Developing a User Interface for Sharing Federated Genomic and Phenotypic Data Using the Beacon v2 protocol

Dionne So, Ricky Nguyen, Jessica Do, Zoe Kamarinos

Children's Cancer Institute

Australian BioCommons

What is the Purpose?

Human genomics research problem

- A major obstacle in human genomics research is the **paucity of accessible, well-annotated genomic** and **phenotypic data,** especially when it comes to data on rare cancer subtypes.

- There is currently no way to access aggregated data on high-risk children cancer, thus the **availability of data is imperative** for personalised medicines, inclusive diagnostics, prognostic and therapeutic strategies.

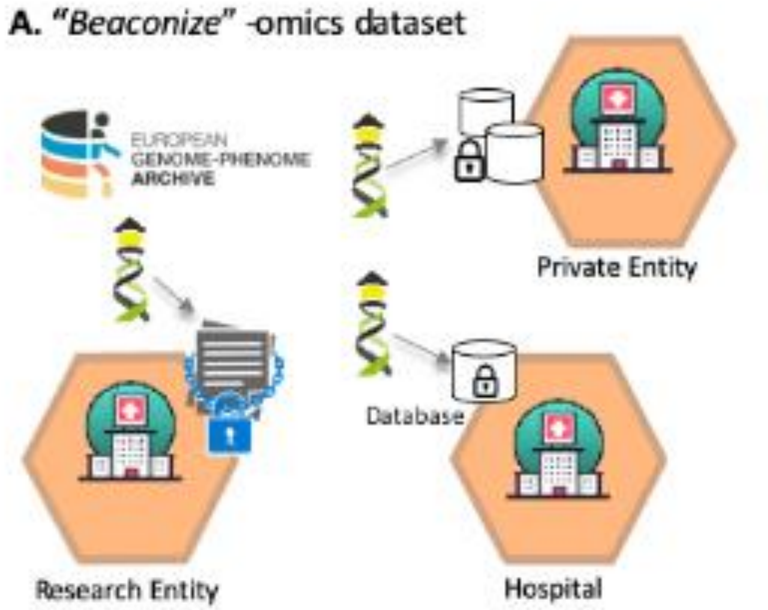
Children's Cancer Institute

Australian BioCommons

Beacon Network Protocol

How does it help solve the problem?

- GA4GH Beacon specification – global standard
- Facilitate querying of aggregated data
- v1 and v2



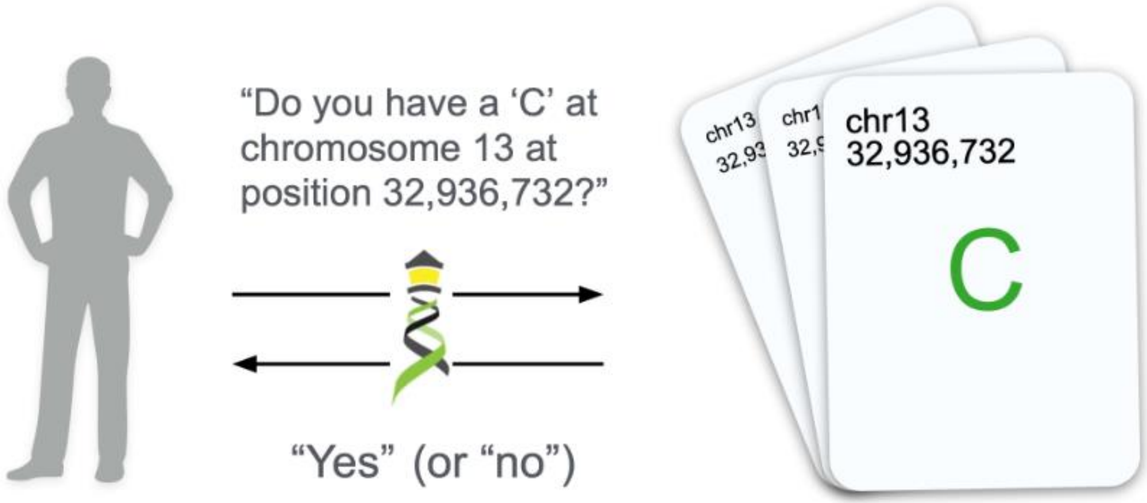
Children's Cancer Institute

Australian BioCommons

The v1 Beacon Network

- Originally, the Beacon protocol (version 0 and 1) allowed researchers to obtain genomic variant data; the ELIXIR Beacon network is an example of an UI that extends this.

- Information about the presence/absence of a given, specific, genomic mutation in a set of data, from patients of a given disease or the population in general.



Children's Cancer Institute

Australian BioCommons

The v2 Beacon Network

- Accepted as GA4GH standard in April 2022.
- Beacon v2 provides greater structured data responses to Beacon queries, introducing new features such as: annotated genomic variations, biosamples, data about individuals and study cohorts.

Beacon V2 Reference Implementation: a Toolkit to enable federated sharing of genomic and phenotypic data

Get access

Manuel Rueda ✉, Roberto Ariosa, Mauricio Moldes, Jordi Rambla

Bioinformatics, btac568, https://doi.org/10.1093/bioinformatics/btac568

Children's Cancer Institute

Australian BioCommons

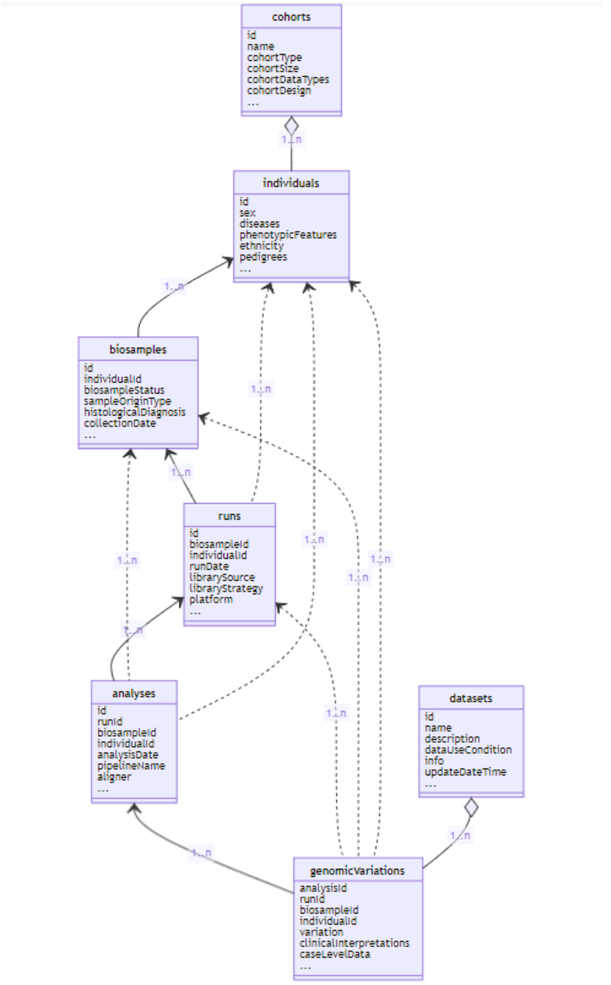
Beacon Network v2

The Model

The Beacon v2 default data model provides a set of schemas for common data entities with a focus on biomedical genomics.

The entities are defined as follows;

- Collections (**Datasets** and **Cohorts**)
- **Genomic variations**
- **Individuals**
- **Biosamples**
- **Analyses & Runs**



Beacon v2 Models entities and their relationships

Children's Cancer Institute

Australian BioCommons

The Beacon Instance

What is a Beacon?

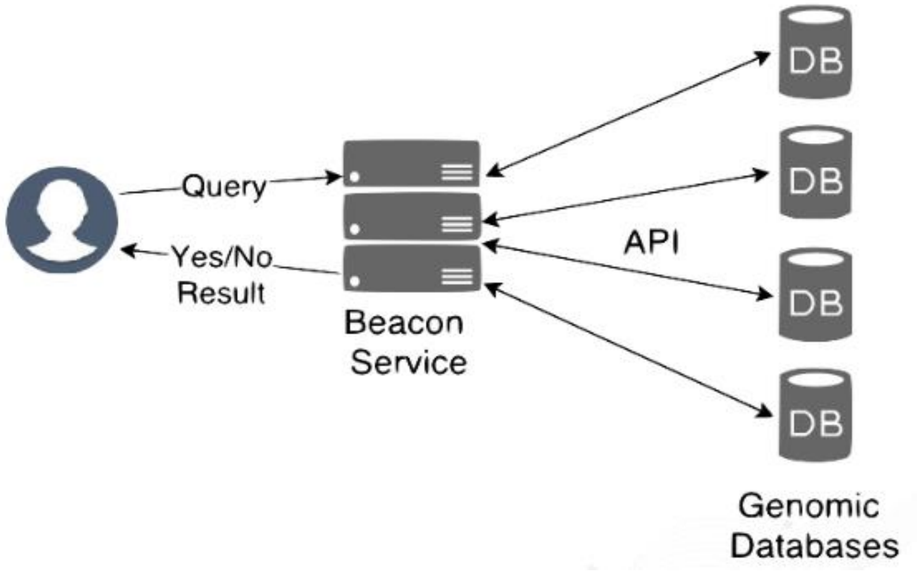
A beacon is a simple yet powerful genomics API that allows for federated data discovery and retrieval.

SIMPLE
Beacons can be implemented on top of any underlying variant or variant annotation datastore

FEDERATED
Beacons can be lit and maintained by individual organizations, and assembled into a distributed network

AGGREGATIVE
Beacon results can be aggregated across an entire populations to support de-identification.

SECURABLE
Beacon access can be restricted using institutional security protocols, and authorization schemes can be implemented to respect consent conditions.



Beacon architecture
M. M. A. Aziz, R. Ghasemi, M. Waliullah, and N. Mohammed, "Aftermath of bustamante attack on genomic beacon service", BMC Medical Genomics, vol. 10, no. S2, 2017.

Children's Cancer Institute

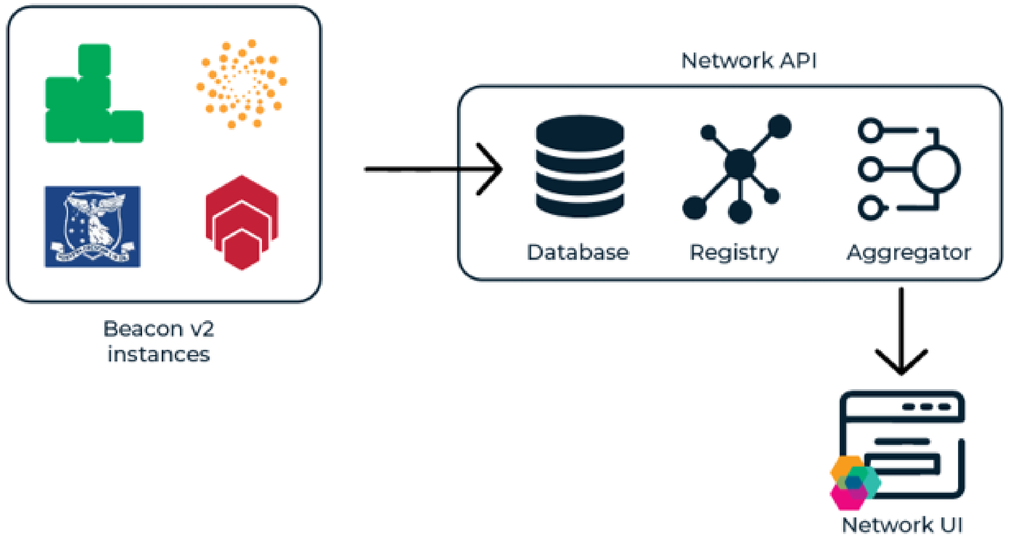
Australian BioCommons

The Beacon Network

The Beacon network is made up of federated data sources (beacons) that can be collectively searched with a single query. The network also contains two components: The Aggregator and the Registry.

The Aggregator

Aggregates return as a single query



The Registry

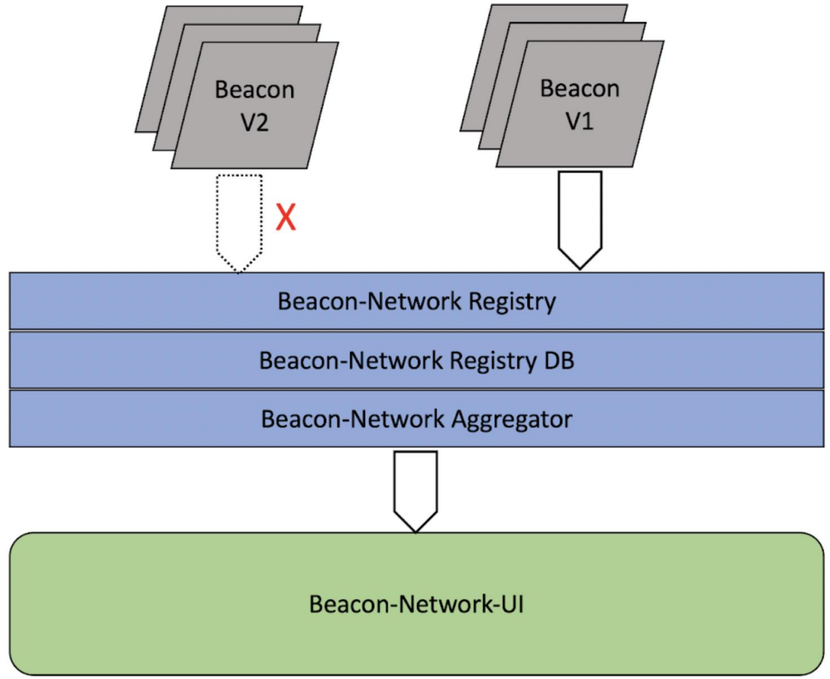
Keeps track of all the Beacon instances

Children's Cancer Institute

Australian BioCommons

Current Issue with the Beacon Network

- Currently most existing Beacon interfaces aggregate data from the v1 Beacon network.
- The current version (v2) greatly increased the complexity and richness of the metadata.
- Necessitates a more sophisticated user interface



Children's Cancer Institute

Australian BioCommons

Very nested metadata


```
"collectionEvents": [
    {
        "eventDataTypes": {
            "availability": true,
            "distribution": {
                "dataTypes": {
                    "blood collected from fasting subject": 51,
                    "survey data": 98
                }
            }
        },
        "eventDate": "2019-04-23",
        "eventEthnicities": {
            "availability": true,
            "availabilityCount": 101,
            "distribution": {
                "ethnicities": {
                    "African": 3,
                    "European": 90,
                    "Latin American": 8
                }
            }
        },
        "eventGenders": {
            "availability": true,
            "availabilityCount": 101,
            "distribution": {
                "genders": {
                    "female": 51,
                    "male": 50
                }
            }
        },
        "eventNum": 1,
        "eventSize": 101
    }
],
```


Cohorts

Term	Description	Type	Properties	Example	Enum
collectionEvents	TBD	array	eventAgeRange, eventCases, eventControls, eventDataTypes, eventDate, eventDiseases, eventEthnicities, eventGenders, eventLocations, eventNum, eventPhenotypes, eventSize, eventTimeline	NA	NA

One term from the Cohort Schema of the Beacon v2 model
Taken from the GA4GH Beacon v2 Documentation

Children's Cancer Institute

Australian BioCommons

Our Solution

Creating a **centralised** and **efficient** method of sharing genomic and clinical data via a user interface fronting the Beacon Network v2.

- Intuitive user interface
- Major advancement in human genomic data sharing in Australia
- Open source
- Enable querying the full catalogue of data by all members of the Beacon network.

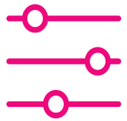
Children's Cancer Institute

Australian BioCommons

Main Requirements

To be able to refine and search queries

A dashboard to display aggregated results



To be able to sort results based on a criteria

Children's Cancer Institute

Australian BioCommons

The User Interface

The HGP Beacon Network

HGP Beacon Network

The project is designed to enhance capability for securely and responsibly sharing human genome research data nationally and internationally, ensuring maximum value can be derived from these valuable assets. It is investigating best practice technologies that have been globally developed for the purposes of human genome data sharing, and deploying Australian first technologies in the form of a 'services toolbox' for improving FAIRness of genomic data at the organisations that hold most human genomes collected for research in Australia.

Login

Sign Up

Get in touch

Do you have a question about our work?
For any enquiries please don't hesitate to contact us.

Contact Us

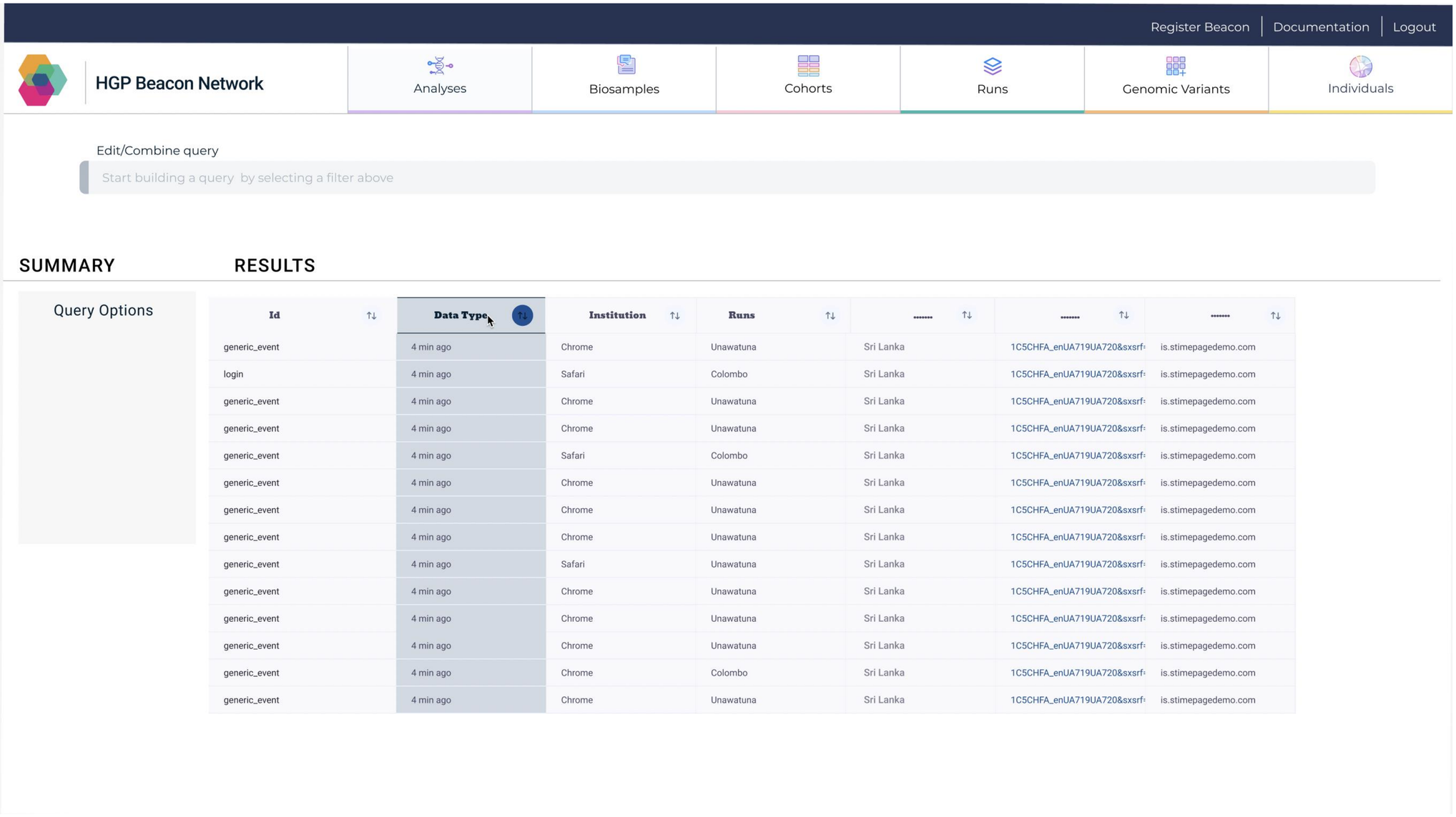
Australian BioCommons

Data Overview

BEACONS	INDIVIDUALS	BIOSAMPLES	ANALYSES	DATASETS
1000	1000	1000	1000	1000

Children's Cancer Institute

Australian BioCommons



Children's Cancer Institute

Australian BioCommons

User Interface FEATURES

Query bar
Displays the query terms as chips when the query option is selected.

Results table
Tabular representation of the filtered information that can be further sorted or paginated.

Summary bar
A total count for the results returned by each query option selected.

Dropdown Menus
Specialised query terms under each query option, which allow for a more specified query.

Input Menus
Prompts the user for a single or double term based on an Id or Label.

Chips
Visually represents the querying options selected.

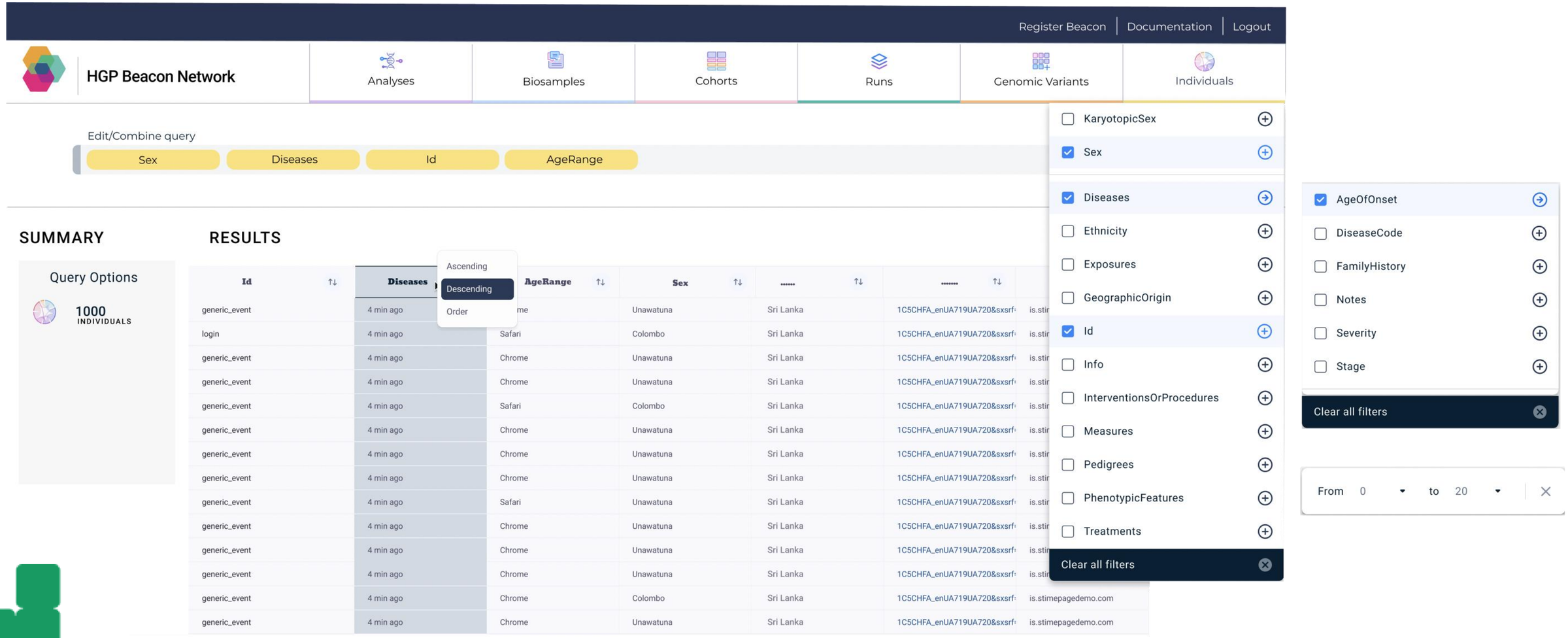
Children's Cancer Institute

Australian BioCommons

Use Cases

As a researcher:

I want to find all paediatric patients (age < 21 yrs)

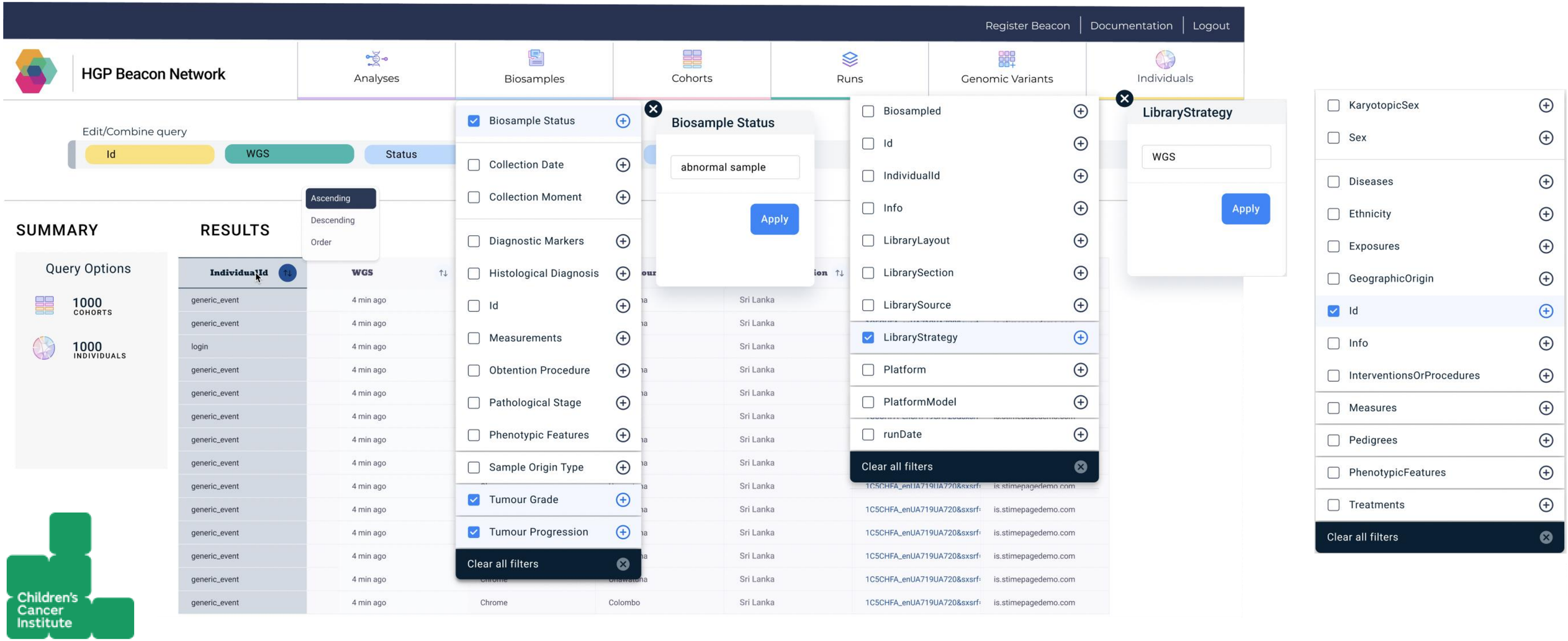


Children's Cancer Institute

Australian BioCommons

As a researcher:

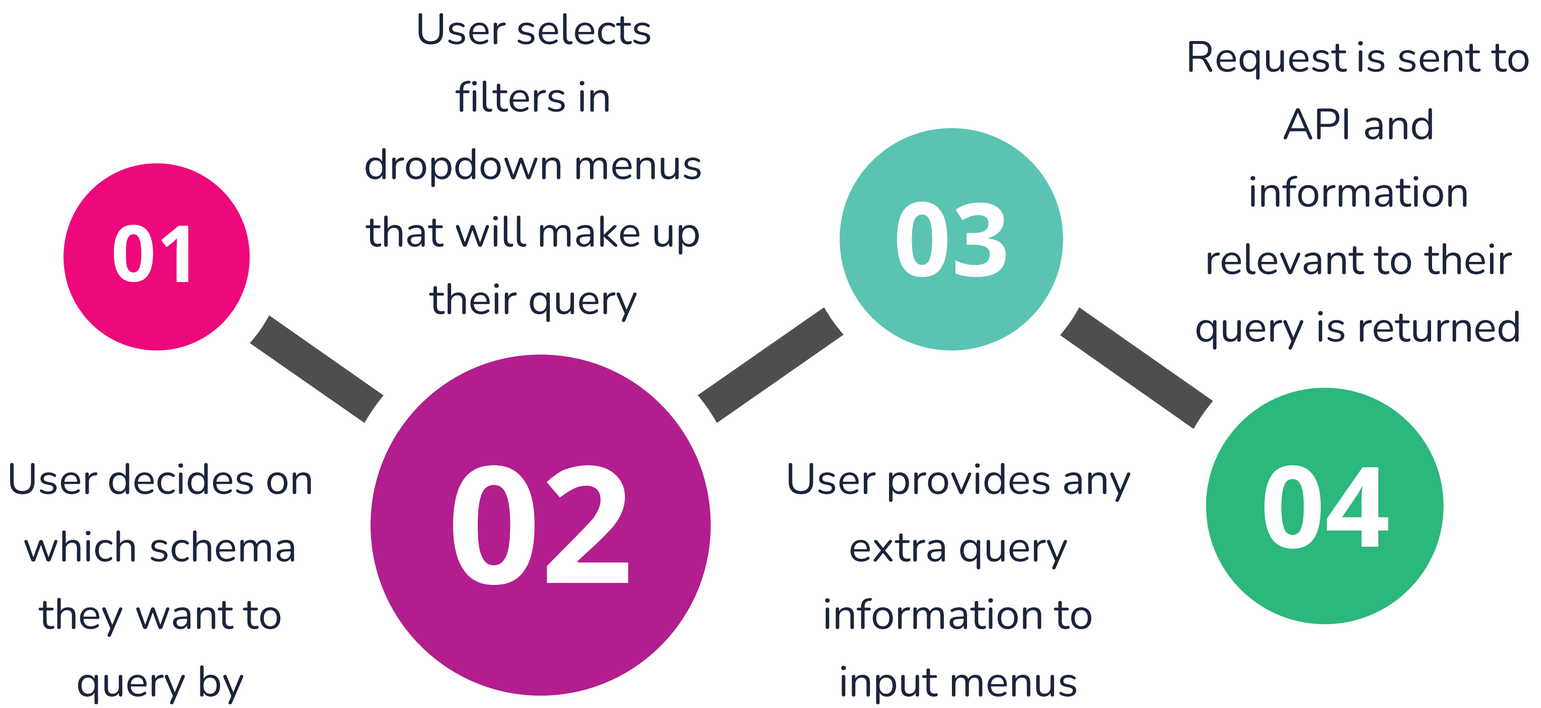
I want to know how many patients have at least two tumours that have undergone WGS



Children's Cancer Institute

Australian BioCommons

USER PROCESS

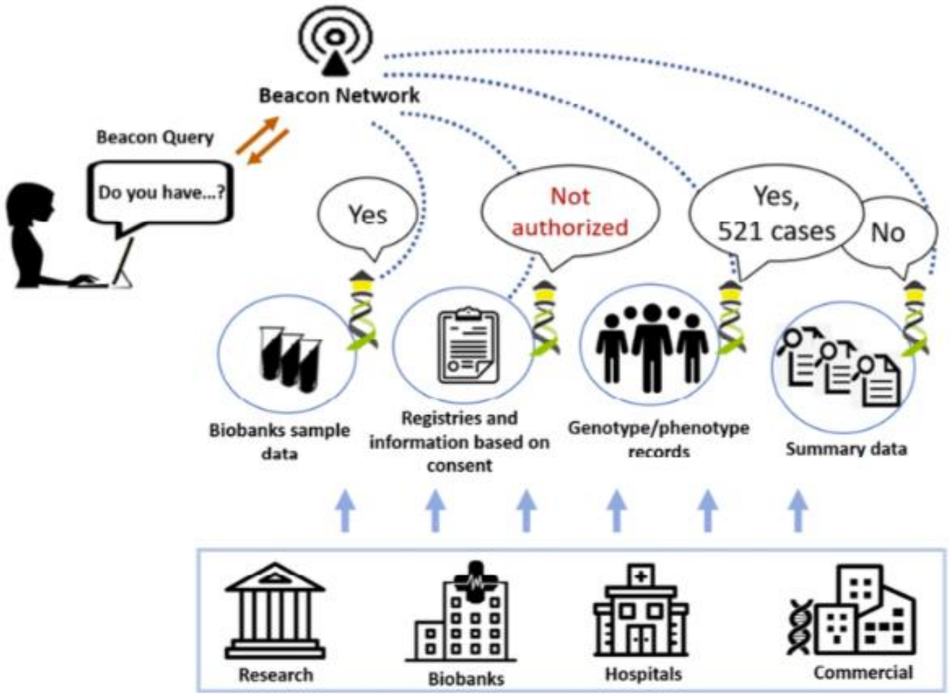


Children's Cancer Institute

Australian BioCommons

End Vision

- Underlying motivation for stems from the limited cohort of available cancer data that each institute has alone.
- Bolster the range of patient data through the updated UI that supports the new v2 Beacon network.
- Provide quick and easy access to query a network of extensive cancer data.
- Facilitate comprehensive cancer research and enable confidence to make statistical inferences when profiling tumors.



Beacon v2 Networks

Children's Cancer Institute

Australian BioCommons

Special Thanks to

Kamile Taouk (6), Angela Lin (6), Marie Wong-Erasmus (6), Mark Cowley (6), Tiffany Boughtwood (3), Jeff Christiansen (1), Joe Copty (7), Shyamsundar Ravishankar (7), Kylie Davies (2), Matthew Downton (9), Kelsey Druken (9), Ben Evans (9), Clara Gaff (8), Andrew Gilbert (5), Christina Hall (1), Matthew Hobbs (7), Oliver Hofmann (11), Jessica Holliday (1), Warren Kaplan (9), Ross Koufariotis (10), Sarah Kummerfeld (7), Conrad Leonard (10), Andrew Lonie (1), Heath Marks (2), Siobhann McCafferty (4), David Monro (9), Andrew Patterson (11), John Pearson (10), Bernard Pope (1), Shyamsundar Ravishankar (7), Florian Reisinger (11), Andrew Robinson (9), Victor San Kho Lin (11), John Scullen (2), Marion Shadbolt (1), Andrew Treloar (4), Jingbo Wang (9), Scott Wood (10), Mustafa Syed (6).

1. Australian BioCommons 2. Australian Access Federation (AAF) 3. Australian Genomics 4. Australian Research Data Commons (ARDC) 5. Bioplatforms Australia 6. Children's Cancer Institute (CCI) and the Zero Childhood Cancer Program (ZERO) 7. Garvan Institute of Medical Research 8. Melbourne Genomics Health Alliance 9. National Computational Infrastructure (NCI) 10. QIMR Berghofer Medical Research Institute (QIMRB) 11. The University of Melbourne Centre for Cancer Research (UMCCR).

Children's Cancer Institute

Australian BioCommons

Thanks!

Any questions?

For further inquiries email: ktaouk@ccia.org.au

Children's Cancer Institute

Australian BioCommons