1492)	



ICD Morphologies

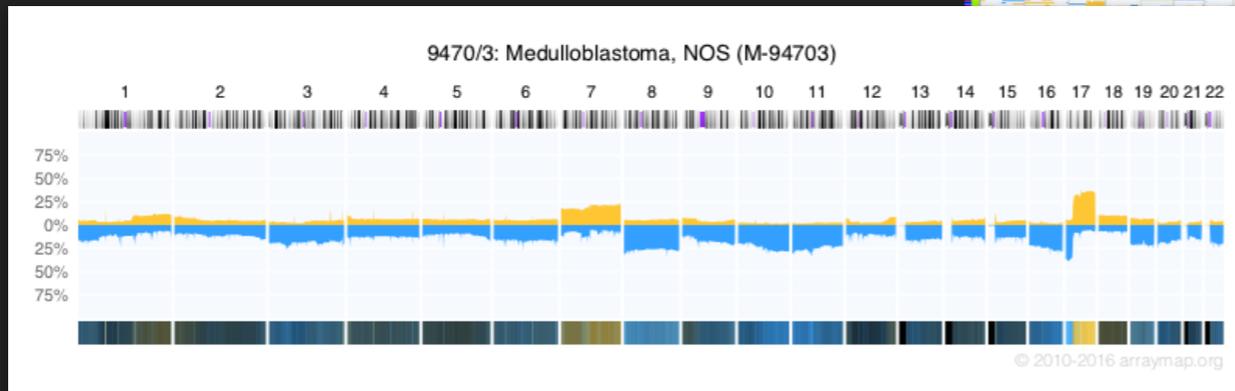
64485 samples from arraymap have an associated "ICDMORPHOLOGYCODE" label.

31902 samples from progenetix have an associated "ICDMORPHOLOGYCODE" label.

400 subsets of type ICDMORPHOLOGYCODE will be parsed.

	Subsets	arraymap
<input type="checkbox"/>	8140/3: adenocarcinoma, nos	9469
<input type="checkbox"/>	0000/0: not classified in icd-o 3 [e.g. non-neoplastic or benign]	8814
<input type="checkbox"/>	8500/3: invasive carcinoma of no special type	8188
<input type="checkbox"/>	9861/3: acute myeloid leukemia, nos	2831
<input type="checkbox"/>	8070/3: squamous cell carcinoma, nos	2443
<input type="checkbox"/>	9440/3: glioblastoma, nos	2294
<input type="checkbox"/>	9823/3: b-cell chronic lymphocytic leukemia/small lymphocytic lymphoma	2114
<input type="checkbox"/>	9470/3: medulloblastoma, nos	2052
<input type="checkbox"/>	9835/3: acute lymphoblastic leukemia, nos	1789
<input type="checkbox"/>	8010/3: carcinoma, nos	1430
<input type="checkbox"/>	8720/3: malignant melanoma, nos	1405
<input type="checkbox"/>	9500/3: neuroblastoma, nos	1333
<input type="checkbox"/>	small cell carcinoma	1195
<input type="checkbox"/>	cell renal cell carcinoma	1095
<input type="checkbox"/>	se large b-cell lymphoma, nos	1044
<input type="checkbox"/>	ular lymphoma, nos	867

UID	SERIESID	PMID	ICDMORPHOLOGYCODE	ICDTOPOGRAPHYCODE
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



ARRAYMAP (META) DATA

“PIPELINE”

NCBI GEO - Accession Display | GSE102468

Array comparative genomic hybridization data from 313 CLL specimens to study immune activation

Overall design: Two condition experiment, test CLL specimens vs. reference human genome DNA equivalent of normal tissue and normal lymphocytes

Contributor(s): Houlbronn J, Thielke A, Thoma K, von K, et al. Genomic imbalance defines three prognostic groups for risk stratification of patients with chronic lymphocytic leukemia. *Leukemia* 2014 Apr;28(4):820-6. PMID: 24474179

Submission date: Sep 12, 2012
Last updated date: Sep 19, 2013

informa
informa
HealthLine

ORIGINAL ARTICLE RESEARCH

Genomic imbalance defines three prognostic groups for risk stratification of patients with chronic lymphocytic leukemia

Jane Houlbronn¹, Anna Gurtajski¹, Veronika Thielke¹, Kay-Ju von K¹, Gerd Mandlhardt¹, Tania Zdzienicka¹, Gerd Hoelzl¹, Ingrid Chesi¹, Ulfert Pfeifer¹, Anthony Meyer¹, Jennifer Braun¹, Karsten Kuhlmann¹, A. K. C. Chaganti²

¹Cancer Genetics, Inc., Juelicherhof 44, 04109 Jena, Germany; ²Department of Molecular Biology and Genetics, Harvard Medical School, Boston, MA, USA

Abstract
Array comparative genomic hybridization (aCGH) has not been fully exploited in prognostic settings in chronic lymphocytic leukemia (CLL). Genomic imbalance was assessed in 388 CLL specimens using a targeted array. Based on 28 alterations, a prognostic model of 288 treatment-naïve specimens was established into a group with poor outcome (PO) and a group with intermediate outcome (IO). The IO group was further divided into two subgroups (IO1 and IO2) based on the presence of a specific set of alterations (IO1: 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100).

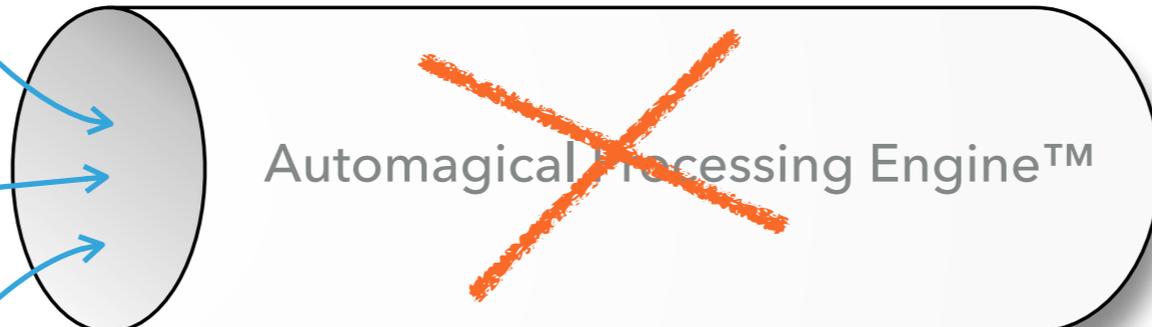
ArrayExpress

E-MTAB-988 - Comparative genomic hybridization by array of human peripheral T-cell lymphoma clinical samples to study their genomic aberration profile

Organization: Houlbronn J, Thielke A, Thoma K, von K, et al. Genomic imbalance defines three prognostic groups for risk stratification of patients with chronic lymphocytic leukemia. *Leukemia* 2014 Apr;28(4):820-6. PMID: 24474179

Platform: Affymetrix 10K

File: Investigation description: GSE102468.tar.gz (1.4 GB)
Raw data: GSE102468.tar.gz (1.4 GB)
Raw data: GSE102468.tar.gz (1.4 GB)



arrayMap 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 easy point for meta-analysis and systems level data integration of high-resolution oncogenic CNA data. The current data reflects:

- 65042 genomic copy number arrays
- 956 experimental series
- 333 array platforms
- 253 ICD-O cancer entities
- 716 publications (PubMed entries)

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 Progenix 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	5791 samples	[?]
BREAST CANCER	8504 samples	[?]
COLORECTAL CANCER	3470 samples	[?]
PROSTATE CANCER	1366 samples	[?]
STOMACH CANCER	1457 samples	[?]

2016-04-11: Sorting cancer subset tables
2015-03-23: SIB Profile 2015
More news ...

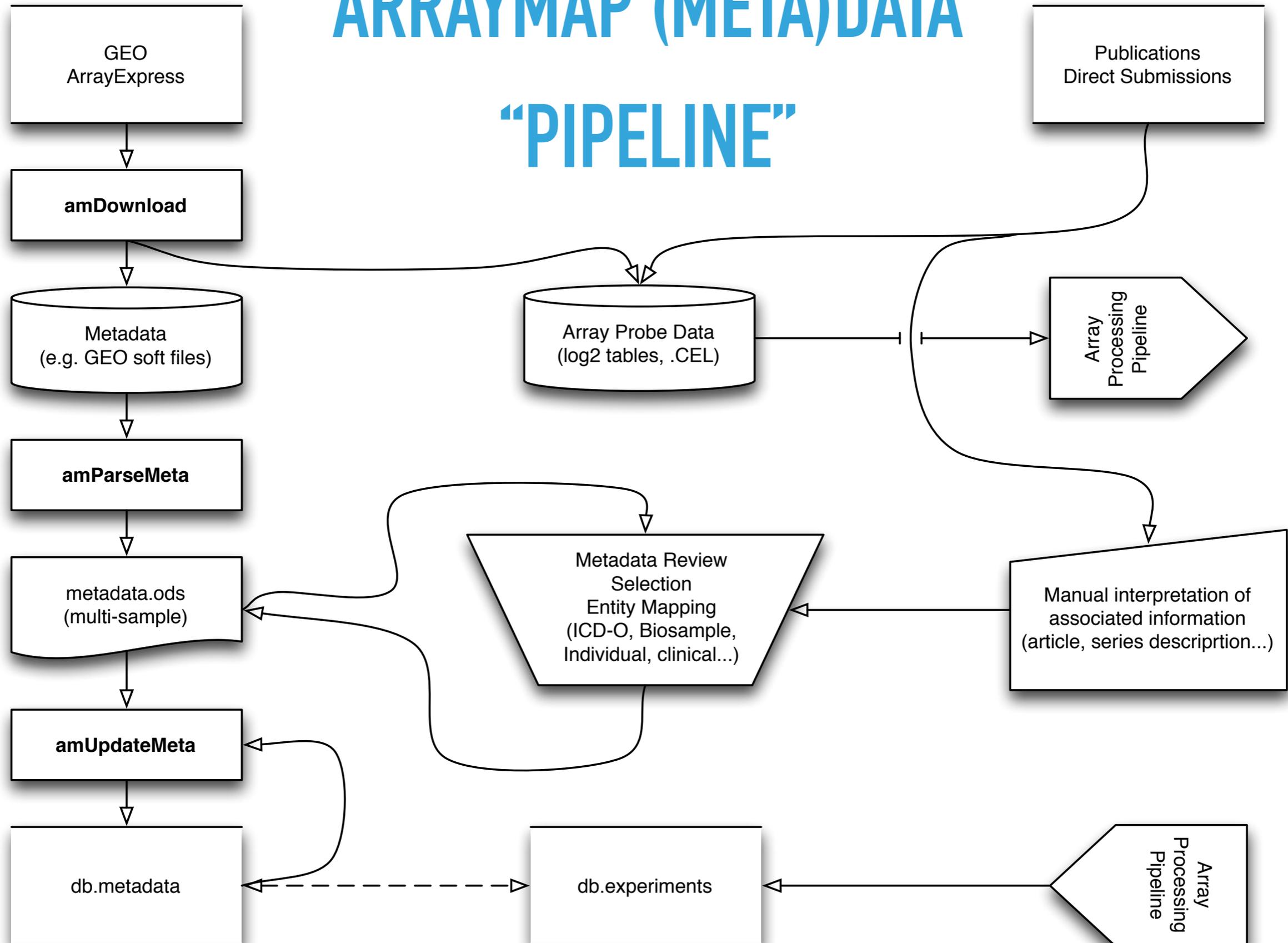
arrayMap ICD Morphologies

64485 samples from arraymap have an associated "ICDMORPHOLOGYCODE" label.
31902 samples from progenix have an associated "ICDMORPHOLOGYCODE" label.
400 subsets of type ICDMORPHOLOGYCODE will be parsed.

Subsets	arraymap	progenix
00000: not classified in ICD-O 3 (e.g. non-neoplastic or benign)	8814	370
80000: neoplasm, malignant	11	1
80100: epithelial tumor, benign	10	11
80102: carcinoma in situ, nos	20	11
80103: carcinoma, nos	1430	396
80103: large cell carcinoma, nos	48	54
80103: large cell neuroendocrine carcinoma	48	80
80200: carcinoma, undifferentiated type, nos	3	2
80210: carcinoma, sarcomatous type, nos	4	41
80220: pleomorphic carcinoma	1	1
80310: giant cell carcinoma	4	3
80320: spindle cell carcinoma	1	1
80330: sarcomatous carcinoma	2	7
80410: small cell carcinoma, nos	132	148
80450: non-small cell carcinoma	1195	184
80500: papillary carcinoma, nos	16	1
80500: squamous cell papilloma	1	130
80701: perianth squamous epithelium, nos	45	162
80702: squamous cell carcinoma in situ, nos	65	16
80703: squamous cell carcinoma, nos	2443	2087
80710: squamous cell carcinoma, keratocystic, nos	11	12
80710: squamous cell carcinoma, acantholytic	1	2
80710: squamous intraepithelial neoplasia, grade II	136	22
80800: undifferentiated neuroepithelial carcinoma	12	203
80900: basal cell carcinoma, nos	28	15
81000: transitional cell carcinoma in situ	10	10
81000: transitional cell carcinoma, nos	310	423
81001: urothelial papilloma, nos	184	39
81002: papillary transitional cell carcinoma, non-invasive	1	56
81002: papillary transitional cell carcinoma	2	6
81400: adenoma, nos	385	361
81401: atypical adenoma	88	88
81402: adenocarcinoma in situ	11	11
81403: adenocarcinoma, nos	8409	3248
81440: adenocarcinoma, intestinal type	167	206
81450: carcinoma, diffuse type	7	36
81460: glandular intraepithelial neoplasia, grade II	1	18
81500: leiomyoma, nos	1	8
81500: leiomyosarcoma, nos	1	18
81500: leiomyosarcoma, nos	1	18

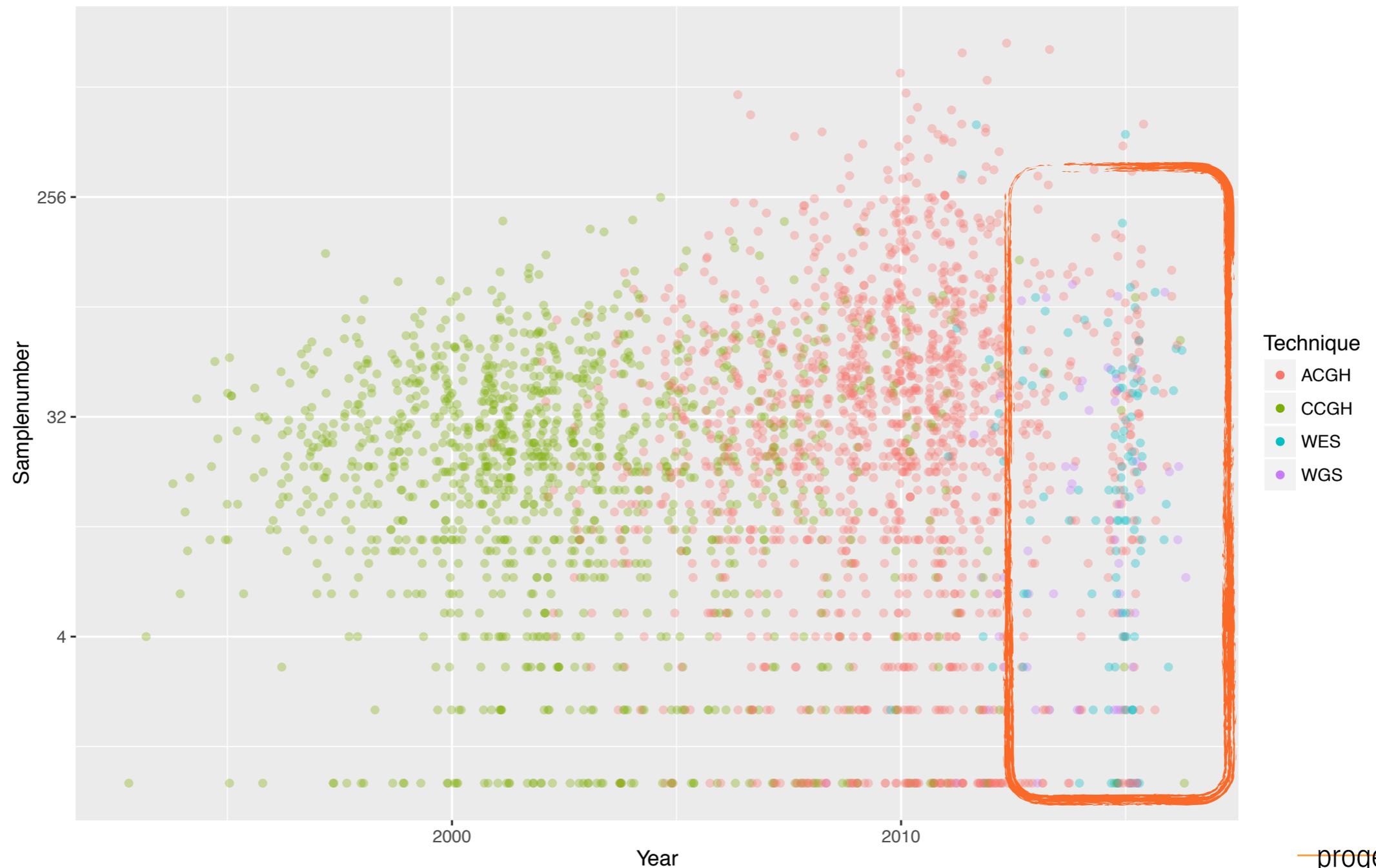
BIOCURATION

ARRAYMAP (META) DATA “PIPELINE”



SHIFT TO SEQUENCING BASED TECHNIQUES LEADS TO SEVERELY LIMITED DATA ACCESSIBILITY

Cancer Samples per Publication for Different Techniques
[129417 samples from 2747 publications]





Genomics API

Learn how the Genomics API is advancing information sharing for DNA data providers and consumers on a global scale and get involved in further development.

→ [Genomics API](#)

Our Work

The diverse members of the Global Alliance are working together to create interoperable approaches to catalyze projects that will help unlock the great potential of genomic data. Our four [Working Groups](#) advance [Initiatives](#) that develop key [Work Products](#).

GA4GH API promotes sharing



ATTTATCTGCTCTCGTTG
GAAGTACAAAATTCATTAAT
GCTATGCACAAAATCTGTAG
CTAGTGTCCCATCTATTT

The mission of the Global Alliance for Genomics and Health is to accelerate progress in human health by helping to establish a **common framework** of harmonized approaches to enable **effective and responsible** sharing of **genomic and clinical data**, and by catalyzing data sharing projects that drive and demonstrate the value of data sharing.

GA4GH DATA ANNOTATION PRINCIPLES

▶ Ontologies

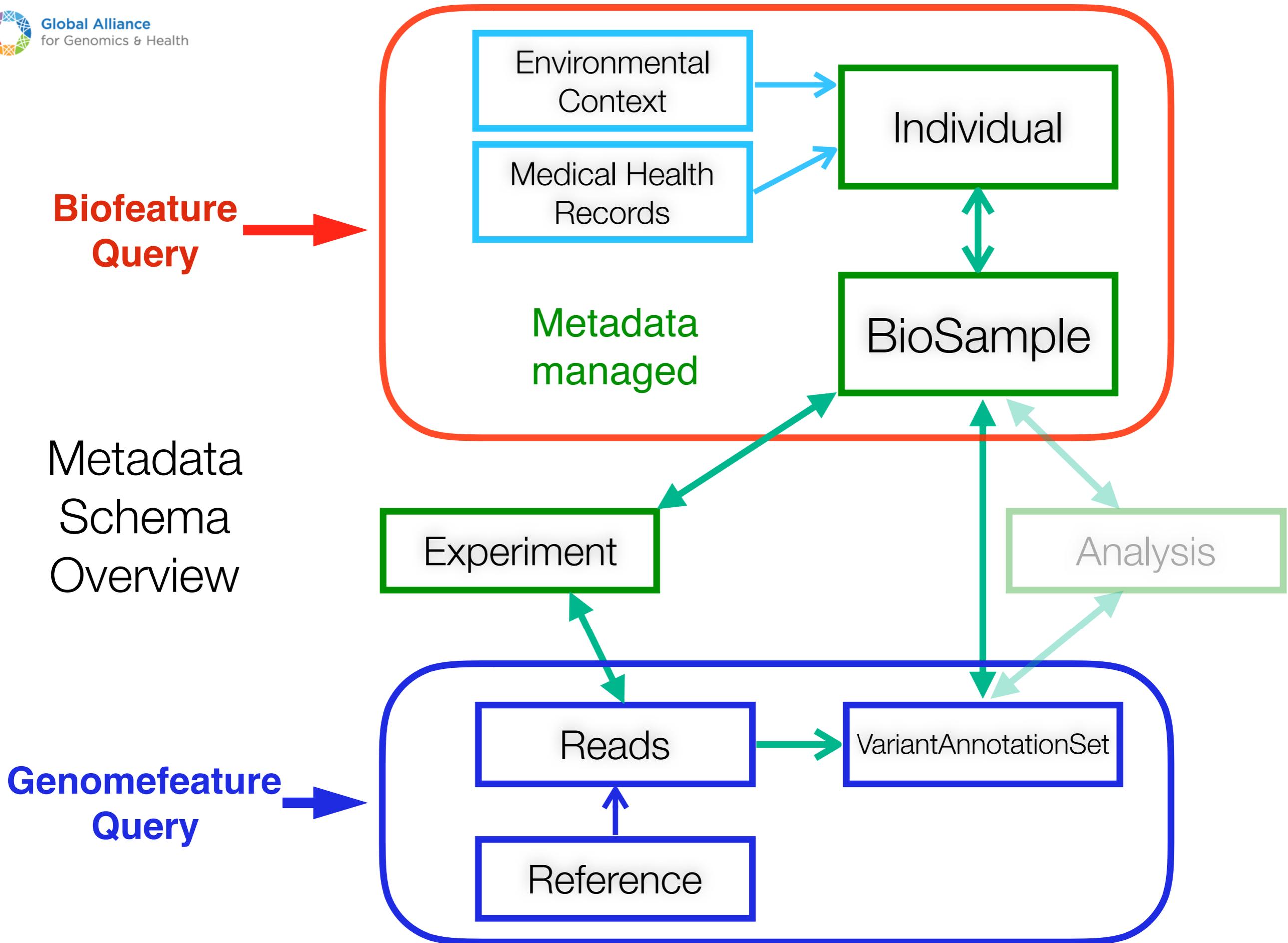
- ▶ limited use of named attributes
- ▶ “OntologyTerm” object type as core for biological and experimental features

▶ External standards

- ▶ ISO8601 time formats

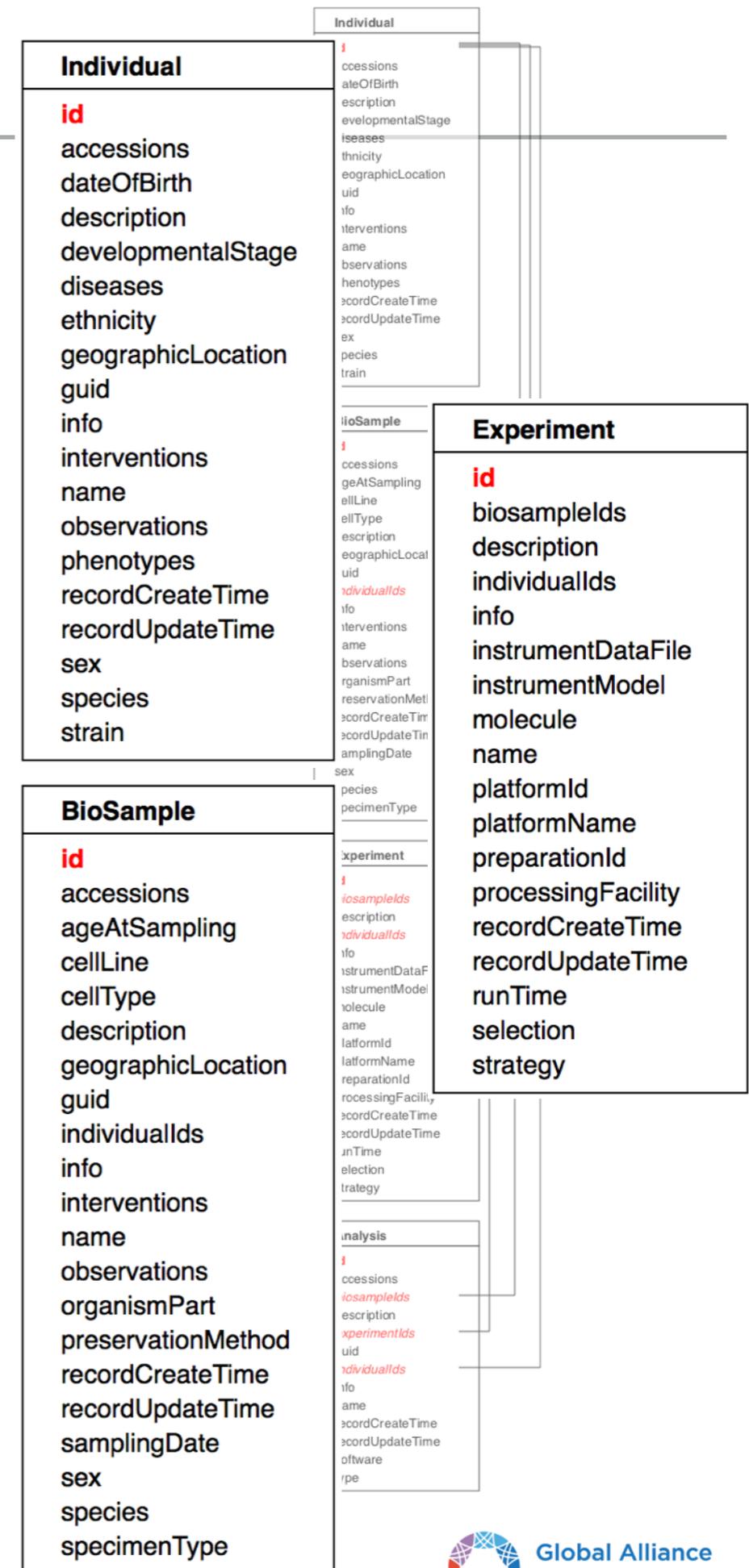
▶ GA4GH managed standards

- ▶ maintenance of core sequence file and data formats (VCF, BAM ...) through affiliate partners/members of the Data Working Group



METADATA SCHEMA – “BIODATA” OBJECTS

- ▶ *Individual* (i.e., basic biological entity) and *BioSample* (e.g. micro dissected part of a tissue biopsy, environmental sample) are the basic “Bioobjects”
- ▶ *Experiment* describes the technical procedures used in the analysis of (an aliquot) of the *BioSample*
- ▶ *Analysis* may be used to store “interpreted” results of an experiment
- ▶ *VariantAnnotationSet* and other low-level analyses reference *Individual*, *Biosample*, *Experiment* for feature inheritance



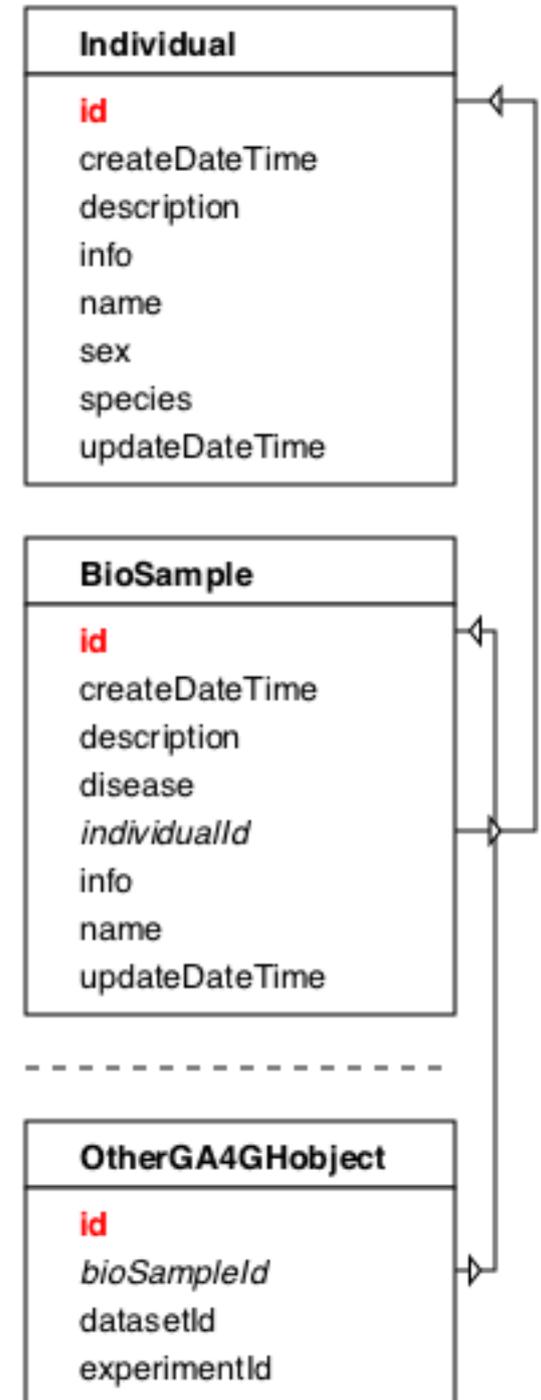
ONTOLOGYTERM IN THE GA4GH SCHEMA - CURRENT STATUS

```
/**
An ontology term describing an attribute. (e.g. the phenotype attribute
'polydactyly' from HPO)
*/
record OntologyTerm {
  /**
  Ontology source identifier - the identifier, a CURIE (preferred) or PURL for an
  ontology source e.g. http://purl.obolibrary.org/obo/hp.obo
  It differs from the standard GA4GH schema's "id" in that it is a URI pointing to
  an information resource outside of the scope of the schema or its implementation.
  */
  string id;

  /* Ontology term - the representation the id is pointing to.*/
  union { null, string } term = null;

  /**
  Ontology source name - the name of ontology from which the term is obtained
  e.g. 'Human Phenotype Ontology'
  */
  union { null, string } sourceName = null;

  /**
  Ontology source version - the version of the ontology from which the OntologyTerm
  is obtained; e.g. 2.6.1. There is no standard for ontology versioning and some
  frequently released ontologies may use a datestamp, or build number.
  */
  union { null, string } sourceVersion = null;
}
```



BIOFEATURE OBJECTS AS ONTOLOGY WRAPPERS?

```
"bioFeatures": [  
  {  
    "description": "squamous cell carcinoma, base of tongue, stage 2",  
    "ageAtObservation": "P56Y6M",  
    "timeOfObservation": "2015-03-24T15:23:00Z",  
    "updateDateTime": "2016-04-14T09:02:00Z",  
    "ontologyTerms": [  
      {  
        "ontologyId": "http://purl.obolibrary.org/obo/DOID_0050865",  
        "term": "tongue squamous cell carcinoma",  
        "sourceName": "disease_ontology",  
        "sourceVersion": "2016-01-25"  
      },  
      {  
        "ontologyId": "http://purl.obolibrary.org/obo/UBERON_0006919",  
        "term": "tongue squamous epithelium",  
        "sourceName": "Uberon multi-species anatomy ontology",  
        "sourceVersion": "2016-01-25"  
      },  
      {  
        "ontologyId": "http://purl.obolibrary.org/obo/UBERON_0010033",  
        "term": "posterior part of tongue",  
        "sourceName": "Uberon multi-species anatomy ontology",  
        "sourceVersion": "2016-01-25"  
      },  
    ],  
  },  
]
```

We're sorry but something has gone wrong. We have been notified of this error.

ONTOLOGIES YOU CAN TRUST?

[NPEx](#) [NLMC](#) [THIS](#) [X-Lab](#)



SNOMED CT Browser

UK Clinical Edition
April 2016

[Concept Search](#)

[About SNOMED-CT](#)

You have searched for: medulloblastoma

[Go back to search results](#)

- [Disorder of brain](#)
- [Glioma \(disorder\)](#)
- [Neuroendocrine tumor \(disorder\)](#)
- [Primary malignant neoplasm](#)

Name: Medulloblastoma [See more descriptions.](#)
Concept ID: 443333004
Read Code: XUjPT
ICD-10 Codes: C716

- [Classic medulloblastoma](#)
- [Medulloblastoma of cerebellum \(disorder\)](#)

- [Medulloblastoma of cerebellum \(disorder\)](#)
- [Classic medulloblastoma](#)

ONTOLOGIES YOU CAN TRUST?

The screenshot displays the SNOMED CT Browser interface. At the top, there are navigation links: NPEx, NLMC, THIS, and X-Lab. The main header area includes the NPE logo and the text 'SNOMED CT Browser UK Clinical Edition April 2016'. Below this, there are links for 'Concept Search' and 'About SNOMED-CT'. The search results section shows a search for 'medulloblastoma'. A sidebar on the left lists categories: 'Disorder of brain', 'Glioma (disorder)', 'Neuroendocrine tumor (disorder)', and 'Primary malignant neoplasm of brain'. The main content area shows the search results for 'Neoplastic disease'. A blue box highlights the details for 'Neuroendocrine tumor (disorder)'. Below this, a list of related terms is shown, including 'Apudoma (disorder)', 'Benign neuroendocrine tumor (disorder)', 'Carcinoid tumour', 'Extra-adrenal paraganglioma (disorder)', 'Gastrointestinal hormone-secreting endocrine tumor (disorder)', 'Malignant neuroendocrine tumor (disorder)', 'Malignant pheochromocytoma (disorder)', 'Medulloblastoma', 'Neuroblastoma', 'Neuroendocrine neoplasm of larynx', 'Neuroendocrine neoplasm of lung (disorder)', 'Olfactory neuroblastoma', 'Paraganglioma (disorder)', 'Primitive neuroectodermal tumour', 'Retinoblastoma', and 'Somatostatinoma (disorder)'. The bottom of the page shows a partial view of the same list of terms.

NPEx NLMC THIS X-Lab

You have searched for: medulloblastoma
Go back to search results

NPEx NLMC THIS X-Lab

NPE

SNOMED CT Browser
UK Clinical Edition
April 2016

Concept Search
About SNOMED-CT

Disorder of brain
Glioma (disorder)
Neuroendocrine tumor (disorder)
Primary malignant neoplasm of brain

Name: Medulloblastoma
Concept ID: 255046005
Read Codes: X78dr
ICD-10 Codes: D357 C755 D355 C759 C751 D440 C754 D351 D446 D448 C798 D356 D358 C753 C752 D352 C750 D443 D359 D354 C758 D445 D093 D353 D350 D442 D444 D441 D449 D447

Neoplastic disease

Name: Neuroendocrine tumor (disorder) See more descriptions.
Concept ID: 255046005
Read Code: X78dr
ICD-10 Codes: D357 C755 D355 C759 C751 D440 C754 D351 D446 D448 C798 D356 D358 C753 C752 D352 C750 D443 D359 D354 C758 D445 D093 D353 D350 D442 D444 D441 D449 D447

- + Apudoma (disorder)
- + Benign neuroendocrine tumor (disorder)
- + Carcinoid tumour
- + Extra-adrenal paraganglioma (disorder)
- + Gastrointestinal hormone-secreting endocrine tumor (disorder)
- + Malignant neuroendocrine tumor (disorder)
- Malignant pheochromocytoma (disorder)
- + Medulloblastoma
- + Neuroblastoma
- Neuroendocrine neoplasm of larynx
- + Neuroendocrine neoplasm of lung (disorder)
- Olfactory neuroblastoma
- + Paraganglioma (disorder)
- + Primitive neuroectodermal tumour
- Retinoblastoma
- Somatostatinoma (disorder)

Somatostatinoma (disorder)
Retinoblastoma
+ Primitive neuroectodermal tumour
+ Paraganglioma (disorder)
Olfactory neuroblastoma
+ Neuroendocrine neoplasm of lung (disorder)

ONTOLOGIES YOU CAN TRUST?

SNOMED CT Browser
UK Clinical Edition April 2016

Name: Medulloblastoma
Concept ID: 44333004
Read Code: X78dr
ICD-10 Codes: C716

Name: Glioma (disorder) See more descriptions.
Concept ID: 39384001
Read Code: .B1M5
ICD-10 Codes: C719

- + Anaplastic astrocytoma of central nervous system (disorder)
- + Dysembryoplastic neuroepithelial tumor (disorder)
- + Ependymoma
- + Glioblastoma multiforme (disorder)
- + Glioma of spinal cord (disorder)
- + Intracranial glioma (disorder)
- + Malignant glioma of central nervous system (disorder)
- + Malignant glioma of eye
- + Medulloblastoma
- + Mixed glioma
- + Oligodendroglioma
- + Optic glioma
- + Primary malignant astrocytoma of central nervous system (disorder)
- + Primitive neuroectodermal tumour

Medulloblastomas are

- embryonal
- neuroepithelial
- brain
- neoplasias

... BUT NOT GLIOMAS & MOST CERTAINLY NOT A NEUROENDOCRINE TUMOURS

ONTOLOGIES ARE RARELY USED IN CASE REPORTING

- ▶ Medical practice relies on established, slow moving classification systems.
- ▶ Medical diagnoses consist of an abundance of observations and classification items.
- ▶ We do not have (never will?) enough ontology concepts for detailed disease descriptions (Where to stop?)
- ▶ Relationships may help - but how to do them uniformly?

ONTOLOGIES ARE RARELY USED IN CASE REPORTING

- ▶ Medical practice relies on established, slow-moving classification systems.
- ▶ Medical diagnoses consist of an abundance of observations and classification items.
- ▶ We do not have (or ever will?) enough ontology concepts for detailed disease descriptions (Where to stop?)
- ▶ Relationships may help - but how to do them uniformly?

THE GOLD STANDARD FOR A MEDICAL DIAGNOSIS IS STILL WELL WRITTEN PROSE.

THE GOLD STANDARD

CORE PROBLEMS AND CONCEPTS TO BE ADDRESSED

- ▶ suitability of **ontologies as core of metadata** features for federated data mining
- ▶ mapping ontologies: **WHO** and **HOW**
- ▶ identification of essential non-OT attributes and stable definition using internationally accepted standards
- ▶ development of a strategy for implementation of **ontology based data annotations for reference data resources**, e.g. Elixir, EBI, SIB ...

DRIVING GA4GH METADATA SCHEMA



▶ arrayMap for GA4GH

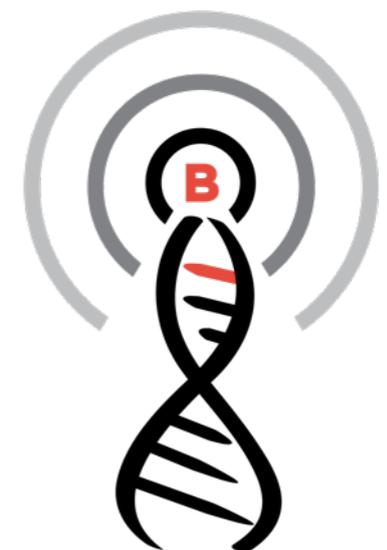
- metadata schema development through implementation of arrayMap resource data
- testing of OntologyTerm object for covering biodata
- implementation using  ontology services



DRIVING BEACON DEVELOPMENT

▶ Beacon⁺

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



DRIVING GA4GH METADATA SCHEMA



▶ arrayMap for GA4GH

- metadata schema development through implementation of arrayMap resource data
- testing of OntologyTerm object for covering biodata
- implementation using  ontology services



DRIVING BEACON DEVELOPMENT



▶ Beacon⁺

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



MAPPING ARRAYMAP TO GA4GH TRANSITIONAL

```
[
  {
    "individuals" : [
      {
        "createDateTime" : "2015-01-01T12:00:00Z",
        "id" : "AM_IND__18769486_MB0262",
        "sex" : {
          "ontologyId" : "http://purl.obolibrary.org/obo/PATO_0020001",
          "sourceName" : "Ontology: PATO (Phenotypic quality)"
        },
        "species" : {
          "ontologyId" : "http://purl.obolibrary.org/obo/NCBITaxon_9606",
          "sourceName" : "NCBITaxon Ontology"
        },
        "updateDateTime" : "2016-04-13T18:51:01Z"
      }
    ]
  },
  {
    "biosamples" : [
      {
        "ageAtSampling" : "P14Y",
        "createTime" : "2015-01-01T12:00:00Z",
        "diagnosis" : {
          "description" : "medulloblastoma [classic]",
          "ontologyTerms" : [
            {
              "ontologyId" : "http://purl.bioontology.org/ontology/SNMI/M-94703",
              "sourceName" : "Systematized Nomenclature of Medicine, International Version",
              "term" : "Medulloblastoma, NOS"
            }
          ]
        }
      }
    ]
  }
]
```

Beacon ArrayMap

First Prototype of a [Beacon v0.2](#) implementation for [ArrayMap](#).

See [documentation](#) and [open questions](#).

Chromosome

Position

Reference

Dataset

Variant Class

<http://beacon-arraymap.vital-it.ch/v0.2/query?chromosome=11&position=34439881&reference=GRCh38&dataset=8070/3&variantClass=DEL>

<http://beacon-arraymap.vital-it.ch/info>

<http://beacon-arraymap.vital-it.ch/v0.3/query?referenceName=11&start=34439881&assemblyId=GRCh38&datasetIds=8070/3&variantClass=DEL>

```
{
  "query": {
    "referenceName": "11",
    "start": "34439881",
    "assemblyId": "GRCh38",
    "datasetIds": "8070/3"
  },
  "response": {
    "exists": "overlap",
    "info": "ok",
    "error": null,
    "observed": 57,
    "NOT_BEACON_ARRAYMAP_DEBUG_INFO": {
      "matchedSegments": [
        {
          "matchedSampleUID": "GSM675751",
          "matchedDataSet": "8070/3",
          "matchedSegment": {
            "SEGSTART": 232718,
            "SEGSOURCE": "aCGH",
            "CHRO": "11",
            "SEGSIZE": 134462435,
            "SEGVALUE": -0.2637,
            "SEGTYPE": -1,
            "SEGSTOP": 134695153
          }
        }
      ]
    },
    ...
  },
  "NOT_BEACON_totalInDataSet": 0
}
```

Heinz Stockinger, Séverine Duvaud & SIB Technology Group

BAUDISGROUP @ UZH

**NI AI
MICHAEL BAUDIS
(HAOYANG CAI)
PAULA CARRIO CORDO
LINDA GROB
SAUMYA GUPTA
(NITIN KUMAR)
ALESSIO MILANESE**

GA4GH DWG

**ANTHONY BROOKES
MARK DIEKHANS
MELISSA HAENDEL
SARAH HUNT
STEPHEN KEENAN
SUZY LEWIS
DAVID LLOYD
MICHAEL MILLER
HELEN PARKINSON
ELEANOR STANLEY
DAVID STEINBERG**

SIB

**HEINZ STOCKINGER
SÉVERINE DUVAUD
VASSILIOS IOANNIDIS
DANIEL TEIXEIRA**

ELIXIR & CRG

**JORDI RAMBLA DE ARGILA
SABELA DE LA TORRE PERNAS
SUSANNA REPO**



RANDOM LINKS

- ▶ arraymap.org
- ▶ beacon-arraymap.vital-it.ch
- ▶ github.com/ga4gh/schemas/tree/metadata
- ▶ www.progenetix.org/publications/
- ▶ wiki.progenetix.org
- ▶ ga4ghdata.org
- ▶ ga4ghdata.org/Sites/github/ga4gh/biosample/html/api/biodata.html