



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

# Beacon API v2

# Structural Variants Scout Team

Michael Baudis | GA4GH Connect March 2020

# HOW it looks like? *Like a REST API*



Global Alliance  
for Genomics & Health

## Allow for entity queries

- Lists
  - .../**gVariants**
  - .../**samples**
  - .../**individuals**
- Individual entities
  - .../gVariants/{**id**}
  - .../samples/{**id**}
  - .../individuals/{**id**}

## Filtering

.../**gvariants?**

**chr=1**

**&pos=1001**

**&alt=T**

.../**individuals?**

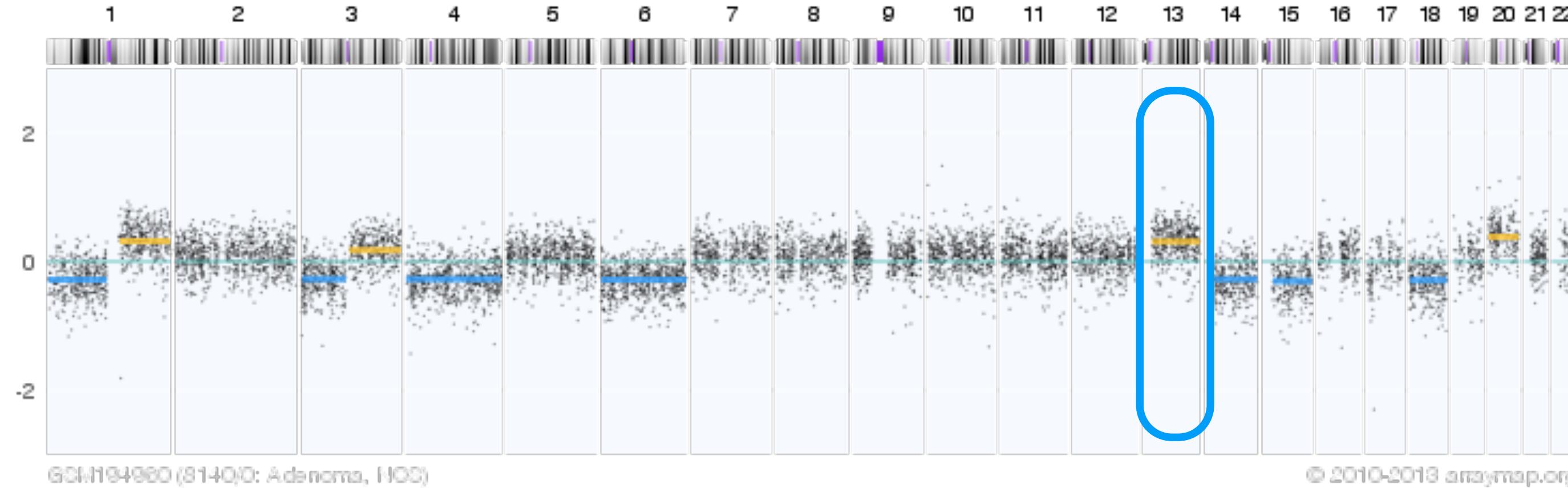
**filters=...** (see later)



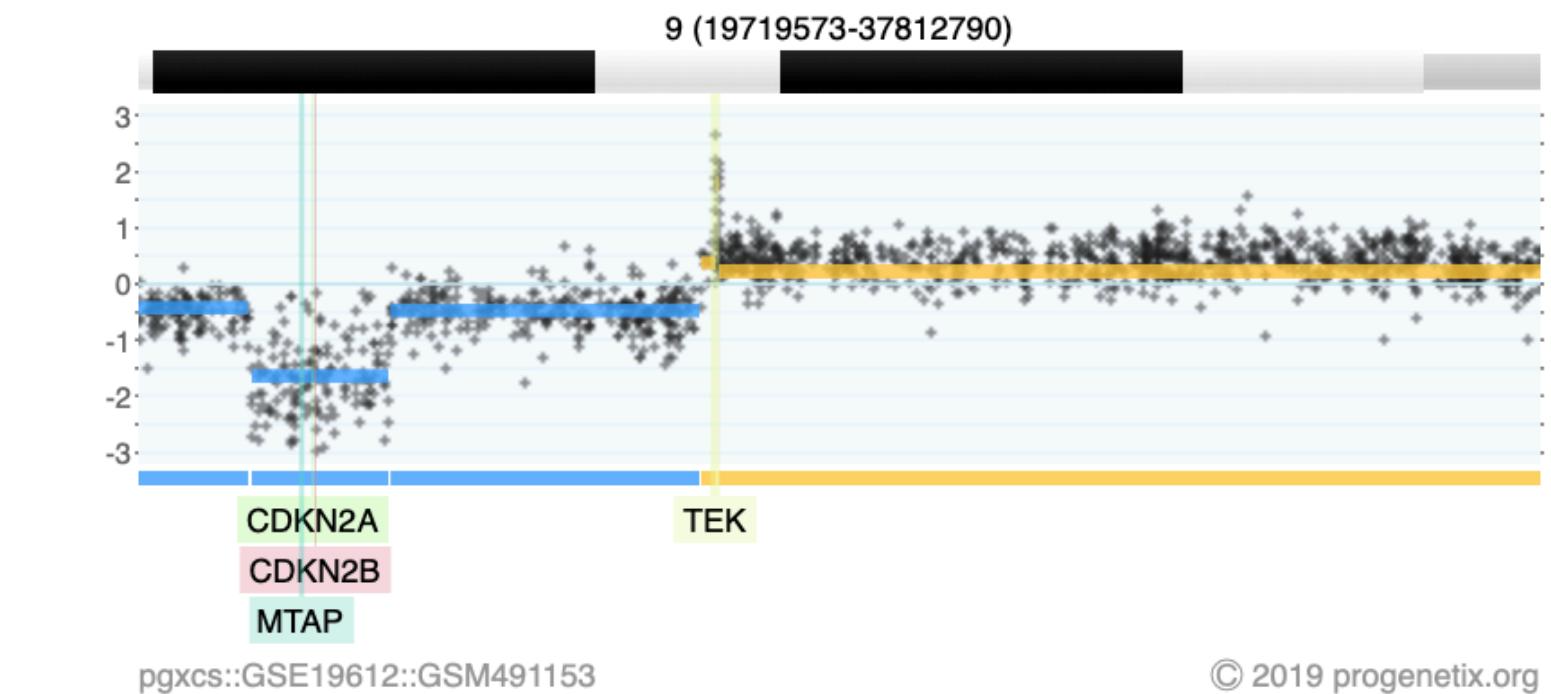
Fundación Progreso y Salud  
CONSEJERÍA DE SALUD



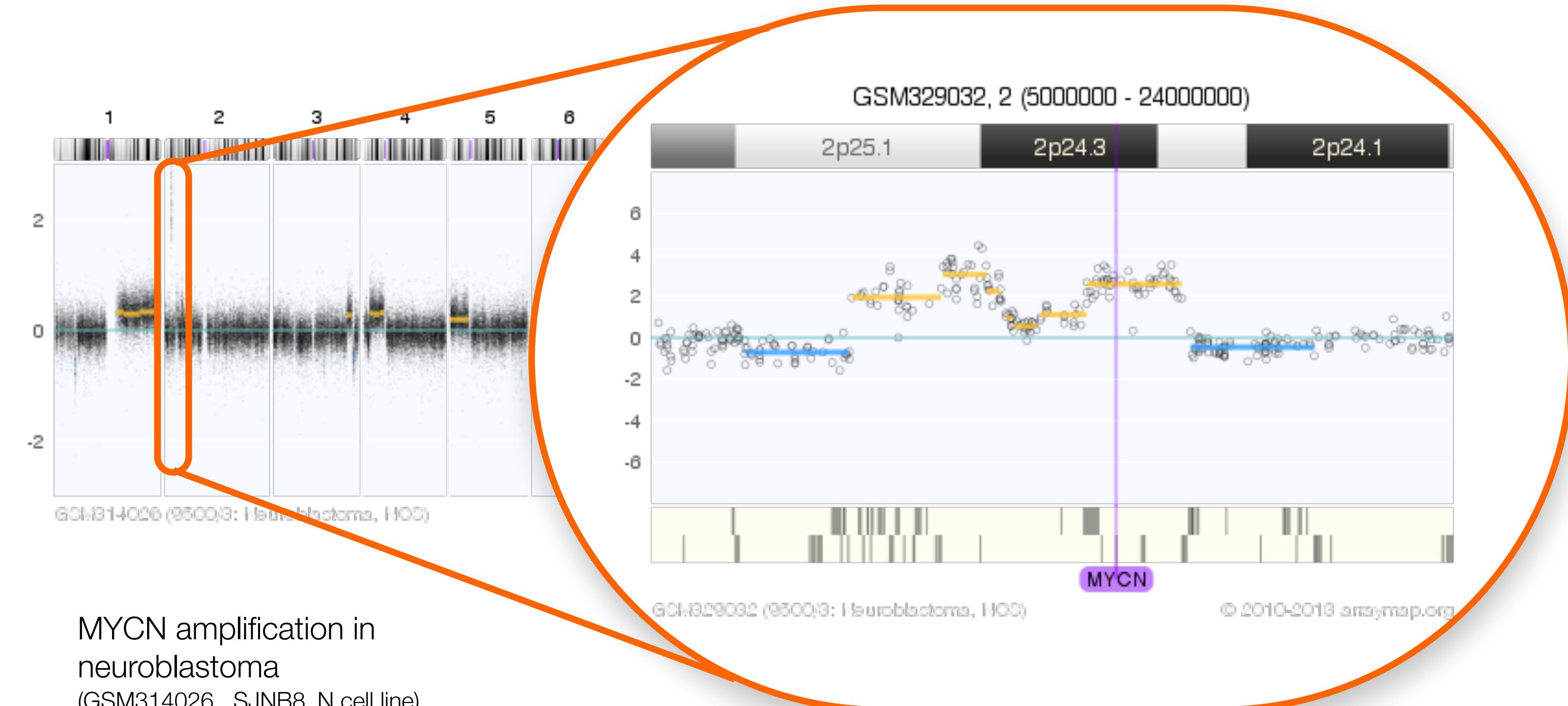
# Somatic Copy Number Variations



Gain of chromosome arm 13q in colorectal carcinoma



2-event, homozygous deletion in a Glioblastoma



MYCN amplification in neuroblastoma  
(GSM314026, SJNB8\_N cell line)

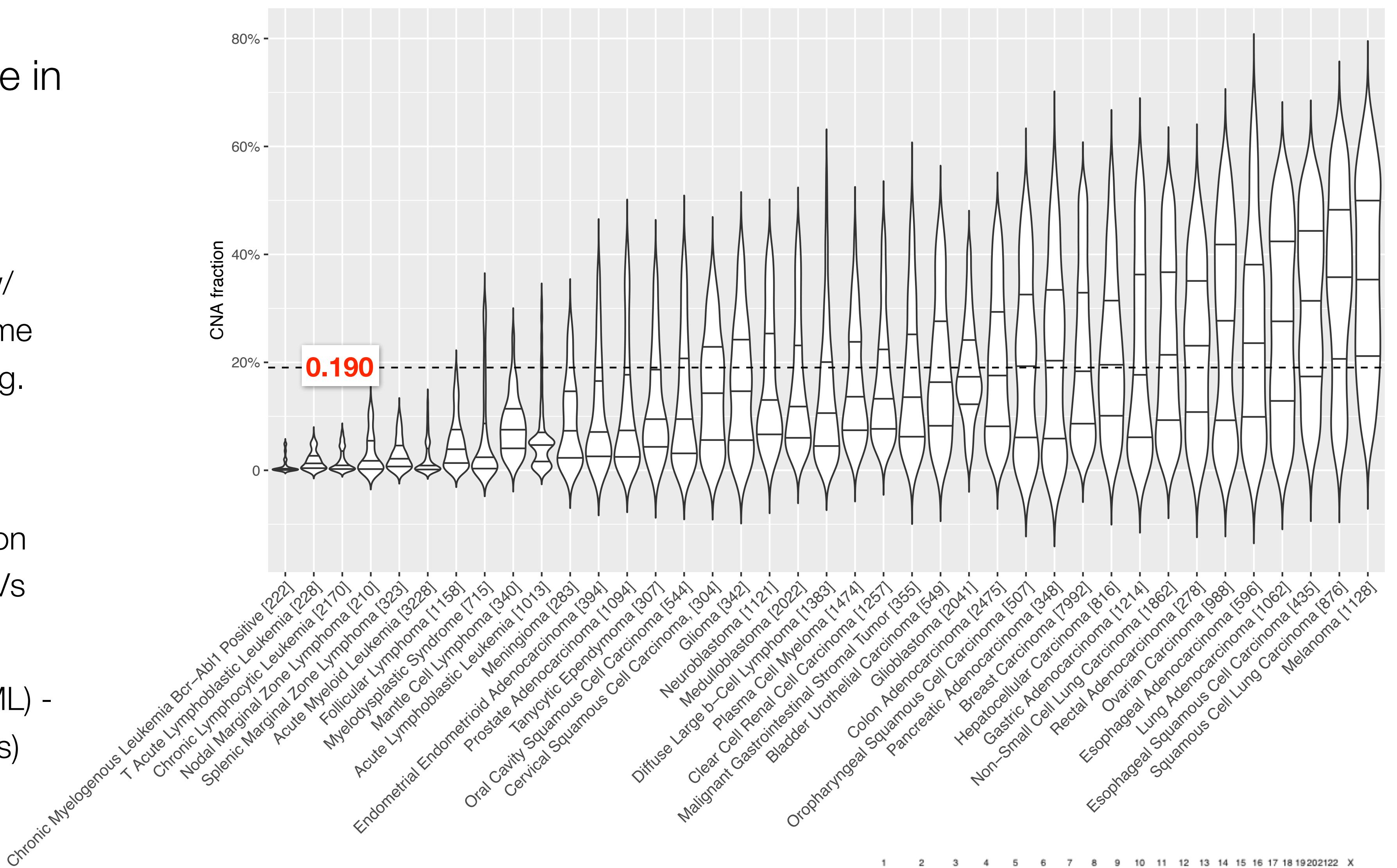
**low level/high level** copy number alterations (CNAs)

arrayMap

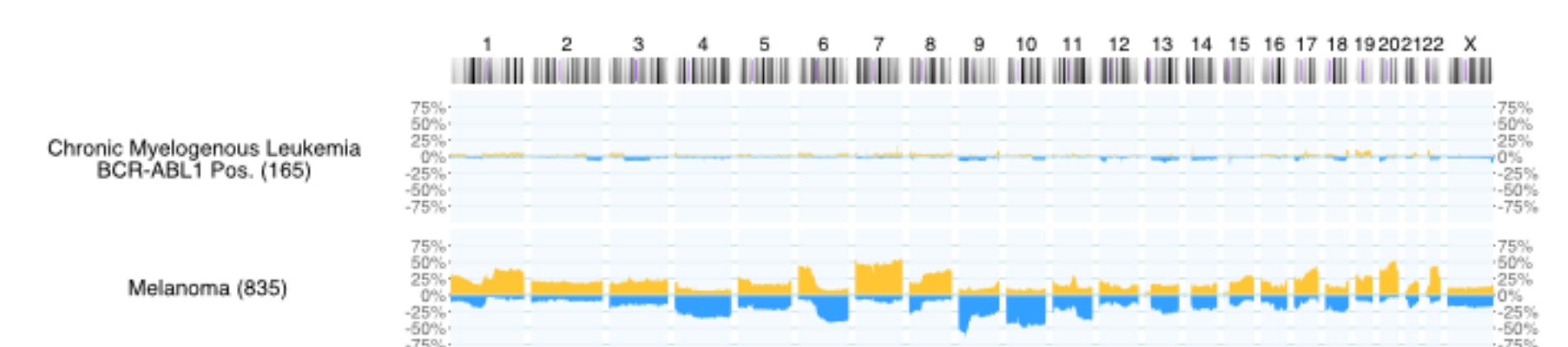


# Genome CNV coverage in Cancer Classes

- 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 =>



# CNV queries in v1.n

- 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](http://schemablocks.org)

[beacon.progenetix.org/ui/](http://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

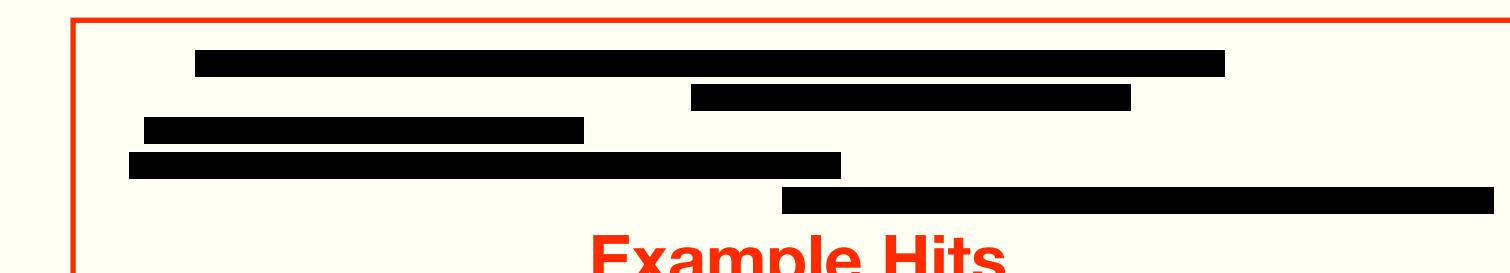
Progenetix datasets

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

startMin - startMax

CDKN2A CDR

endMin - endMax



Example Hits

Filters

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

# Beacon v2 Scout Team "Structural Variants"



Global Alliance  
for Genomics & Health

- SV Scout Team evaluates “structural” types of genomic variants

- ▶ selection
- ▶ representation
- ▶ query

- intersects with
  - ▶ ELIXIR h-CNV
  - ▶ GA4GH GKS

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

**News**  
[Specification & Roadmap](#)  
[Beacon Networks](#)  
[Events](#)  
[Examples, Guides & FAQ](#)  
[Contributors & Teams](#)  
[Contacts](#)  
[Meeting Minutes](#)

**Related Sites**  
[ELIXIR BeaconNetwork](#)  
[Beacon @ ELIXIR](#)  
[GA4GH](#)  
[beacon-network.org](#)  
[Beacon+](#)  
[GA4GH::SchemaBlocks](#)  
[GA4GH::Discovery](#)

**Github Projects**  
[Beacon API and Tools](#)  
[SchemaBlocks](#)

**Tags**  
[CNV](#) [EB](#) [FAQ](#) [SV](#) [VCF](#) [beacon](#)

**Teams: Beacon Structural Variants**

The Structural Variants Team evaluates “structural” types of genomic variants with respect to their representation & query through current & future versions of the Beacon API. This scope intersects with work being pursued in the [ELIXIR h-CNV community](#) and considers standards under development by the GA4GH GKS workstream.

**Members**

- Babita Singh (chair)
- David Salgado
- Tony Brooks
- Michael Baudis

**Links**

- CNV Annotation Formats at ELIXIR
- BeaconCnvRequest prototype schema
- meeting minutes 2020

@mbaudis 2020-01-15

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

**News**  
[Specification & Roadmap](#)  
[Beacon Networks](#)  
[Events](#)  
[Examples, Guides & FAQ](#)  
[Contributors & Teams](#)  
[Contacts](#)  
[Meeting Minutes](#)

**Related Sites**  
[ELIXIR BeaconNetwork](#)  
[Beacon @ ELIXIR](#)  
[GA4GH](#)  
[beacon-network.org](#)  
[Beacon+](#)  
[GA4GH::SchemaBlocks](#)  
[GA4GH::Discovery](#)

**Github Projects**  
[Beacon API and Tools](#)  
[SchemaBlocks](#)

**Tags**  
[CNV](#) [EB](#) [FAQ](#) [SV](#) [VCF](#)  
[beacon](#) [clinical](#) [code](#)  
[compliance](#) [contacts](#) [definitions](#)  
[developers](#) [development](#)  
[events](#) [filters](#) [minutes](#) [network](#)  
[press](#) [proposal](#) [queries](#)  
[releases](#) [roadmap](#) [specification](#)  
[teams](#) [v2](#) [versions](#) [website](#)

**BeaconCnvRequest [beacon ↗]**

{S}[B] Status [i]	proposed
Provenance	<ul style="list-style-type: none"><li>• Beacon API</li></ul>
Used by	<ul style="list-style-type: none"><li>• Beacon</li><li>• Progenetix database schema (Beacon+ backend)</li></ul>
Contributors	<ul style="list-style-type: none"><li>• Michael Baudis</li><li>• Beacon developers...</li></ul>
Source (v1.1.0)	<ul style="list-style-type: none"><li>• raw source [JSON]</li><li>• Github</li></ul>

**Attributes**  
**Type:** object  
**Description:** CNV request against Beacon resources.  
TODO: The documented queries will fail current tests due to the lack of the `referenceBases` parameter.

**Properties**

Property	Type
assemblyId	string
datasetIds	array of string
endMax	integer (int64)
endMin	integer (int64)
includeDatasetResponses	string
referenceName	<a href="https://beacon-project.io/schemas/beacon/v1.1.0/Chromosome.json">https://beacon-project.io/schemas/beacon/v1.1.0/Chromosome.json</a> [SRC] [HTML]
startMax	integer (int64)
startMin	integer (int64)
variantType	string

# Beacon v2 SVST: First Conclusions



Global Alliance  
for Genomics & Health

- There is some confusion about variant types which are result of a functional attribution in contrast to "sequence change type" (ME? CON? ...)
- Some "named VCF variants" ambiguous, difficult to "beaconify" and/or duplication of alternative representations (e.g. "INDEL" - precise or structural? how to disambiguate in query for e.g. copy number effect?)
- Structural Variants will be addressed by a number of "typed queries"
- Starting with **focus on** sequence variations with **unanimous** "find by position(s) and variant type" (DUP, DEL, TL ...)
- Provide a **general "anything here" query type**, which will allow the matching of different variants (powerful when combined with responses detailing matches)
- Add well scoped, specific query types during v2 => v2.n development
- While VCF presents a basis for scoping, Beacon development will not try to "recapitulate everything VCF"

# 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



elxir

[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.



HUMAN GENOMIC VARIATION SOCIETY

COPY NUMBER VARIATION - DATA COLLECTION & ANALYSIS

Göteborg, Sweden

Friday 14th June | satellite to ESHG





{B} {H}

BIOHACKATHON-EUROPE

PARIS 2019 November 18th - 22th

@BIOHACKTHONEU



## People

- ▶ Babita Singh (chair)
- ▶ David Salgado
- ▶ Tony Brooks
- ▶ Michael Baudis
- ▶ Diana Lemos

**Please get involved!**

SV Beacon Scouts document and minutes - [link](#)

GA4GH Beacon working group meeting minutes - [link](#)

h-CNV community page - [link](#)