

Cancer Genomics and Implementation of Data Driven Standards for Genomic Data Exchange

CNV Databases :: Variant Representation & Query Formats :: ELIXIR Beacon :: GA4GH



University of
Zurich^{UZH}



1992



Heidelberg

Student of medicine | doctoral thesis in molecular cytogenetics @ DKFZ (Peter Licher) | 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.



Swiss Institute of
Bioinformatics

Structural Genome Variants in Cancer: Research & Resources

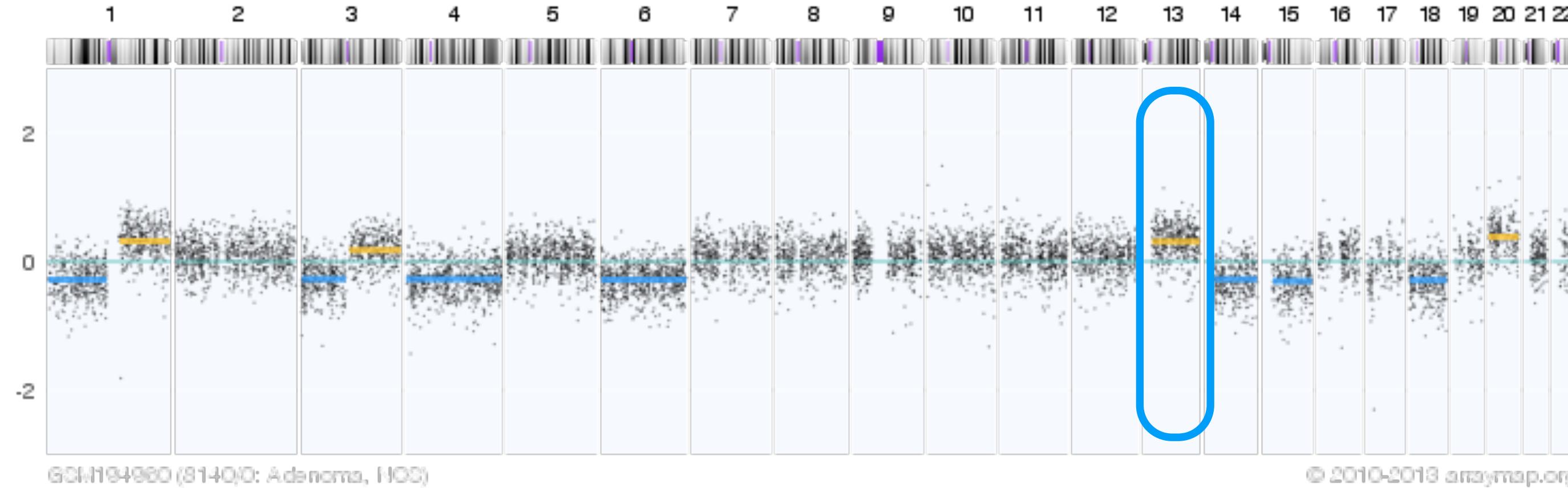
sCNV Frequencies and Patterns :: CNV Databases :: Bioinformatics Tools



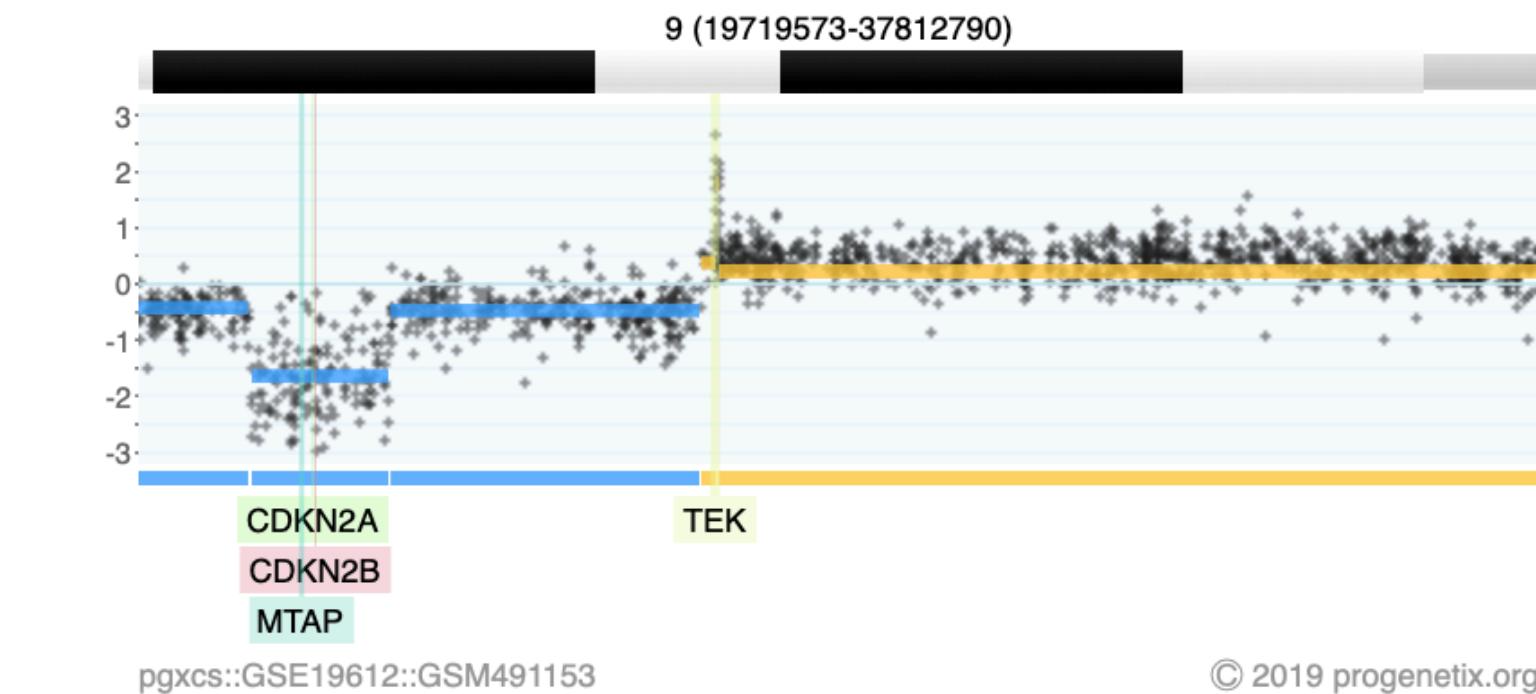
University of
Zurich^{UZH}



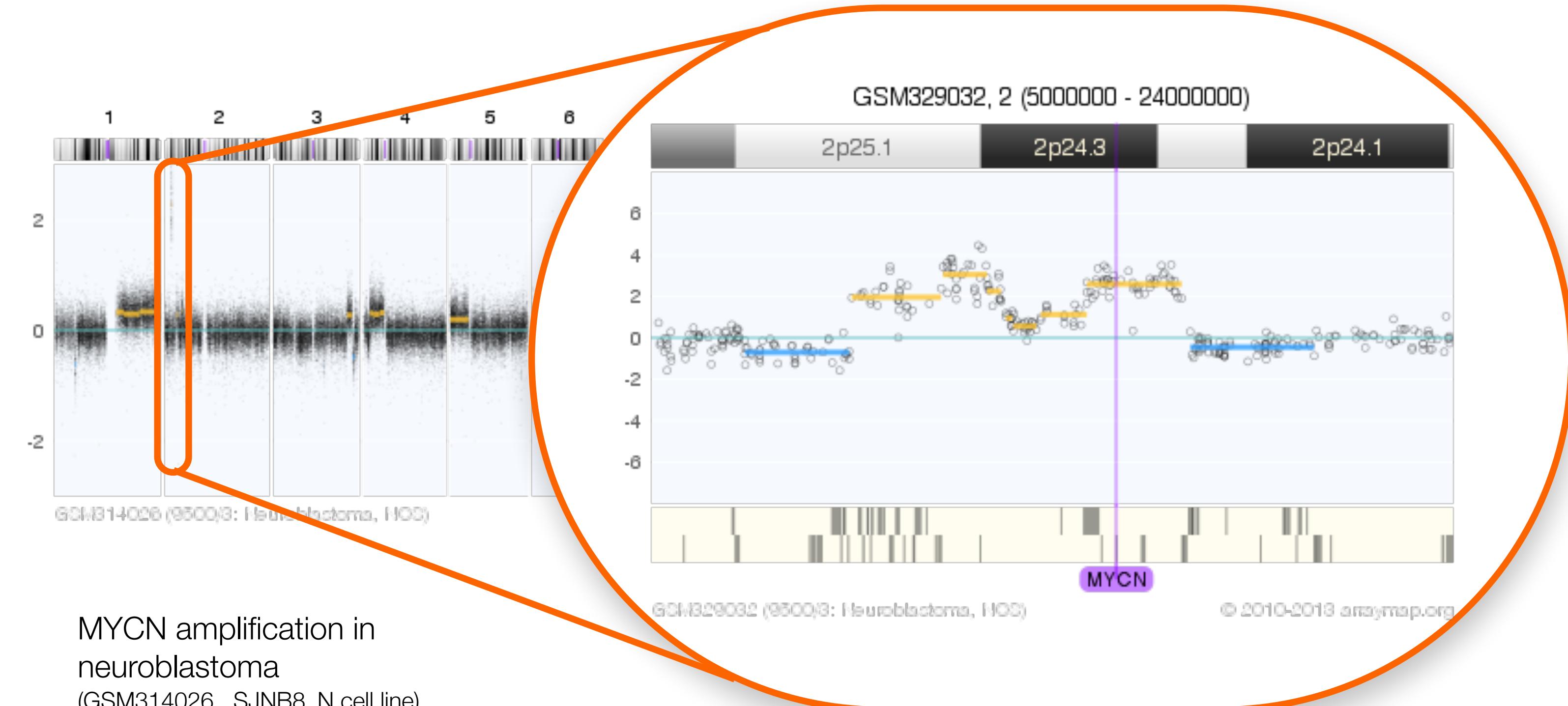
Somatic Copy Number Variations



Gain of chromosome arm 13q in colorectal carcinoma



2-event, homozygous deletion in a Glioblastoma

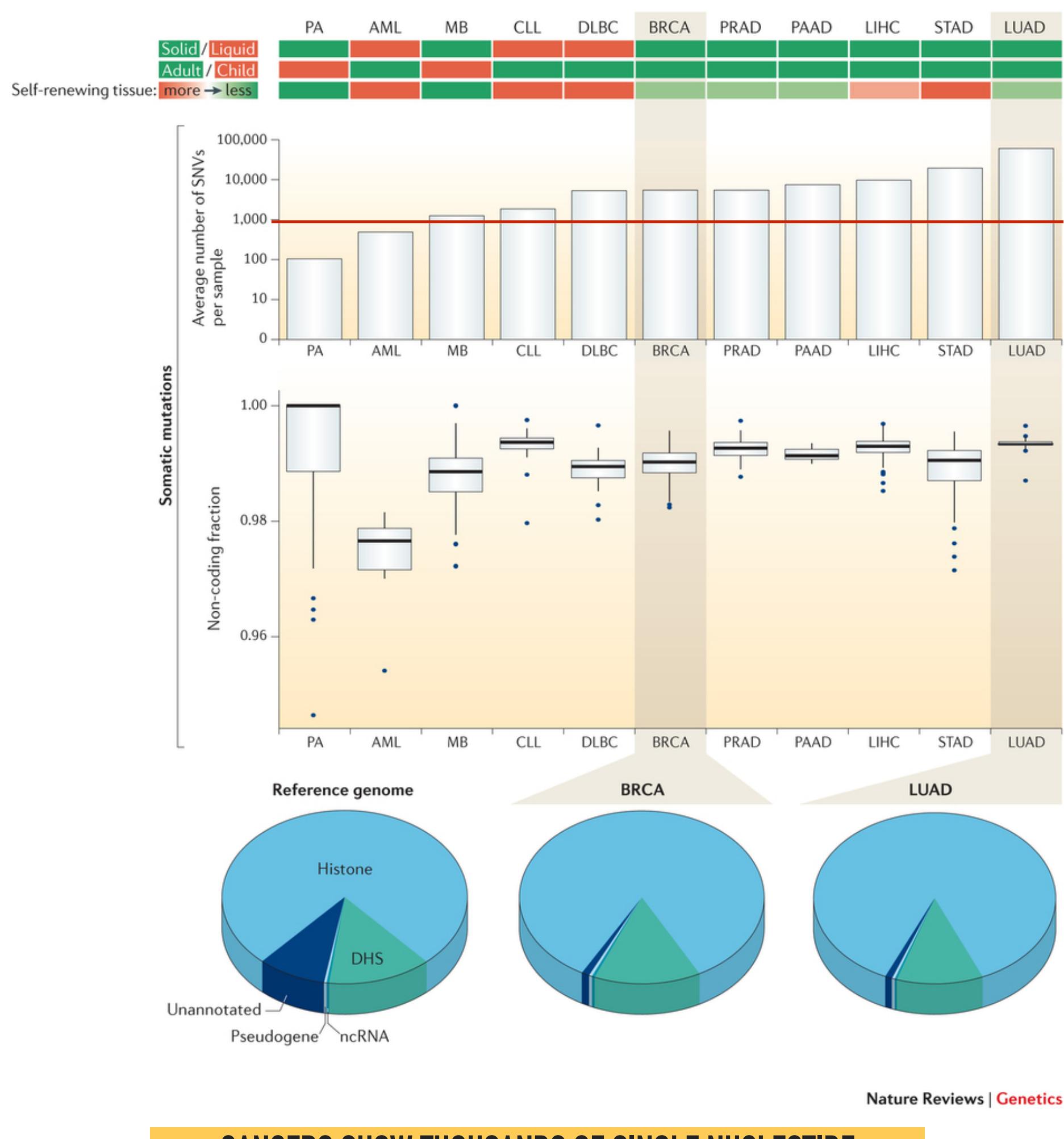


low level/high level copy number alterations (CNAs)

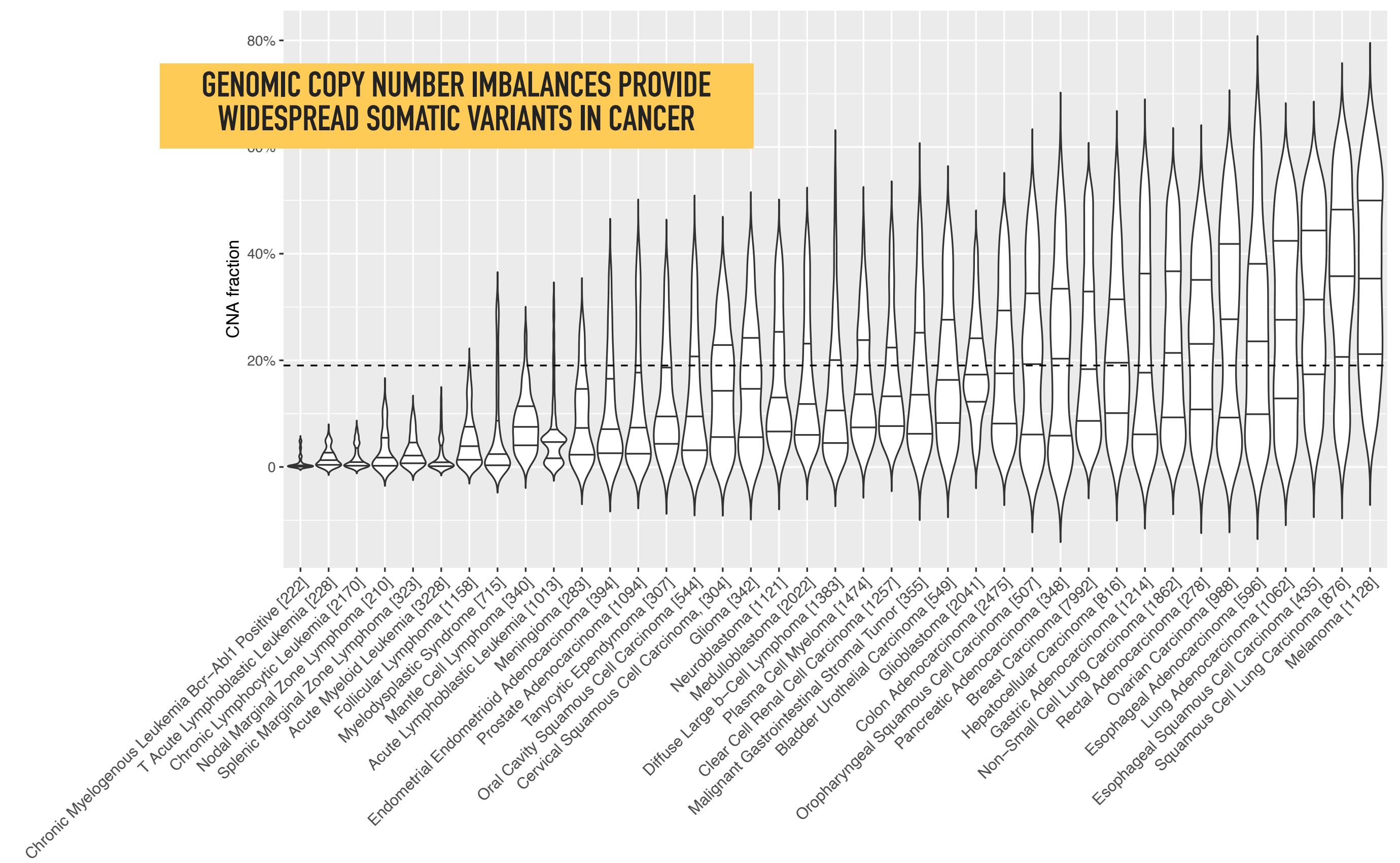
arrayMap



Quantifying Somatic Mutations In Cancer



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

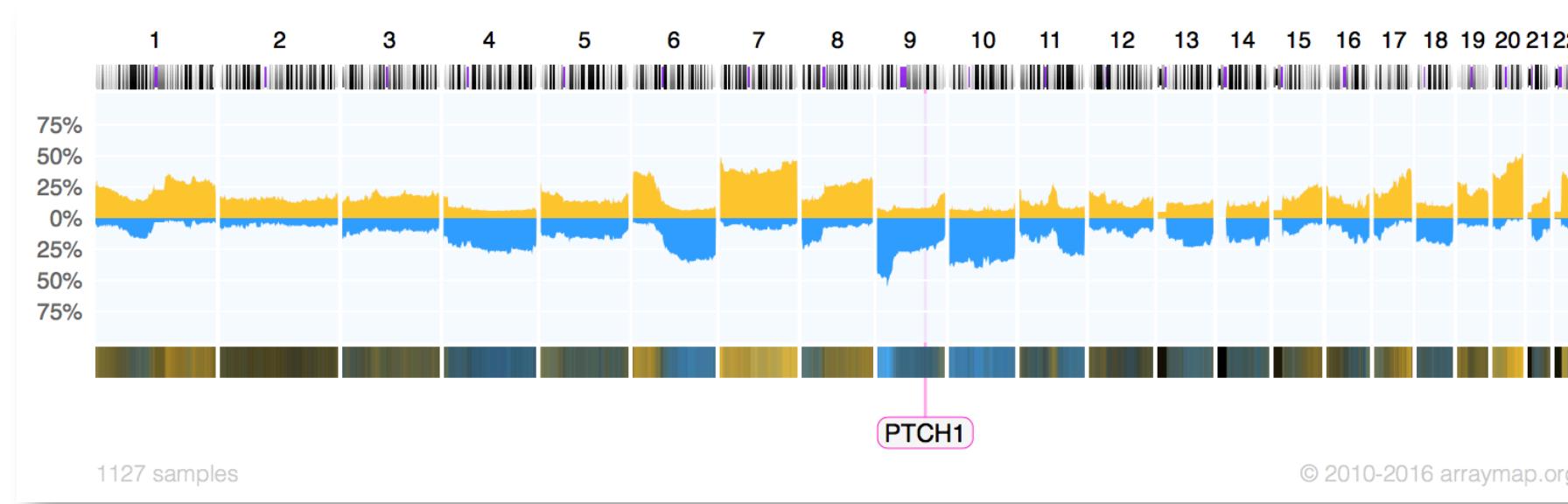
Rare CNV Events & Hidden Therapeutic Options?

Example: PTCH1 deletions in malignant melanomas

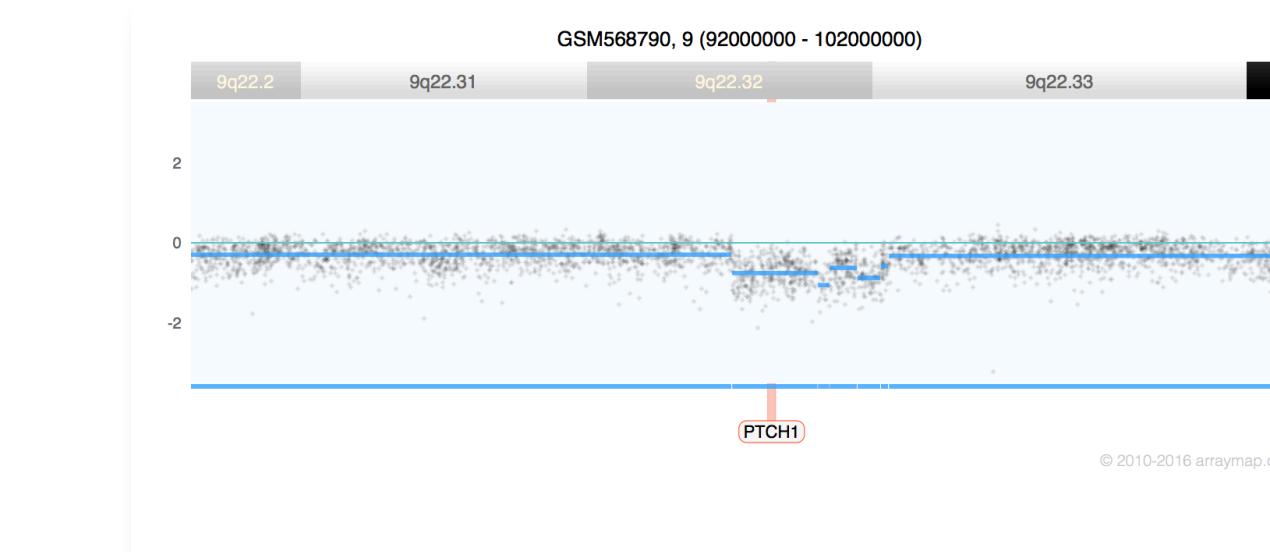
PTCH1 is a actionable tumor suppressor gene, which has been demonstrated in e.g. basalomas and medulloblastomas

analysis of 1127 samples from 26 different publications could identify **focal** deletions in 4 samples

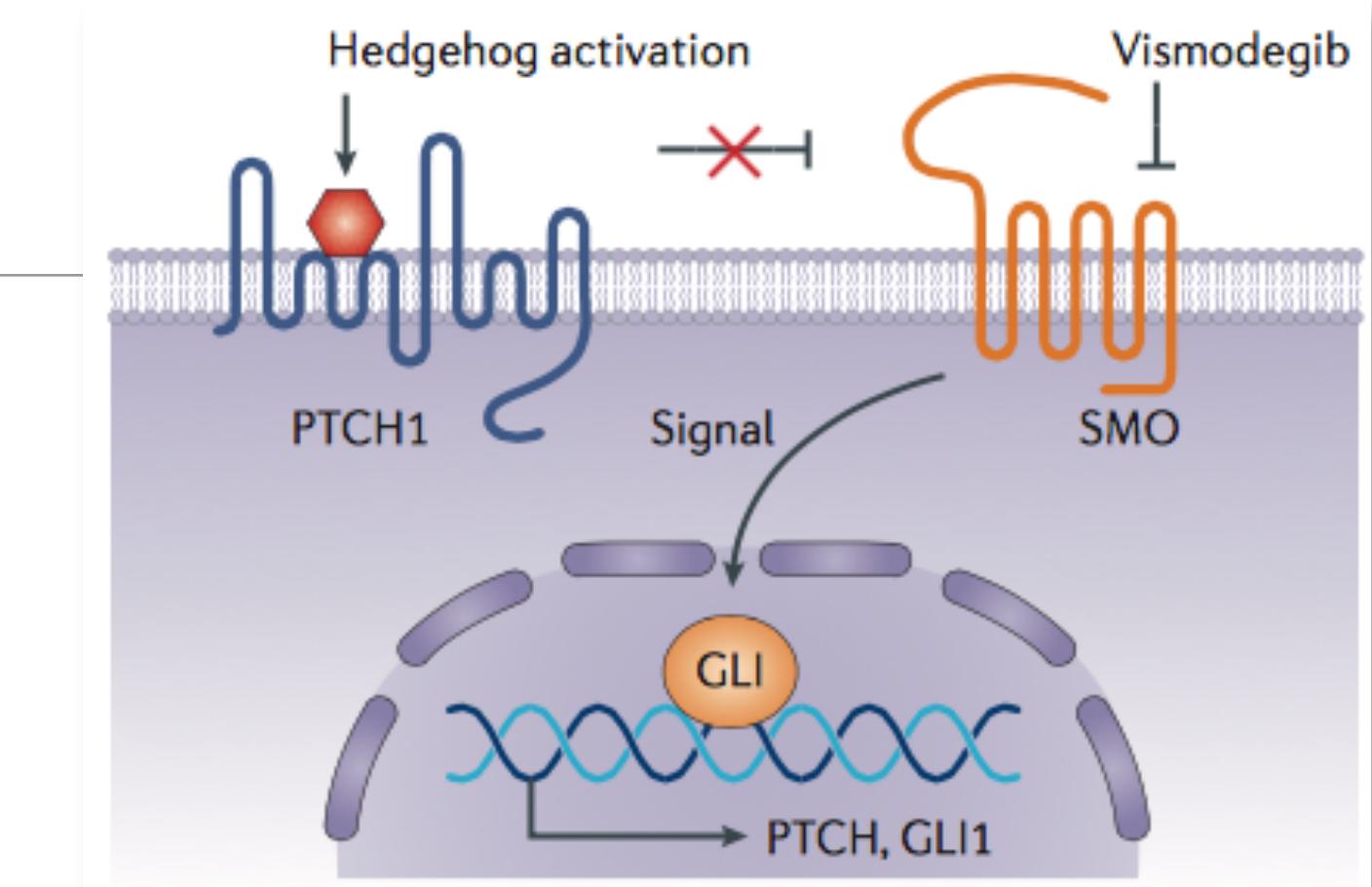
a current project addresses the focal involvement of all mapped genes, in >50'000 cancer genome profiles



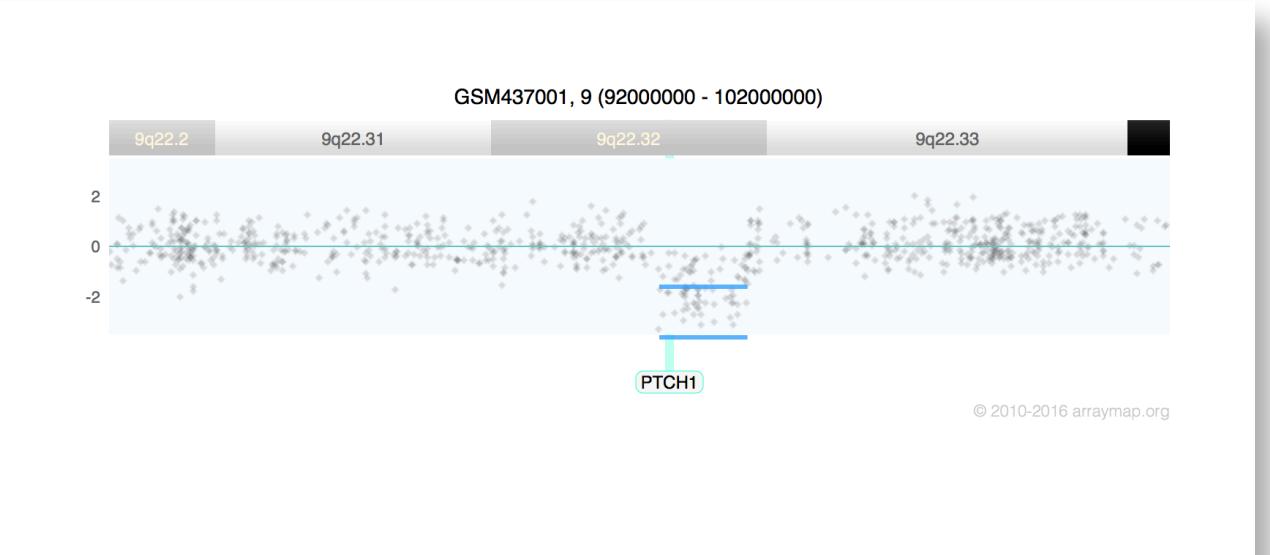
Summary of somatic copy number aberrations from the analysis of 1127 genome profiles of malignant melanomas, collected in our arraymap.org cancer genome resource. While PTCH1 does not represent a deletion hotspot, the genomic locus is part of larger deletions in ~25% of melanoma samples.



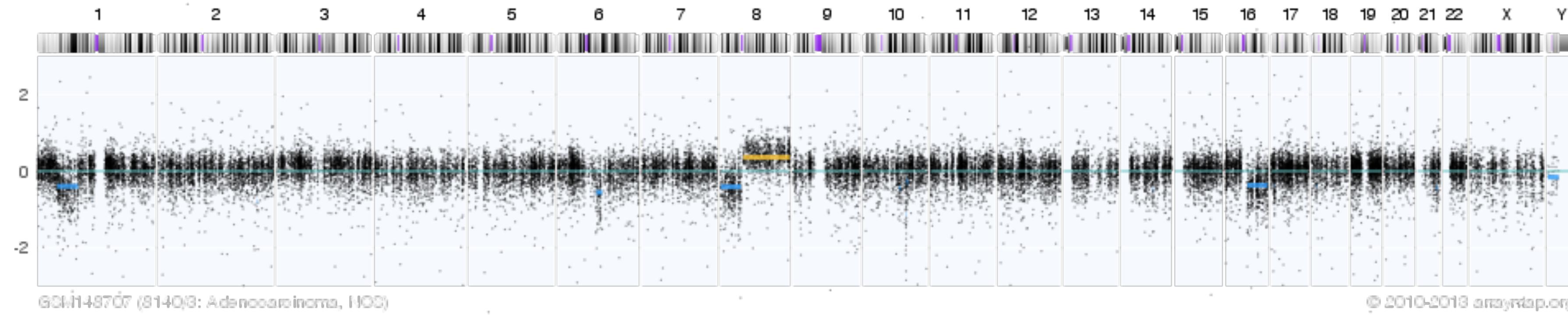
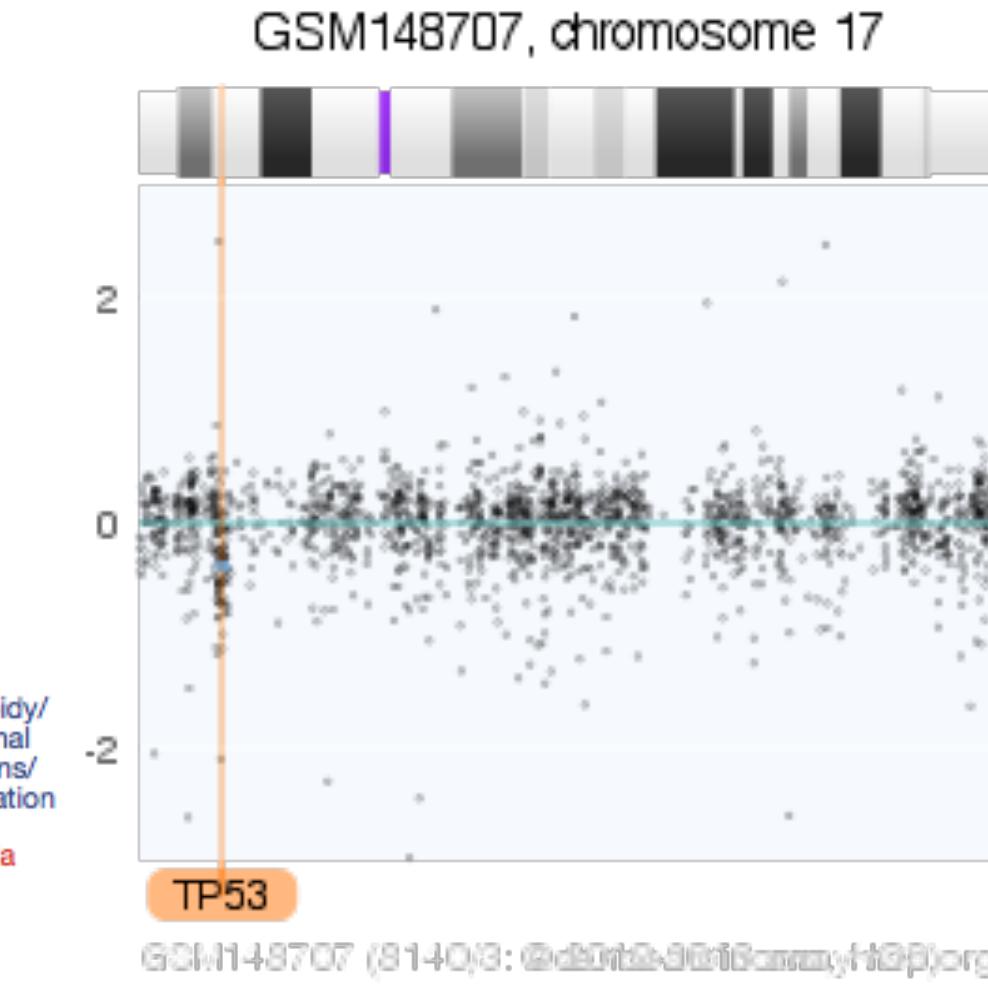
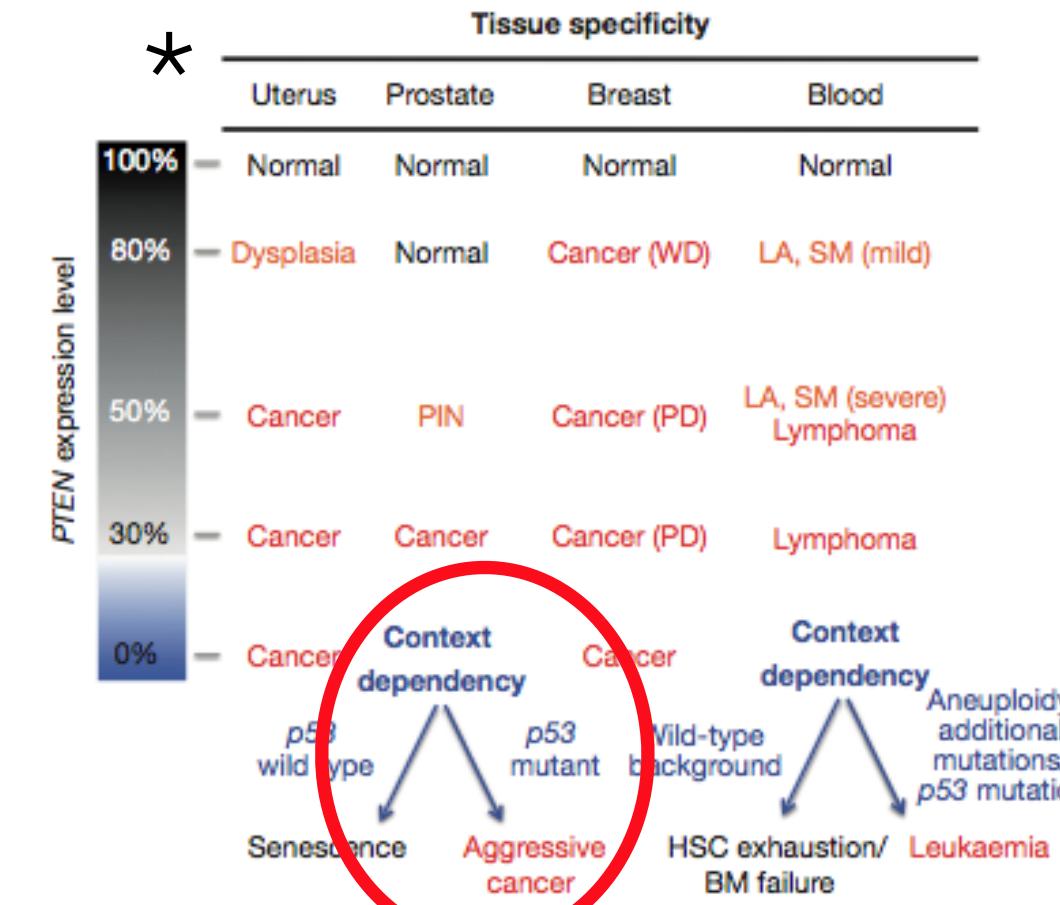
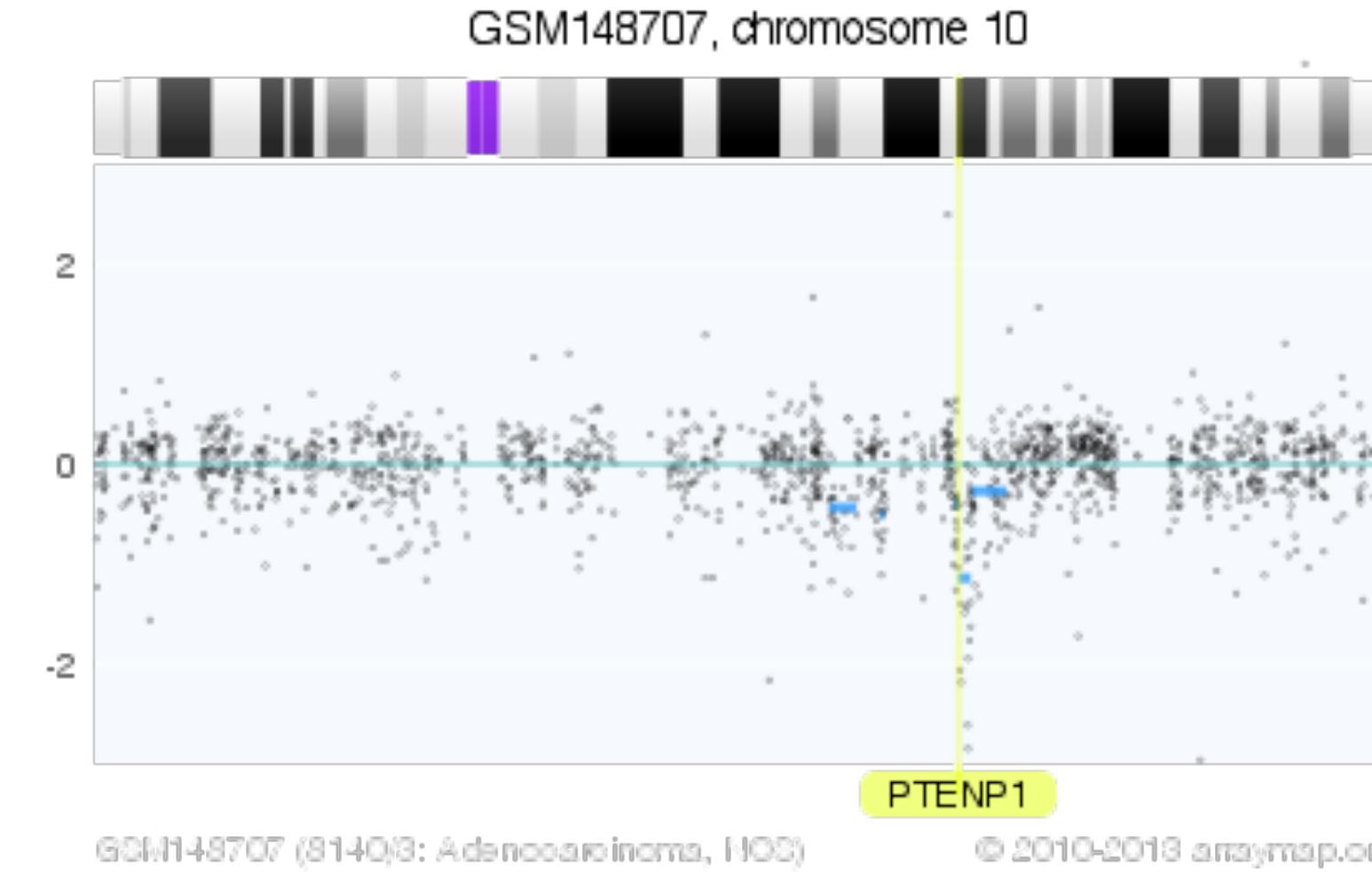
Examples of focal / homozygous PTCH1 deletions detected in the analysis of 1127 genomic array datasets. Focal somatic imbalance events are considered an indicator for oncogenic involvement of the affected target genes.



In its normal function, PTCH1 is a tumor suppressor gene in the sonic hedgehog pathway and inhibits SMO driven transcriptional activation. A loss of PTCH1 function (mutation, deletion) can be mitigated through drugs antagonistic to SMO activation.



Gene dosage phenomena beyond simple on/off effects



Combined heterozygous deletions involving *PTEN* and *TP53* loci in a case of prostate adenocarcinoma
(GSM148707, PMID 17875689, Lapointe et al., CancRes 2007)

* A. H. Berger, A. G. Knudson, and P. P. Pandolfi, "A continuum model for tumour suppression," *Nature*, vol. 476, no. 7359, pp. 163–169, Aug. 2011.

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 >93000 single sample CNV tracks from >1500 publications
- **aCGH**, cCGH, WES, WGS
- additionally tracking and annotating of publications reporting compatible original data (more than 3200 articles as of 2019)



cancer genome data @ progenetix.org

The Progenetix database provides an overview of copy number abnormalities in human cancer from currently **93640** array and chromosomal Comparative Genomic Hybridization (CGH) experiments, as well as Whole Genome or Whole Exome Sequencing (WGS, WES) studies. The cancer profile data in Progenetix was curated from **1547** articles and represents **402** and **469** different cancer types, according to the International classification of Diseases in Oncology (ICD-O) and NCIt "neoplasm" classification, respectively.

Additionally, the website attempts to identify and present all publications (currently **3230** articles), referring to cancer genome profiling experiments. The database & software are developed by the group of Michael Baudis at the University of Zurich.

Progenetix: Oropharyngeal Squamous Cell Carcinoma (ncit:C8181)

© 2019 progenetix.org

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. [\[PubMed\]](#)

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

Baudis, M. 2007. Genomic imbalances in 5918 malignant epithelial tumors: An explorative meta-analysis of chromosomal CGH data. *BMC Cancer* 7:226. [\[PubMed\]](#)

Baudis, M. 2006. Online database and bioinformatics toolbox to support data mining in cancer cytogenetics. *Biotechniques* 40, no. 3: 296-272. [\[PubMed\]](#)

Baudis, M, and ML Cleary. 2001. Progenetix.net: an online repository for molecular cytogenetic aberration data. *Bioinformatics* 12, no. 17: 1228-1229. [\[PubMed\]](#)

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.

© 2000 - 2019 Michael Baudis, refreshed 2019-10-03T13:27:34Z in 3.00s 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-O
Locus
NCIt
?

arrayMap

Accessing Probe-Level Genomic Array Data in Cancer



Search Samples

Search Publications

Progenetix



Citation & Licensing

User Guide

People

Beacon+



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:

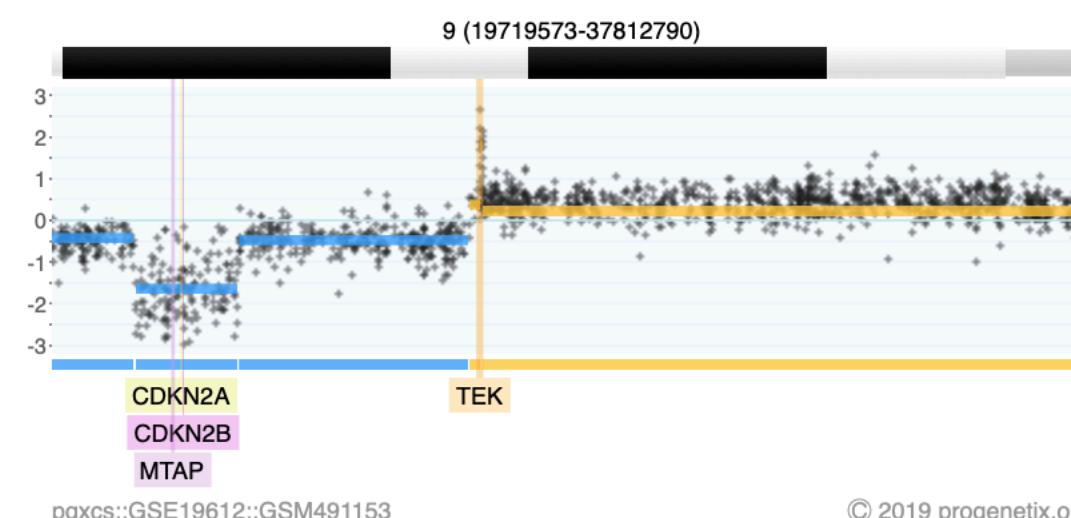
72724 genomic array profiles

898 experimental series

257 array platforms

341 ICD-O cancer entities

795 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.

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.

© 2000 - 2019 Michael Baudis, refreshed 2019-06-12T21:00:19Z in 6.00s 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.

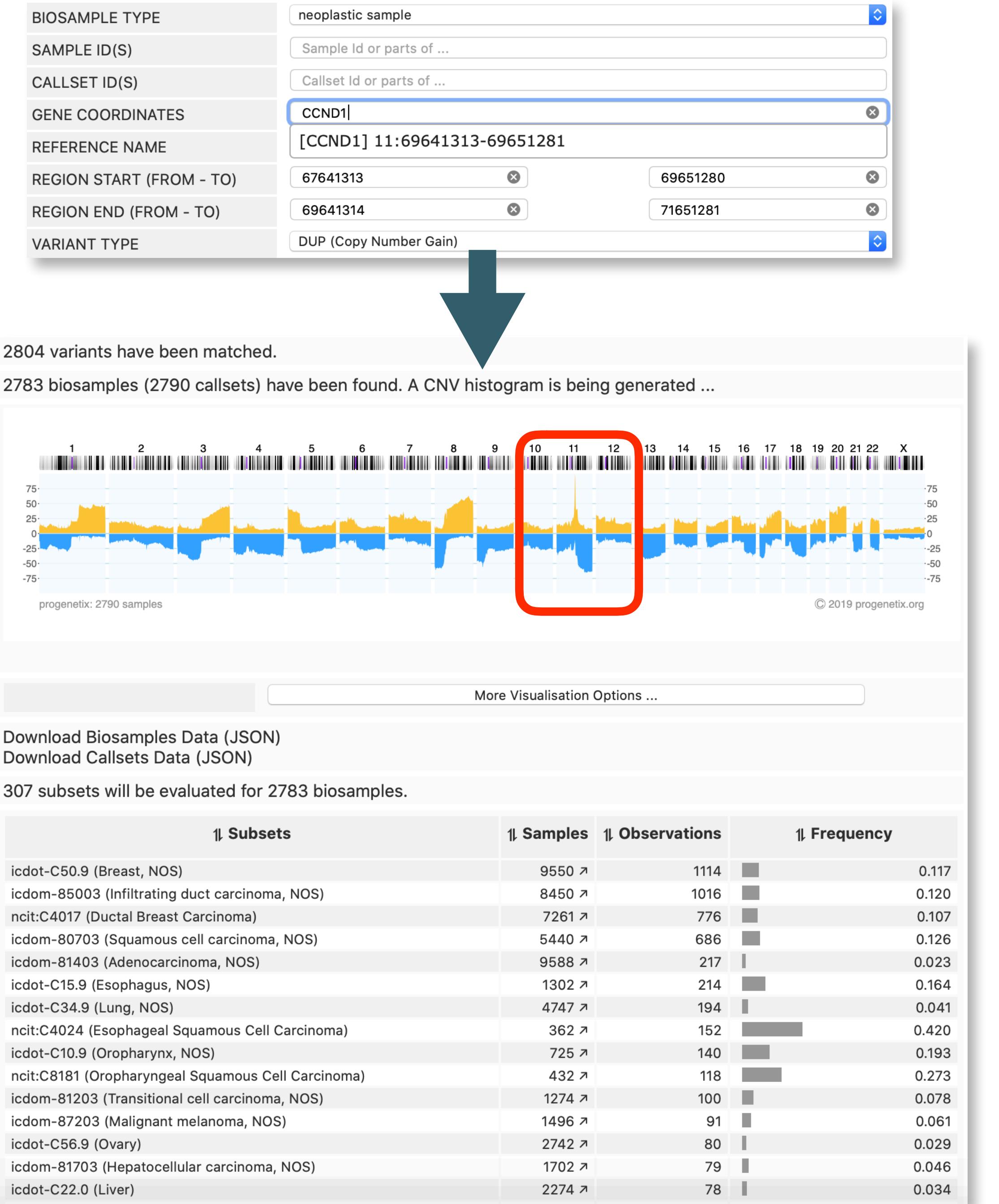


arrayMap



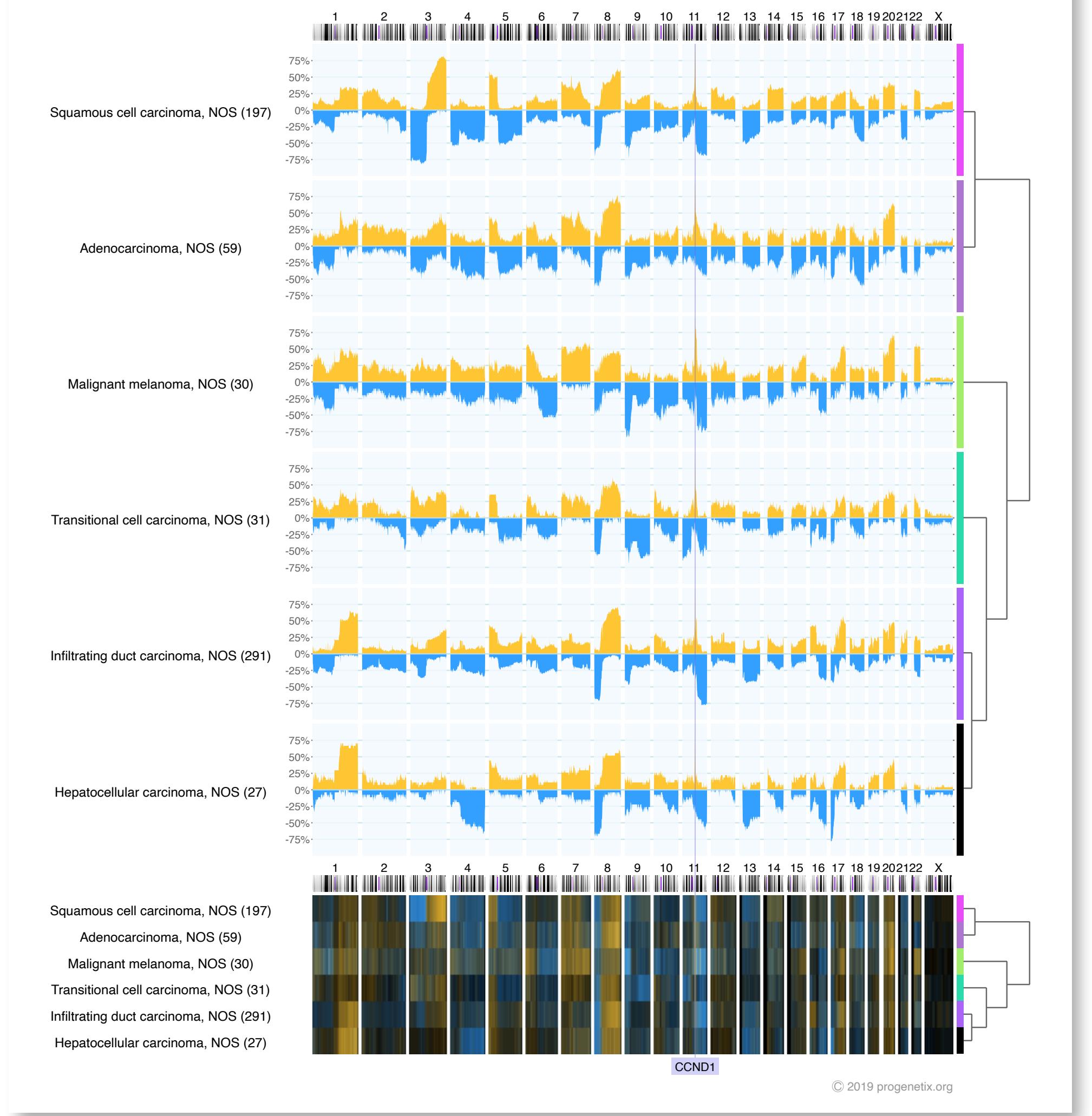
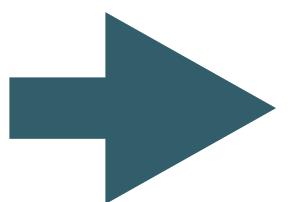
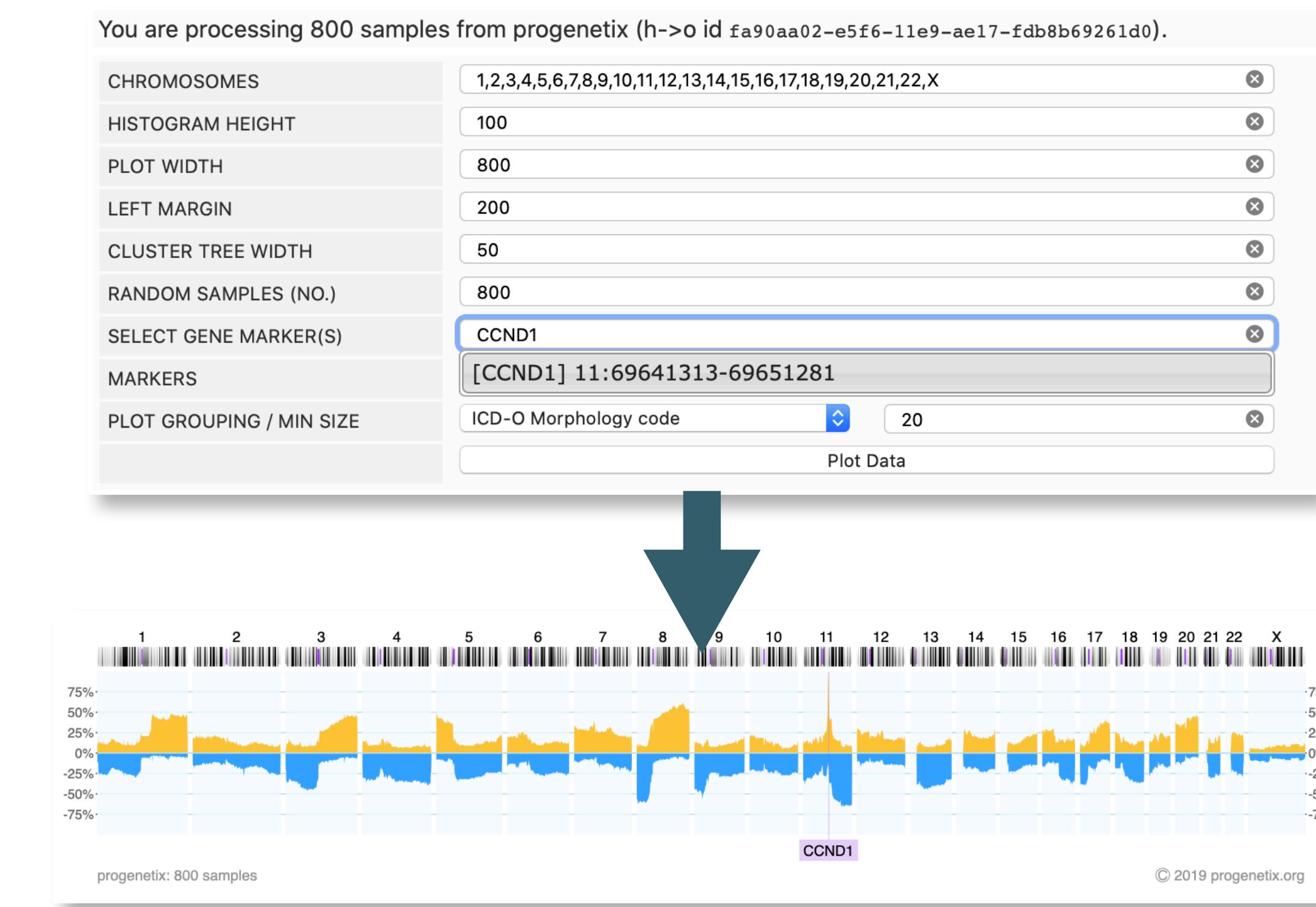
Progenetix - Reference Resource for Oncogenomic Profiling Data

- 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?")



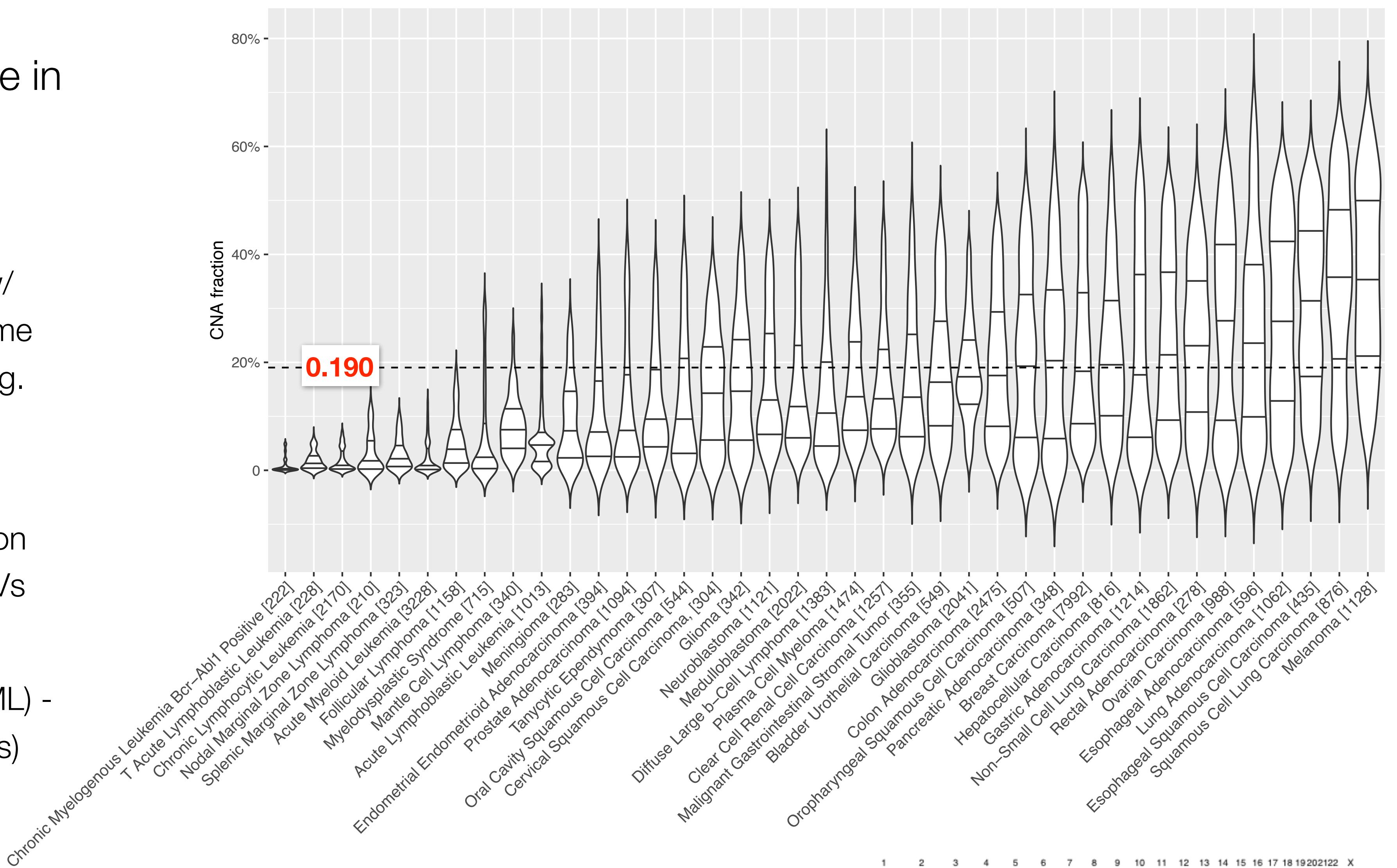
Progenetix - Reference Resource for Oncogenomic Profiling Data

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

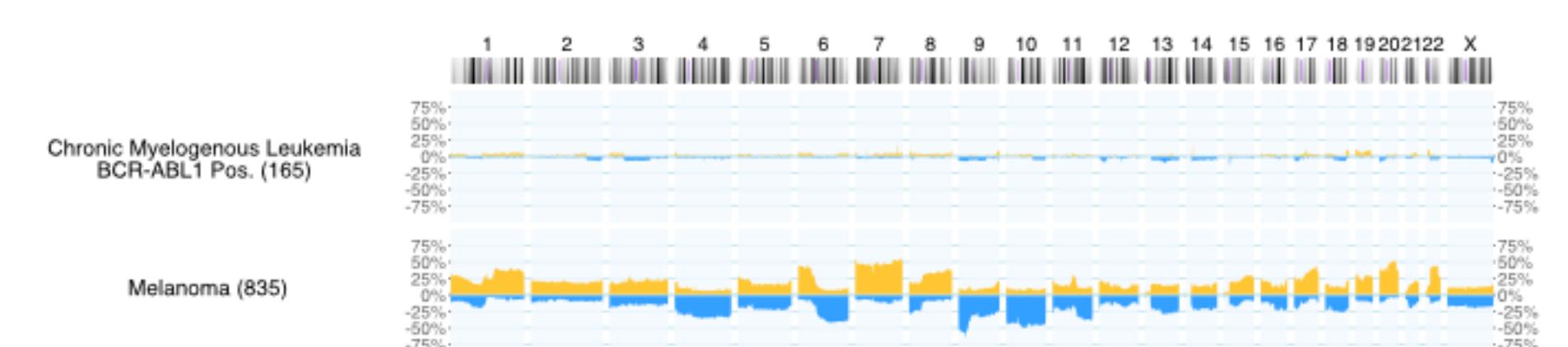


Genome CNV coverage in Cancer Classes

- 43654 out of 93640 CNV profiles; filtered for entities w/ >200 samples (removed some entities w/ high CNV rate, e.g. sarcoma subtypes)
- Single-sample CNV profiles were assessed for the fraction of the genome showing CNVs (relative gains, losses)
- range of medians 0.001 (CML) - 0.358 (malignant melanomas)



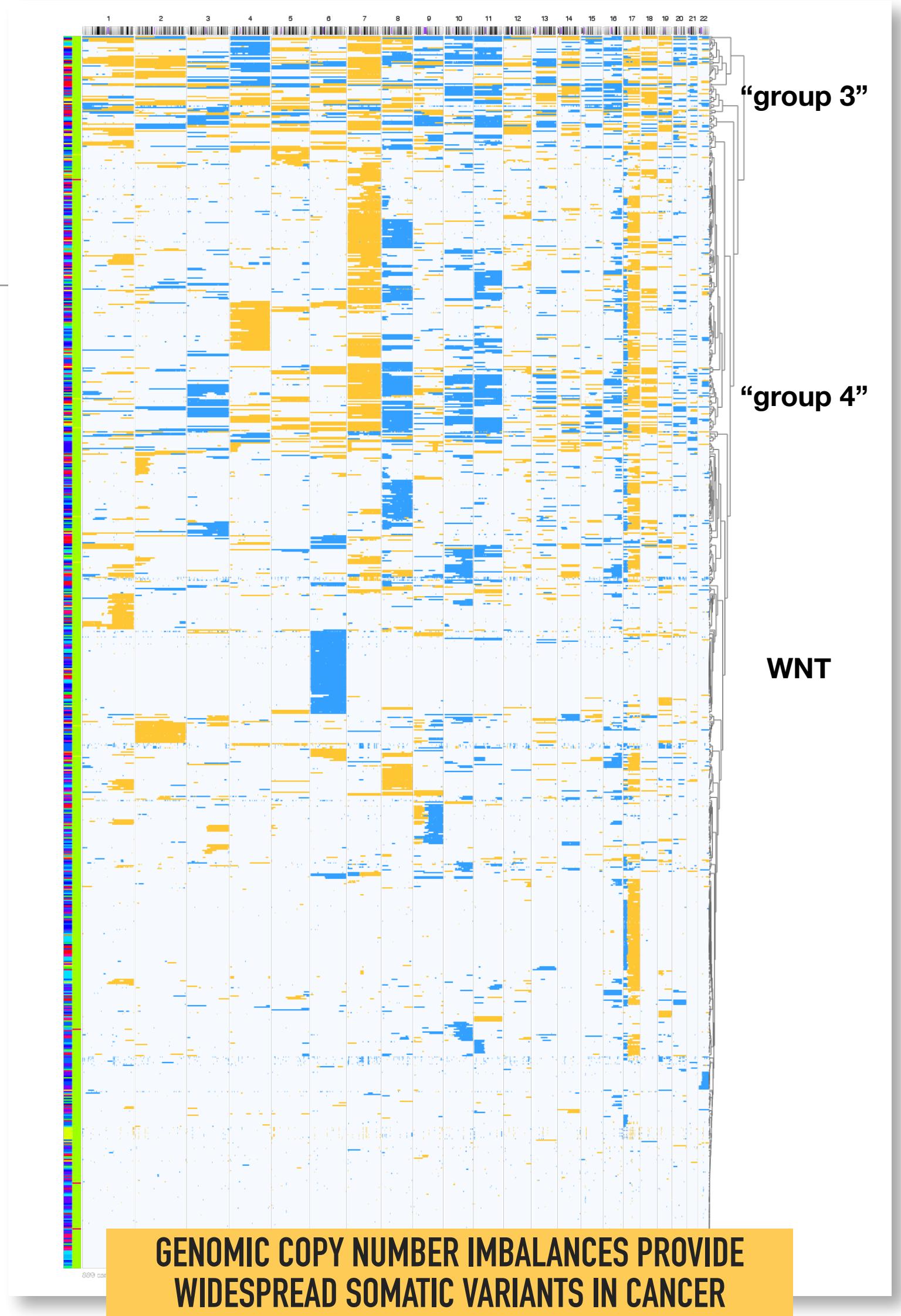
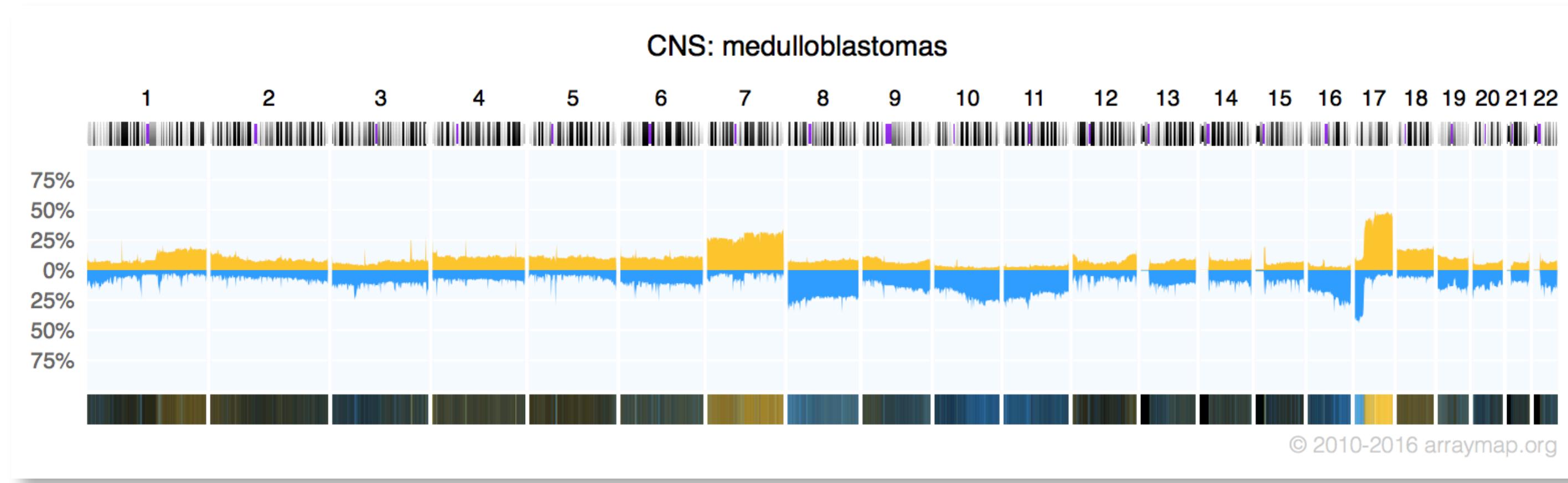
Lowest / Highest CNV fractions =>



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?



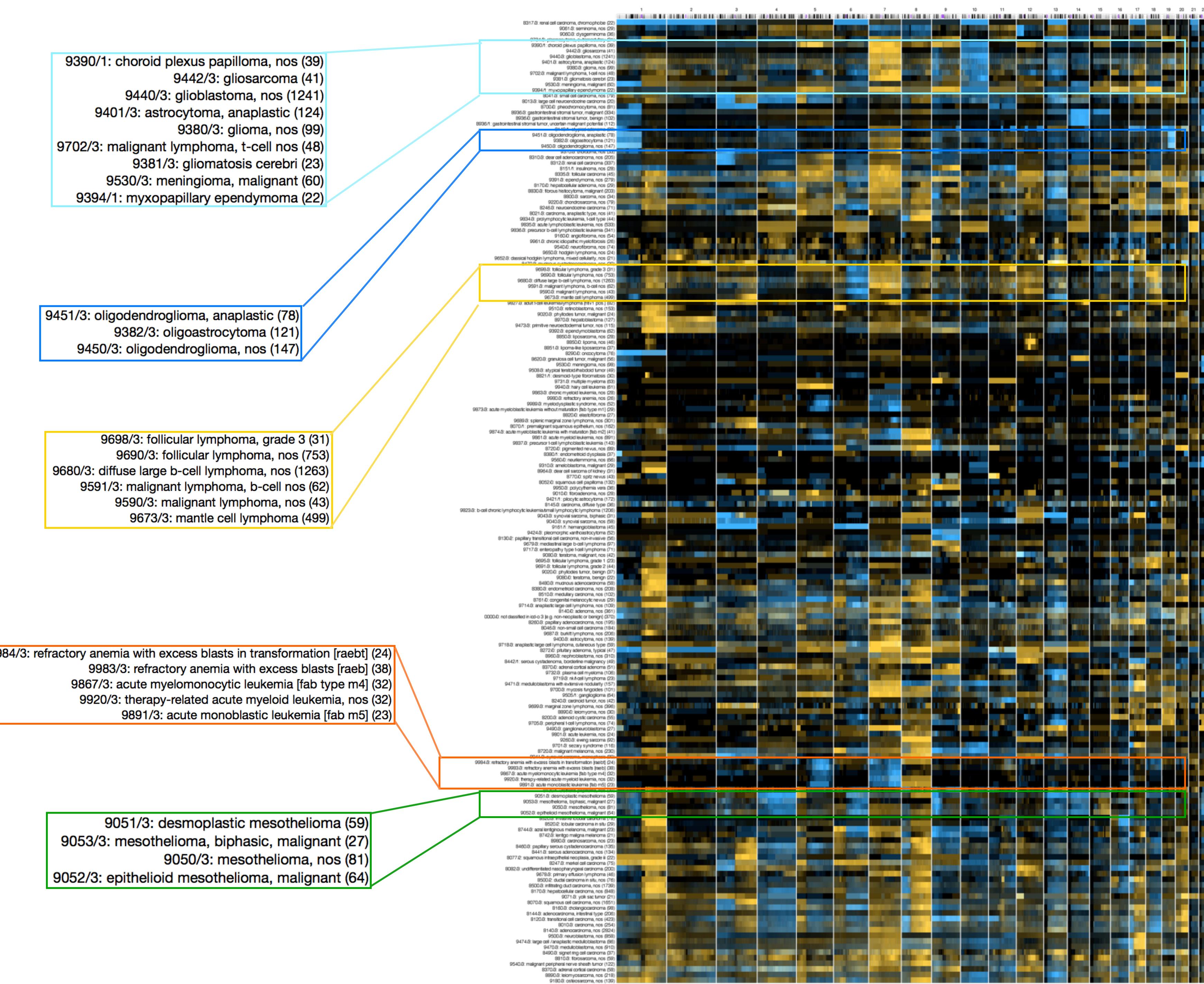
A genomic copy number histogram for malignant medulloblastomas, the most frequent type of pediatric brain tumors, displaying regions of genomic duplications and deletions. These can be decomposed into individual tumor profiles which segregate into several clusters of related mutation patterns with functional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



Somatic Mutations In Cancer: Patterns

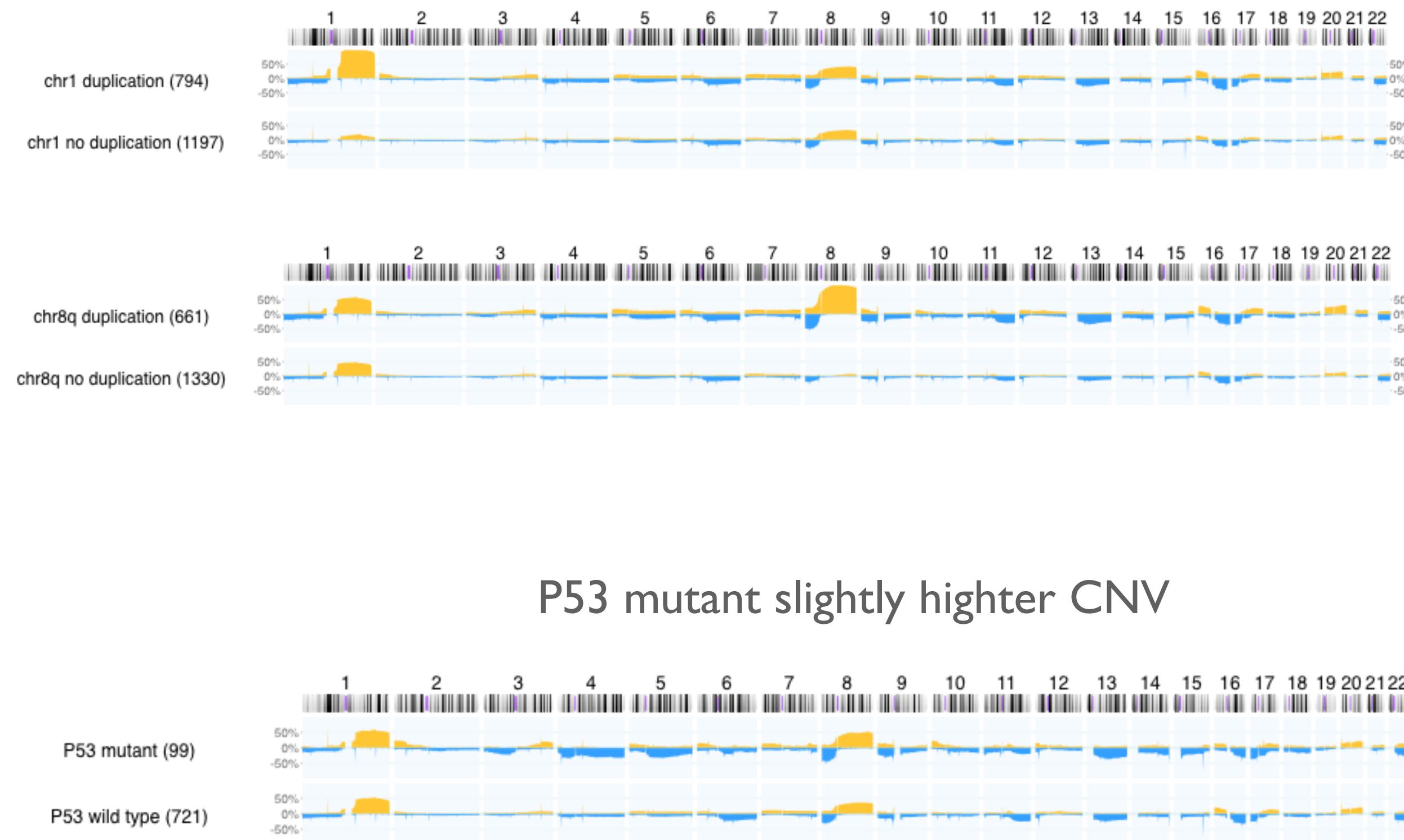
Making the case for genomic classifications

Some related cancer entities show similar copy number profiles

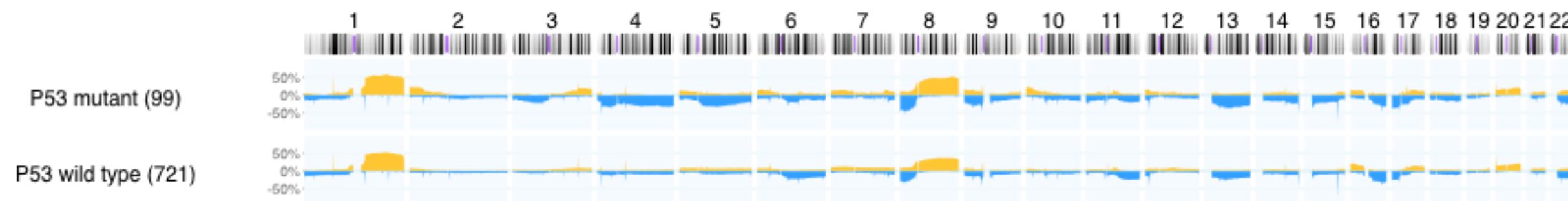


Interpret CNV by Association to Molecular/Clinical Information

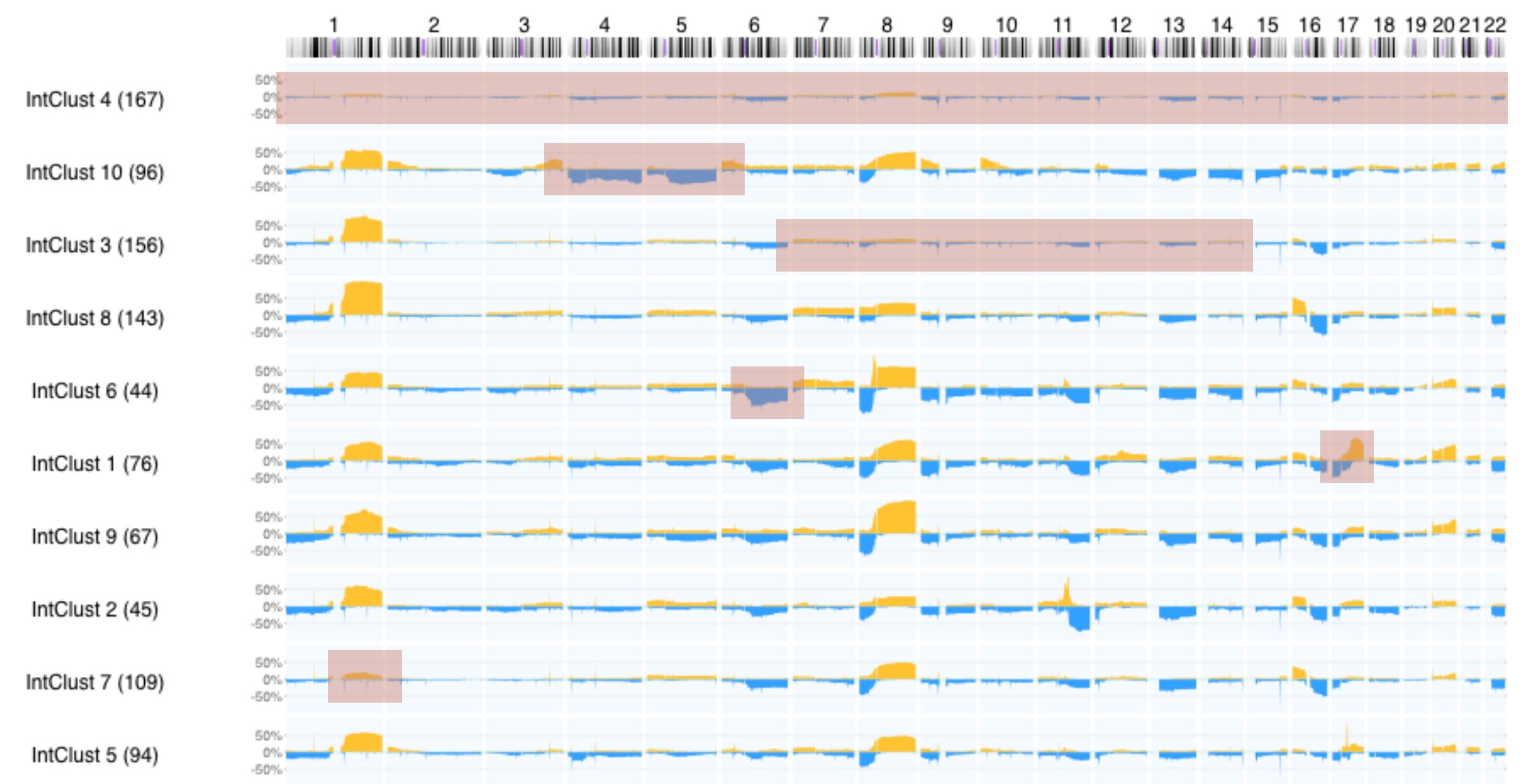
Chr1q and chr8q duplication unlinked



P53 mutant slightly higher CNV



Some IntClust groups define Distinctive CNV patterns



METABRIC

Interpret CNV by Association to Molecular/Clinical Information

ER status

ER neg (440)

ER pos (1508)

not specified (44)

HER2 level

HER2 level 0 (5)

HER2 level 2 (27)

HER2 level 3 (121)

not specified (1168)

HER2 level 1 (671)

Pam50 category

→ Pam50 Normal (202)

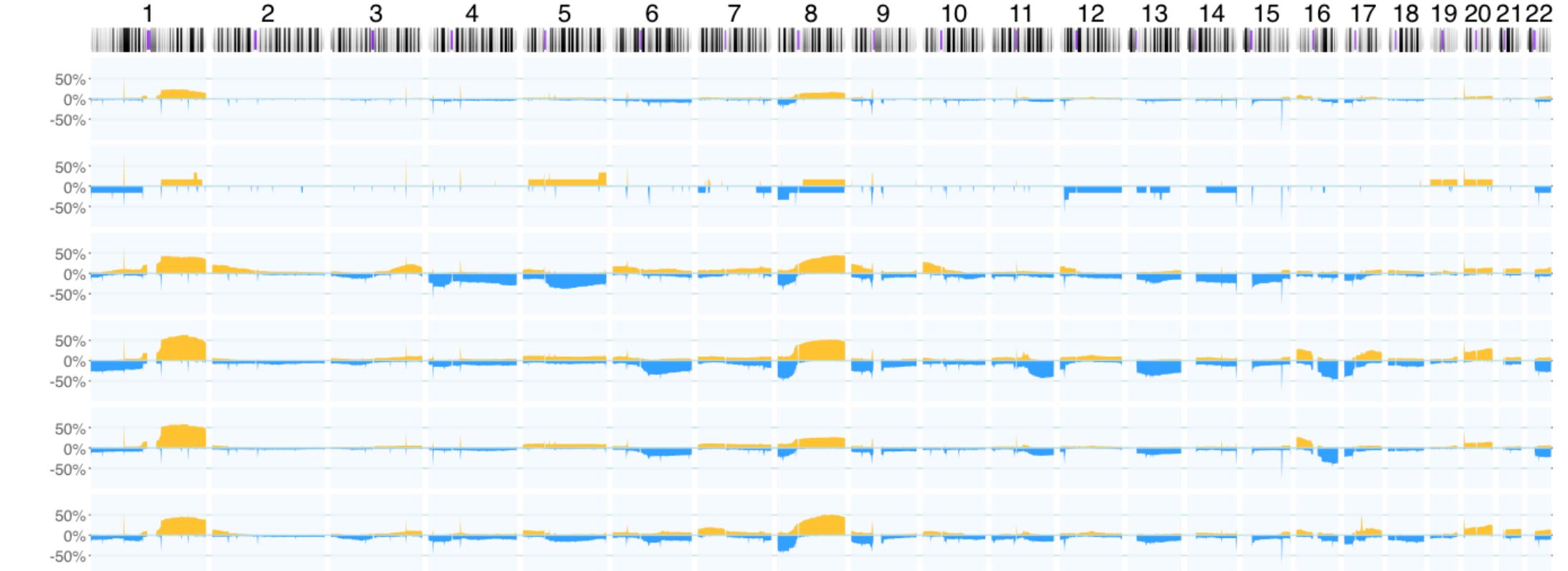
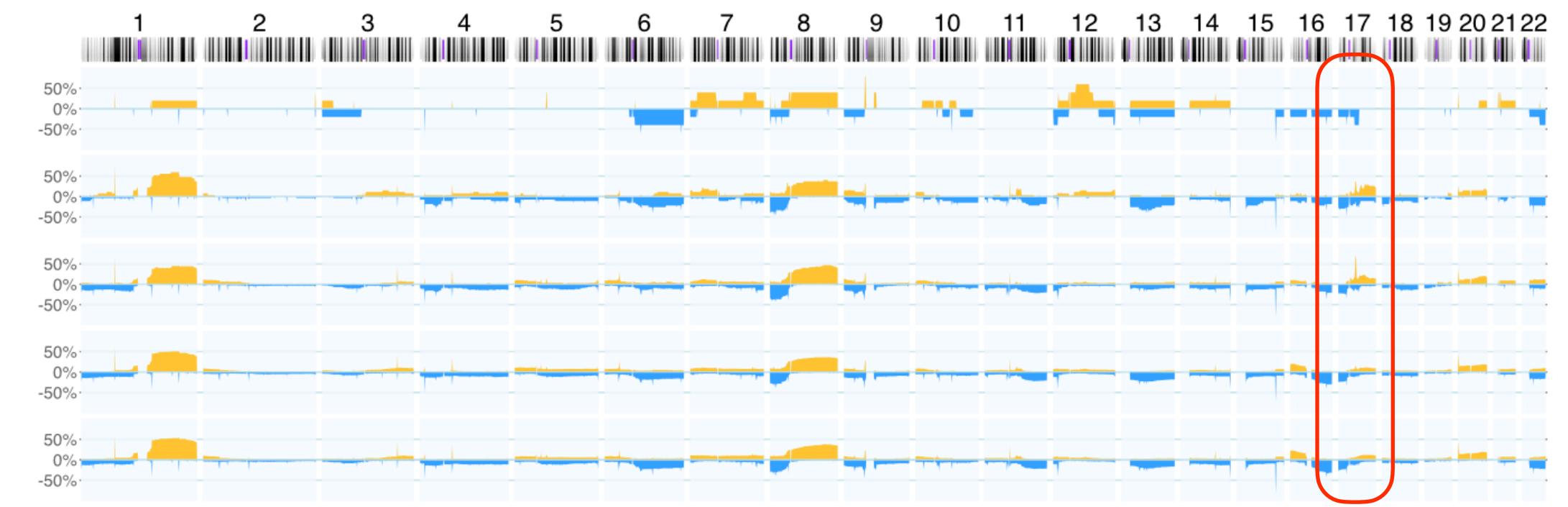
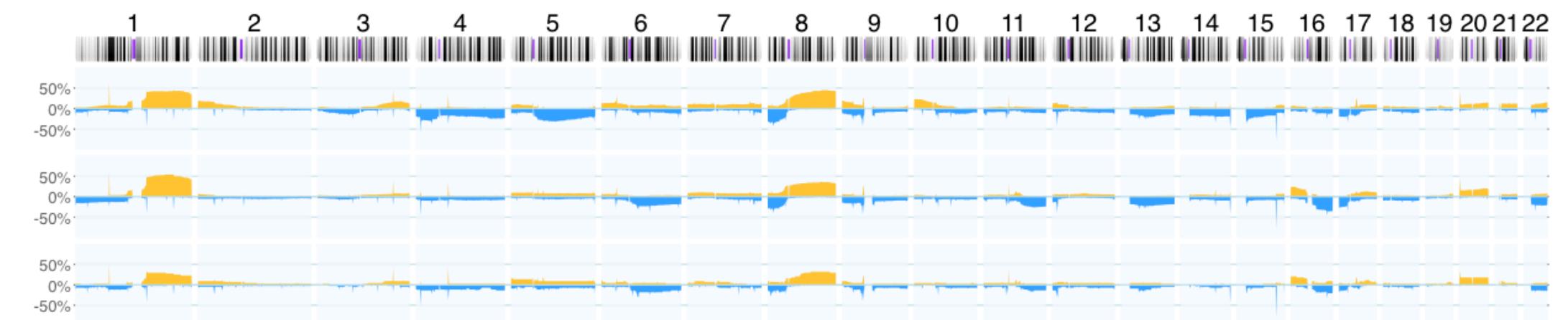
not specified (6)

Pam50 Basal (331)

Pam50 LumB (492)

Pam50 LumA (721)

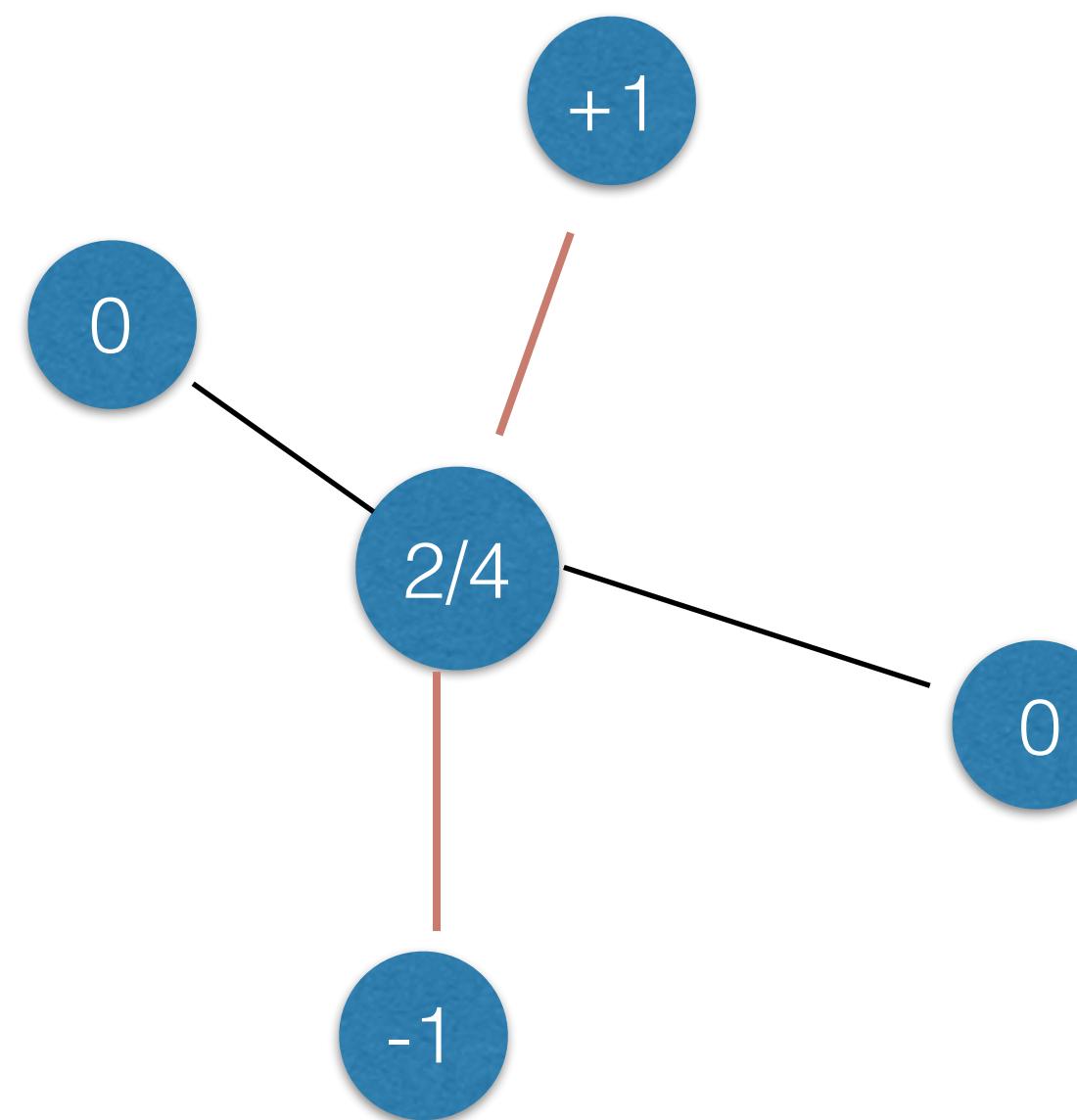
Pam50 Her2 (240)



Interpret Gene-Level CNV With Protein Networks

Highly Connected Genes Have Similar CNV Patterns as Canonical Driver Genes?

Gene CNV score



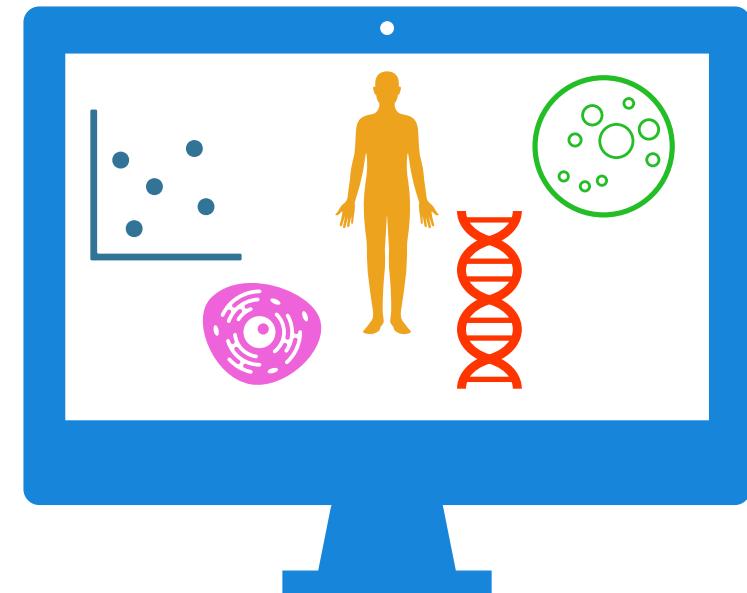
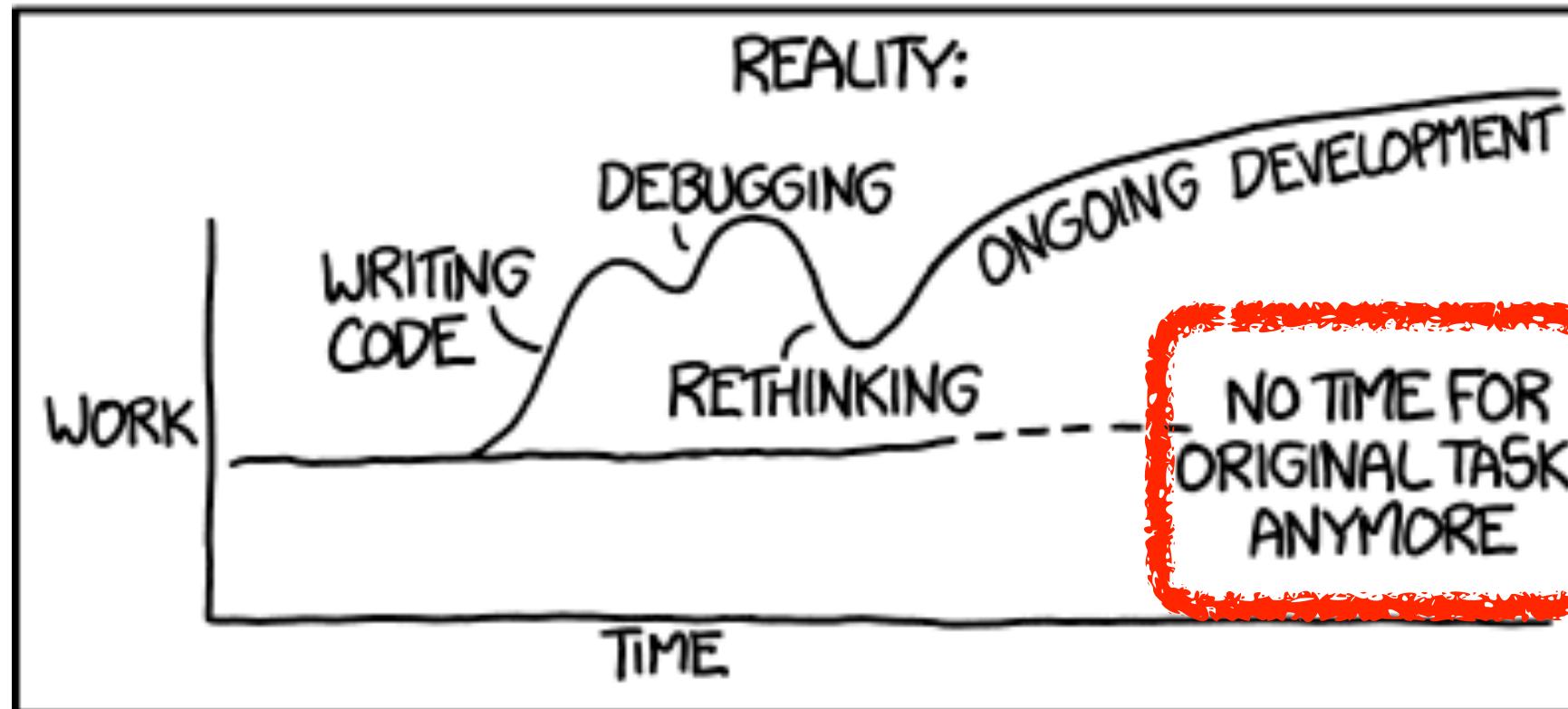
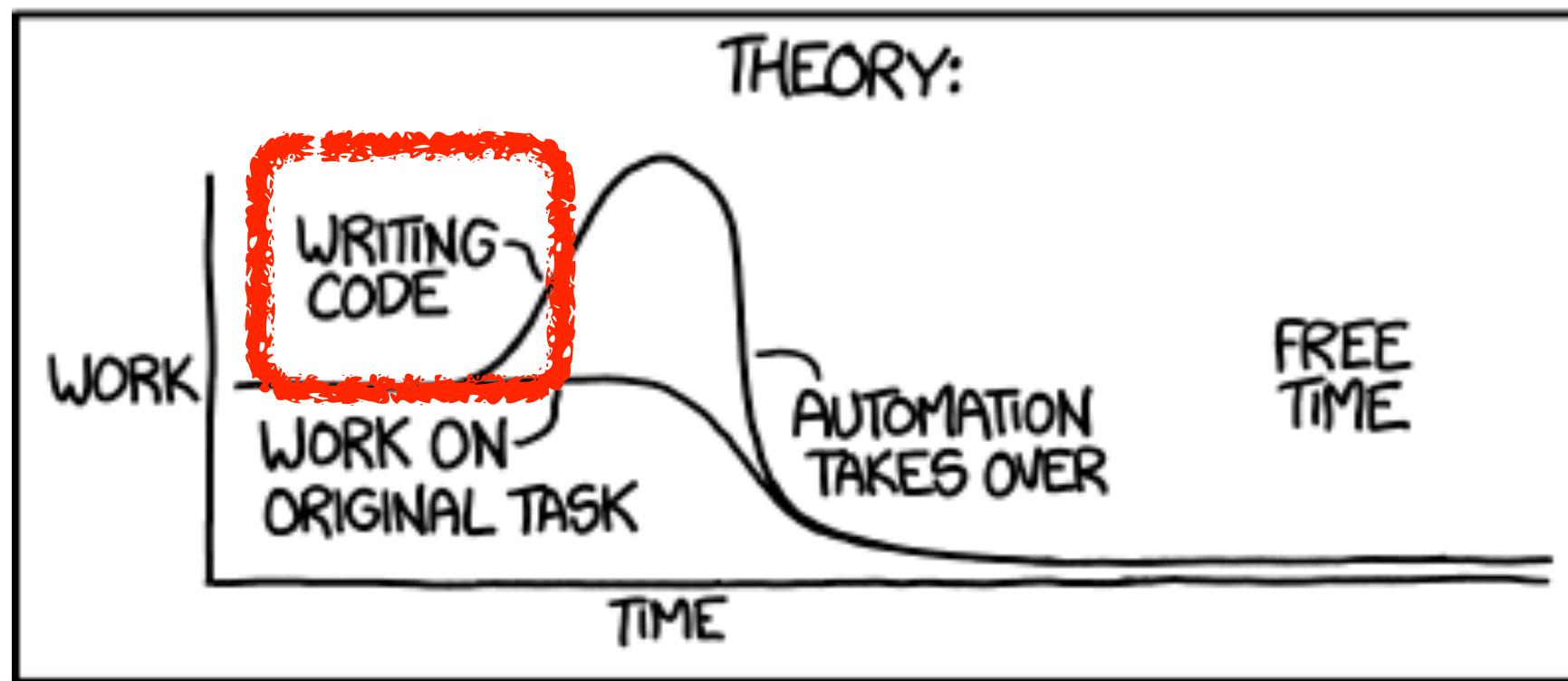
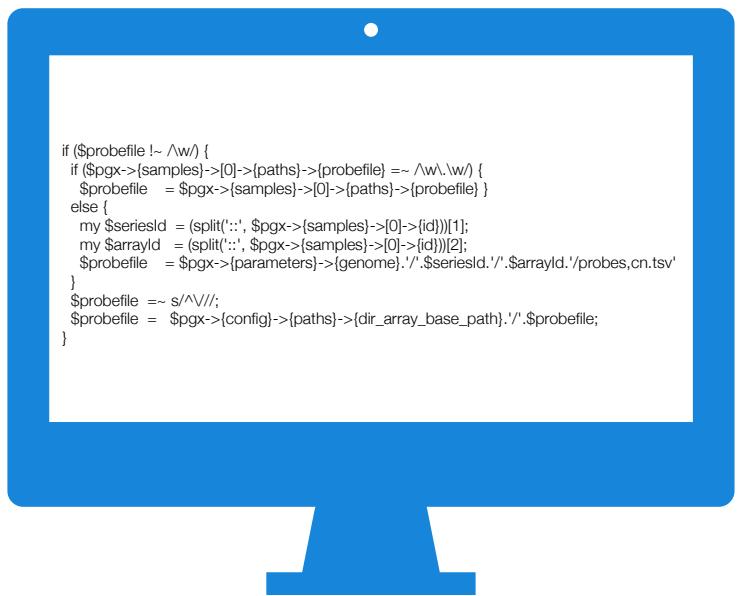
- Incorporate protein network
- Define gene CNV score by partner CNV status
- Driver genes distinguish from rest
- Correct for connectivity → effect is gone

{bio_informatics_science}

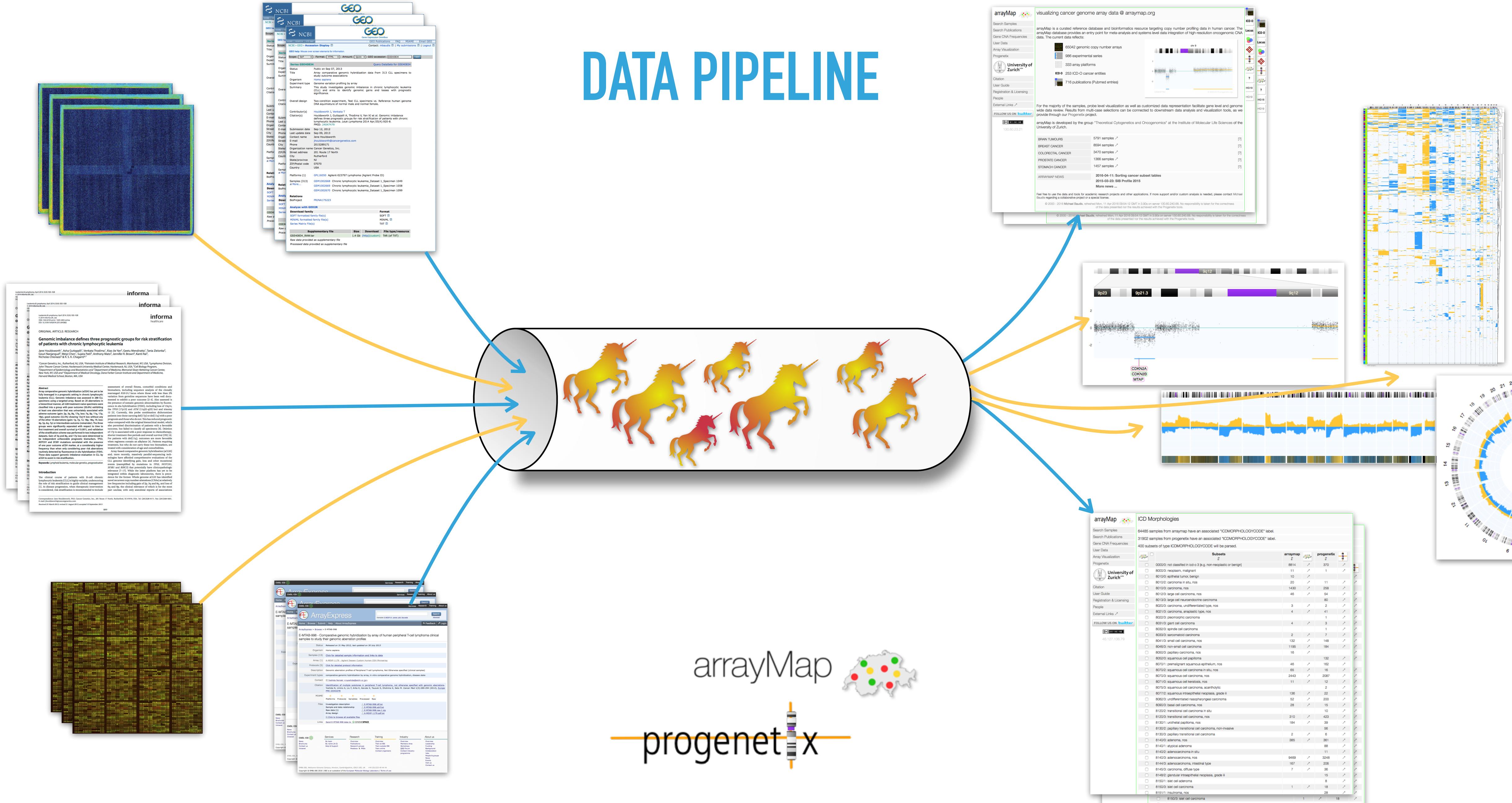


{bio_informatics_science}

"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



DATA PIPELINE



DATA PIPELINE

BIOCURATION BIOINFORMATICS



NCBI GEO Accession Display

Series GSE640034 Public on Sep 07, 2013

Organism: Human

Experiment type: Genomic variation profiling by array

Summary: This study investigates genomic variation in chronic lymphocytic leukemia (CLL) specimens with prognostic significance.

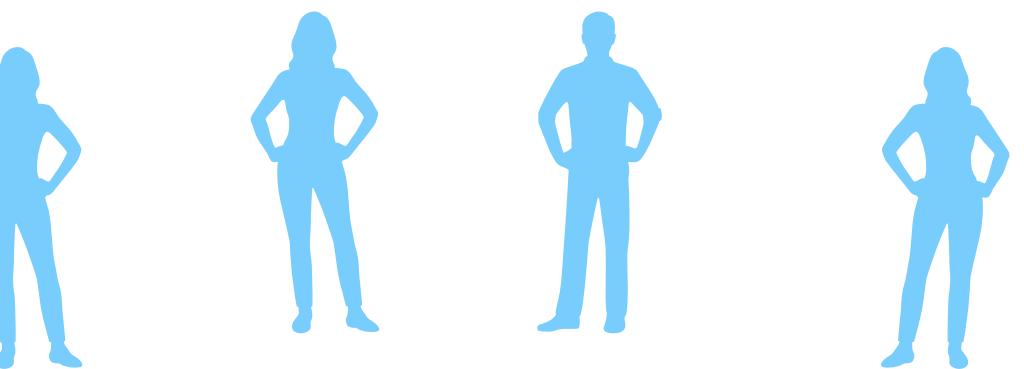
Overall design: Overall design experiment, Test vs. Specimens vs. Reference human genome

Contributor(s): Houldsworth J, Venkata T, Guttagji A, Thoduri V, Yan XI et al.

Sample ID: GSE640034

Platform: Agilent G1317P Lymphoma (Agilent Probe ID)

Supplementary file: GSE640034.RAW.tar



Informa Healthcare

Original Article Research

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

Jane Houldsworth¹, Asha Guttapalli¹, Venkata Thoduri¹, Xiao Jie Yan¹, Geeta Mendiratta¹, Tamja Zelenka², Gouri Nangisetty², Wei Chen², Supratik Pati², Anthony Mato², Jennifer R. Brown², Kari Rair²

¹Cancer Genetics, Inc., Rutherford, NJ, USA; ²Weinstein Institute of Medical Research, Manhattan, NY, USA; ³Lymphoma Division, Department of Epidemiology and Biostatistics and ⁴Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA; ⁵Department of Pathology, ⁶Department of Oncology, David Helfer Cancer Institute and Department of Medicine, Harvard Medical School, Boston, MA, USA

Abstract: Aberrant chromosomal imbalances have been fully leveraged in a prognostic setting in chronic lymphocytic leukemia (CLL). However, the prognostic value of genomic variation has not been fully explored. We used array-based genomic variation profiling to identify novel prognostic markers. We analyzed 20 CLL specimens using a targeted array. Based on 20 aberrations in each specimen, we identified a group with low risk and a group with high risk. The high-risk group had a significantly higher first treatment and overall survival ($p < 0.001$), and validation of this group in an independent cohort showed similar results. Genes of 9q loss and 17q gain were determined to be prognostic. In addition, we found that patients with TP53 (17q10) and ATM (11q13) gain and intensity of deletion at 11q13 and 17q10 were associated with overall survival. TP53 and ATM mutations correlated with the presence of 9q gain. TP53 mutations were associated with a higher frequency when regions contain an allele. Patients requiring chemotherapy had a significantly lower overall survival compared with those not requiring chemotherapy. These data support genomic imbalance evaluation in CLL by assessment of several genes, constitutional conditions and biomarkers, including sequence analysis of the clinically relevant genes ATM and TP53. The prognostic value of genomic variation from genomic sequence has been well documented. Our findings indicate that genomic variation may be a valuable tool for CLL prognosis.

ArrayExpress

E-MTAB-998 Comparative genomic hybridization array of human peripheral T-cell lymphoma clinical samples to study their genomic aberration profiles

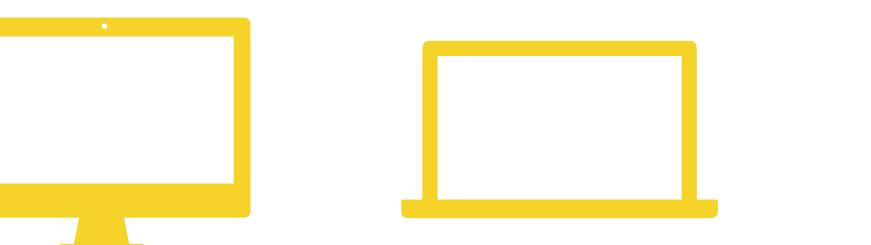
Organism: Homo sapiens

Experiment type: Comparative genomic hybridization array of human peripheral T-cell lymphoma, not otherwise specified (clinical sample)

Platform: Agilent G1317P Lymphoma (Agilent Custom Human CLL Microarray)

Sample ID: E-MTAB-998

Platform ID: E-MTAB-998



arrayMap
progenetix

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 entry point for meta-analysis and systems level integration of high-resolution oncogenomic DNA data. The current data reflects:

- 65024 genomic copy number arrays
- 985 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 Prognetic project.

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

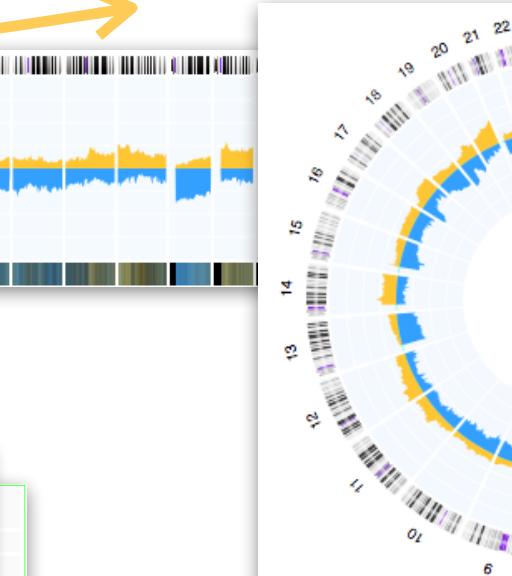
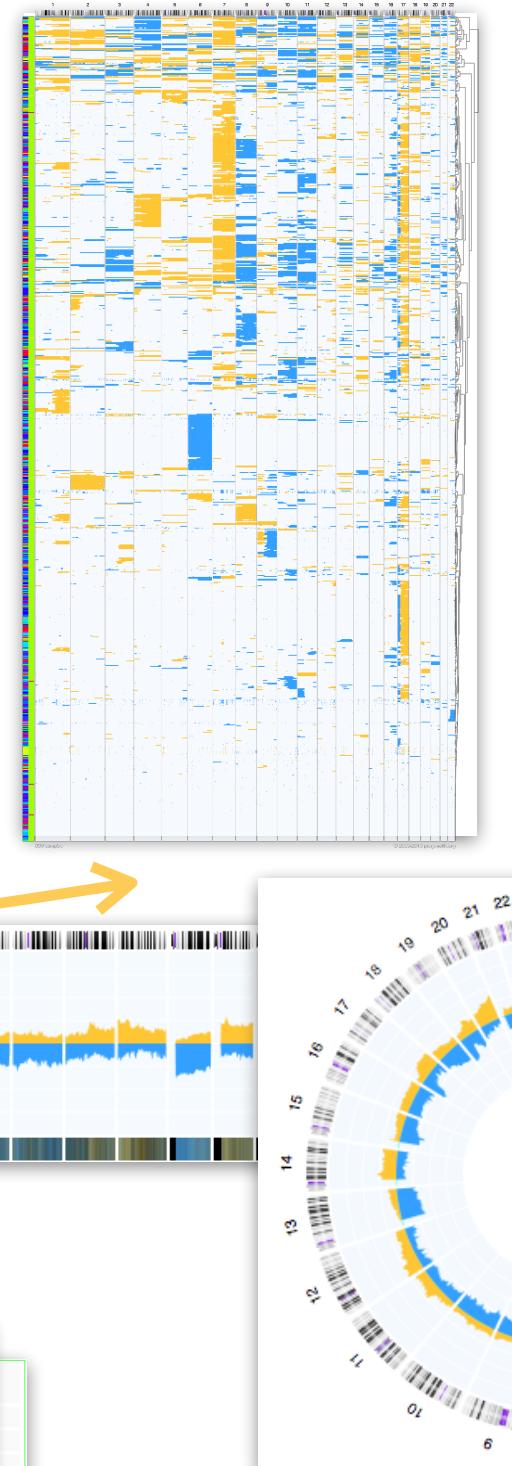
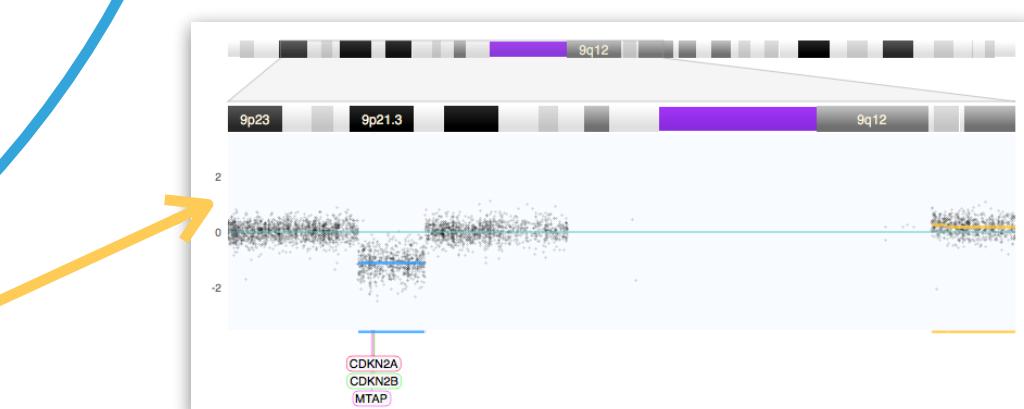
ICD-O

ICD-O-0

Locus

HG18

HG19



arrayMap

ICD Morphologies

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

31922 samples from progenetix have an associated "ICDMORPHOLOGYCODE" label.

400 subsets of type ICDMORPHOLOGYCODE will be parsed.

Subsets

	arraymap	progenetix
00000: not classified in icd-3 (e.g. non-neoplastic or benign)	8614	370
00003: neoplasm, malignant	11	1
00100: epithelial tumor, benign	10	1
00102: carcinoma, nos	20	11
00120: large cell carcinoma, nos	1430	258
00200: squamous cell carcinoma, nos	46	54
00203: carcinoma, undifferentiated type, nos	3	2
00210: carcinoma, anaplastic type, nos	4	41
00220: giant cell carcinoma	4	1
00303: spindle cell carcinoma	1	1
00333: sarcomatoid carcinoma	2	7
00413: small cell carcinoma, nos	132	148
00500: mesothelioma, nos	1195	184
00503: papillary carcinoma, nos	16	1
00701: meningothelial squamous epithelium, nos	46	162
00702: squamous cell carcinoma, nos	65	16
00703: squamous cell carcinoma, nos	2443	2087
00707: squamous cell carcinoma, nos	11	12
00750: squamous cell carcinoma, acantholytic	136	22
00800: peritoneal mesothelioma, epithelial grade ii	52	200
00900: basal cell carcinoma, nos	28	15
01200: transitional cell carcinoma, nos	310	423
01300: uterine papilloma, nos	184	39
01302: papillary transitional cell carcinoma, non-invasive	56	1
01303: papillary transitional cell carcinoma	2	6
01400: basal cell carcinoma, nos	385	361
01401: acral melanoma	88	1
01402: adenocarcinoma, nos	11	1
01403: adenocarcinoma, in situ	9469	3248
01443: adenocarcinoma, intestinal type	167	206
01450: carcinoma, diffuse type	7	36
01500: breast cell adenoma	15	1
01501: breast carcinoma	8	1
01510: breast carcinoma, nos	1	18
01511: breast carcinoma	1	28
01512: insular carcinoma	29	29

University of Zurich

Citation

User Guide

Registration & Licensing

People

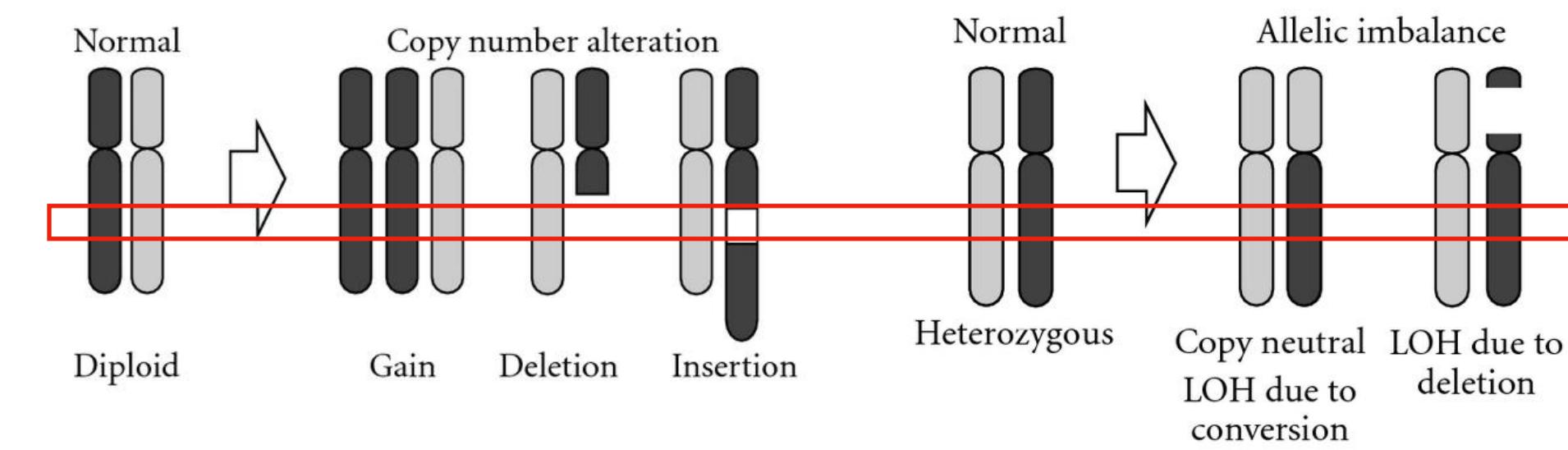
External Links

FOLLOW US ON [twitter](#)

Signal noise in copy number profiling

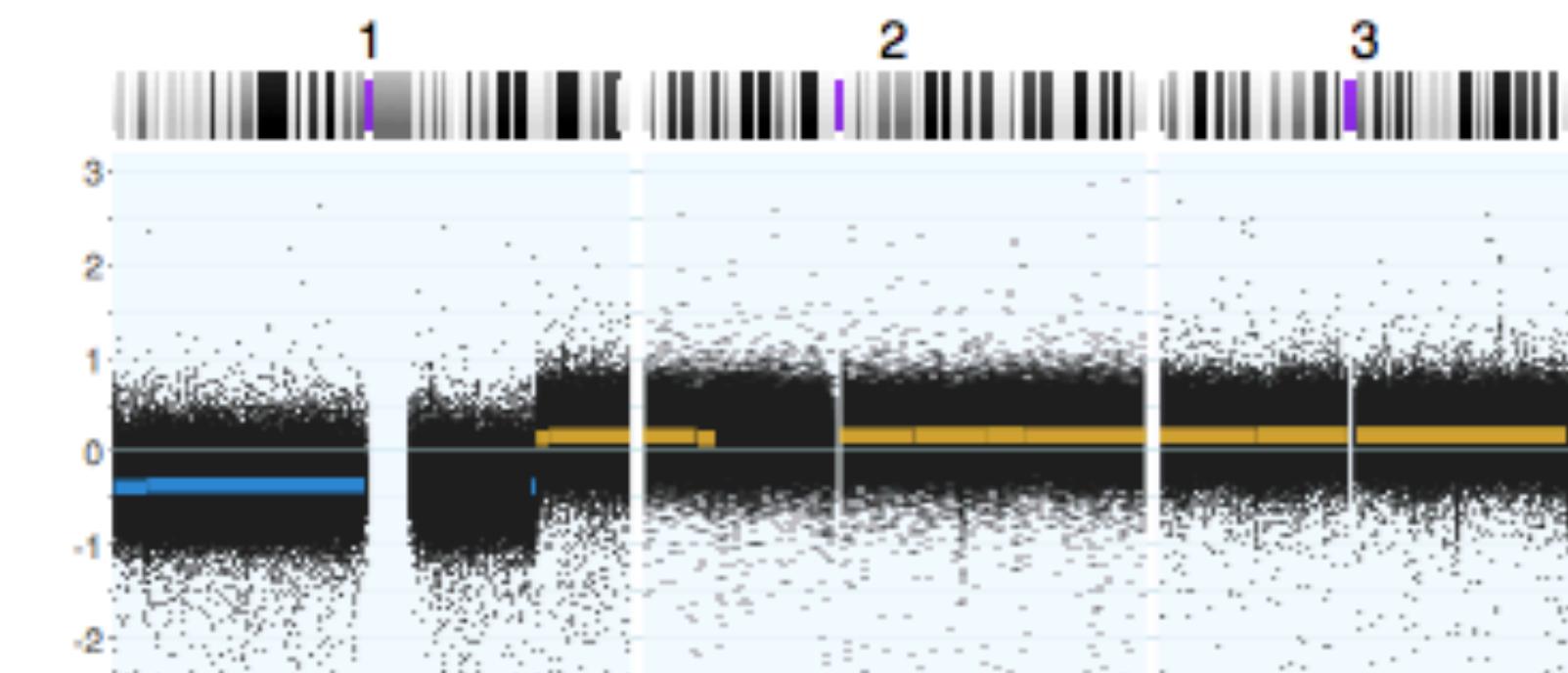
- Actual DNA copy numbers per cell/clone are integers.
- Cancer copy number profiles frequently cannot be interpreted with a simple "integer" model.
- CN profiles difficult to compare:
 - Every sample has its own noise level
 - Every sample has its own signal scale
 - Ambiguity (aneuploidy, subclones)

Regardless mutation types, the copy number of DNA segment should always be an integer

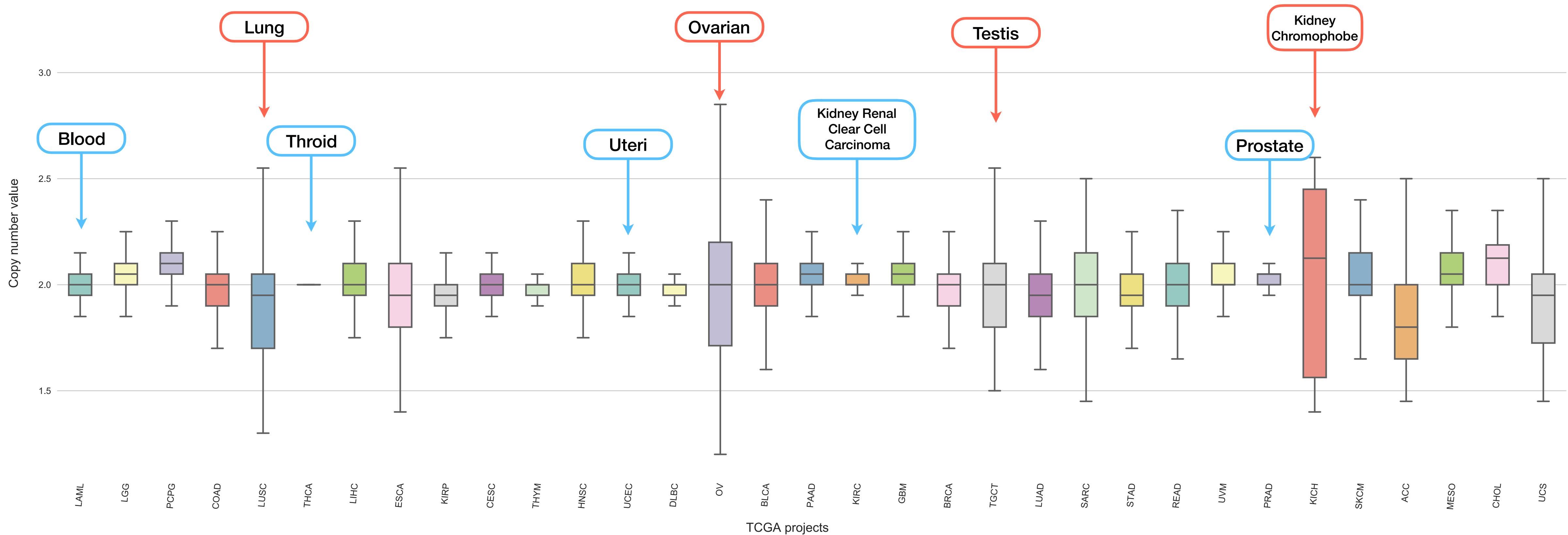


Source: Gibb, Ewan A., et al. "Deciphering squamous cell carcinoma using multidimensional genomic approaches." *Journal of skin cancer* 2011 (2011).

Copy number profile of a gastrointestinal stromal tumor (GSM2443442)



Baseline deviation in TCGA data



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

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

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

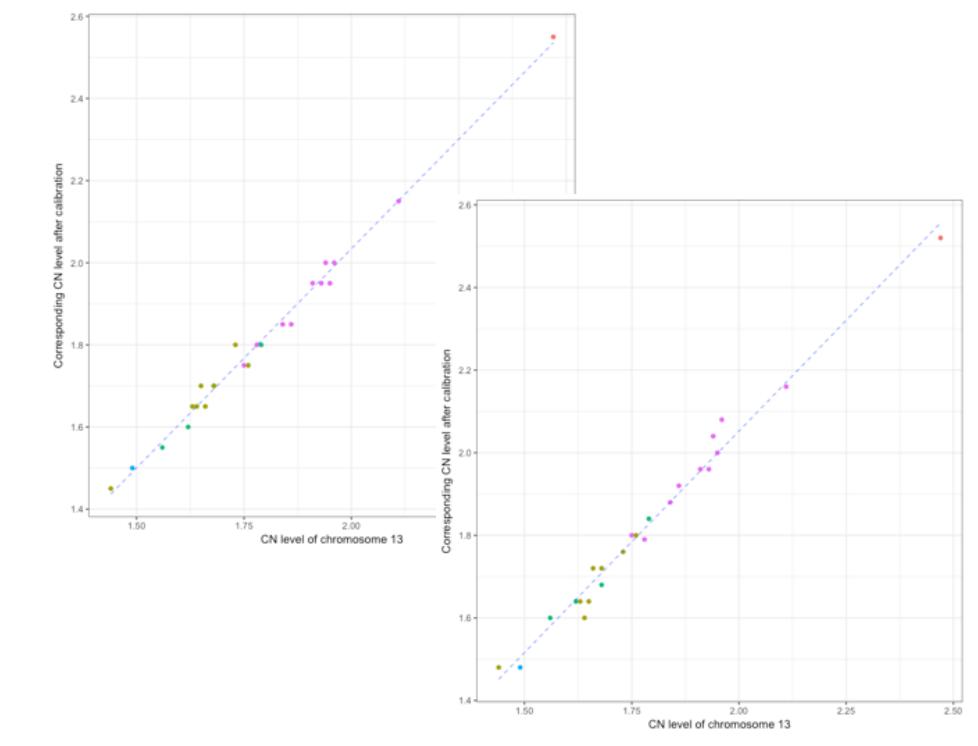
$$= aN_i + bT_i + cS_i + E_i$$

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

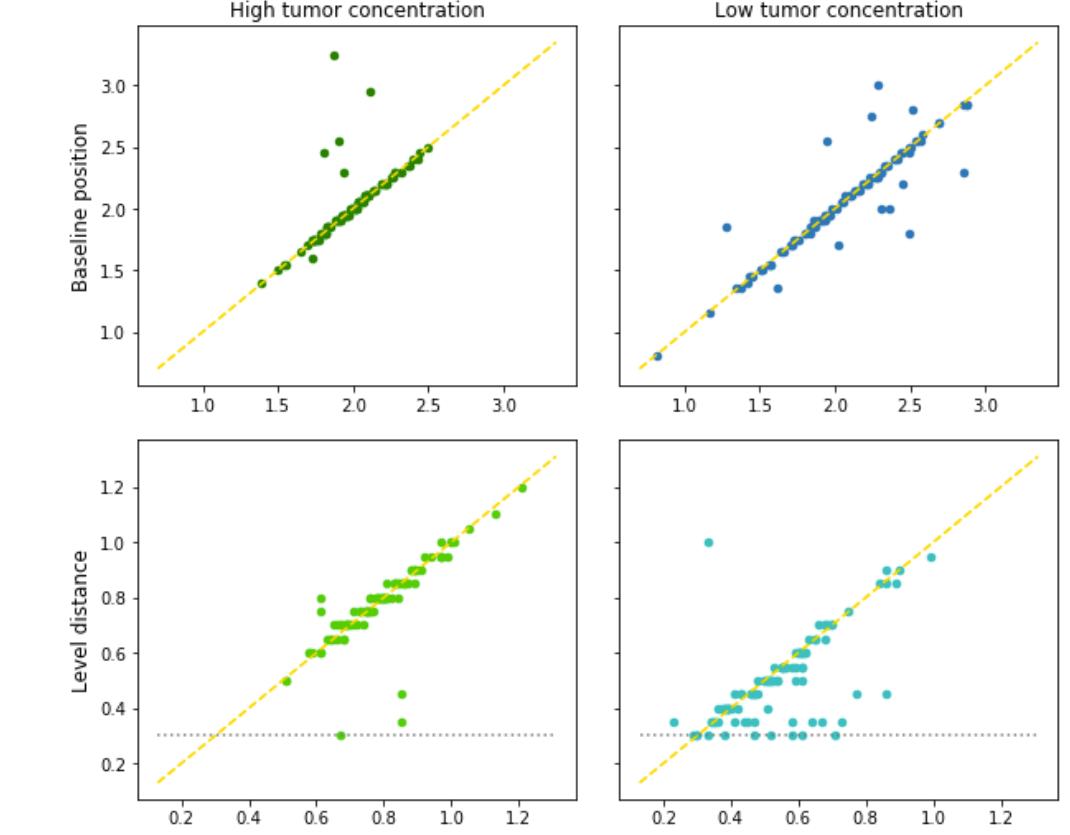
$$R(i, j, k) = \frac{b(T_i - T_k) + c(S_i - cS_k) + E_{i,k}}{b(T_i - T_j) + c(S_i - cS_j) + E_{i,j}}$$

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

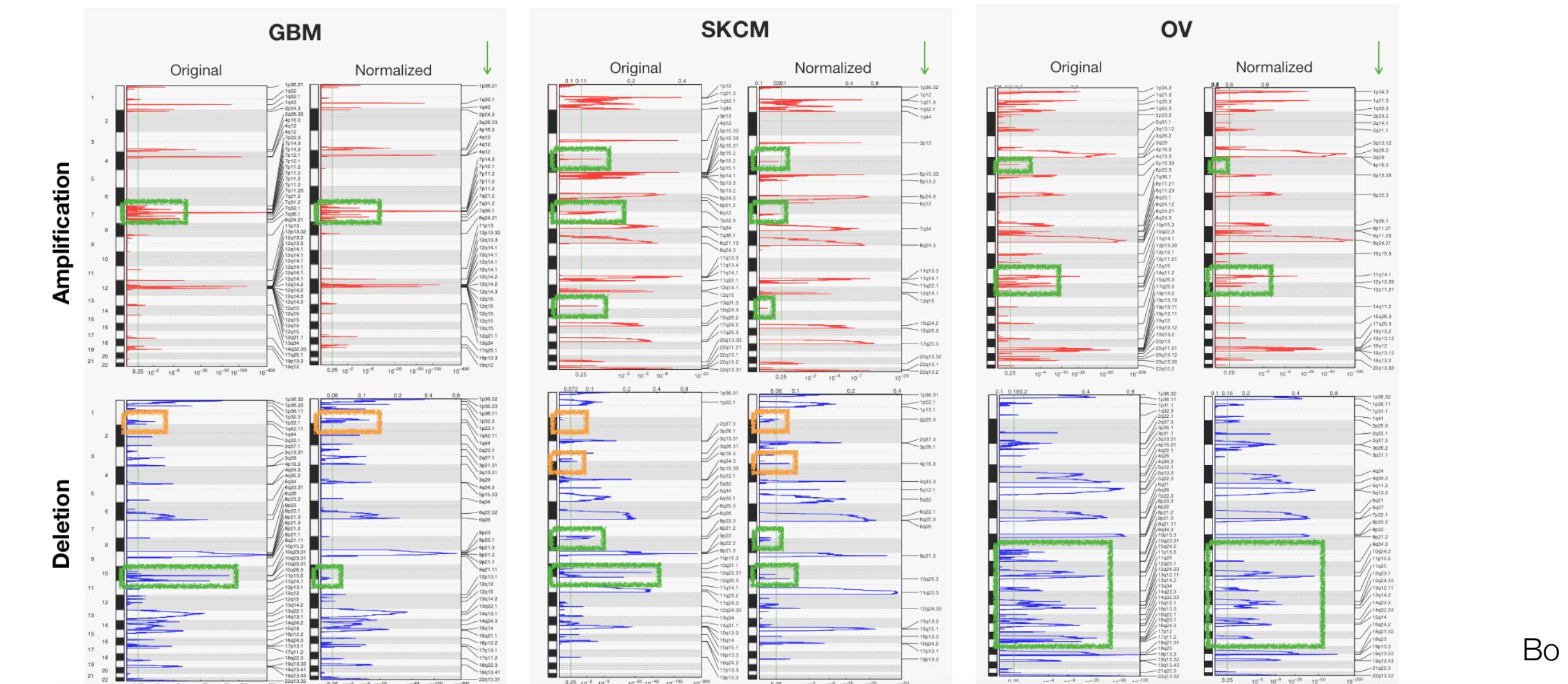
Benchmarking against karyotyping and ABSOLUTE



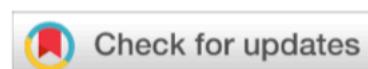
Benchmarking on simulation data



Application on GISTIC analysis of TCGA data



Bo Gao



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

The difficulties in copy number segment liftover

Challenge

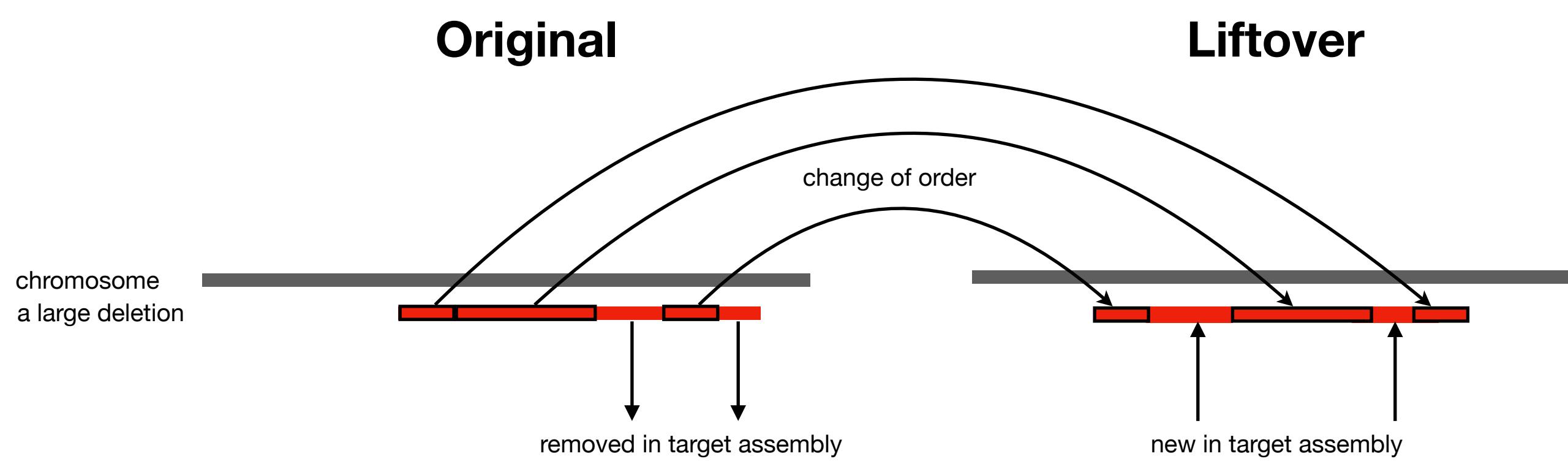
1. Keep the integrity of copy number segments after Liftover.
2. 10% data lost from straight Liftover.
3. 1TB segment and probe data.

Solution

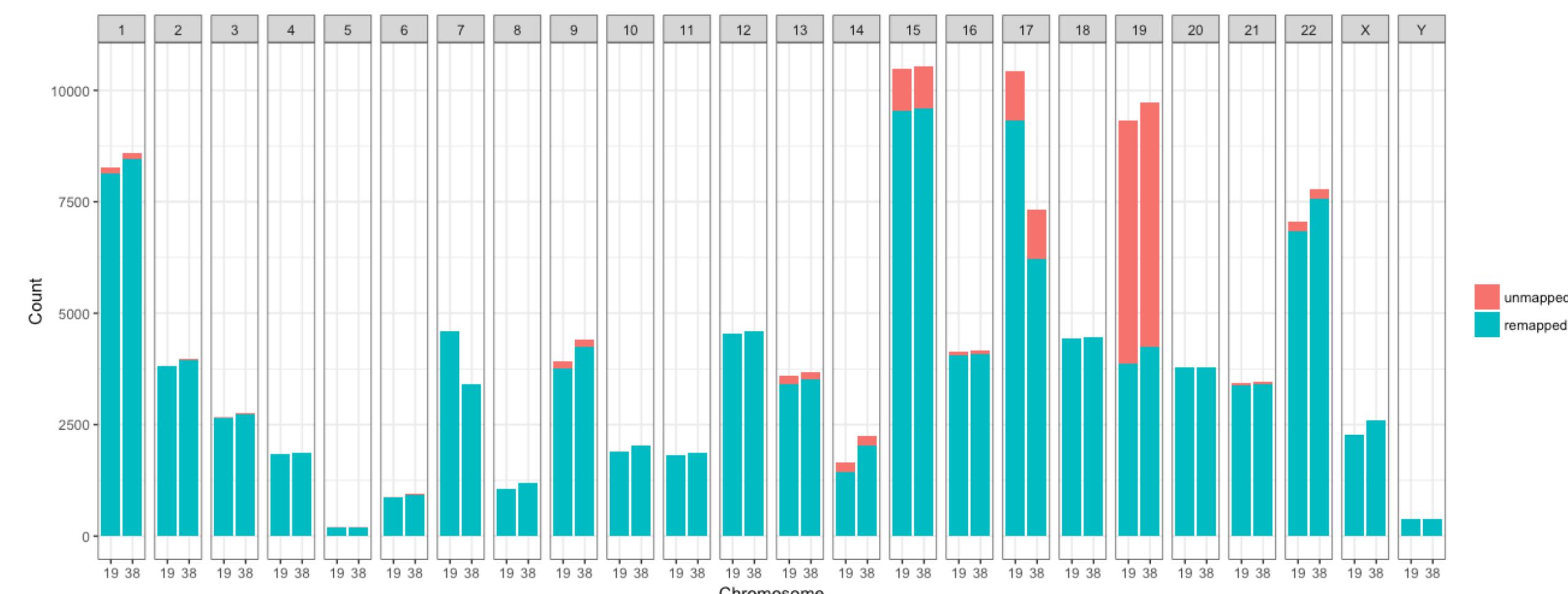
1. Algorithm to lift copy number segments.
2. Algorithm for fuzzy remapping.
3. Parallel processing and failure recovery mechanism.

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%**



Results of segmentLiftover on our data



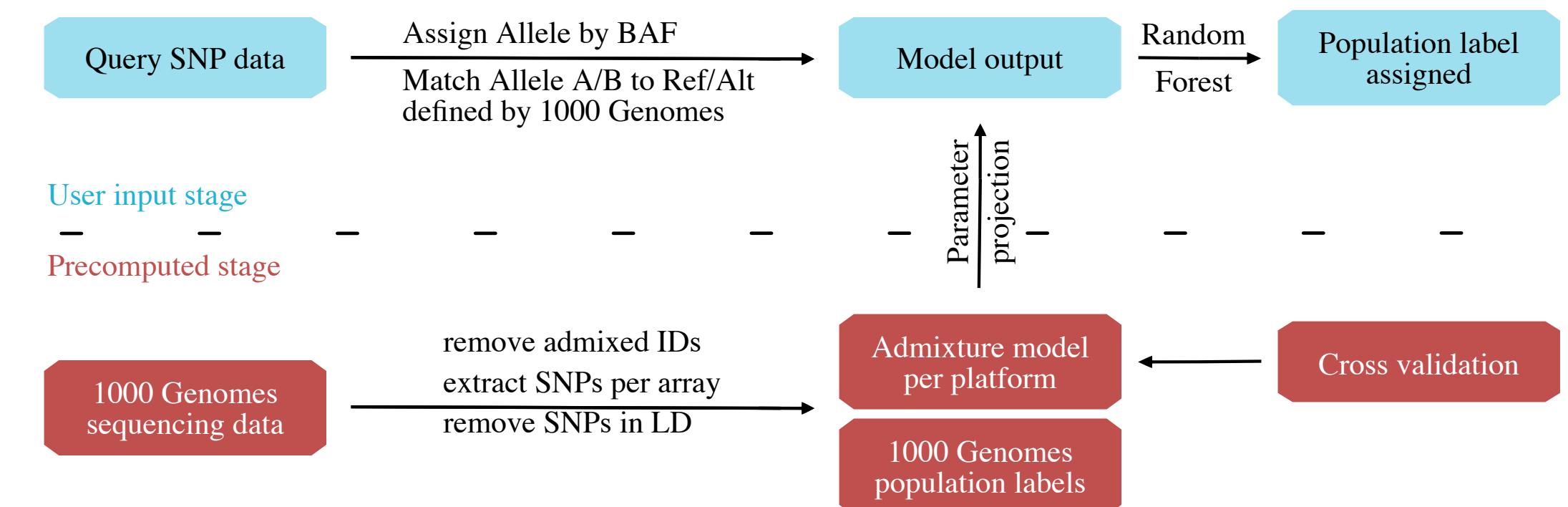
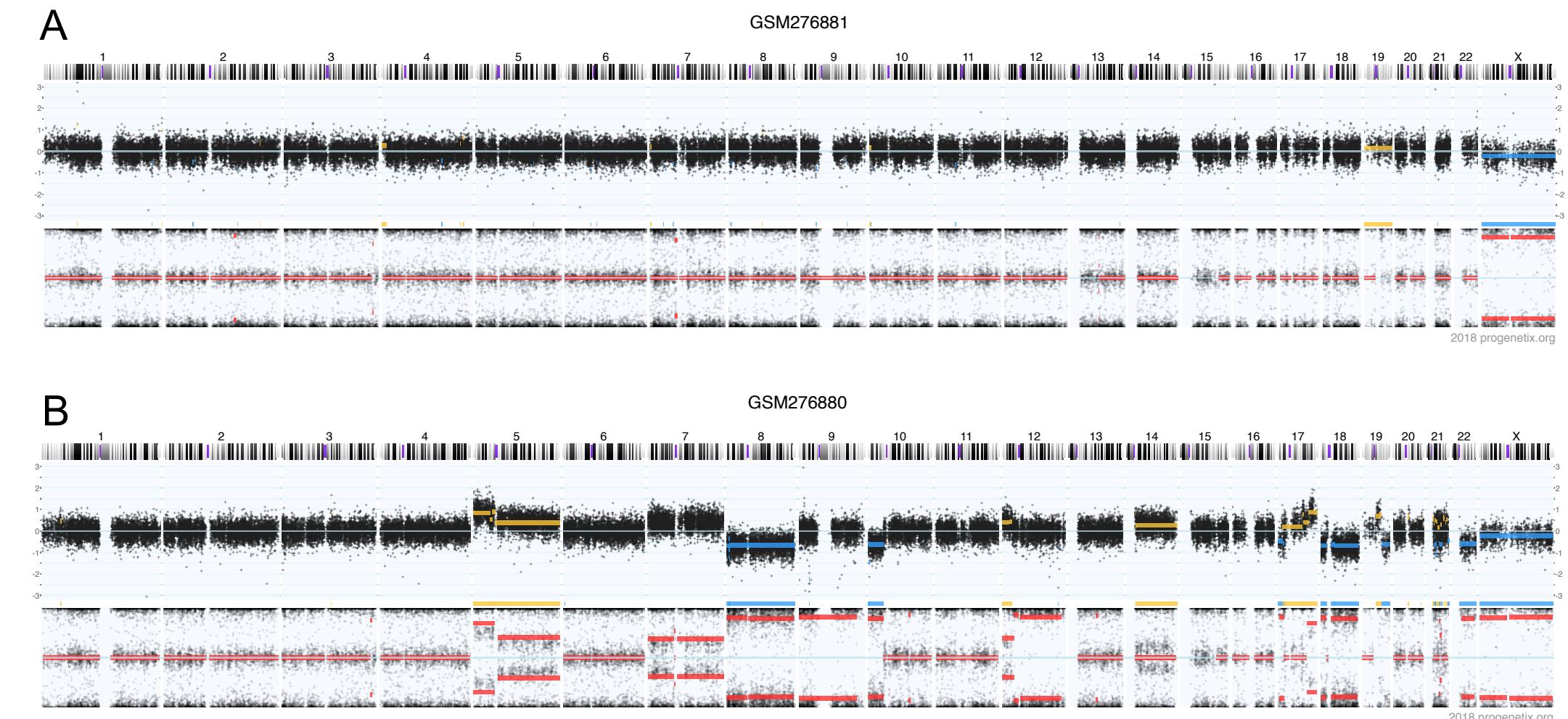
Population stratification in cancer samples based on SNP array data

- 2504 genome profiles from 1000 Genome project phase 1 as reference
- 5 (or 26) superpopulations: South Asia, Europe, South America, East Asia and Africa.
- SNP positions used in 9 Affymetrix SNP arrays are extracted to train a population admixture model.

arrayMap 

Enabling population assignment from cancer genomes with SNP2pop

Qingyao Huang^{1,2} and Michael Baudis^{1,2✉}



Qingyao Huang

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

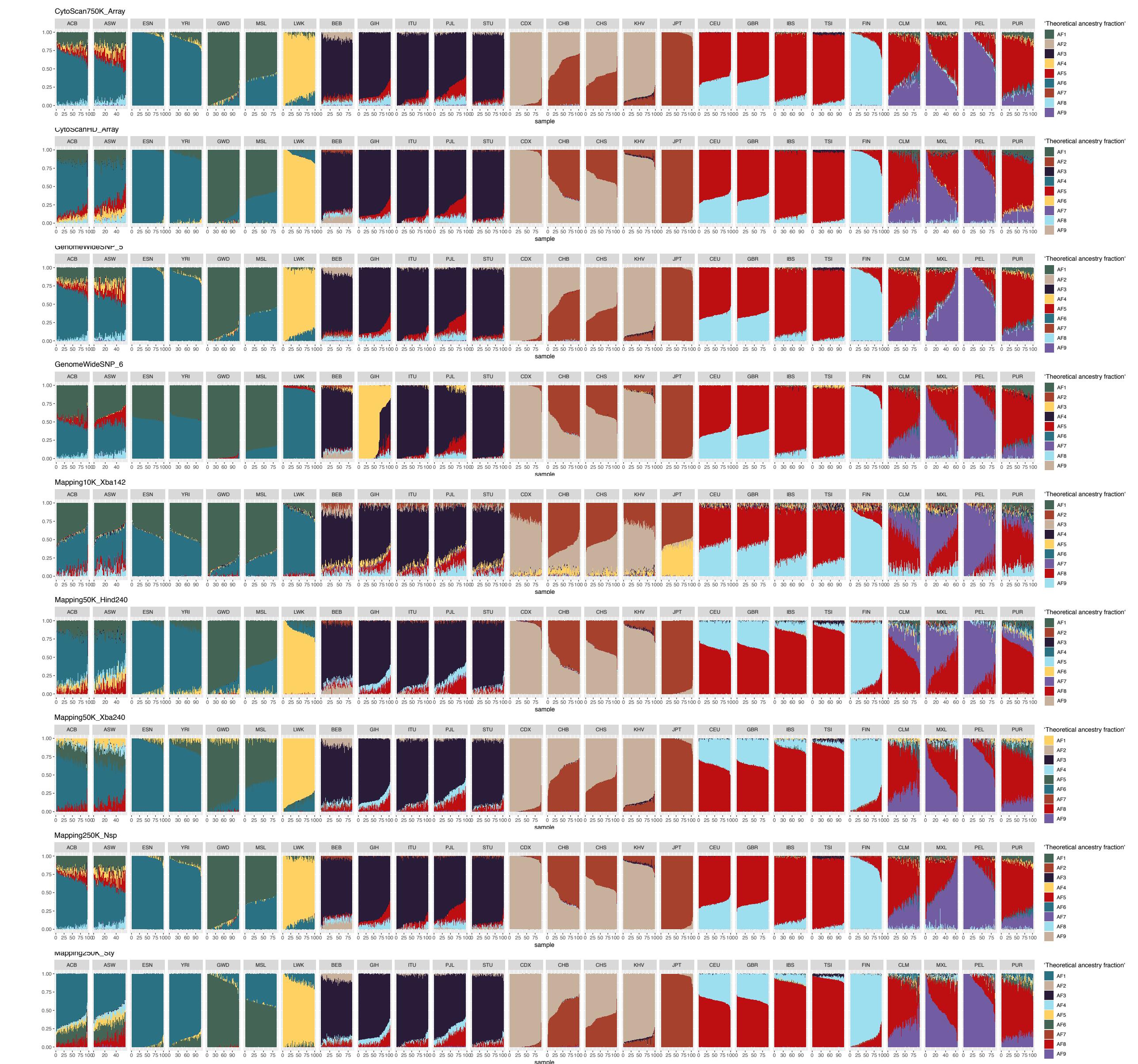


Figure S1 The fraction or contribution of theoretical ancestors ($k=9$) in reference individuals from 1000 Genomes Project with regard to nine SNP array platforms. The x-axis are individual samples, grouped by their respective population. Groups belonging to the same continent/superpopulation are placed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

Progenetix - Cancer CNV Information Resource

- launched online in 2001 as *progenetix.net*
- **curation of published CNV profiling data**
 - originally cCGH and CNV extraction from Mitelman database
 - + aCGH, WES, WGS; - karyotype data
- increasingly focused on representing the "publication landscape" of cancer genome screening - What? Where?
- **Genomes:**
 - 93640 CNV profiles (cCGH, aCGH, WES, WGS) from 469 cancer types (NCIt & ICD-O mapping)
 - 6'817'645 "CNVs" (i.e. called segments)
- **Articles:**
 - 3229 registered articles
 - geographic mapping
 - "cancer type" labelling
 - represent 174'530 reported samples

Progenetix :: Info

Structural Cancer Genomics Resource
Documentation and Example Pages

[News](#)
[About...](#)
[Documentation](#)
[Publications](#)
[Data Pages](#)

Related Sites

[arrayMap](#)
[Baudisgroup @ UZH](#)
[Beacon+](#)
[SchemaBlocks {S}\[B\]](#)
[ELIXIR Beacon](#)
[Baudisgroup Internal](#)

Github Projects

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

Tags

[API](#) [article](#) [code](#) [documentation](#)
[licensing](#) [maps](#) [statistics](#) [tools](#)

Progenetix Publication Collection

The current page lists publications of whole genome screening experiments in cancer, registered in the Progenetix publication collection.

This page is a *beta* version, intended to replace the [original publications](#) page.

Show entries

Publication	Samples			
	cCGH	aCGH	WES	WGS
Harada K, Okamoto W, Mimaki S, Kawamoto Y, Bando et al. (2019): Comparative sequence analysis of patient-matched primary colorectal cancer, metastatic, and recurrent metastatic tumors ... BMC Cancer 19(1), 2019 (30898102) 	0	0	4	0
Lavrov AV, Chelysheva EY, Adilgereeva EP, Shukhov et al. (2019): Exome, transcriptome and miRNA analysis don't reveal any molecular markers of TKI efficacy in primary CML ... BMC Med Genomics 12(Suppl 2), 2019 (30871622) 	0	0	62	0
Zandberg DP, Tallon LJ, Nagaraj S, Sadzewicz LK, Zhang et al. (2019): Intratumor genetic heterogeneity in squamous cell carcinoma of the oral cavity. Head Neck, 2019 (30869813) 	0	0	5	0
Heinrich MC, Patterson J, Beadling C, Wang Y, Debiec-Rychter et al. (2019): Genomic aberrations in cell cycle genes predict progression of KIT-mutant gastrointestinal stromal tumors ... Clin Sarcoma Res 9, 2019 (30867899) 	0	0	29	0
Jiao J, Sagnelli M, Shi B, Fang Y, Shen Z, Tang T, Dong et al. (2019): Genetic and epigenetic characteristics in ovarian tissues from polycystic ovary syndrome patients with irregular ... BMC Endocr Disord 19(1), 2019 (30866919) 	0	0	20	0
Mueller S, Jain P, Liang WS, Kilburn L, Kline C, Gupta et al. (2019): A pilot precision medicine trial for children with diffuse intrinsic pontine glioma - PNOC003: a report from the Pacific ... Int. J. Cancer, 2019 (30861105) 	0	0	14	14
Xie SN, Cai YJ, Ma B, Xu Y, Qian P, Zhou JD, Zhao et al. (2019): The genomic mutation spectra of breast fibroadenomas in Chinese population by whole exome sequencing ... Cancer Med, 2019 (30851086) 	0	0	12	0

Showing 1 to 50 of 3,232 entries



Progenetix - Cancer CNV Information Resource

- Progenetix literature collection contains information about articles reporting genome profiling experiments (aCGH, cCGH, WES, WGS) in cancer samples
- continuous collection
- annotation of metadata extracted from the articles
 - ▶ cancer type
 - ▶ geographic location (by author or from text)
 - ▶ sample numbers per technology
 - ▶ contact information

Progenetix :: Info

Structural Cancer Genomics Resource
Documentation and Example Pages

[News](#)
[About...](#)
[Documentation](#)
[Publications](#)
[Data Pages](#)

Related Sites

[arrayMap](#)
[Baudisgroup @ UZH](#)
[Beacon+](#)
[SchemaBlocks {S}\[B\]](#)
[ELIXIR Beacon](#)
[Baudisgroup Internal](#)

Github Projects

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

Tags

[API](#) [article](#) [code](#) [documentation](#)
[licensing](#) [maps](#) [statistics](#) [tools](#)



Progenetix Publication Collection

The current page lists publications of whole genome screening experiments in cancer, registered in the Progenetix publication collection.

This page is a *beta* version, intended to replace the [original publications](#) page.

Show entries

Search:

Samples

cCGH	aCGH	WES	WGS
------	------	-----	-----

Publication				
 Harada K, Okamoto W, Mimaki S, Kawamoto Y, Bando et al. (2019): Comparative sequence analysis of patient-matched primary colorectal cancer, metastatic, and recurrent metastatic tumors ...	BMC Cancer 19(1), 2019 (30898102)		0	0 4 0
 Lavrov AV, Chelysheva EY, Adilgereeva EP, Shukhov et al. (2019): Exome, transcriptome and miRNA analysis don't reveal any molecular markers of TKI efficacy in primary CML ...	BMC Med Genomics 12(Suppl 2), 2019 (30871622)		0 0 62 0	
 Zandberg DP, Tallon LJ, Nagaraj S, Sadzewicz LK, Zhang et al. (2019): Intratumor genetic heterogeneity in squamous cell carcinoma of the oral cavity.	Head Neck, 2019 (30869813)		0 0 5 0	
 Heinrich MC, Patterson J, Beadling C, Wang Y, Debiec-Rychter et al. (2019): Genomic aberrations in cell cycle genes predict progression of KIT-mutant gastrointestinal stromal tumors ...	Clin Sarcoma Res 9, 2019 (30867899)		0 0 29 0	
 Jiao J, Sagnelli M, Shi B, Fang Y, Shen Z, Tang T, Dong et al. (2019): Genetic and epigenetic characteristics in ovarian tissues from polycystic ovary syndrome patients with irregular ...	BMC Endocr Disord 19(1), 2019 (30866919)		0 0 20 0	
 Mueller S, Jain P, Liang WS, Kilburn L, Kline C, Gupta et al. (2019): A pilot precision medicine trial for children with diffuse intrinsic pontine glioma - PNOC003: a report from the Pacific ...	Int. J. Cancer, 2019 (30861105)		0 0 14 14	
 Xie SN, Cai YJ, Ma B, Xu Y, Qian P, Zhou JD, Zhao et al. (2019): The genomic mutation spectrums of breast fibroadenomas in Chinese population by whole exome sequencing ...	Cancer Med, 2019 (30851086)		0 0 12 0	

Showing 1 to 50 of 3,232 entries



Publication Landscape of Cancer CNV Profiling

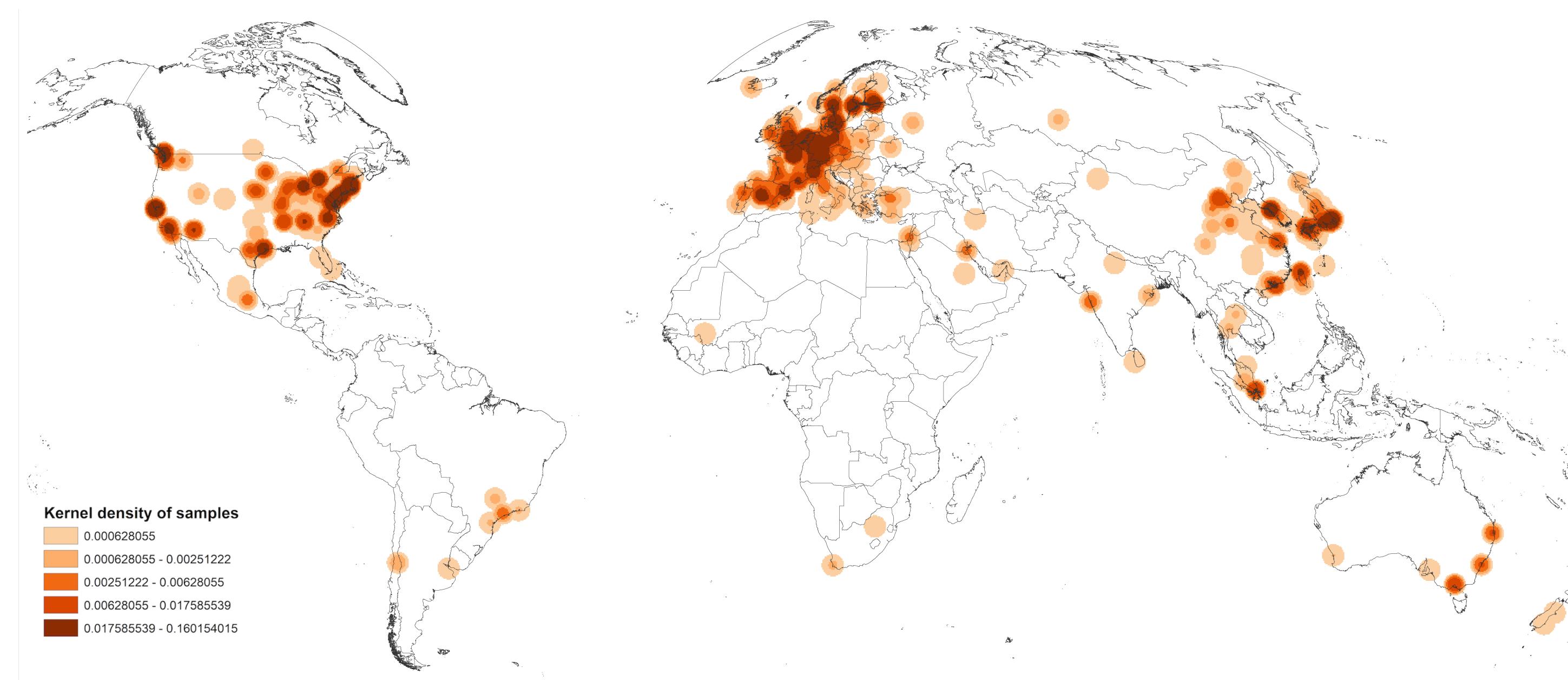


Publication statistics for cancer genome screening studies. The graphic shows our assessment of publications reporting whole-genome screening of cancer samples, using molecular detection methods (chromosomal CGH, genomic array technologies, whole exome and genome sequencing).

For the years 1993-2018, we found 3'229 publications reporting 174'530 individual samples in single series from 1 to more than 1000 samples. Y-axis and size of the dots correspond to the sample number; the color codes indicate the technology used.

Map of the geographic distribution (by first author affiliation) of the 104'543 genomic array, 36'766 chromosomal CGH and 15'409 whole genome/exome based cancer genome datasets.

The numbers are derived from the 3'240 publications registered in the Progenetix database.



GA4GH API promotes sharing

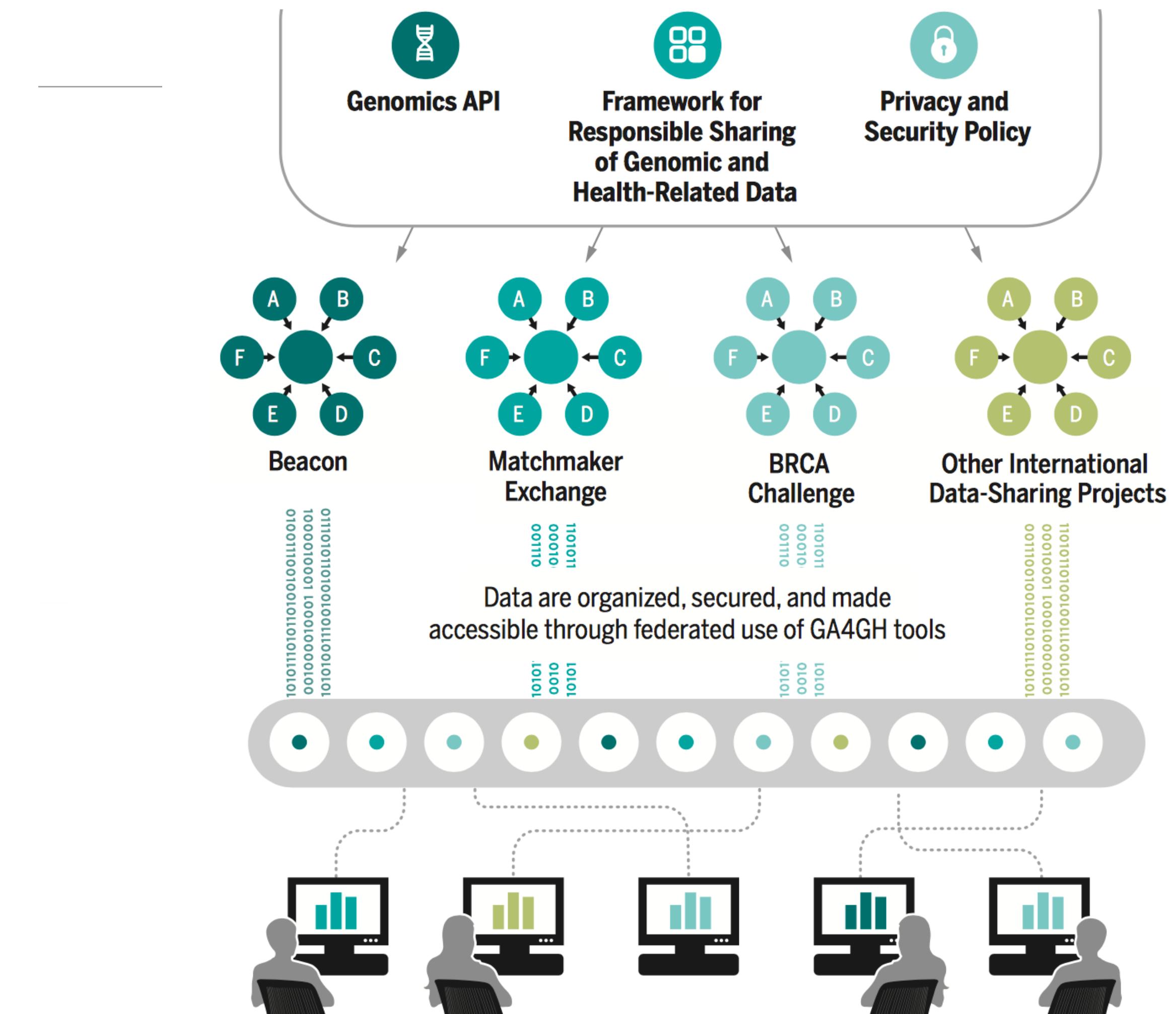
A federated data ecosystem. To share genomic data globally, this approach furthers medical research without requiring compatible data sets or compromising patient identity.



GENOMICS

A federated ecosystem for sharing genomic, clinical data

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





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](#)

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

Global Learning for Health



Global Alliance
for Genomics & Health

Interoperable
APIs, standards &
frameworks to
support global
data sharing



VICC

Variant Interpretation
for Cancer Consortium



**Genomic
Knowledge
Exchanges**



Matchmaker
Exchange



Beacon Project

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



Beacon Network

Search Beacons

Search all beacons for allele

GRCh37 ▾ 10:118969015 C / CT Search

Response All None
 Found 16
 Not Found 27
 Not Applicable 22

Organization All None
 AMPLab, UC Berkeley
 BGI
 BioReference Laborato...
 Brazilian Initiative on ...
 BRCA Exchange
 Broad Institute
 Centre for Genomic R...
 Centro Nacional de A...
 Curoverse
 EMBL European Bio...
 Global Alliance for G...
 Google
 Institute for Systems ...
 Instituto Nacional de ...

Response	All	None
<input checked="" type="checkbox"/> Found	16	
<input type="checkbox"/> Not Found	27	
<input type="checkbox"/> Not Applicable	22	

BioReference BioReference Hosted by BioReference Laboratories Found

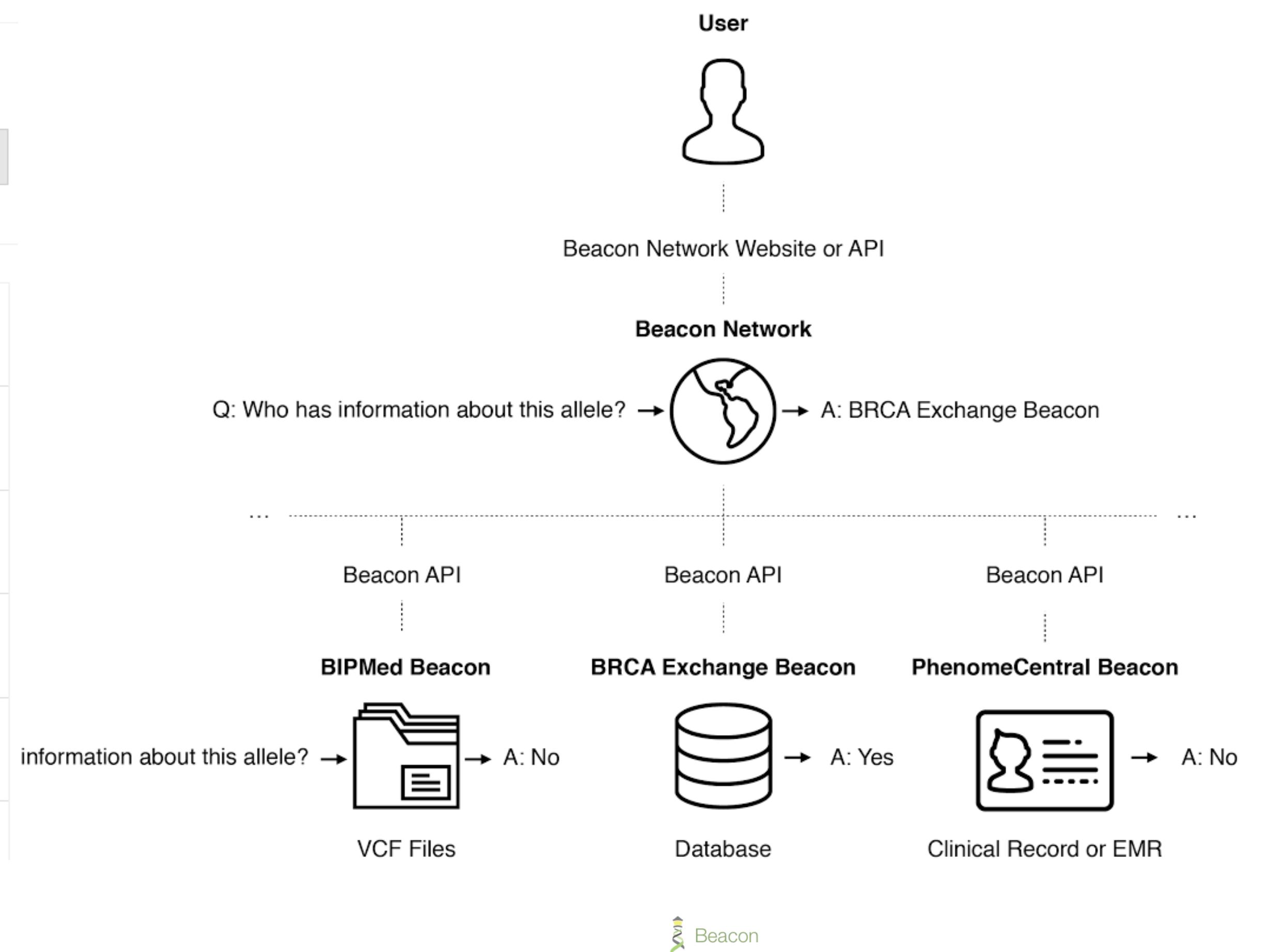
Catalogue of Somatic Mutations in Cancer Catalogue of Somatic Mutations in Cancer Hosted by Wellcome Trust Sanger Institute Found

Cell Lines Cell Lines Hosted by Wellcome Trust Sanger Institute Found

Conglomerate Conglomerate Hosted by Global Alliance for Genomics and Health Found

COSMIC COSMIC Hosted by Wellcome Trust Sanger Institute Found

dbGaP: Combined GRU Catalog and NHLBI Exome Seq... dbGaP: Combined GRU Catalog and NHLBI Exome Seq... Found



Date	Tag	Title
2018-01-24	v0.4.0	Beacon
2016-05-31	v0.3.0	Beacon

Data Driven Standards for Genomic Data Exchange

ELIXIR Beacon :: Beacon⁺ :: SchemaBlocks {S}[B]



University of
Zurich^{UZH}



Simplify the way people search for and request access to potentially identifiable data in international and national genomic data resources



Working towards GA4GH standards, APIs and toolkits to be used throughout ELIXIR Nodes for human data discovery and access => GA4GH in Europe

ELIXIR Members



ELIXIR Observers



TBC

15/23 Nodes connected

61 connections

Clinical & Phenotypic
Data Capture

Large Scale
Genomics

Genomic Knowledge
Standards

Discovery

Cloud

Data Use &
Researcher
Identities (DURI)

Regulatory
& Ethics

Data Security



Global Alliance
for Genomics & Health

ELIXIR Beacon Project

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

The screenshot shows the 'Driver Projects' section of the GA4GH website. It features a red circular icon with a white rocket ship. Below it, the text 'Driver Projects' is displayed. A detailed description follows: 'GA4GH Driver Projects are real-world genomic data initiatives that help guide our development efforts and pilot our tools. Stakeholders around the globe advocate, mandate, implement, and use our frameworks and standards in local contexts.' To the right, there is a box for the 'ELIXIR Beacon' project, which includes the ELIXIR logo, the URL www.elixir-europe.org, the text 'Europe', and 'Champions: Serena Scollen, Ilkka Lappalainen, Michael Baudis'.

Beacon forward



- **structural variations** (DUP, DEL) in addition to SNV
 - ... more structural queries (translocations/fusions...)
- **filters** for bio- & technical 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)

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/

Beacon+



This example shows the query for CNV deletion variants overlapping the CDKN2A gene's coding region with at least a single base, but limited to "focal" hits (here i.e. <= ~4Mbp in size). The query is against the arrayMap collection and can be modified e.g. through changing the position parameters or data source.

CNV Example SNV Range Example SNV Example BND Example

Dataset* arraymap
progenetix
tcga
dipg
beacon_test

Dataset Responses All Selected Datasets

Reference name* 9

Genome Assembly* GRCh38 / hg38

(structural) variantType DEL (Deletion)

Gene Coordinates CDKN2A

Start min Position* 18000000

Start max Position 21975098

End min Position 21967753

End max Position 26000000

Bio-ontology no selection
icdom-94423: Gliosarcoma (9)
icdom-94403: Glioblastoma, NOS
icdot-C16: Stomach (133)
icdot:C40.1: Short bones of up
icdot-C55+: Uterus, NOS (89)

Biosample Type (no selection)

Beacon Query

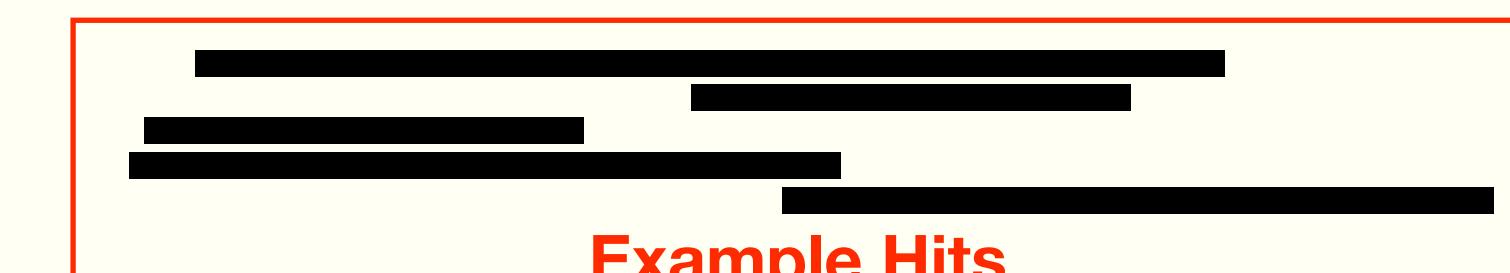
Progenetix datasets

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

startMin - startMax

CDKN2A CDR

endMin - endMax



Filters

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

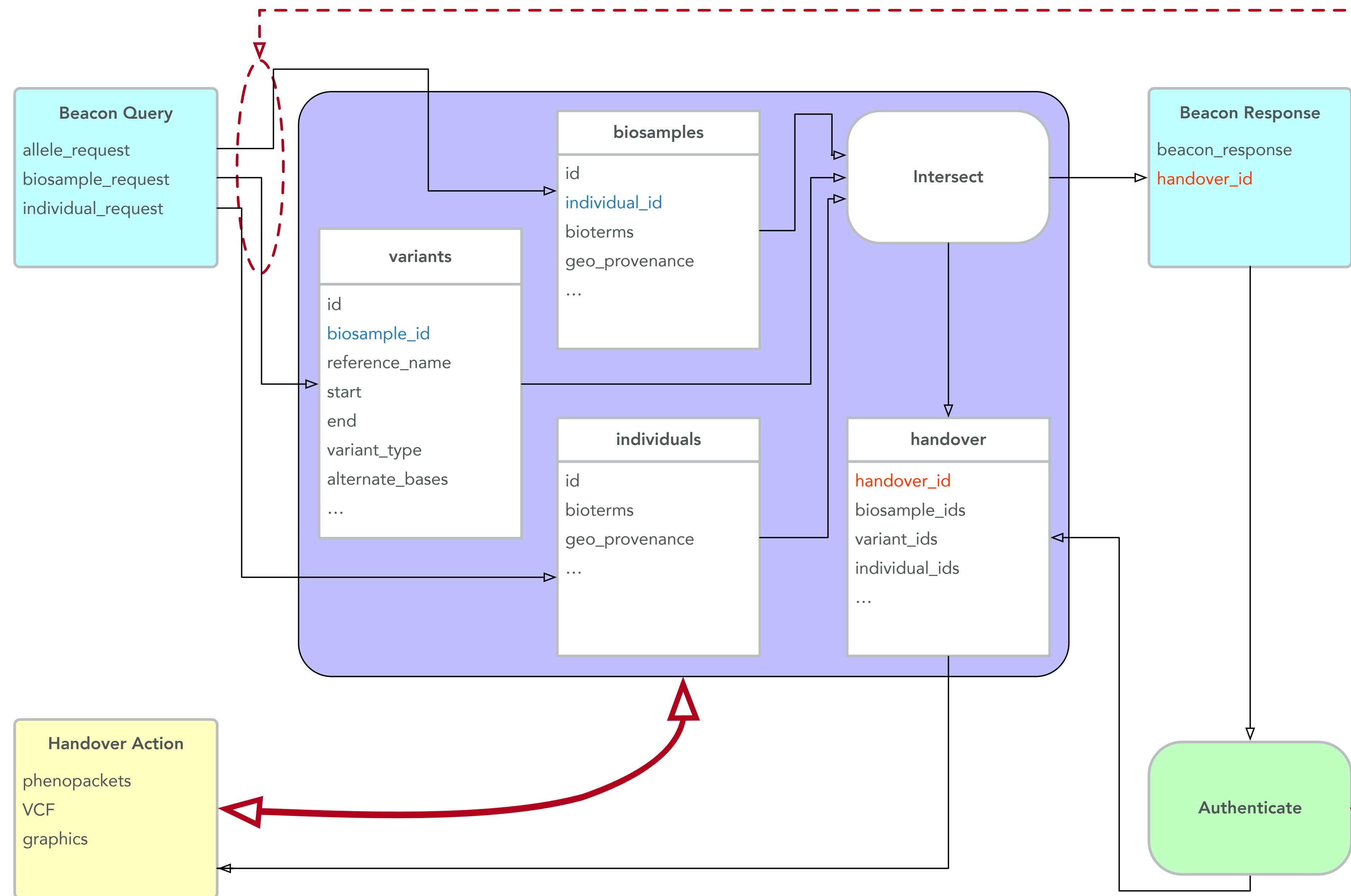
Beacon & Handover

Beacons support connection of **data delivery services**

Beacon I/O

Handover

Authentication





This example shows the query for CNV deletion variants overlapping the CDKN2A gene's coding region with at least a single base, but limited to "focal" hits (here i.e. < ~4Mbp in size). The query is against the arrayMap collection and can be modified e.g. through changing the position parameters or data source.

CNV Example SNV Range Example SNV Example BND Example

Dataset*	arraymap progenetix tcga dipg beacon_test
Dataset Responses	All Selected Datasets
Reference name*	9
Genome Assembly*	GRCh38 / hg38
(structural) variantType	DEL (Deletion)
Gene Coordinates	CDKN2A
Start min Position*	18000000
Start max Position	21975098
End min Position	21967753
End max Position	26000000
Bio-ontology	no selection icdom-94423: Gliosarcoma (9) icdom-94403: Glioblastoma, NOS icdot-C16: Stomach (133) icdot:C40.1: Short bones of upper limb (1) icdot-C55+: Uterus, NOS (89)
Biosample Type	(no selection)
Beacon Query	

<https://beacon.progenetix.org/cgi-bin/beaconresponse.cgi?datasetIds=arraymap&datasetIds=dipg&datasetAlleleResponses=ALL&referenceName=9&assemblyId=GRCh38&variantType=DEL&startMin=18000000&startMax=21975098&endMin=21967753&endMax=26000000&filters=icdom-94403>

query

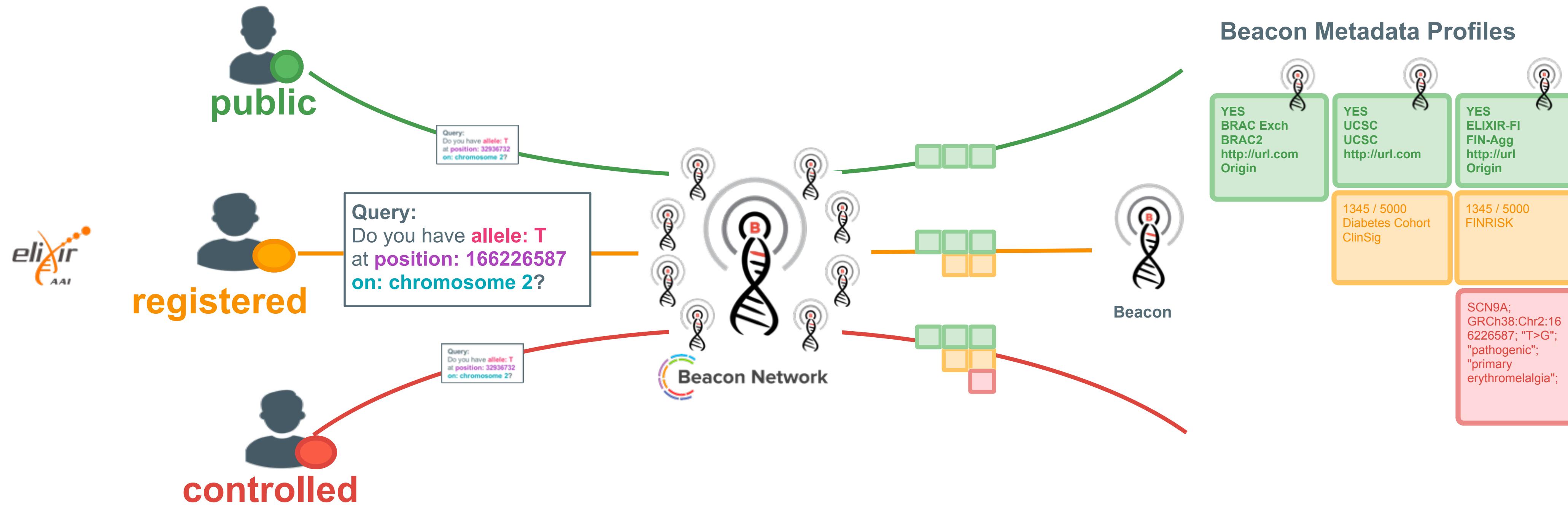
Response									
Dataset	Assembly	Chro	Position Start Range	Ref Alt Type	Bio Query	Variants Calls Samples	f_alleles	Response Context	
arraymap	GRCh38	9	18000000 - 21975098 21967753 - 26000000	N DEL	icdom-94403	588 588 588	0.0081	JSON UCSC [H->O] Biosamples [H->O] Callsets Variants [H->O] CNV Histogram [H->O] Progenetix Interface [H->O] Variants	
dipg	GRCh38	9	18000000 - 21975098 21967753 - 26000000	N DEL	icdom-94403	0 0 0	0	JSON UCSC	

```
"datasetHandover" : [
    {
        "url" : "https://progenetix.org/cgi-bin/beacondeliver.cgi?do=biosamplesdata&accessid=d1ffd548-e68e-11e9-87ed-fcb07b51aec4",
        "description" : "retrieve data of the biosamples matched by the query",
        "handoverType" : {
            "label" : "Biosamples",
            "id" : "pgx:handover:biosamplesdata"
        }
    },
}
```

handover

A Beacon "datasetAlleleResponses" can provide **handover** objects to initiate further actions (data retrieval, visualization ...) outside of the Beacon protocol.

Integrating permissions and discovery

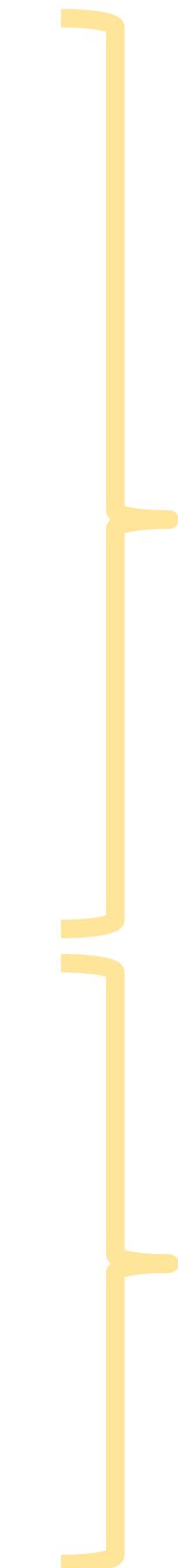


<https://www.youtube.com/watch?v=LyfmvAs7LtQ&feature=youtu.be>



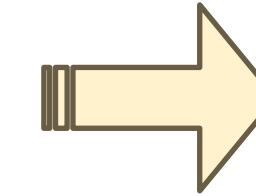
Beacon v2.n - Areas of Change

- One query for each of the existing types
 - SNPs
 - Structural Variants
 - Region
 - ...
- Access levels
- Filters
- New types of queries:
 - By sample
 - By patient
 - ...
- Schema versions



v2.0

Tested and already
implemented by real
Beacons



GA4GH
approval
process

Ongoing



ELIXIR Genome Beacons

A Driver Project of the Global Alliance for Genomics and Health

About...

News & Press

Contributors

Events

Examples, Guides & FAQ

Specification

Roadmap

Beacon Networks

Meeting Minutes

Contacts

Related Sites

[Beacon @ ELIXIR](#)

[GA4GH](#)

[Beacon+](#)

[beacon-network.org](#)

[GA4GH::SchemaBlocks](#)

[GA4GH::Discovery](#)

[GA4GH::CLP](#)

[GA4GH::GKS](#)

Github Projects

[ELIXIR Beacon](#)

[SchemaBlocks](#)

Tags

[EB](#) [FAQ](#) [contacts](#) [definitions](#)

[developers](#) [development](#)

[minutes](#) [network](#) [press](#)

[proposal](#) [queries](#) [releases](#)

[specification](#) [versions](#) [website](#)



Roadmap

The ELIXIR Beacon Roadmap delineates short-, mid- and long-term objectives, to expand functional scope and reach of Beacon as a protocol and genomic data ecosystem.

Beacon Flavours

Beacons may be able to increase their functionality through the development of distinct **flavours**, which can extend the core Beacon concept for specific use cases.

@mbaudis 2018-10-24: [more ...](#)

Bio-metadata Query Support

Future Beacon API versions will support querying for additional, non-sequence related data types.

@mbaudis 2018-10-18: [more ...](#)

EvidenceBeacon Notes - GA4GHconnect 2019

The topic of "EvidenceBeacon" was discussed with many different attendants during the speed dating session and beyond, leading to some clearer picture about the (widening) extent & next steps.

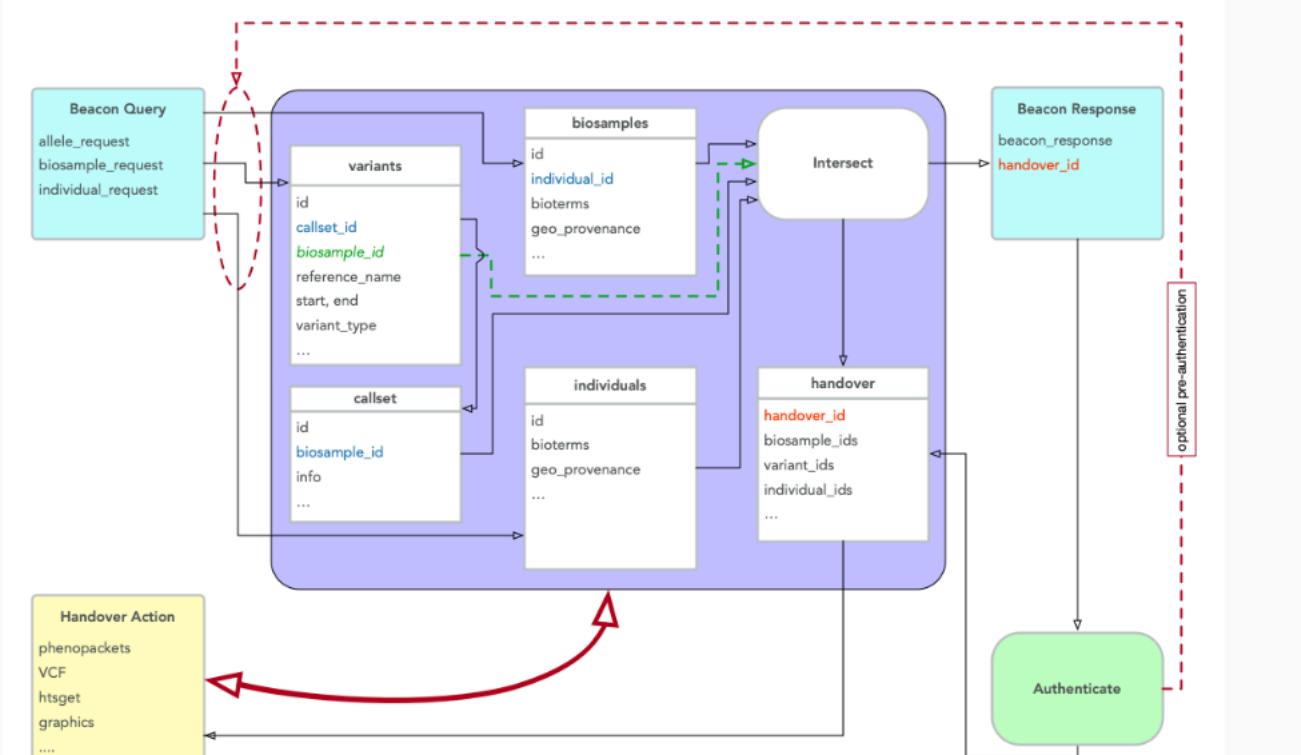
@mbaudis 2019-04-30: [more ...](#)

[H→O] Beacon Handover for Data Delivery

While the Beacon response should be restricted to aggregate data (yes/no, counts, frequencies ...), the usage of the protocol could be greatly expanded by providing an access method to data elements matched by a Beacon query.

As part of the mid-term product strategy, the ELIXIR Beacon team is evaluating the use of a "handover" protocol, in which rich data content (e.g. variant data, phenotypic information, low-level sequencing results) can be provided from linked services, initiated through a Beacon query (and possibly additional steps like protocol selection, authentication...). A discussion of the topic can e.g. be found in the Beacon developer area on Github (issue #114).

As of 2018-11-13, the **handover** concept has become part of the [ongoing code development](#).



beacon-project.io



Search or jump to...

Pull requests Issues Marketplace Explore



Beacon

Beacon Project, Global Alliance for Genomics & Health.

<http://beacon-project.io>

Repositories 7

People 15

Teams 2

Projects 1

Settings

Customize pinned repositories

Pinned repositories

[ga4gh-beacon.github.io](#)

Website of ELIXIR Beacon - A GA4GH Driver Project

HTML ★ 3 ⚡ 2

[specification](#)

GA4GH Beacon specification.

★ 28 ⚡ 23

Find a repository...

Type: All ▾

Language: All ▾

New

beacon-elixir

Elixir Beacon Reference Implementation

Java ⚡ 4 ★ 9 ⚡ 3 ⚡ 0 Updated 21 hours ago



Top languages

JavaScript Java HTML

PLpgSQL

ga4gh-beacon.github.io

Website of ELIXIR Beacon - A GA4GH Driver Project

website beacon ga4gh

HTML Apache-2.0 ⚡ 2 ★ 3 ⚡ 15 ⚡ 1 Updated 9 days ago



Most used topics

beacon ga4gh

Manage

specification

GA4GH Beacon specification.

openapi beacon ga4gh

Apache-2.0 ⚡ 23 ★ 28 ⚡ 41 ⚡ 7 Updated on May 9



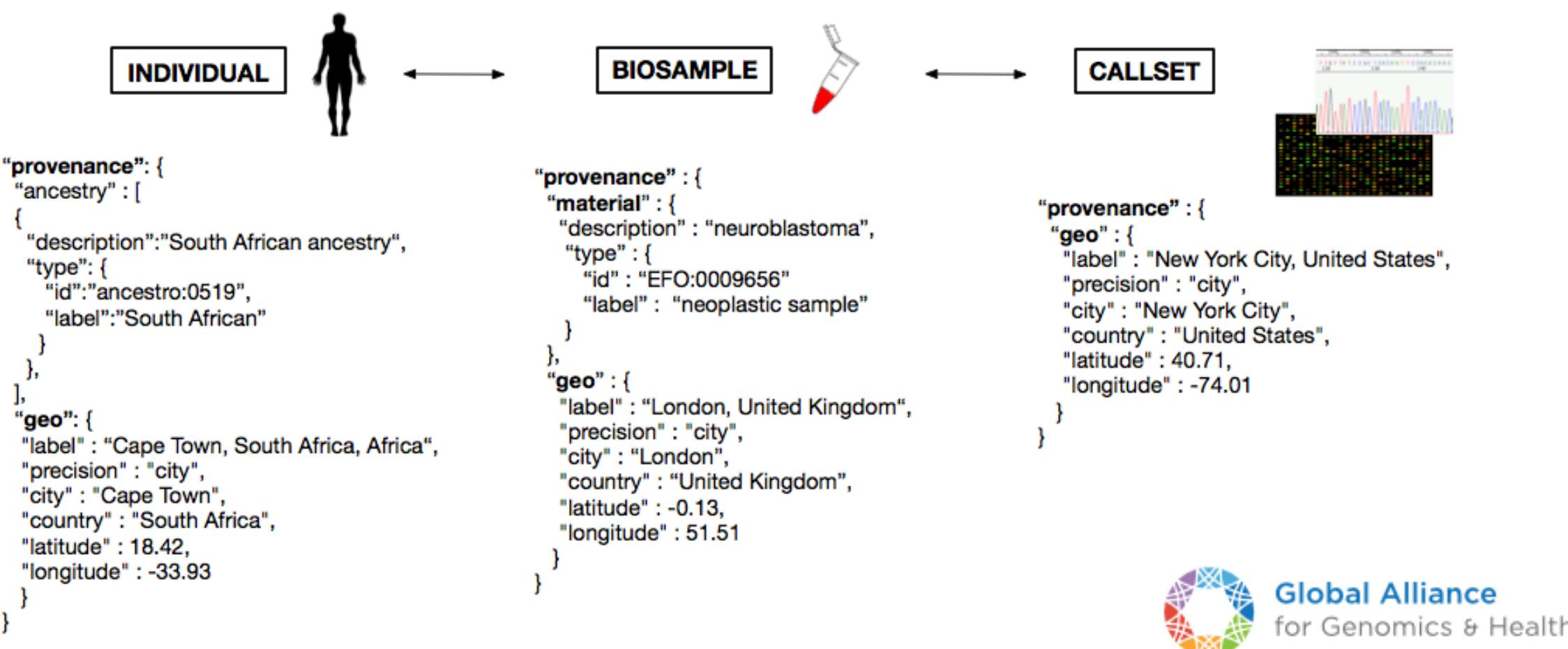
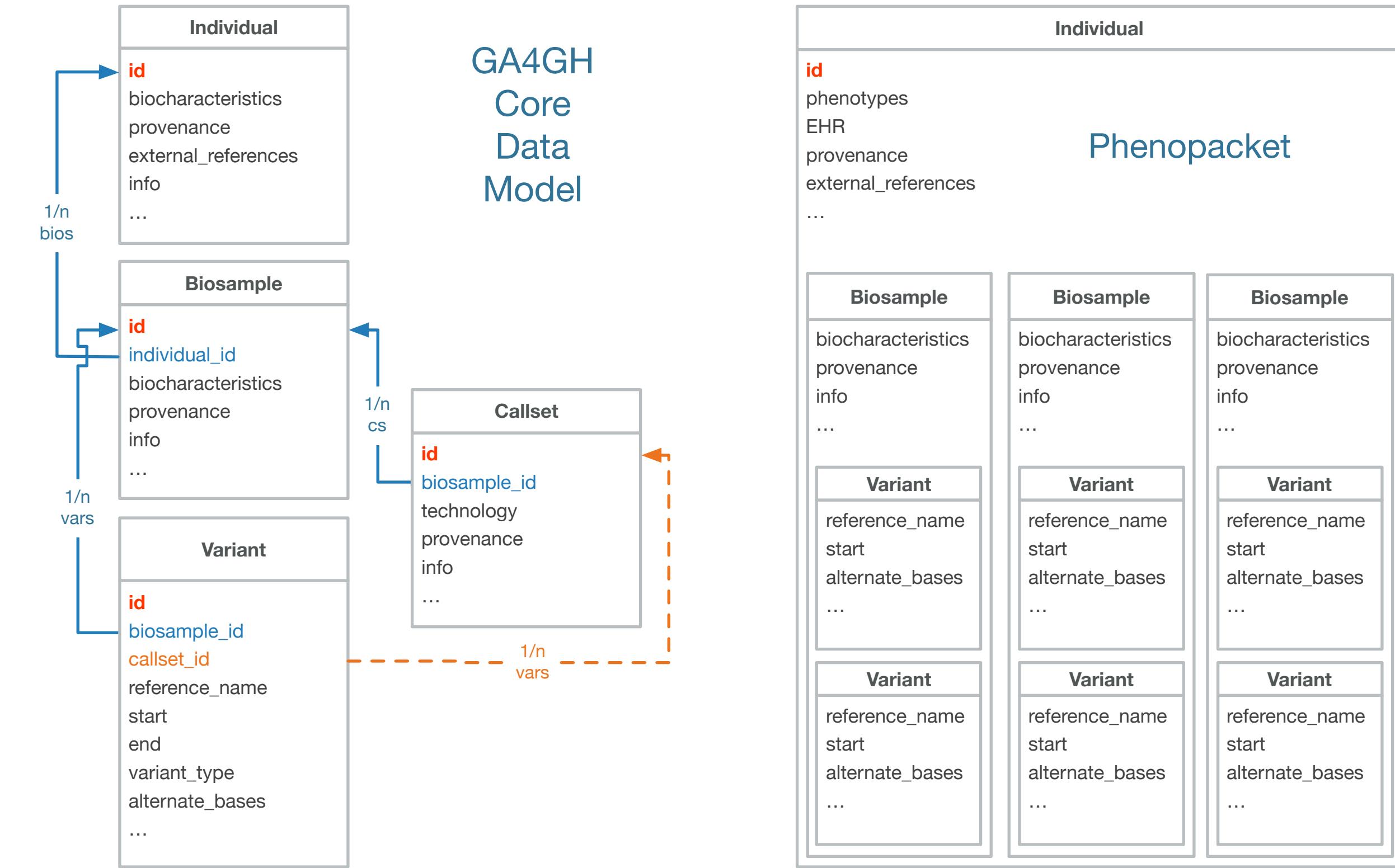
People



github.com/ga4gh-beacon/

Standardized Data Model for Consistent Schema Development

- A consistent high-level data model is essential for the development of reliable schemas and tools for
 - genomic and clinical, metadata storage
 - development of genomic query and data delivery APIs
 - distributed/federated access across separate (geographic, logistic) data repositories using consistent logical structure:
 - "**BRCA1 variant** in **germline sample** from a male **individual** with a diagnosis of breast carcinoma ([ncit:C5214](#))
- The abstract data model can be expressed in different types of implementations
 - Phenopackets data exchange standard
 - Progenetix database model
 - schema-derived object storage datacollections for individuals, biosamples, callsets and variants



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”



GA4GH :: SchemaBlocks

An Initiative by Members of the Global Alliance for Genomics and Health

About {S}[B]

News

Participants

Data Formats

Data Schemas

Examples, Guides & FAQ

Meeting minutes

Contacts

Related Sites

[GA4GH::Discovery](#)

[GA4GH::CLP](#)

[GA4GH::GKS](#)

[ELIXIR Beacon](#)

[Phenopackets](#)

[GA4GH](#)

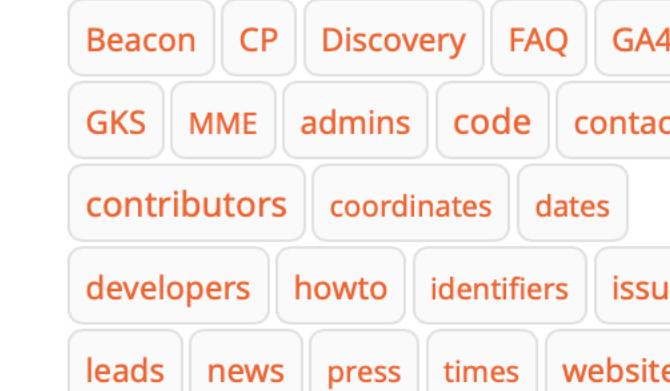
[Beacon+](#)

Github Projects

[SchemaBlocks](#)

[ELIXIR Beacon](#)

Tags



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

schemablocks.org

GA4GH Data Model

Recommendation (DRAFT)

The GA4GH data model recommends the use of a default object hierarchy in standard and product design processes. While it reflects concepts from the original [GA4GH schema](#), it provides mostly a structural guideline for API and data store design, but is not thought to provide a set of absolute implementation requirements.

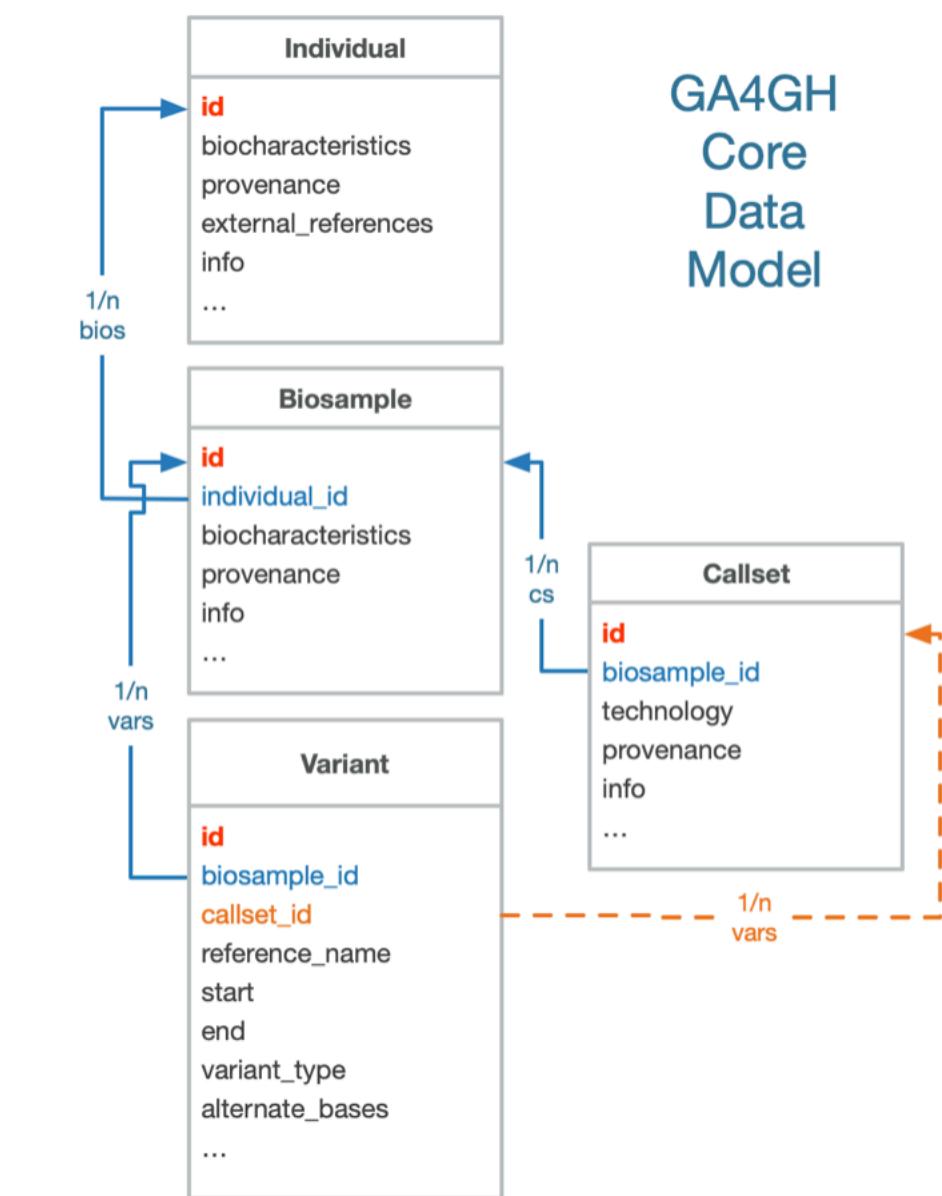
Contributors

- [@mcourtot](#)
- [@mbaudis](#)

Summary

The GA4GH data model for genomics recommends the use of a principle object hierarchy, consisting of

- **variant**
 - a single molecular observation, e.g. a genomic variant observed in the analysis of the DNA from a biosample
- **callset**
 - the entirety of all variants, observed in a single experiment on a single sample
 - a **callset** can be compared to a data column in a **VCF** variant annotation file
 - **callset** has an optional position in the object hierarchy, since **variants** describe biological observations in a biosample
- **biosample**
 - a reference to a physical biological specimen on which analyses are performed
- **individual**
 - in a typical use a human subject from which the biosample(s) was/were extracted



A graph showing recommended basic objects and their relationships. The names and attributes are examples and may diverge in count and specific wording (e.g. "subject" instead of "individual") in specific implementations.

These basic definitions will be detailed further on.

Additional concepts (e.g. *dataset*, *study* ...) may be added in the future.



{S}[B] SchemaBlocks - Future Directions

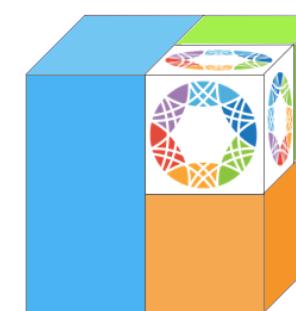


- Receive continuous contributions from WS in form of “blocks” and documentation through interaction w/ different development teams
 - Variant annotation types and models from **GKS**
 - Ontology, phenotype format & recommendations from **C/P (Phenopackets...)**
 - Search components from **Discovery & Beacon**, usage conditions (**DUO/ADAM**)...
- Formalise approval levels & governance model
- Become part of GA4GH product approval process (TASC proposal)
 - products in approval process demonstrate document awareness of SchemaBlocks through
 - Contribution of code or documentation
 - Use of existing code or formats
 - (Or Statement about lack of applicability...)



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 / kategesis / chromoplexis...)
- Annotation and exchange formats have to move towards extensible models
 - referring reference genome positions, w/ remapping, provenance
 - technology agnostic (but provenance...)
- Search and exchange APIs have to accommodate distributed and/or federated data access models
 - modular object design, independent from backend structure
 - common interfaces/service APIs/registries



ELIXIR h-CNV Community

- 1st community meeting September 2018 - Hinxton
- h-CNV community approved in February 2019
- Implementation study accepted: start 2019-06-01
- community workshop ELIXIR All-Hands Lisbon
- chaired by Christophe Béroud and David Salgado (INSERM Marseille)

Node	Name of PI
ELIXIR-FR	Christophe Béroud, David Salgado, Marc Hanauer, Victoria Dominguez
ELIXIR-CH	Michael Baudis
ELIXIR-DE	Jan Korbel
EMBL-EBI	Thomas Keane, Fiona Cunningham
ELIXIR-ES	Joaquin Dopazo, Alfonso Valencia, Salvador Capella, Sergi Beltran, Steven Laurie, Gemma Bullich, Laura I. Furlong, Janet Piñero
ELIXIR Hub	John Hancock, Gary Saunders, Kathi Lauer, Leyla Garcia
ELIXIR-NL	Bauke Ylstra, Daoud Sie, Leon Mei, Morris Swertz (UMCG), Lennart Johansson
ELIXIR-NO	Eivind Hovig, Pubudu Samarakoon
ELIXIR-HU	Attila Gyenessei ,Katalin Monostory
ELIXIR-SI	Brane Leskošek, Polonca Ferk, Marko Vidak
ELIXIR-UK	Krzysztof Poterlowicz


[ABOUT US](#) [SERVICES](#) [PLATFORMS](#) [COMMUNITIES](#) [EVENTS](#) [NEWS](#) [INTRANET](#)

Home » Communities »

COMMUNITIES

[Human Data](#)

[Rare Diseases](#)

[Marine Metagenomics](#)

[Plant Sciences](#)

[Metabolomics](#)

[Proteomics](#)

[Galaxy](#)

[Human Copy Number Variation](#)

Human Copy Number Variation

Most human cells divide by mitosis, where one parent cell divides to produce two daughter cells. Just before a cell divides it makes a copy of its own DNA, but this copy may not have the same gene sequence as the parent DNA. One gene, for example, may be copied twice into the new DNA, or not copied at all. This phenomenon is called **Copy Number Variation (CNV)**.

It is thought that these Copy Number Variations are vital for evolution, but they also play important roles in disease. Despite the fact that Copy Number Variations are the most prevalent genetic mutation type, identifying and interpreting them is still a major challenge. The ELIXIR human Copy Number Variation (hCNV) Community aim to implement processes to make the detection, annotation and interpretation of these variations easier.



{B}H{H}

BIOHACKATHON-EUROPE

PARIS 2019 November 18th - 22th

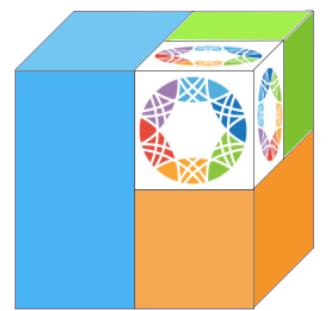
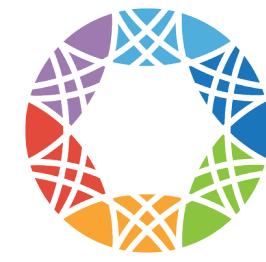
@BIOHACKTHONEU





Michael's Hopes for this Visit

- Awareness of **standards**, projects supported by **GA4GH**, **ELIXIR**
 - **Beacon** v1.1 => v2.n as the standard for distributed "variant plus" queries
 - Active engagement as **stakeholders & developers** in GA4GH / ELIXIR projects
- **h-CNV** is, again, hot topic to contribute to:
 - Engage with the emerging **ELIXIR h-CNV** community (standards, technical definitions...)
 - Contributing data from cancer genome screening experiments to the **Progenetix** resource to increase exposure & improve representation (**rare cancers**, **population** effects ...)
 - Collaborative **projects** in cancer CNV landscape analyses?
- Increase **cross-talk**, knowledge **exchange** between **groups** and **organizations**



BAUDISGROUP @ UZH

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

SIB

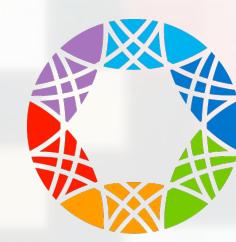
AMOS BAIROCH
HEINZ STOCKINGER
DANIEL TEIXEIRA

@WORLD

MATTHIAS ALTMAYER
THOMAS EGGERMANN
ROSA NOGUERA
REINER SIEBERT
CAIUS SOLOVAN



University of
Zurich^{UZH}



Global Alliance
for Genomics & Health

GA4GH

LARRY BABB
ANTHONY BROOKES
MELANIE COURTOT
MELISSA HAENDEL
MICHAEL MILLER
HELEN PARKINSON
GUNNAR RÄTSCH
ANDY YATES

ELIXIR & CRG

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

H-CNV

CHRISTOPHE BÉROUD
DAVID SALGADO



University of
Zurich^{UZH}



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



Global Alliance
for Genomics & Health



