Bioinformatics and data exchange for genomics in an international context

CNV Databases :: ELIXIR Beacon :: GA4GH :: DataFormats :: SchemaBlocks :: SPHN

3rd St. Gallen Radiation Oncology Informatics Meeting
Michael Baudis | St. Gallen | 2019-11-08

University of Zurich[UZH] elixir SIB Global Alliance for Genomics & Health SPHN

1992	Heidelberg	Student of medicine | doctoral thesis in molecular cytogenetics @ DKFZ (Peter Lichter) | resident in clinical hematology/oncology | data, clinical studies & cancer systematics
2001	Stanford	Post-doc in hemato-pathology (Michael Cleary) | molecular mechanisms of leukemogenesis | transgenic models | expression arrays | systematic cancer genome data collection | *Progenetix* website
2003	Gainesville	Assistant professor in paediatric haematology | molecular mechanisms of leukemogenesis | focus on bioinformatics for cancer genome data analysis
2006	Aachen	Research group leader in genetics | genomic array analysis for germline alterations | descriptive analysis of copy number aberration patterns in cancer entities
2007	Zürich	Professor of bioinformatics @ IMLS (2015) | systematic assembly of oncogenomic data | databases and software tools | patterns in cancer genomes | *arraymap* online resource | GA4GH | SPHN

Michael @ SIB, GA4GH, ELIXIR & SPHN

- member GA4GH since 2014
  - co-lead Discovery WS (2017->)
- ELIXIR Beacon project
  - previous co-chair; now responsible GA4GH liaison
- ELIXIR h-CNV project
- Swiss Personalized Health Network (SPHN)
  - SPHN project champion @ GA4GH
- Swiss Institute of Bioinformatics (SIB) group leader

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

elixir BEACON

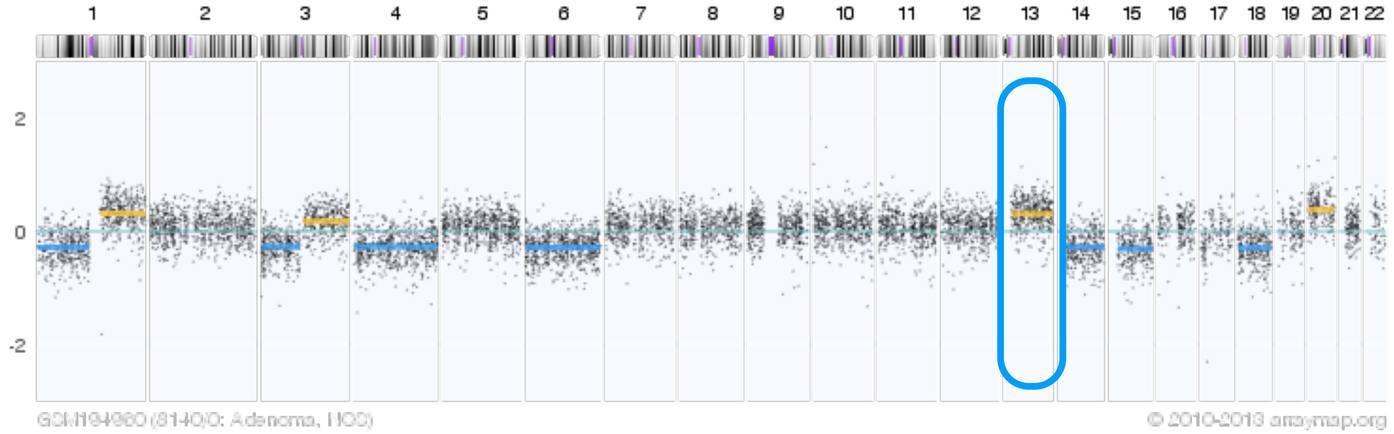
Beacon

elixir

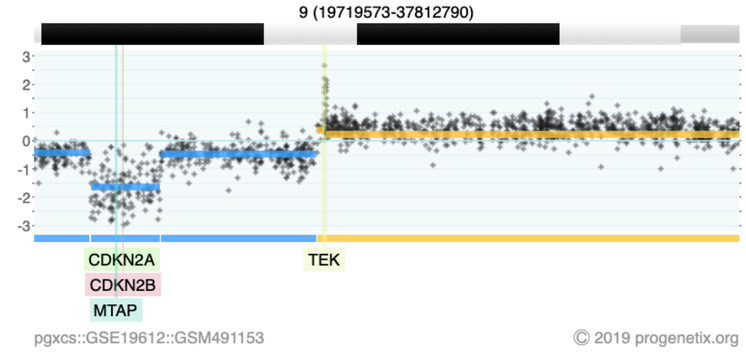
Swiss Personalized Health Network

SIB
Swiss Institute of Bioinformatics

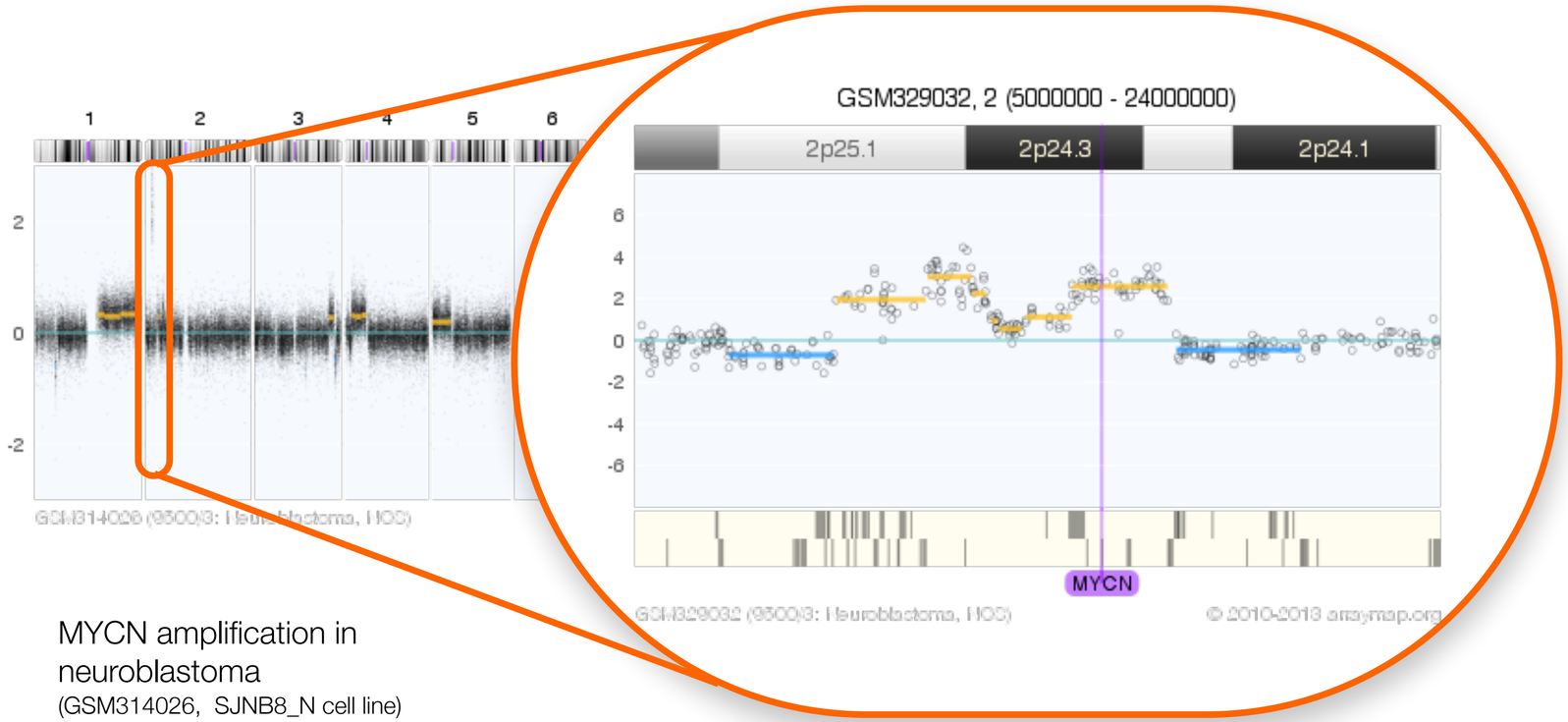
Somatic Copy Number Variations



Gain of chromosome arm 13q in colorectal carcinoma



2-event, homozygous deletion in a Glioblastoma

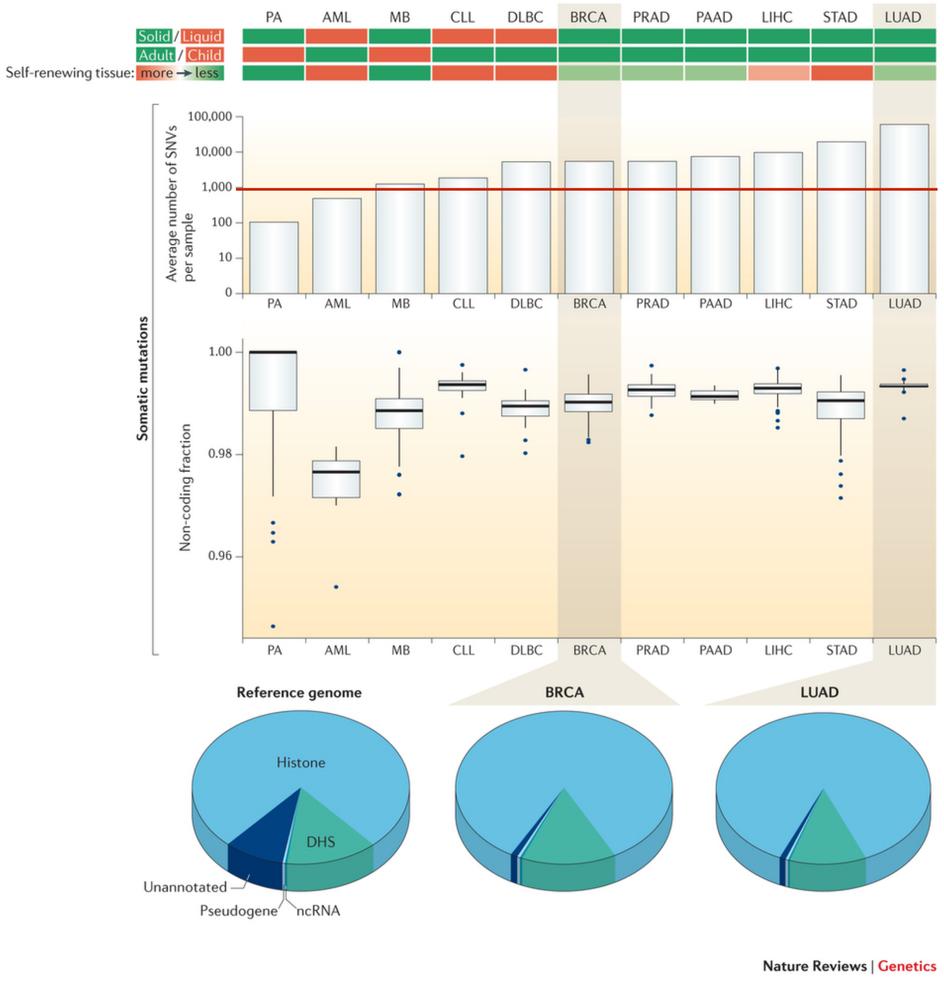


MYCN amplification in neuroblastoma
(GSM314026, SJNB8_N cell line)

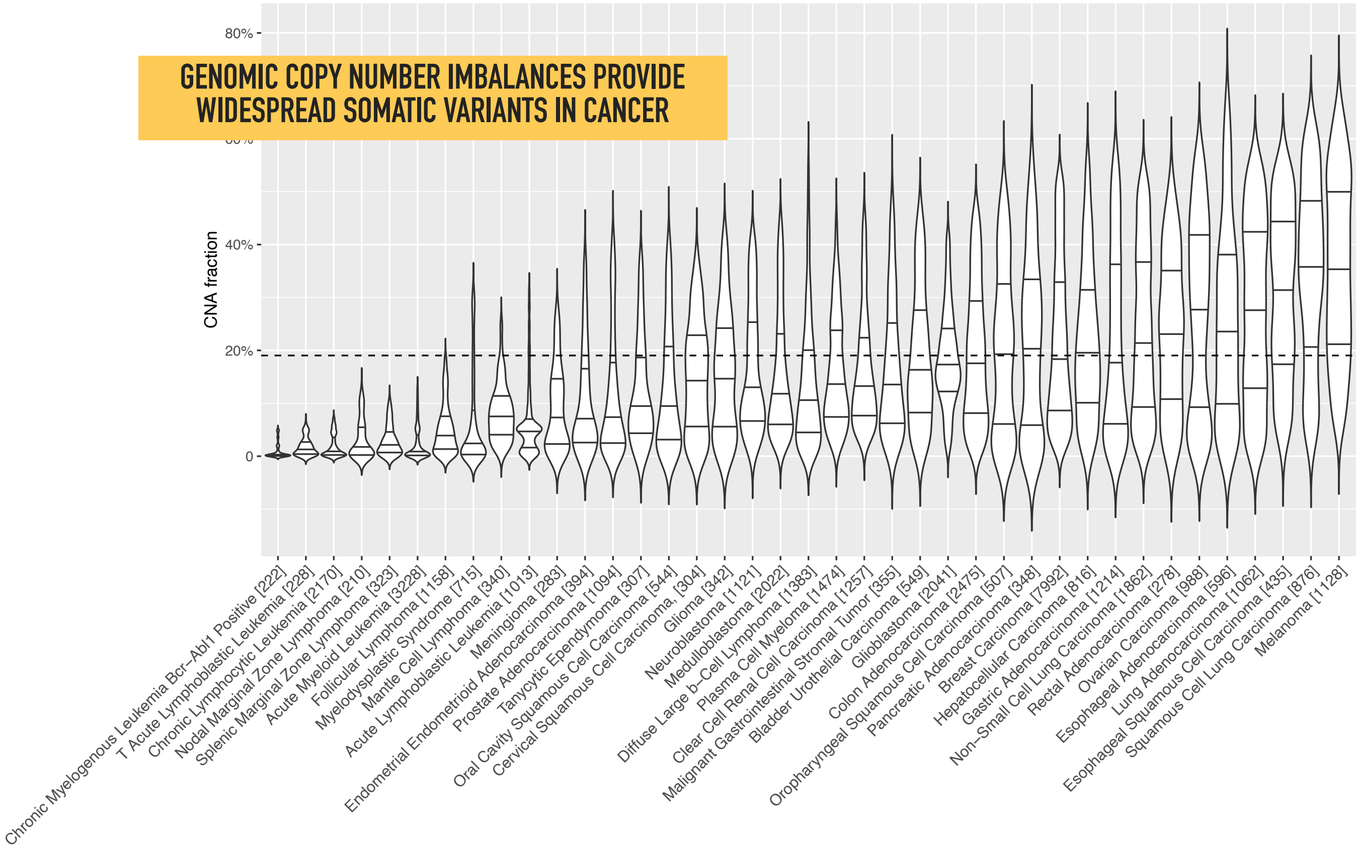
low level/**high level** copy number alterations (CNAs)

arrayMap

Quantifying Somatic Mutations In Cancer



CANCERS SHOW THOUSANDS OF SINGLE NUCLEOTIDE VARIANTS PER SAMPLE, MOSTLY IN NON-CODING REGIONS

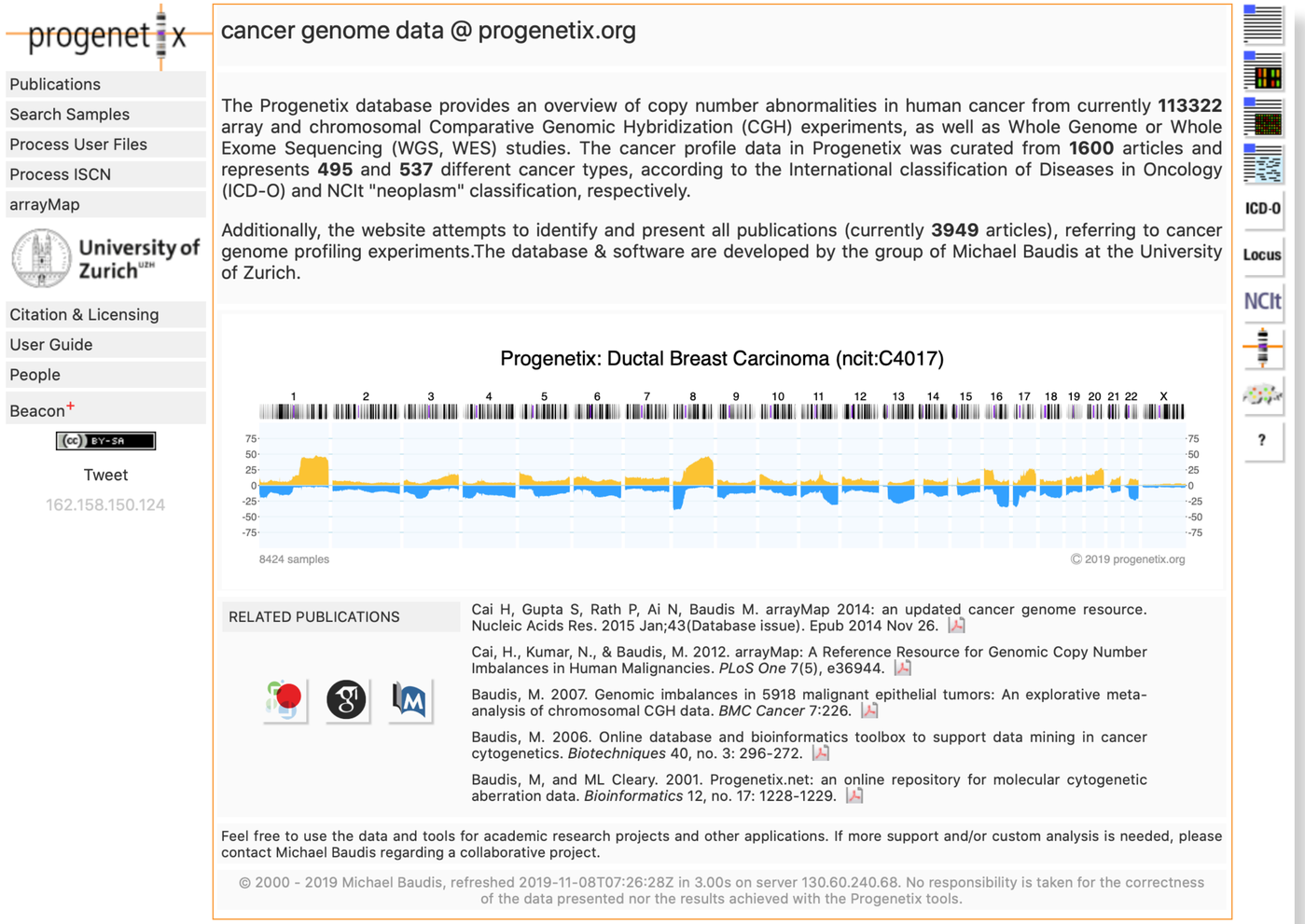


Pan-Cancer Analysis of Whole Genomes (PCAWG) data show widespread mutations in non-coding regions of cancer genomes (Khurana et al., Nat. Rev. Genet. (2016)

On average ~19% of a cancer genome are in an imbalanced state (more/less than 2 alleles); Original data based on 43654 cancer genomes from progenetix.org

Progenetix - Reference Resource for Oncogenomic Profiling Data

- launched in 2001 as progenetix.net with 999 samples (September 2001)
- curated CNV data from chromosomal CGH studies
- now containing >113'000 single sample CNV tracks from 1600 publications
- **aCGH**, cCGH, WES, WGS
- additionally tracking and annotating of publications reporting compatible original data (more than 3200 articles as of 2019)



arrayMap

Accessing Probe-Level Genomic Array Data in Cancer

arrayMap

- Search Samples
- Search Publications
- Progenetix

University of Zurich[UZH]

- Citation & Licensing
- User Guide
- People
- Beacon+

CC BY-SA

Tweet

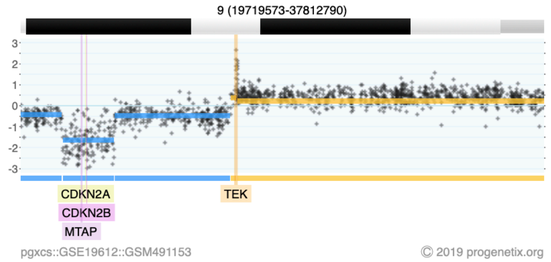
162.158.150.56

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:

- 72724 genomic array profiles
- 898 experimental series
- 257 array platforms
- ICD-O 341 ICD-O cancer entities
- 795 publications (Pubmed entries)



Genomic copy number imbalances on chromosome 9 in a case of Glioblastoma (**GSM491153**), indcating, 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.

RELATED PUBLICATIONS

Cai H, Gupta S, Rath P, Ai N, Baudis M. arrayMap 2014: an updated cancer genome resource. Nucleic Acids Res. 2015 Jan;43(Database issue). Epub 2014 Nov 26.

Cai, H., Kumar, N., & Baudis, M. 2012. arrayMap: A Reference Resource for Genomic Copy Number Imbalances in Human Malignancies. *PLoS One* 7(5), e36944.

Baudis, M. 2007. Genomic imbalances in 5918 malignant epithelial tumors: An explorative meta-analysis of chromosomal CGH data. *BMC Cancer* 7:226.

Baudis, M. 2006. Online database and bioinformatics toolbox to support data mining in cancer cytogenetics. *Biotechniques* 40, no. 3: 296-272.

Baudis, M, and ML Cleary. 2001. Progenetix.net: an online repository for molecular cytogenetic aberration data. *Bioinformatics* 12, no. 17: 1228-1229.

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.

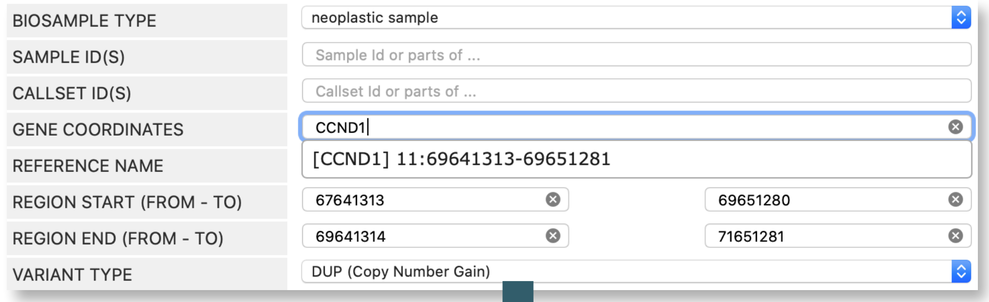
UID	SERIESID	PMID	ICDMORPHOLOGYCODE	ICDTOPOGRAPHYCODE
GSM1000061	GSE36942	23457519	8070/3	C10
GSM1000062	GSE36942	23457519	8070/3	C10
GSM1001316	GSE40777	23571474	8070/3	C53
GSM1001317	GSE40777	23571474	8010/3	C34
GSM1001318	GSE40777	23571474	8070/3	C09
GSM1001319	GSE40777	23571474	8010/3	C34
GSM1002668	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

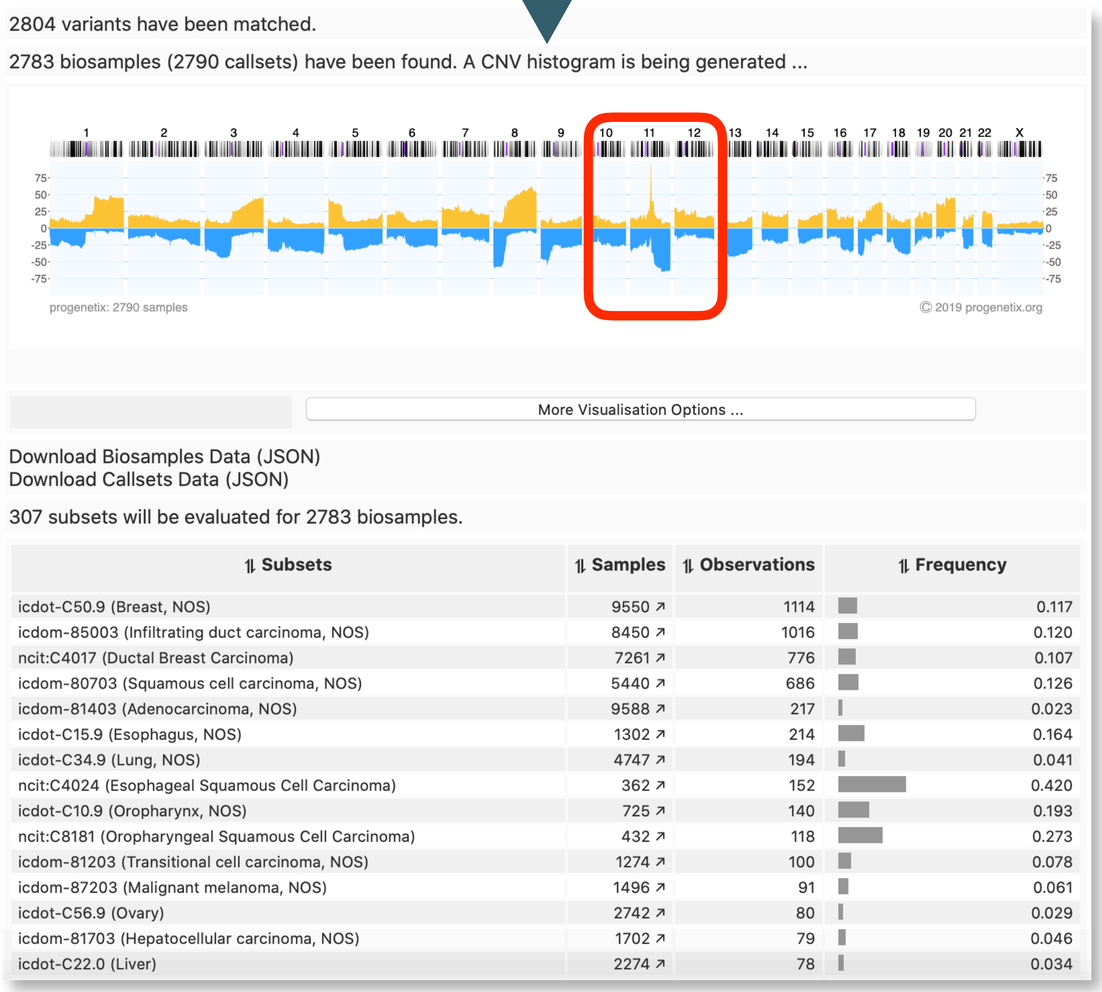
arrayMap

Progenetix - Reference Resource for Oncogenomic Profiling Data

- Progenetix is based on the single-sample CNV tracks of cancer samples from 402/469 (ICD-O/NCIt) diagnostic categories
- typical applications include
  - reference CNV patterns in given diagnoses (e.g. "does my analysis match the diagnosis/ prediction")
  - target gene entity mapping (e.g. "in which tumour type is this gene frequently gained/ lost?")

progenetix

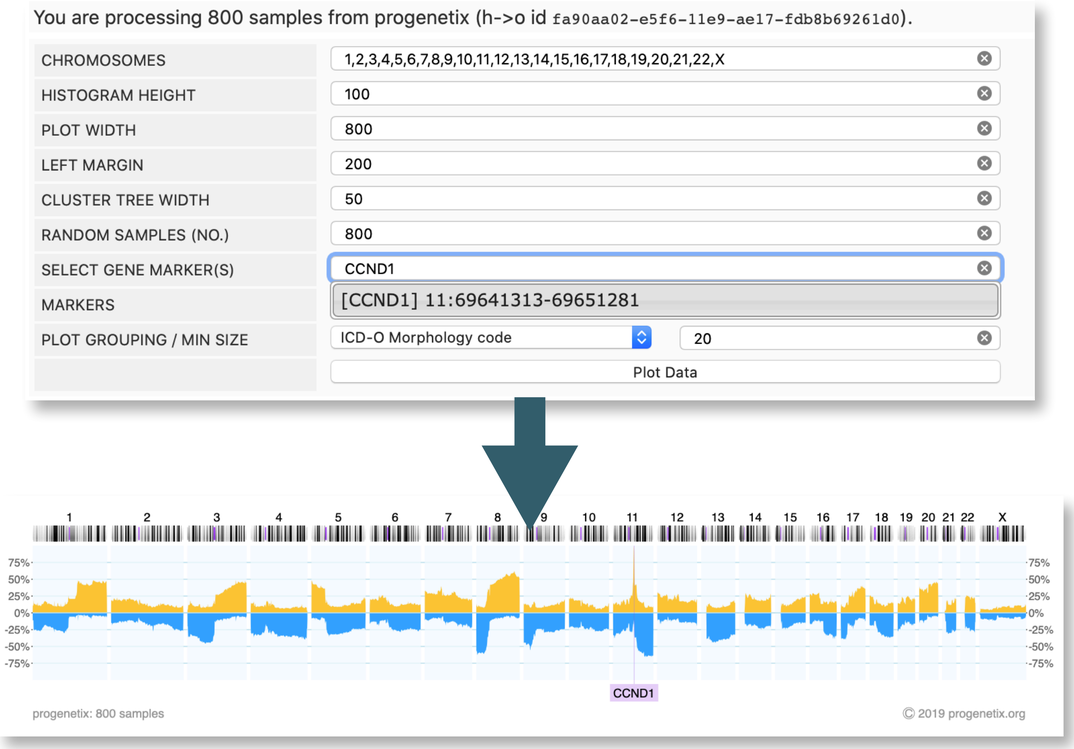


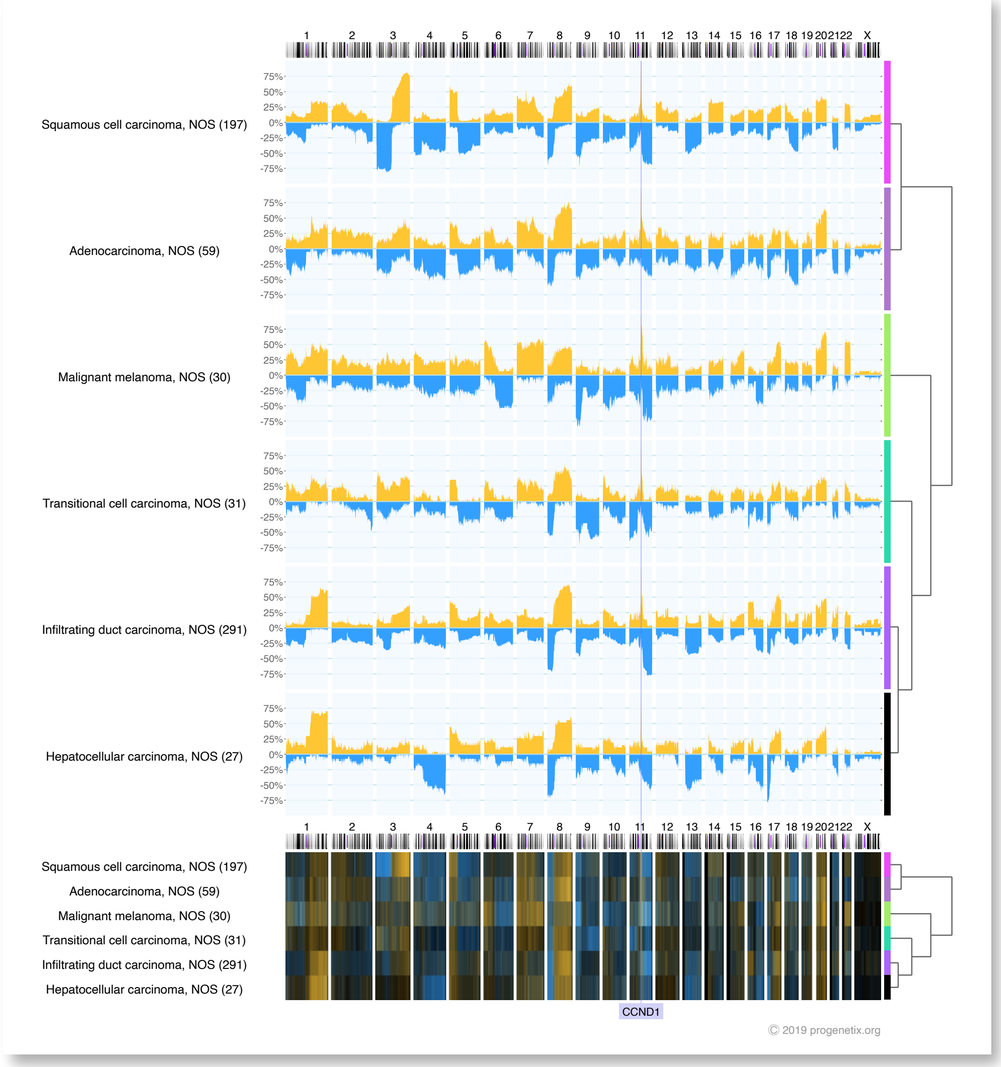


Subsets	Samples	Observations	Frequency
icdot-C50.9 (Breast, NOS)	9550	1114	0.117
icdom-85003 (Infiltrating duct carcinoma, NOS)	8450	1016	0.120
ncit:C4017 (Ductal Breast Carcinoma)	7261	776	0.107
icdom-80703 (Squamous cell carcinoma, NOS)	5440	686	0.126
icdom-81403 (Adenocarcinoma, NOS)	9588	217	0.023
icdot-C15.9 (Esophagus, NOS)	1302	214	0.164
icdot-C34.9 (Lung, NOS)	4747	194	0.041
ncit:C4024 (Esophageal Squamous Cell Carcinoma)	362	152	0.420
icdot-C10.9 (Oropharynx, NOS)	725	140	0.193
ncit:C8181 (Oropharyngeal Squamous Cell Carcinoma)	432	118	0.273
icdom-81203 (Transitional cell carcinoma, NOS)	1274	100	0.078
icdom-87203 (Malignant melanoma, NOS)	1496	91	0.061
icdot-C56.9 (Ovary)	2742	80	0.029
icdom-81703 (Hepatocellular carcinoma, NOS)	1702	79	0.046
icdot-C22.0 (Liver)	2274	78	0.034

Progenetix - Reference Resource for Oncogenomic Profiling Data

- Group histogram and heatmap representation of CNV profiles by external labels (disease codes, publications ...)

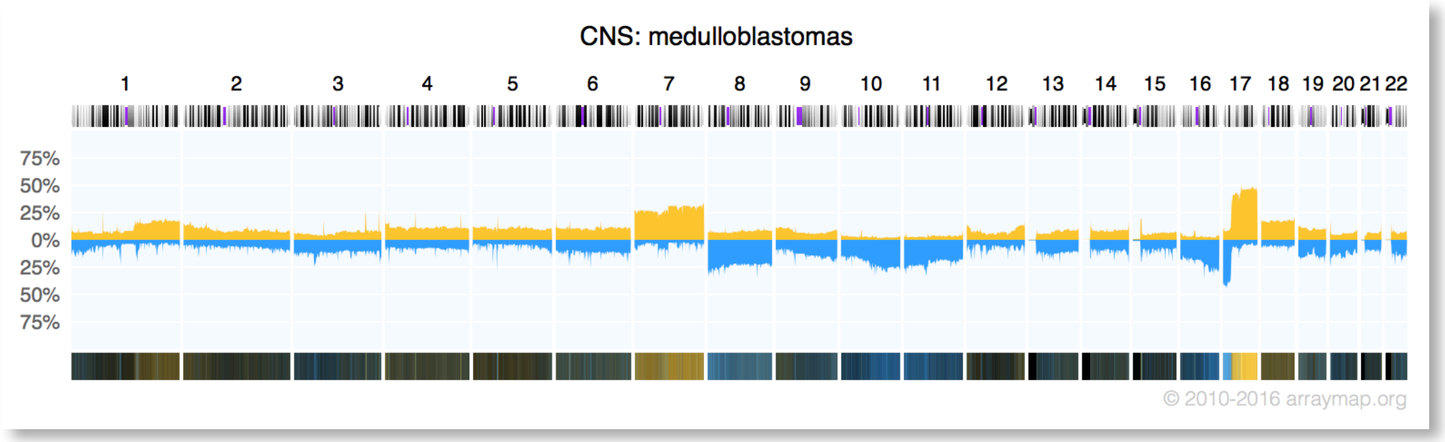


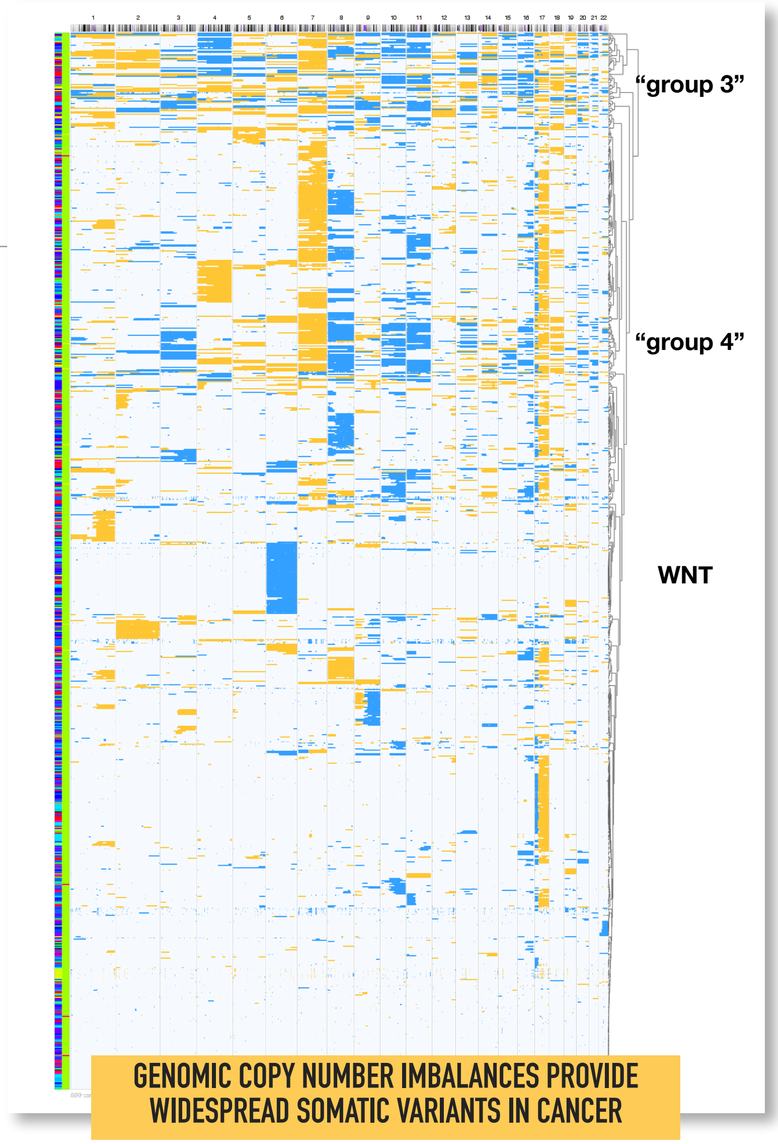


Somatic CNVs In Cancer: Patterns

Many tumor types express **recurrent mutation patterns**

How can those patterns be used for classification and determination of biological mechanisms?





A genomic copy number histogram for malignant medulloblastomas, the most frequent type of pediatric brain tumors, displaying regions of genomic duplications and deletions. These can be decomposed into individual tumor profiles which segregate into several clusters of related mutation patterns with functional relevance and clinical correlation.

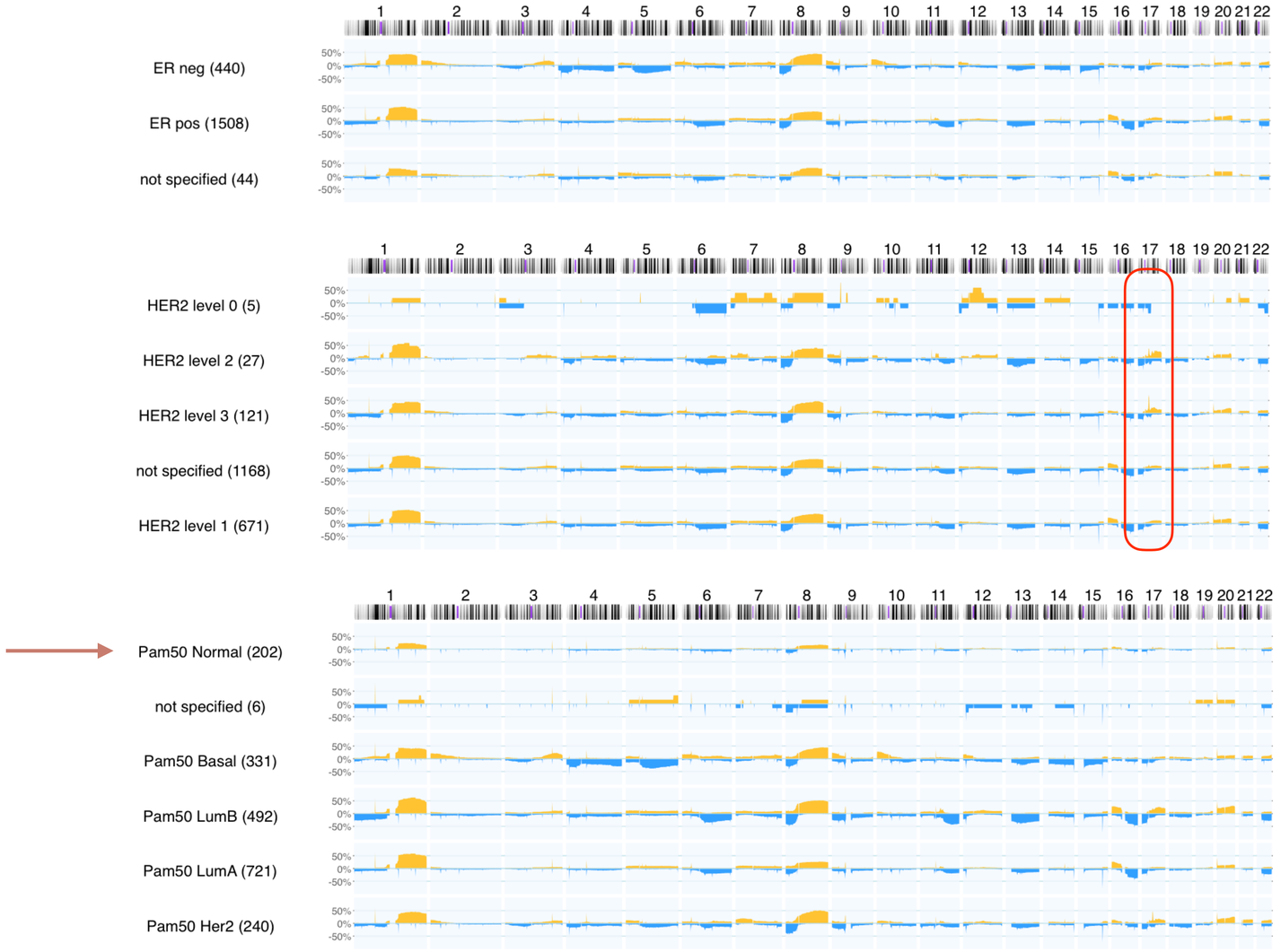
arrayMap

Interpret CNV by Association to molecular/clinical information

ER status

HER2 level

Pam50 category



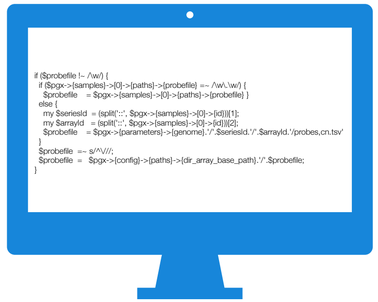
Data source: METABRIC 1992 breast cancer samples, hierarchical clustering, tree not shown

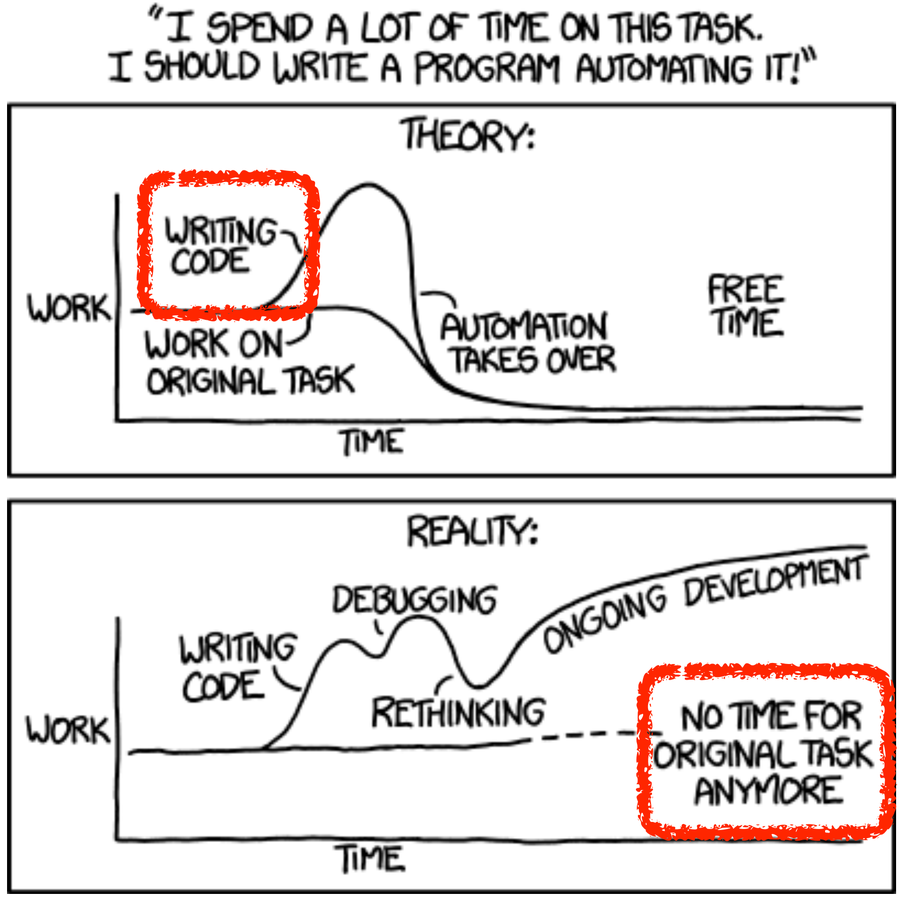
METABRIC
Qingyao Huang

{bio_informatics_science}

```
if ($probefile !~ /\w/) {
  if ($pgx->{samples}->[0]->{paths}->{probefile} =~ /\w\.\w/) {
    $probefile    = $pgx->{samples}->[0]->{paths}->{probefile} }
  else {
    my $seriesId  = (split('::', $pgx->{samples}->[0]->{id}))[1];
    my $arrayId   = (split('::', $pgx->{samples}->[0]->{id}))[2];
    $probefile    = $pgx->{parameters}->{genome}.'/'.$seriesId.'/'.$arrayId.'/
probes,cn.tsv'
  }
  $probefile  =~ s/^\///;
  $probefile  =   $pgx->{config}->{paths}->{dir_array_base_path}.'/'.$probefile;
}
```

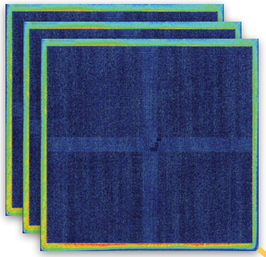
{bio_informatics_science}

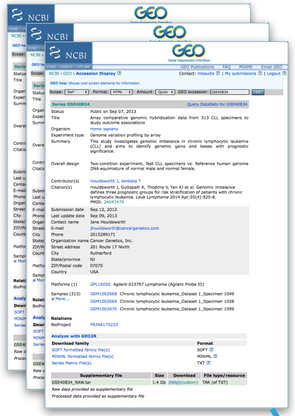


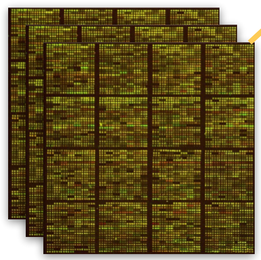


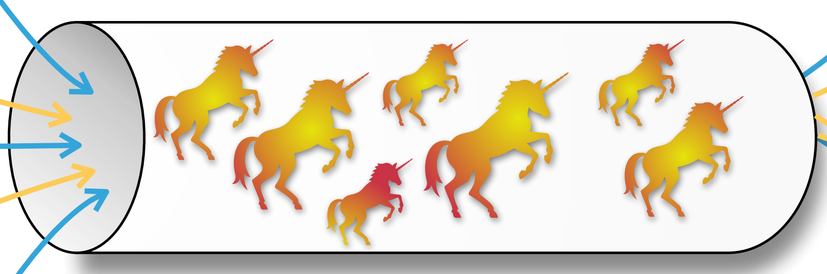
Randall Munroe - XKCD
https://xkcd.com/1319/

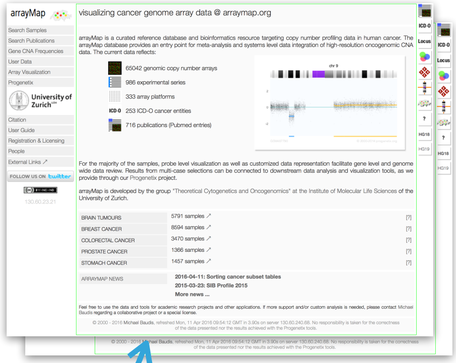
DATA PIPELINE

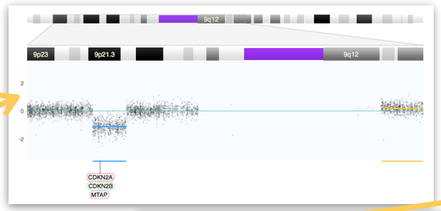


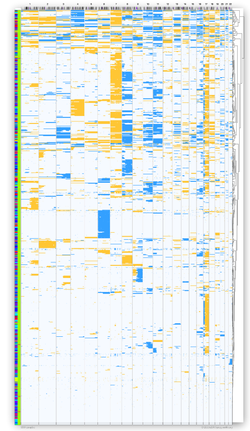


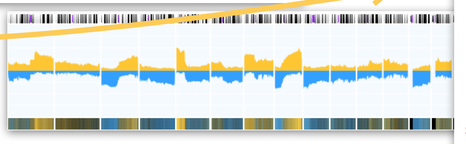


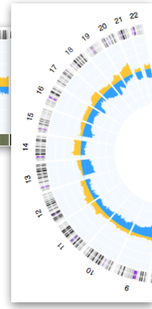


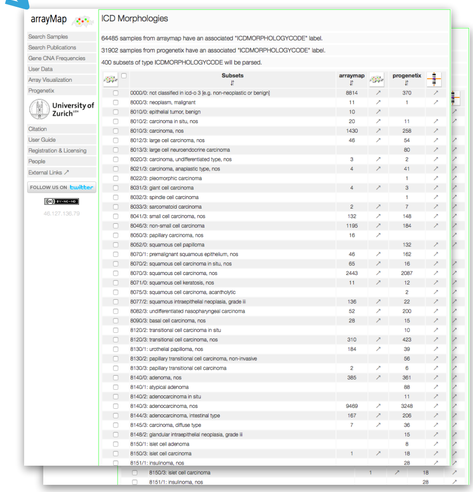








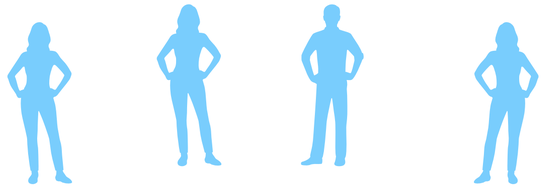




arrayMap

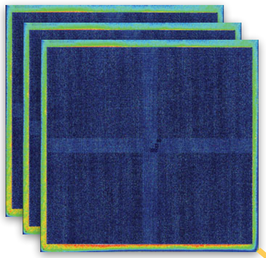
progenetix

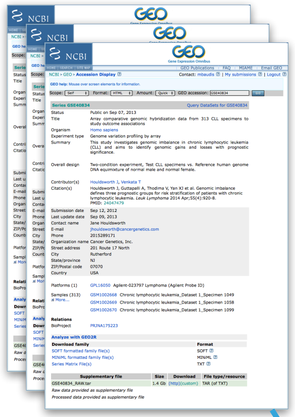
DATA PIPELINE

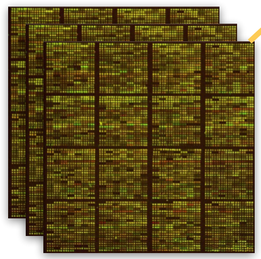


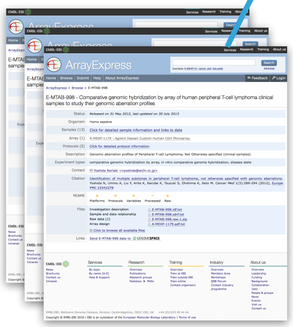
BIOCURATION

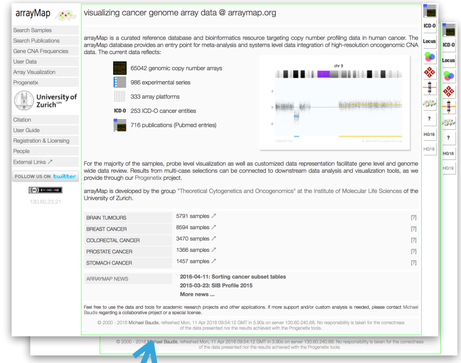
BIOINFORMATICS

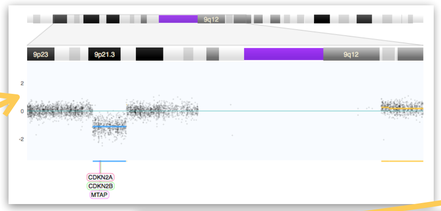


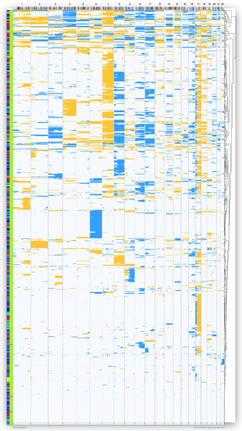


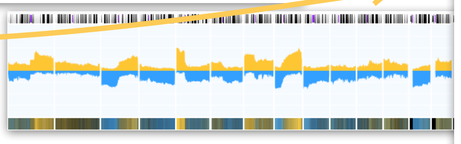


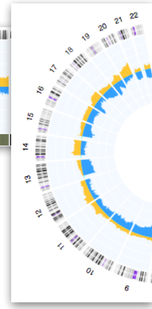


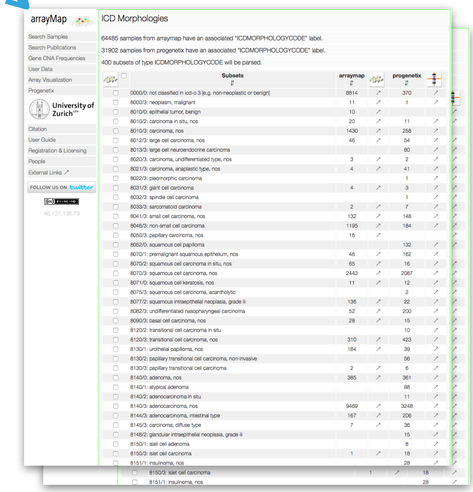












arrayMap

progenetix

bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

New Results

Minimum Error Calibration and Normalization for Genomic Copy Number Analysis

Bo Gao, Michael Baudis

doi: https://doi.org/10.1101/720854

Mecan4CNA:

Minimum Error Calibration and Normalization for Copy Number Analysis

Goal

Calibrate and normalize copy number datasets

Key feature

Without estimating true copy number levels of each sample

Modeling and deduction

$$x_i = (aN_i + bT_i + \sum_{k=1}^{n} c_i^n S_i^n + \sum_{h=1}^{m} E_i^m) \prod_{j=1}^{l} (1 + e_j)$$

$$= (aN_i + bT_i + cS_i + E_i)(1 + e)$$

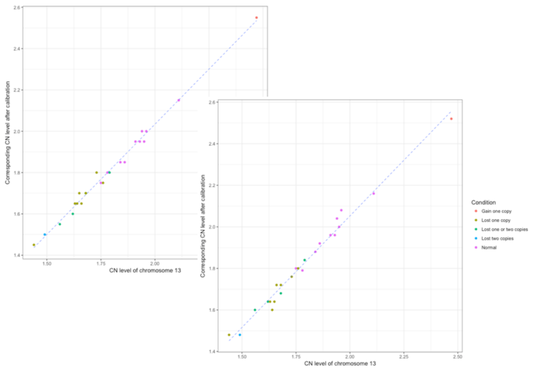
$$= aN_i + bT_i + cS_i + E_i$$

$$R(i,j,k) = \frac{D(i,k)}{D(i,j)}$$

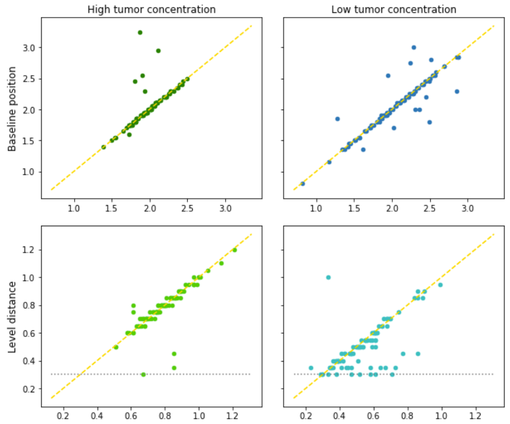
$$R(i,j,k) = \frac{b(T_i - T_k) + c(S_i - cS_k) + E_{i,k}}{b(T_i - T_j) + c(S_i - cS_j) + E_{i,j}}$$

$$= \frac{T_i - T_k}{T_i - T_j}\left(1 + \frac{c(S_i - cS_j) + E_{i,j}}{b(T_i - T_j) + c(S_i - cS_j) + E_{i,j}}\right) + \frac{c(S_i - cS_k) + E_{i,k}}{b(T_i - T_j) + c(S_i - cS_j) + E_{i,j}}$$

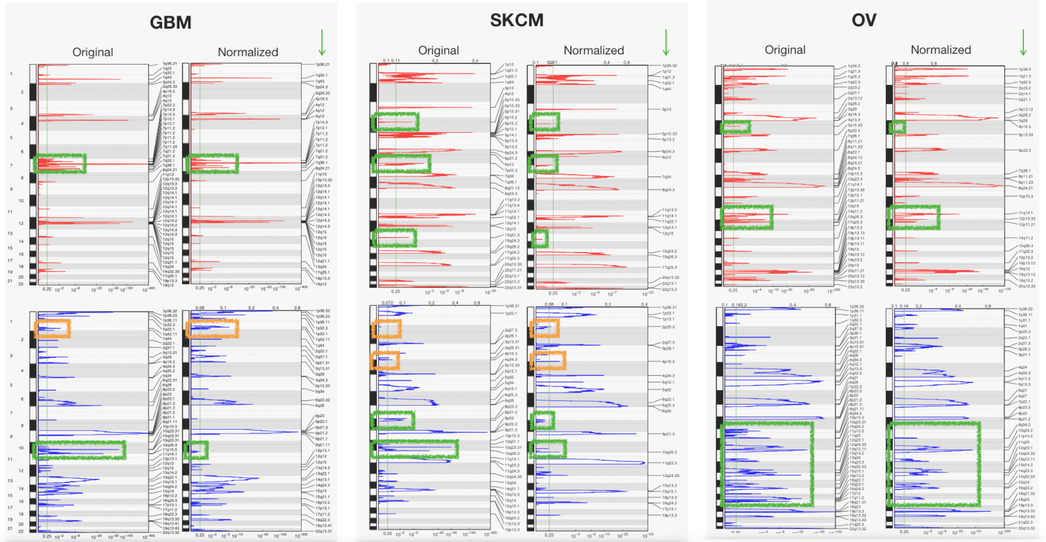
Benchmarking against karyotyping and ABSOLUTE



Benchmarking on simulation data



Application on GISTIC analysis of TCGA data



Bo Gao

F1000Research F1000Research 2018, 7:319 Last updated: 17 MAY 2019

Check for updates

SOFTWARE TOOL ARTICLE

REVISED **segment_liftover : a Python tool to convert segments between genome assemblies [version 2; peer review: 2 approved]**

Bo Gao [1,2], Qingyao Huang[1,2], Michael Baudis [1,2]

segmentLiftover:
A tool to re-map segmental genome data between reference genome editions

Challenge

1. Keep the integrity of copy number segments after Liftover.
2. 10% data lost from straight Liftover.
3. 1TB segment and probe data.

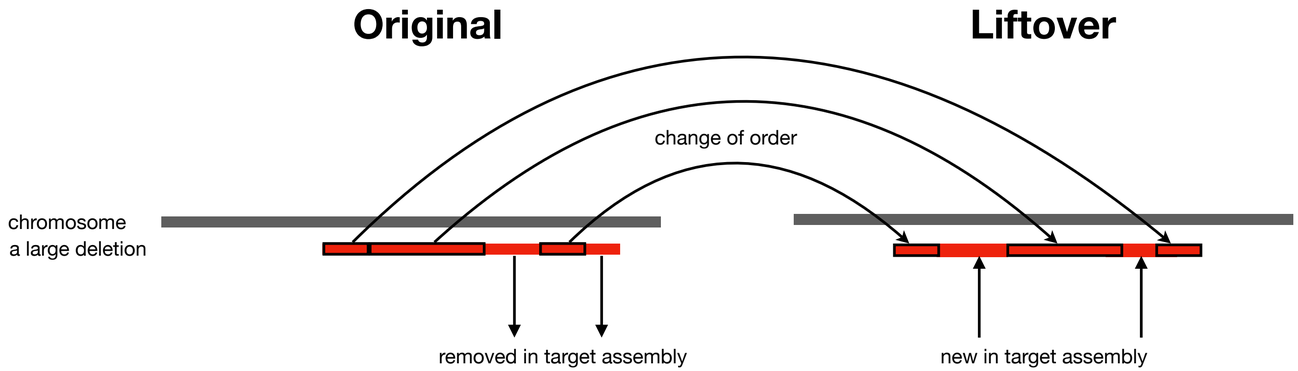
Solution

1. Algorithm to lift copy number segments.
2. Algorithm for fuzzy remapping.
3. Parallel processing and failure recovery mechanism.

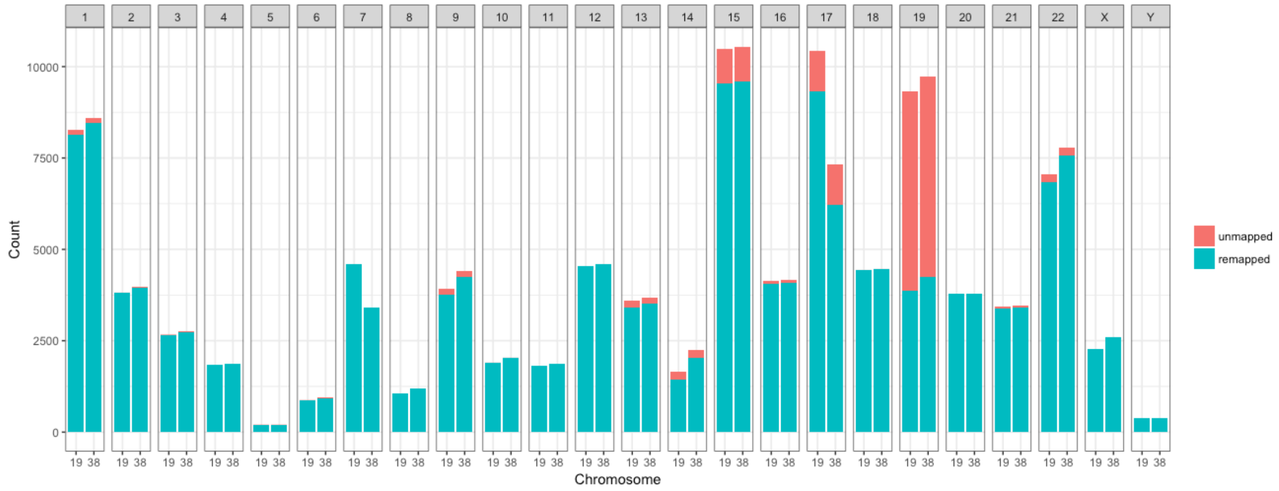
Results

1. Converts hg18 | hg19 | GRCh38
2. Processed 122,788 files, 26,164,205 segments and 28,941,899,671 probes in total
3. straight forward run > 1 week => x4 parallel processes <3 days
4. Reduced data loss: 10% => **0.1%**

The difficulties in copy number segment liftover



Results of segmentLiftover on our data



Population stratification in cancer samples based on SNP array data

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

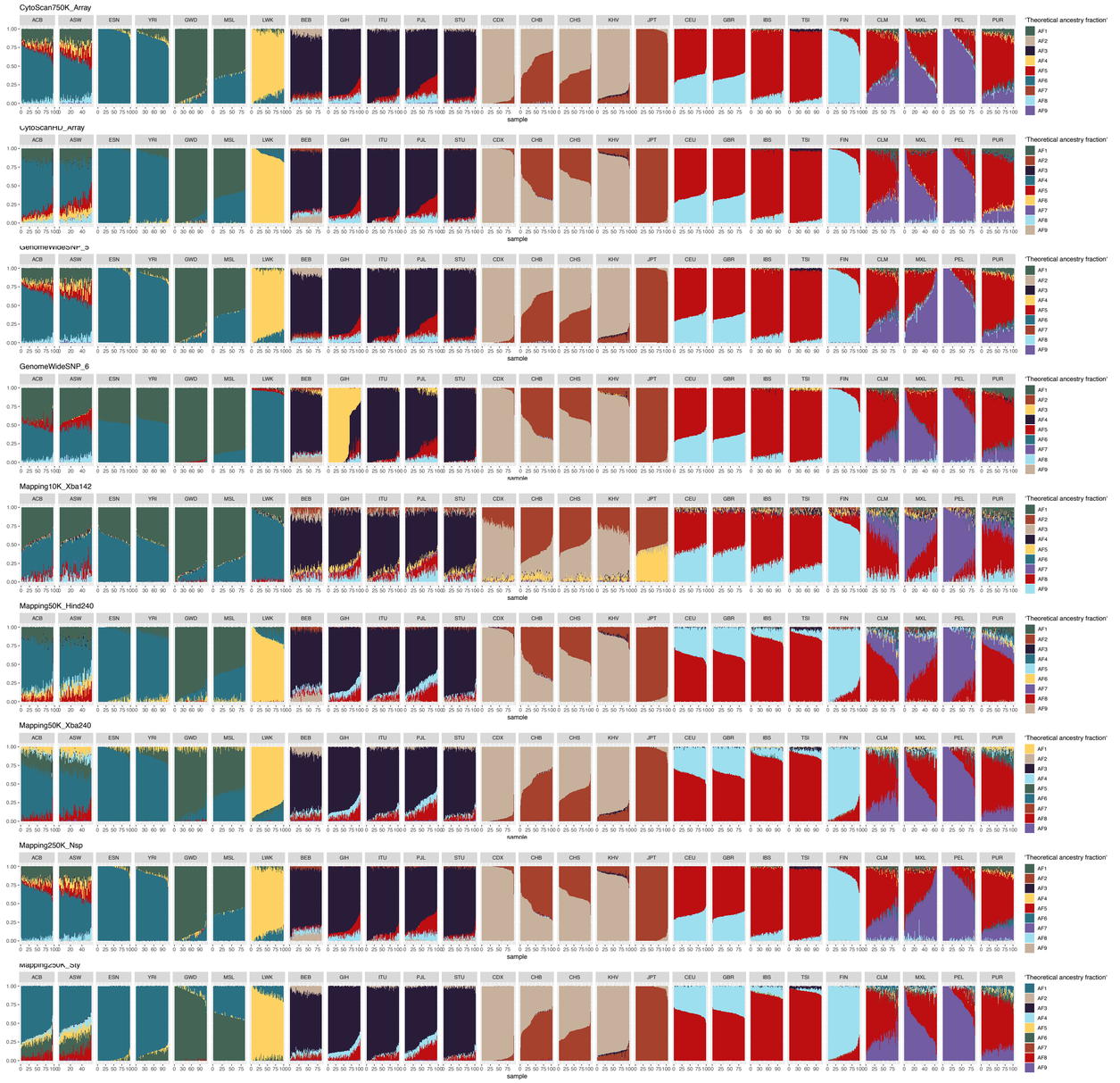


Figure S1 The fraction or contribution of theoretical ancestors (k=9) in reference individuals from 1000 Genomes Project with regard to nine SNP array platforms. The x-axis are individual samples, grouped by their respective population. Groups belonging to the same continent/superpopulation are placed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

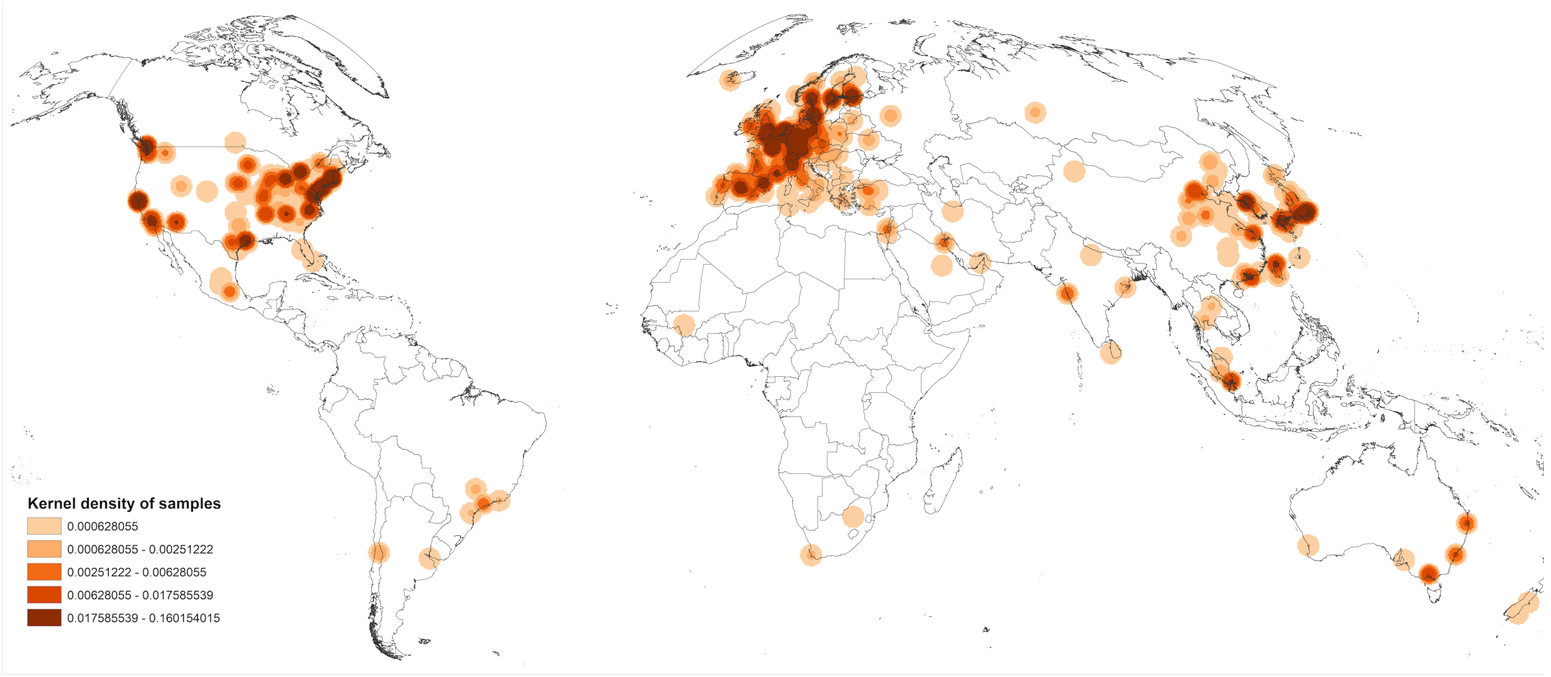
arrayMap

Qingyao Huang

Geographic assessment of cancer genome profiling studies

Paula Carrio Cordo[1,2], **Elise Acheson**[3], **and Michael Baudis**[1,2]✉

[1]Institute of Molecular Life Science, University of Zurich, Winterthurerstrasse 190, 8057, Zurich, Switzerland
[2]SIB, Swiss Institute of Bioinformatics, Winterthurerstrasse 190, 8057, Zurich, Switzerland
[3]Department of Geography, University of Zurich, Switzerland



Enabling genomic data sharing for the benefit of human health

The Global Alliance for Genomics and Health (GA4GH) is a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a human rights framework

Genomic Data Toolkit →

Regulatory & Ethics Toolkit →

Data Security Toolkit →

VIEW OUR LEADERSHIP

MORE ABOUT US

BECOME A MEMBER

Scientists Seek Order to Potential Confusion of Gene Data
Bloomberg - Drew Armstrong & Robert Langreth
June 5, 2013

Q&A: David Altshuler on How to Share Millions of Human Genomes
Science - Jocelyn Kaiser
June 7, 2013

DNA data to be shared worldwide in medical research project
The Guardian - Ian Sample
June 5, 2013

Geneticists push for global data-sharing
Nature - Erika Check Hayden
June 5, 2013

Accord Aims to Create Global Trove of Genetic Data
The New York Times - Gina Kolata
June 5, 2013

Global Alliance
for Genomics & Health

Poking Holes in Genetic Privacy
The New York Times - Gina Kolata
June 16, 2013

New alliance aims to create international system for sharing genomic data
The Globe and Mail - By André Picard
June 5, 2013

White House Open Science 'Champions' Highlights Genomic Data Pioneers
GenomeWeb
June 19, 2013

Our Genes, Their Secrets
The New York Times
June 18, 2013

Une alliance pour partager les données génomiques et cliniques
Le Monde - Sandrine Cabut
June 14, 2013

Global alliance to create framework for sharing genomic data
The Boston Globe - Carolyn Y. Johnson
June 5, 2013

Scientists Seek Order to Potential Confusion of Gene Data
Bloomberg - Drew Armstrong & Robert Langreth
June 5, 2013

Q&A: David Altshuler on How to Share Millions of Human Genomes
Science - Jocelyn Kaiser
June 7, 2013

DNA data to be shared worldwide in medical research project
The Guardian - Ian Sample
June 5, 2013

Geneticists push for global data-sharing
Nature - Erika Check Hayden
June 5, 2013

Accord Aims to Create Global Trove of Genetic Data
The New York Times - Gina Kolata
June 5, 2013

Global Alliance
for Genomics & Health

Poking Holes in Genetic Privacy
The New York Times - Gina Kolata
June 16, 2013

New alliance aims to create international system for sharing genomic data
The Globe and Mail - By André Picard
June 5, 2013

White House Open Science 'Champions' Highlights Genomic Data Pioneers
GenomeWeb
June 19, 2013

Our Genes, Their Secrets
The New York Times
June 18, 2013

Une alliance pour partager les données génomiques et cliniques
Le Monde - Sandrine Cabut
June 14, 2013

Global alliance to create framework for sharing genomic data
The Boston Globe - Carolyn Y. Johnson
June 5, 2013

GA4GH API promotes sharing

GENOMICS

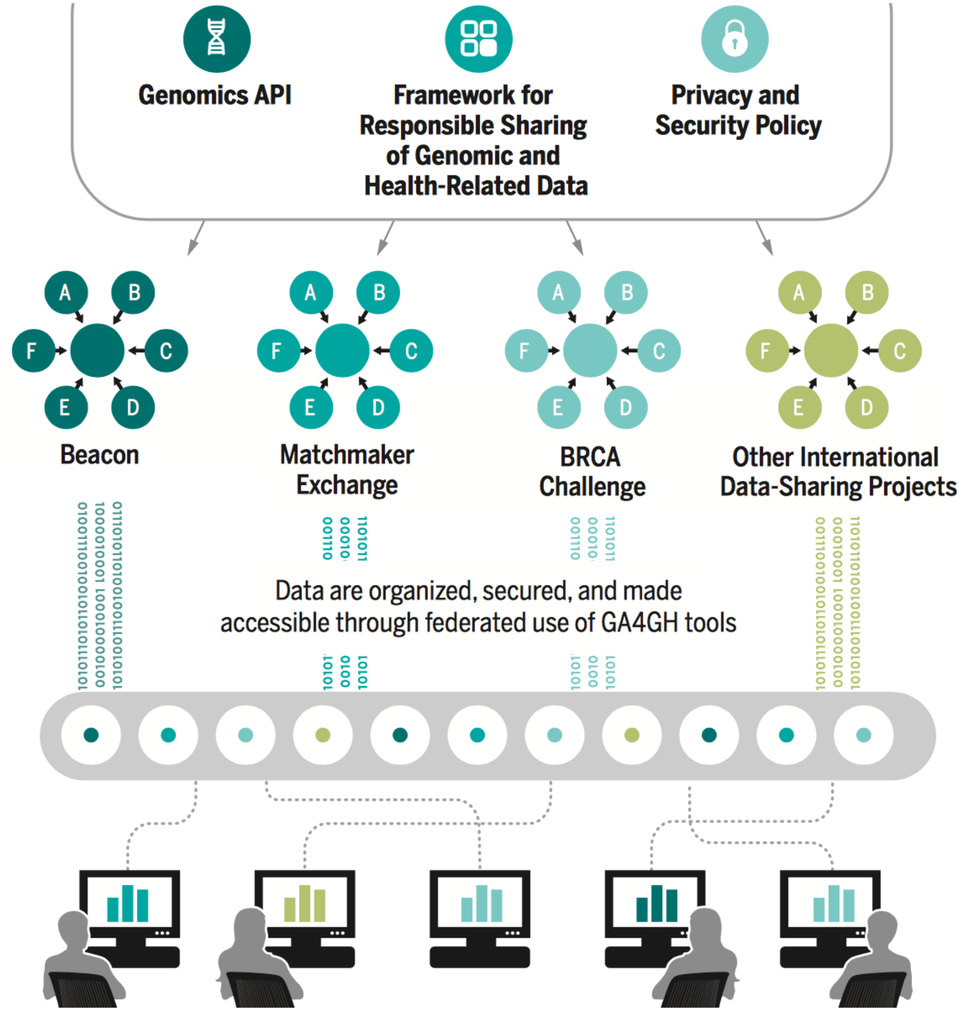
A federated ecosystem for sharing genomic, clinical data

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

The Global Alliance for Genomics and Health*

SCIENCE 10 JUNE 2016 • VOL 352 ISSUE 6291

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



Global Alliance for Genomics & Health

GA4GH HISTORY & MILESTONES

- January 2013 - 50 participants from eight countries
- June 2013 - White Paper, over next year signed by 70 “founding” member institutions (e.g. SIB, UZH)
- March 2014 - Working group meeting in Hinxton & 1st plenary in London
- October 2014 - Plenary meeting, San Diego; interaction with ASHG meeting
- June 2015 - 3rd Plenary meeting, Leiden
- September 2015 - GA4GH at ASHG, Baltimore
- October 2015 - DWG / New York Genome Centre
- April 2016 - Global Workshop @ ICHG 2016, Kyoto
- October 2016 - 4th Plenary Meeting, Vancouver
- May 2017 - Strategy retreat, Hinxton
- October 2017 - 5th plenary, Orlando
- May 2018 - Vancouver
- October 2018 - 6th plenary, Basel
- May 2019 - GA4GH Connect, Hinxton
- October 2019 - 7th Plenary, Boston


GENOMICS

A federated ecosystem for sharing genomic, clinical data

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

The Global Alliance for Genomics and Health*

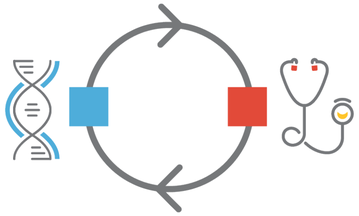
SCIENCE 10 JUNE 2016 • VOL 352 ISSUE 6291

Global Alliance for Genomics & Health

Core Principles of Data Sharing

Global Alliance for Genomics & Health

Enable **international** data sharing



Promote sharing across the **translational continuum**
(discovery research, clinical trials, healthcare systems, diagnostic labs, industry)

Encourages technology-enabled **federated approaches**
(bring analysis to the data)

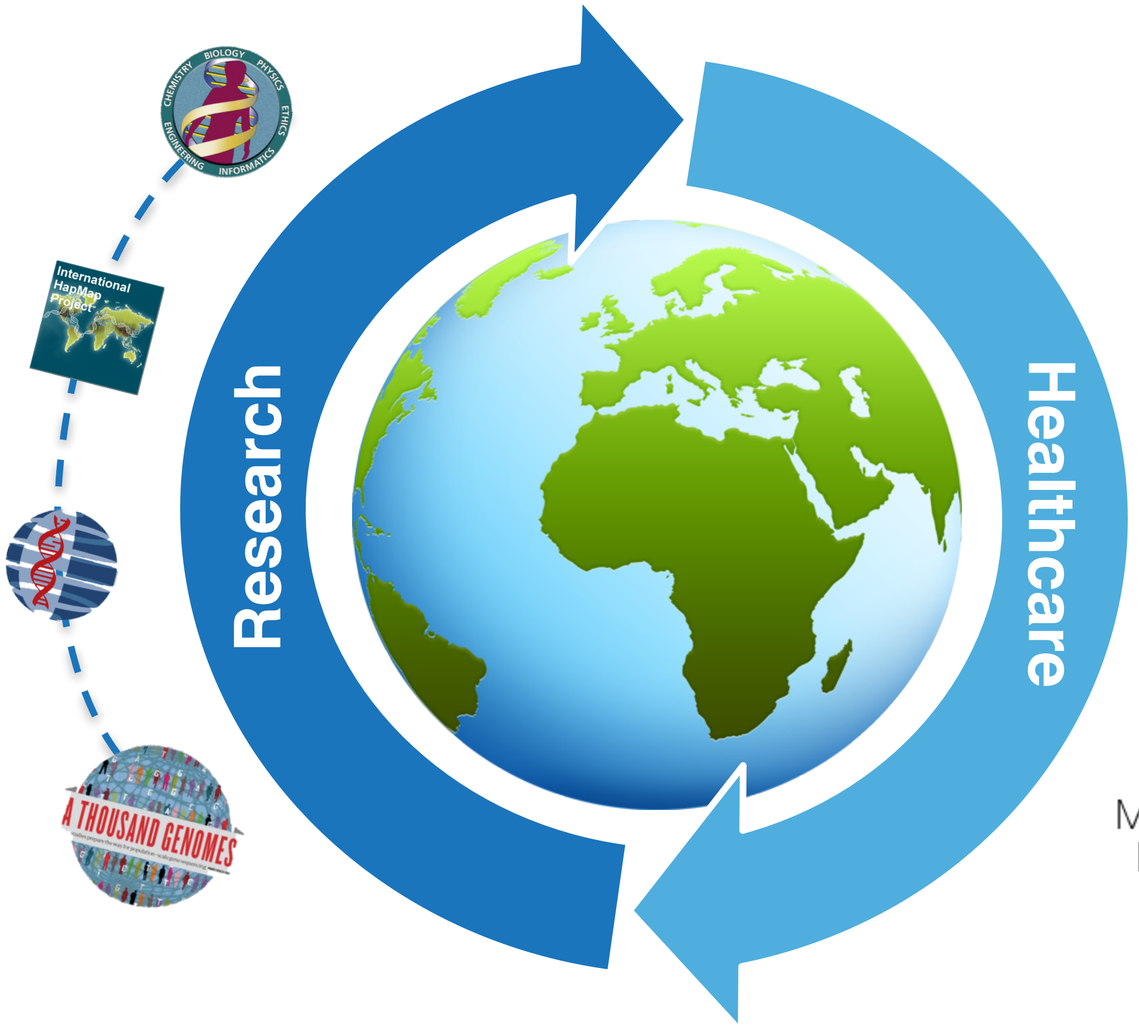
Promote **interoperability**

- Scientific: Standards adoption; transparent documentation
- Technical: Standardized file formats, variant calling protocols, variant & gene annotation
- Ethical: Consent policies to ensure data can be shared internationally

Global Learning for Health

Global Alliance for Genomics & Health

Interoperable APIs, standards & frameworks to support global data sharing



VICC
Variant Interpretation for Cancer Consortium

BRCA CHALLENGE

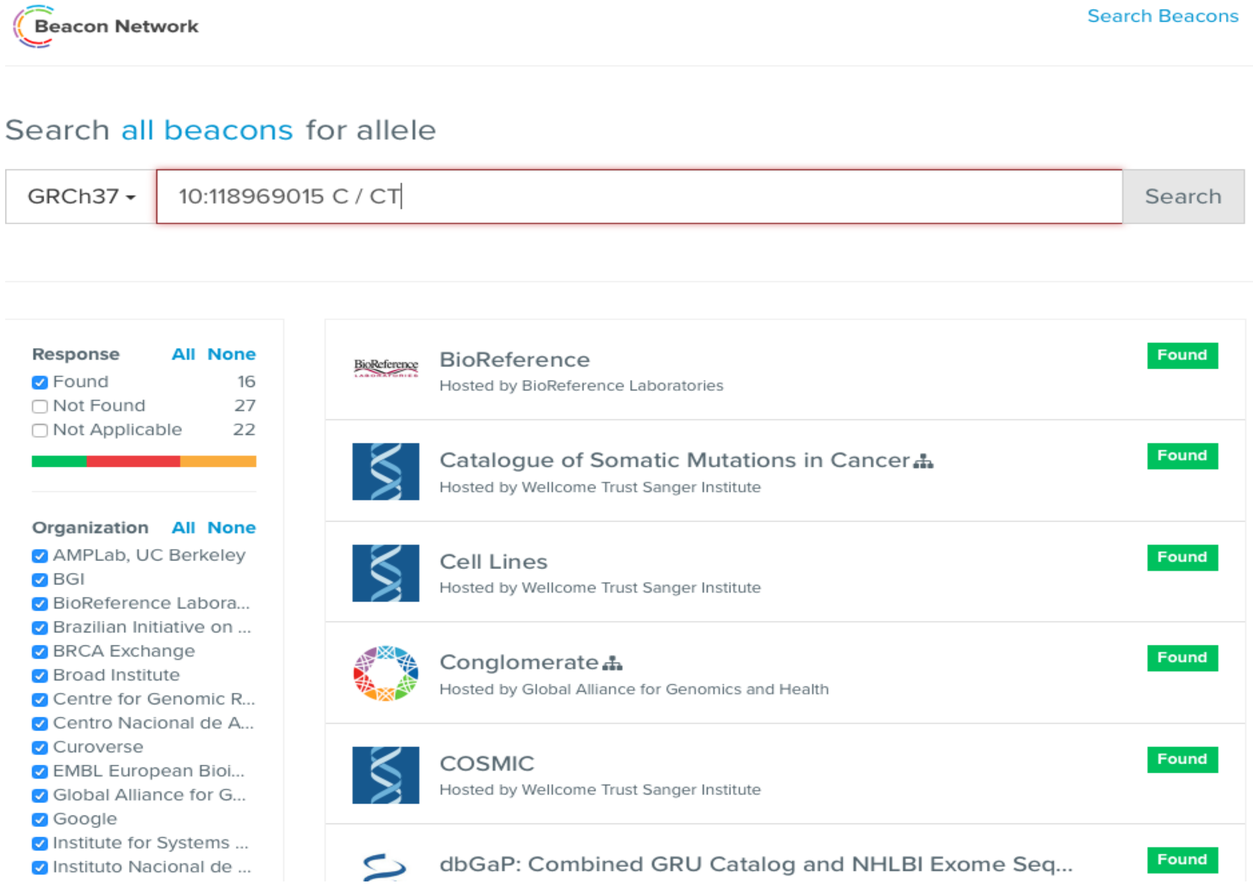
Genomic Knowledge Exchanges

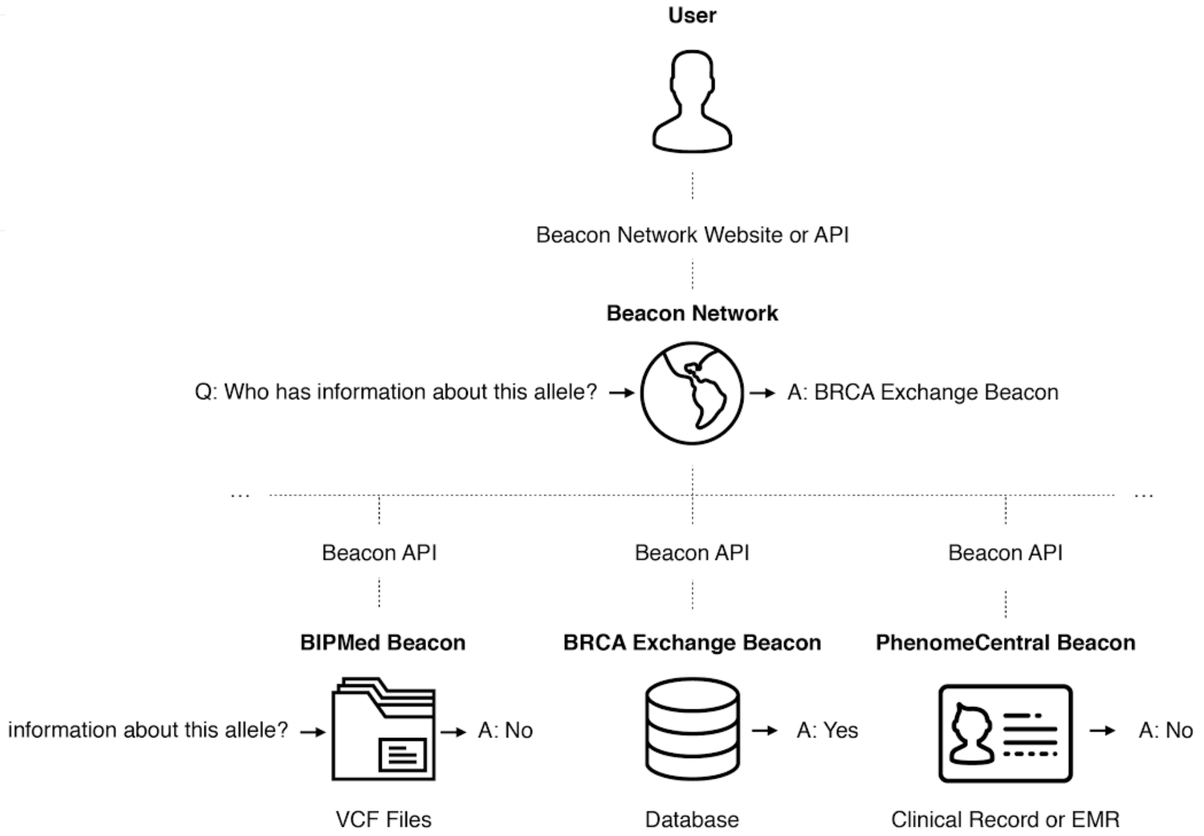
Matchmaker Exchange

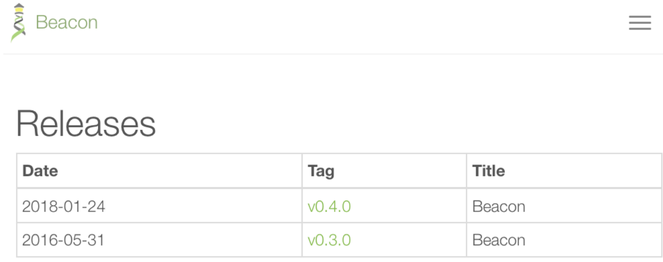
ClinGen
Clinical Genome Resource

Beacon Project

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







Global Alliance for Genomics & Health

35+ Organizations

90+ Beacons

200+ Datasets

100K+ Individuals

ELIXIR Beacon Project

- Driver project on GA4GH roadmap
- aligns with Discovery Work Stream
- strong impact on GA4GH developments as a concrete, funded project

Beacon *forward*

- ☑ **structural variations** (DUP, DEL) in addition to SNV
- … more structural queries (translocations/fusions…)
- **filters** for bio- & techical metadata
- ☑ layered authentication system using **ELIXIR AAI**
- quantitative responses
- ☑ Beacon queries as entry for **data handover** (outside Beacon protocol)
- Ubiquitous **deployment** (e.g. throughout ELIXIR network)

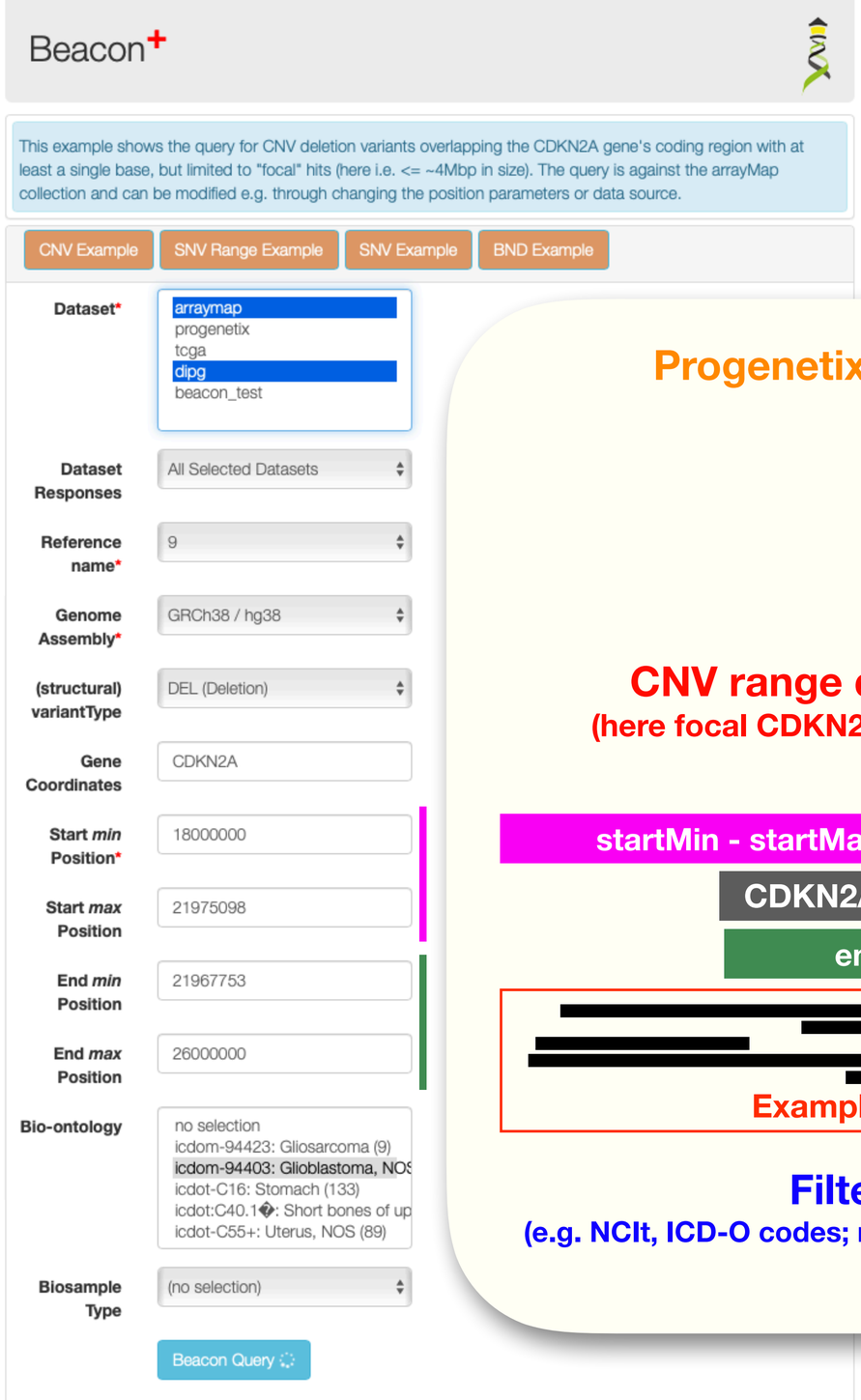
Global Alliance for Genomics & Health

Beacon+ @ UZH

A Beacon Project Technology Demonstrator

- implementing features from roadmap for feasibility testing
  - ***structural variants*** (implemented in v1.0.1)
  - ***handover*** mechanism (implemented in v1.1.0)
  - ***filters*** for phenotypes and other parameters (pre v2)
- runs against complete Progenetix (including TCGA) and arrayMap resources
- backend storage follows GA4GH object model
  - see schemablocks.org

beacon.progenetix.org/ui/



Progenetix datasets

CNV range query example
(here focal CDKN2A/B & MTAP deletion)

startMin - startMax

CDKN2A CDR

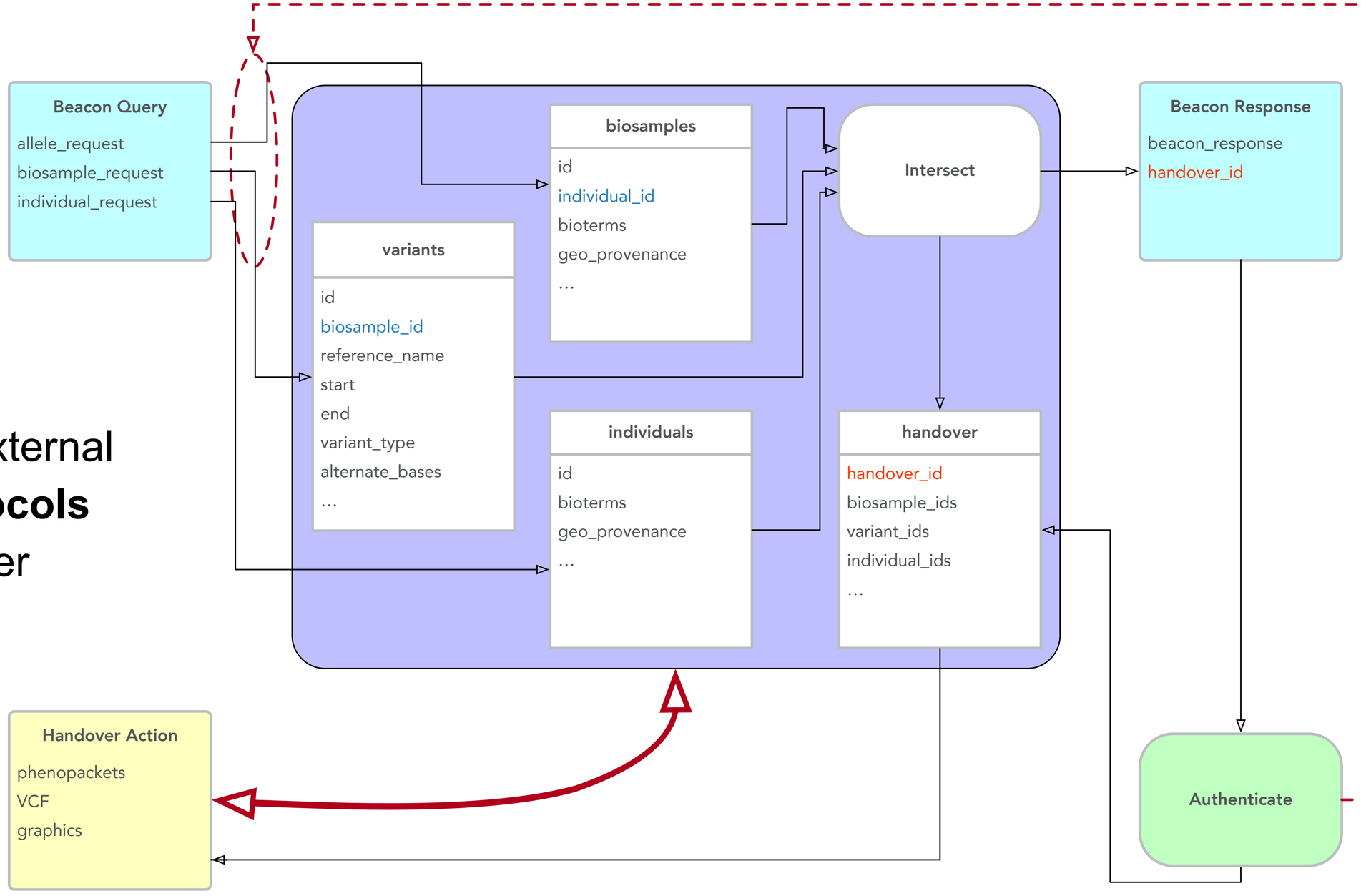
endMin - endMax

Example Hits

Filters
(e.g. NCIt, ICD-O codes; neoplastic/reference ...)

Beacon & Handover

The secure Beacon protocol supports external **data delivery protocols** through the handover mechanism



Beacon I/O

Handover

Authentication

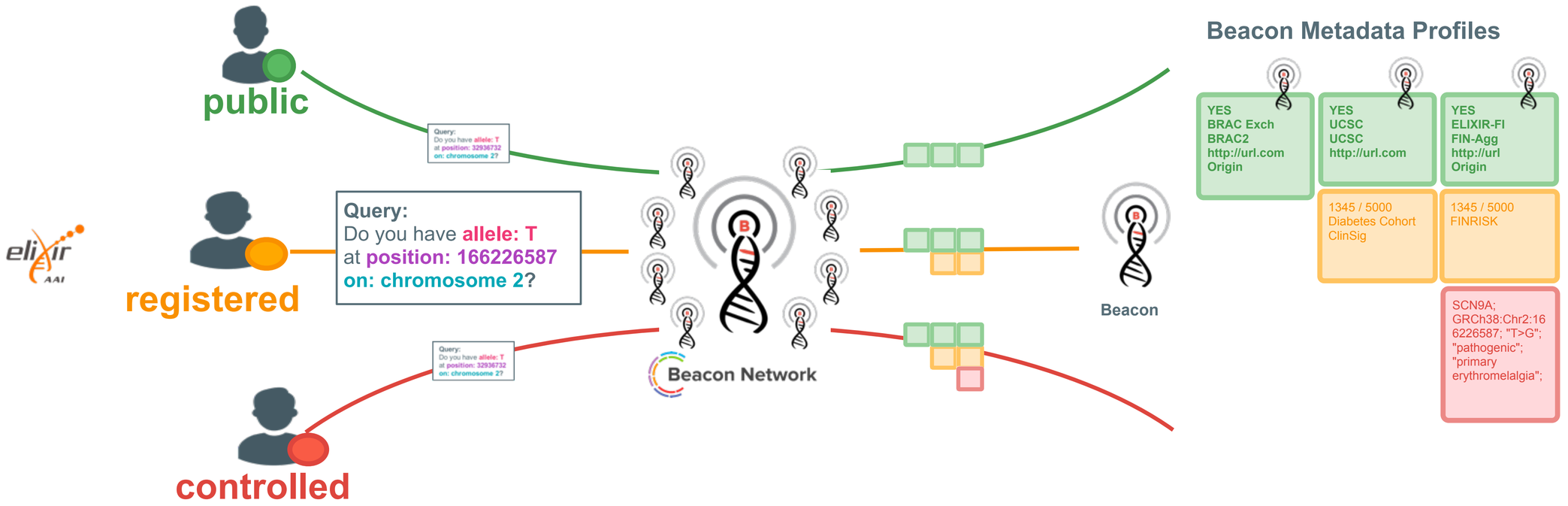
Global Alliance for Genomics & Health

SIB

University of Zurich UZH

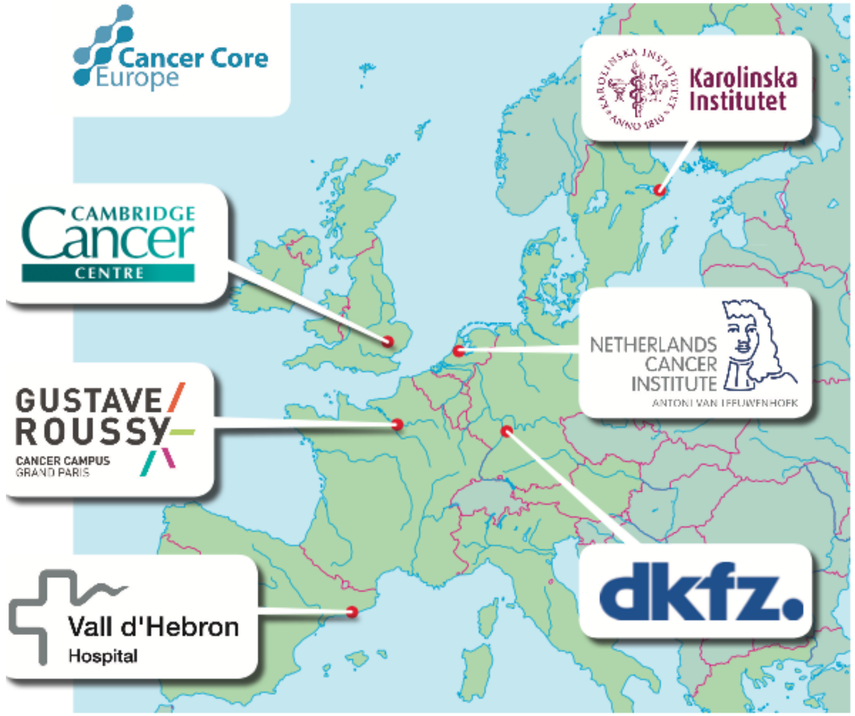
elixir

Integrating permissions and discovery



https://www.youtube.com/watch?v=LyfmvAs7LtQ&feature=youtu.be

ELIXIR Beacon: clinical utility



ELSEVIER

Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

EJC

Position Paper

Cancer Core Europe: A consortium to address the cancer care – Cancer research continuum challenge

Alexander M.M. Eggermont [a,*], Carlos Caldas [b], Ulrik Ringborg [c], René Medema [d], Josep Tabernero [e], Otmar Wiestler [f]

[a] *Gustave Roussy Cancer Campus Grand Paris, Villejuif, France*
[b] *Cambridge Cancer Centre, Cambridge, United Kingdom*
[c] *Karolinska Institutet, Stockholm, Sweden*
[d] *Netherlands Cancer Institute (NKI), Amsterdam, The Netherlands*
[e] *Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain*
[f] *National Center for Tumour Diseases (DKFZ-NCT), Heidelberg, Germany*

Received 31 July 2014; accepted 31 July 2014

KEYWORDS
Cancer care
Research
Continuum
Consortium
Europe

Abstract European cancer research for a transformative initiative by creating a consortium of six leading excellent comprehensive cancer centres that will work together to address the cancer care-cancer research continuum.
Prerequisites for joint translational and clinical research programs are very demanding. These require the creation of a virtual single 'e-hospital' and a powerful translational platform, inter-compatible clinical molecular profiling laboratories with a robust underlying computational biology pipeline, standardised functional and molecular imaging, commonly agreed Standard Operating Procedures (SOPs) for liquid and tissue biopsy procurement, storage and processing, for molecular diagnostics, 'omics', functional genetics, immune-monitoring and other assessments. Importantly also it requires a culture of data collection and data storage that provides complete longitudinal data sets to allow for: effective data sharing and common database building, and to achieve a level of completeness of data that is required for conducting outcome research, taking into account our current understanding of cancers as communities of evolving clones. Cutting edge basic research and technology development serve as an important driving force for innovative translational and clinical studies. Given the excellent track records of the six participants in these areas, Cancer Core Europe will be able to support the full spectrum of research required to address the cancer research- cancer care continuum. Cancer Core Europe also constitutes a unique environment to train the next generation of talents in innovative translational and clinical oncology.

* *Corresponding author*. Address: Gustave Roussy Cancer Campus Grand Paris, 114 Rue Edouard Vaillant, 94805 Villejuif, France. Tel.: +33 1 42 11 40 16; fax: +33 1 42 11 52 52.
E-mail address: alexander.eggermont@gustaveroussy.fr (A.M.M. Eggermont).

http://dx.doi.org/10.1016/j.ejca.2014.07.025
0959-8049/

CORBEL
SHARED SERVICES FOR LIFE-SCIENCE

Cancer Core Europe

- To light Beacons for all 7 Cancer Core Europe participating institutes

ALLEANZA CONTRO IL CANCRO

JUNTA DE ANDALUCIA
CONSEJERÍA DE SALUD

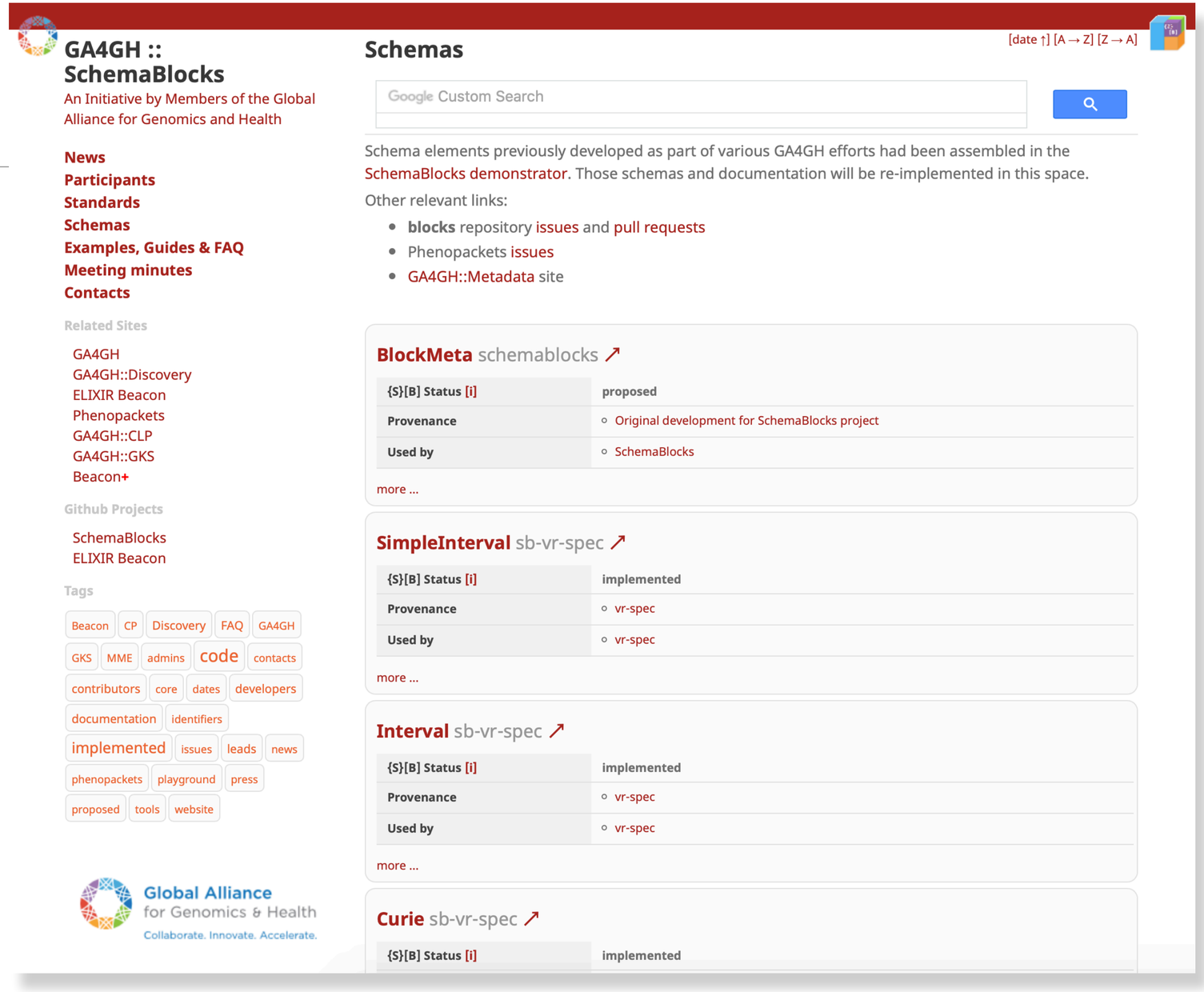
Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

elixir

GA4GH {S}[B]
SchemaBlocks

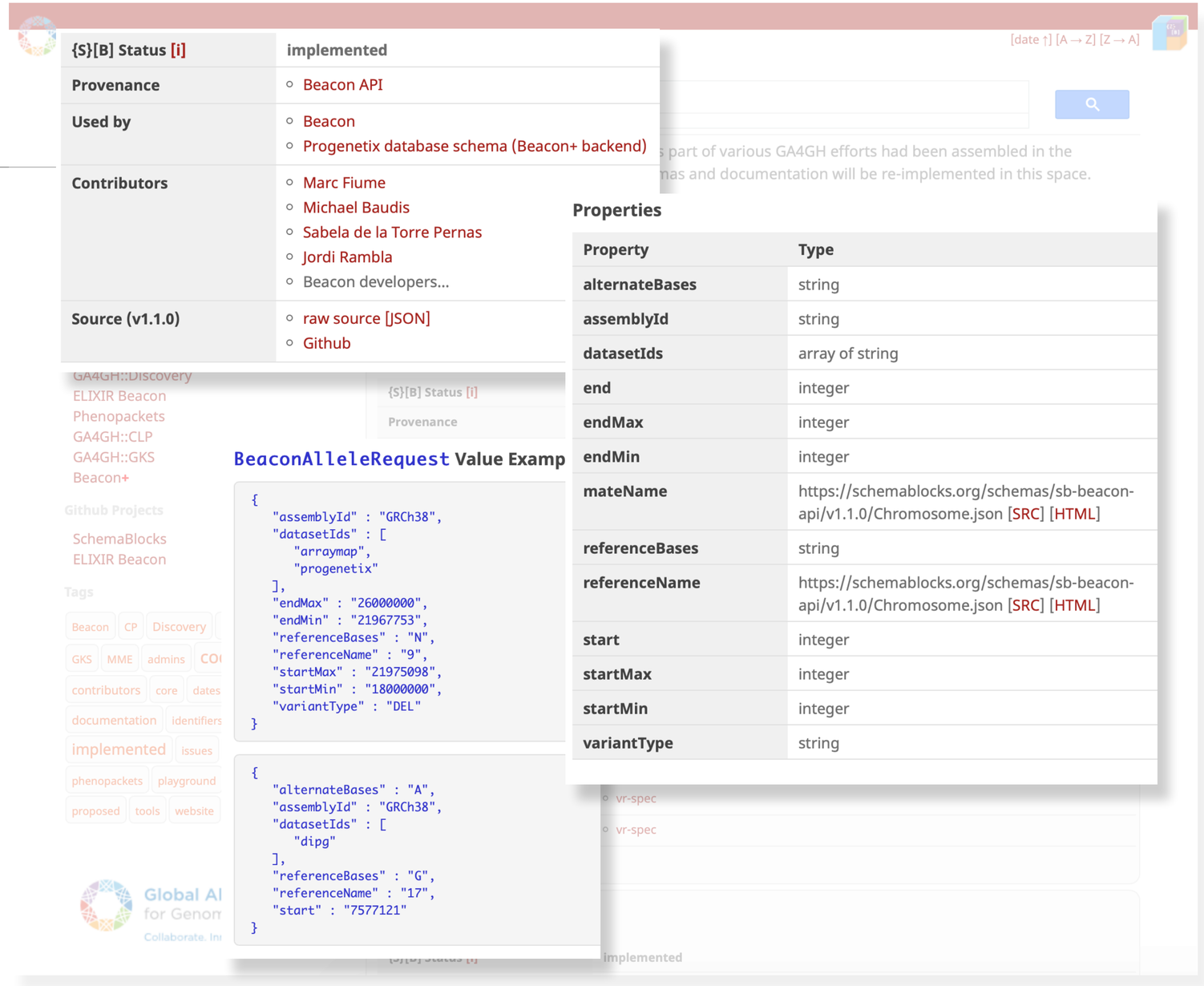
- “cross-workstreams, cross-drivers” initiative to document GA4GH object standards and prototypes, data formats and semantics
- launched in December 2018
- documentation and implementation examples provided by GA4GH members
- no attempt to develop a rigid, complete data schema
- object vocabulary and semantics for a large range of developments
- currently not “authoritative GA4GH recommendations”

schemablocks.org



GA4GH {S}[B]
SchemaBlocks

- “cross-workstreams, cross-drivers” initiative to document GA4GH object standards and prototypes, data formats and semantics
- launched in December 2018
- documentation and implementation examples provided by GA4GH members
- no attempt to develop a rigid, complete data schema
- object vocabulary and semantics for a large range of developments
- currently not “authoritative GA4GH recommendations”



schemablocks.org

SPHN: a federal mandate

MANDATES

SAMS
Swiss Academy of Medical Sciences

SIB
Swiss Institute of Bioinformatics

GOVERNANCE

NSB
National Steering Board

BioMedIT Board

IMPLEMENTATION

Management Office
- Management of the overall initiative
- Daily operations

Personalized Health Informatics Group
- SPHN Data Coordination Centre (DCC)
- BioMedIT project

Expert Advisory Groups
- International Advisory Board (IAB)
- Scientific Expert Board (SEB)
- Ethic Legal Societal Implications Advisory Group (ELSIag)

DCC Working Groups
- Clinical Data Semantic Interoperability
- Bioinformatics and Data Analytics
- BioMedIT Interoperability
- Hospital IT
- IT Security

BioMedIT nodes
1. Core-IT/Vital-IT
2. sciCORE
3. SIS

PARTNERS

Universities, ETHZ, EPFL
University Hospitals
Personalized Health and Related Technologies (PHRT)

COLLABORATIONS

- eHealth Suisse
- Health 2030
- Personalized Health Alliance Zurich Basel (PH ZH-BS)
- swissuniversities
- Swiss Biobanking Platform (SBP)
- Swiss Clinical Trial Organisation (SCTO)
- Swiss Data Science Centre (SDSC)
- Swiss National Science Foundation (SNSF)
- unimedsuisse

SPHN FUNDED PROJECTS (29)

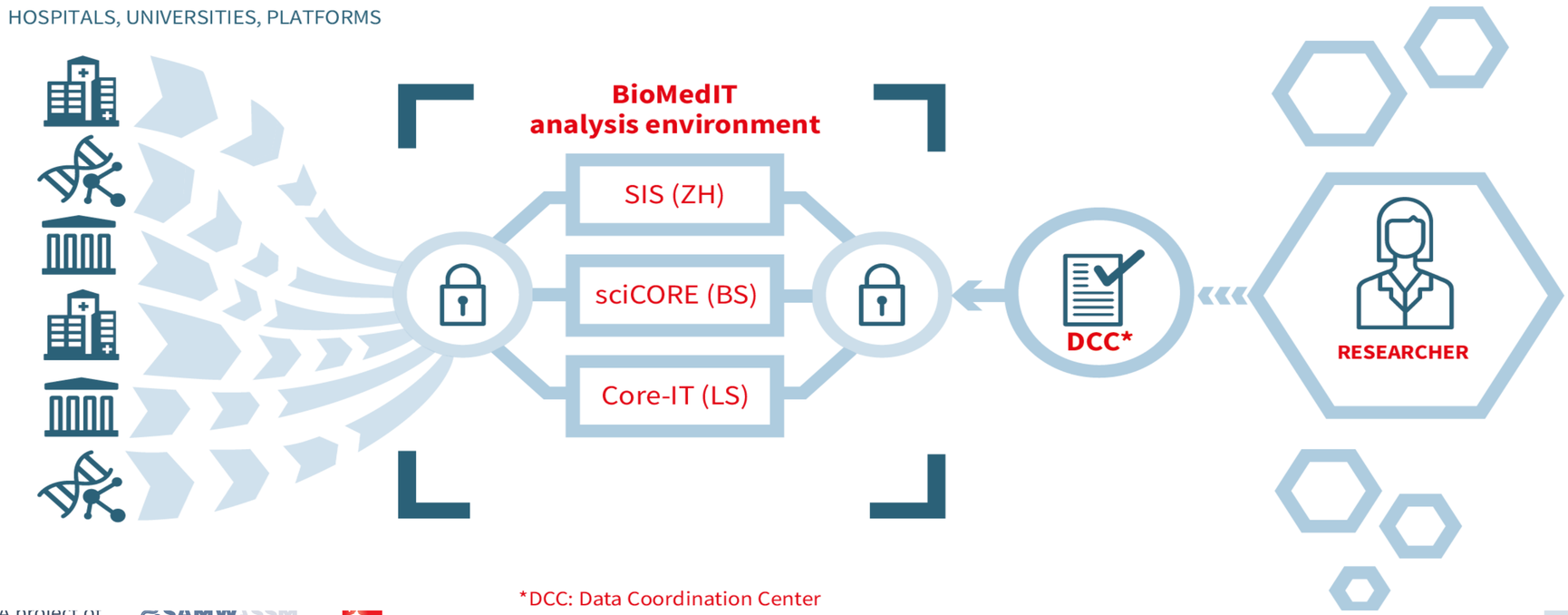
- Collaboration Agreements with University Hospitals
- Infrastructure Development Projects (2 jointly funded with PHRT)
- Driver Projects (4 jointly funded with PHRT)

PHRT FUNDED PROJECTS (56)

- Research Projects (4 jointly funded with SPHN)
- Translational Technology (2 projects jointly funded with SPHN)

Receiving data
The BioMedIT network

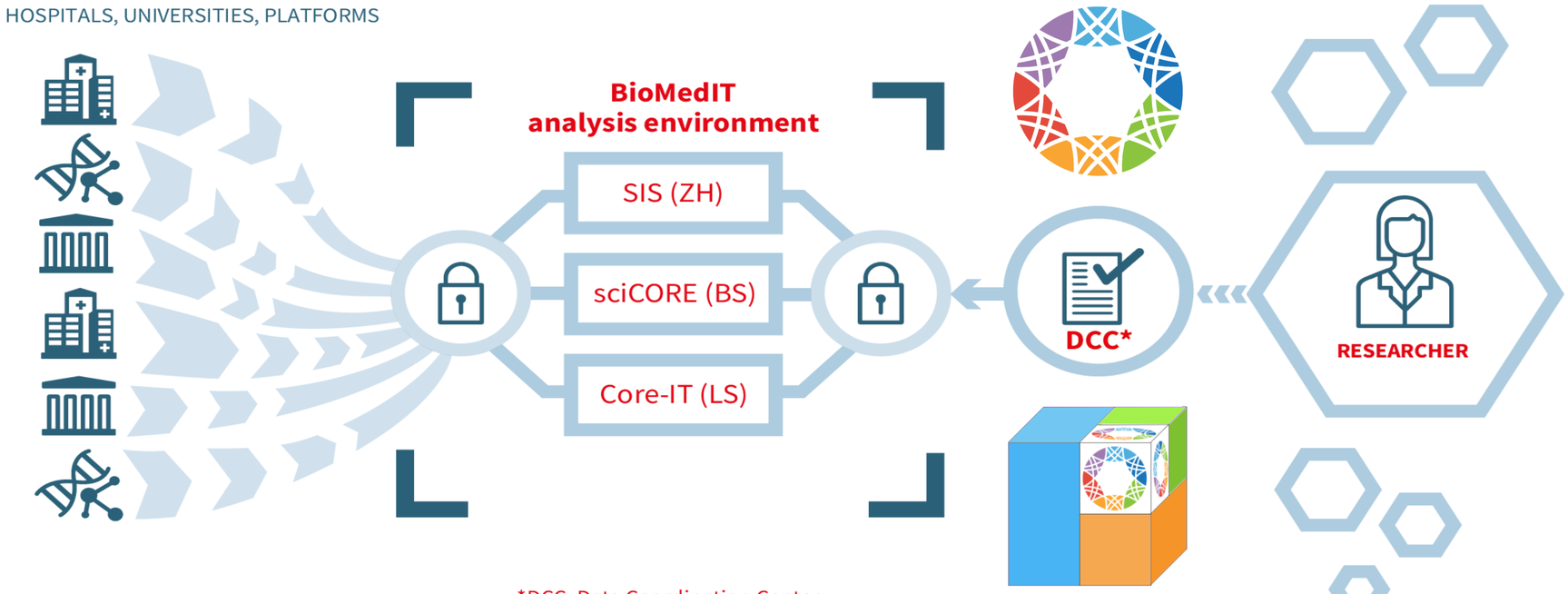
BioMedIT provides researchers with access to a secure and protected computing environment for analysis of sensitive data without compromising data privacy



*DCC: Data Coordination Center

SPHN

- has been a GA4GH Driver project since 2019
- will support Beacon adoption throughout SPHN
- will support the SchemaBlocks initiative and the use of the emerging standards throughout SPHN nodes and resources



Random Thoughts on "Big Data" CNVs for (Cancer) Genomics

- Data accessibility - **quantity**
  - open data w/ "just in time" access & active work to open repositories, archives
  - data curation and long term storage has to be promoted and supported
- New technologies for **qualitatively** new possibilities
  - deep WGS with molecular reconstruction of complex events (chromothripsis...)
- Annotation and exchange formats in **extensible models**
  - referring reference genome positions, w/ remapping, provenance
  - technology agnostic (but provenance...)
- Search and exchange APIs for **distributed** and/or **federated data access**
  - modular object design, independent from backend structure
  - common interfaces/service APIs/registries

BAUDISGROUP @ UZH

(NI AI)
MICHAEL BAUDIS
(HAOYANG CAI)
PAULA CARRIO CORDO
BO GAO
QINGYAO HUANG
(SAUMYA GUPTA)
(NITIN KUMAR)
RAHEL PALOOTS

SIB

AMOS BAIROCH
HEINZ STOCKINGER
DANIEL TEIXEIRA

@WORLD

MATTHIAS ALTMEYER
THOMAS EGGERMANN
ROSA NOGUERA
REINER SIEBERT
CAIUS SOLOVAN

University of Zurich[UZH]

GA4GH

LARRY BABB
ANTHONY BROOKES
MELANIE COURTOT
MELISSA HAENDEL
HELEN PARKINSON
ANDY YATES

ELIXIR & CRG

JORDI RAMBLA DE ARGILA
GARY SAUNDERS
ILKKA LAPPALAINEN
S. DE LA TORRE PERNAS
SERENA SCOLLEN
JUHA TÖRNROOS

SPHN

KATRIN CRAMERI
SABINE ÖSTERLE

H-CNV

CHRISTOPHE BÉROUD
DAVID SALGADO

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

arraymap.org
progenetix.org
info.baudisgroup.org
sib.swiss/baudis-michael
imls.uzh.ch/en/research/baudis
beacon-project.io
schemablocks.org