Progenetix - A cancer genomics reference resource around GA4GH standards

Qingyao Huang, Paula Carrio-Cordo, Rahel Paloots, Bo Gao and Michael Baudis

Department of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, Switzerland

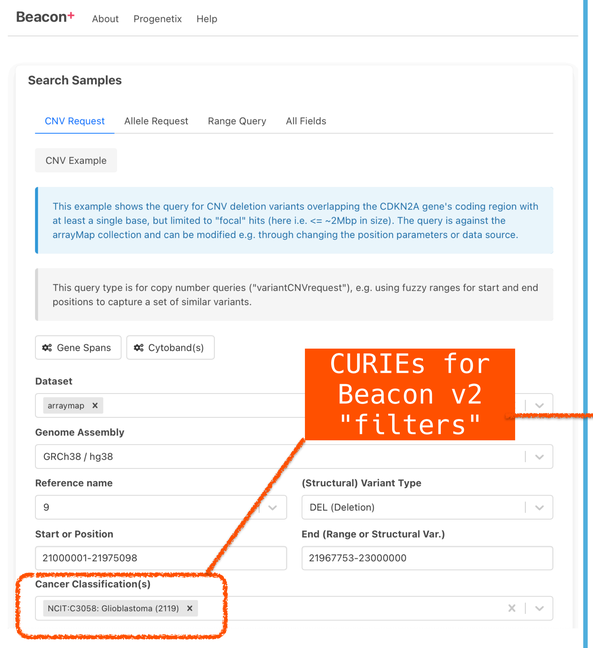
The *Progenetix* Oncogenomics Resource

The Progenetix oncogenomics resource provides sample-specific cancer genome profiling data and biomedical annotations as well as provenance data from thousands of individual cancer studies.

With currently 111'840 sample specific curated genomic copy number number (CNV) profiles from 1600 studies, representing over 780 cancer types (according to NCIt neoplasm core), Progenetix empowers aggregate and comparative analyses which vastly exceed individual studies or single diagnostic concepts.

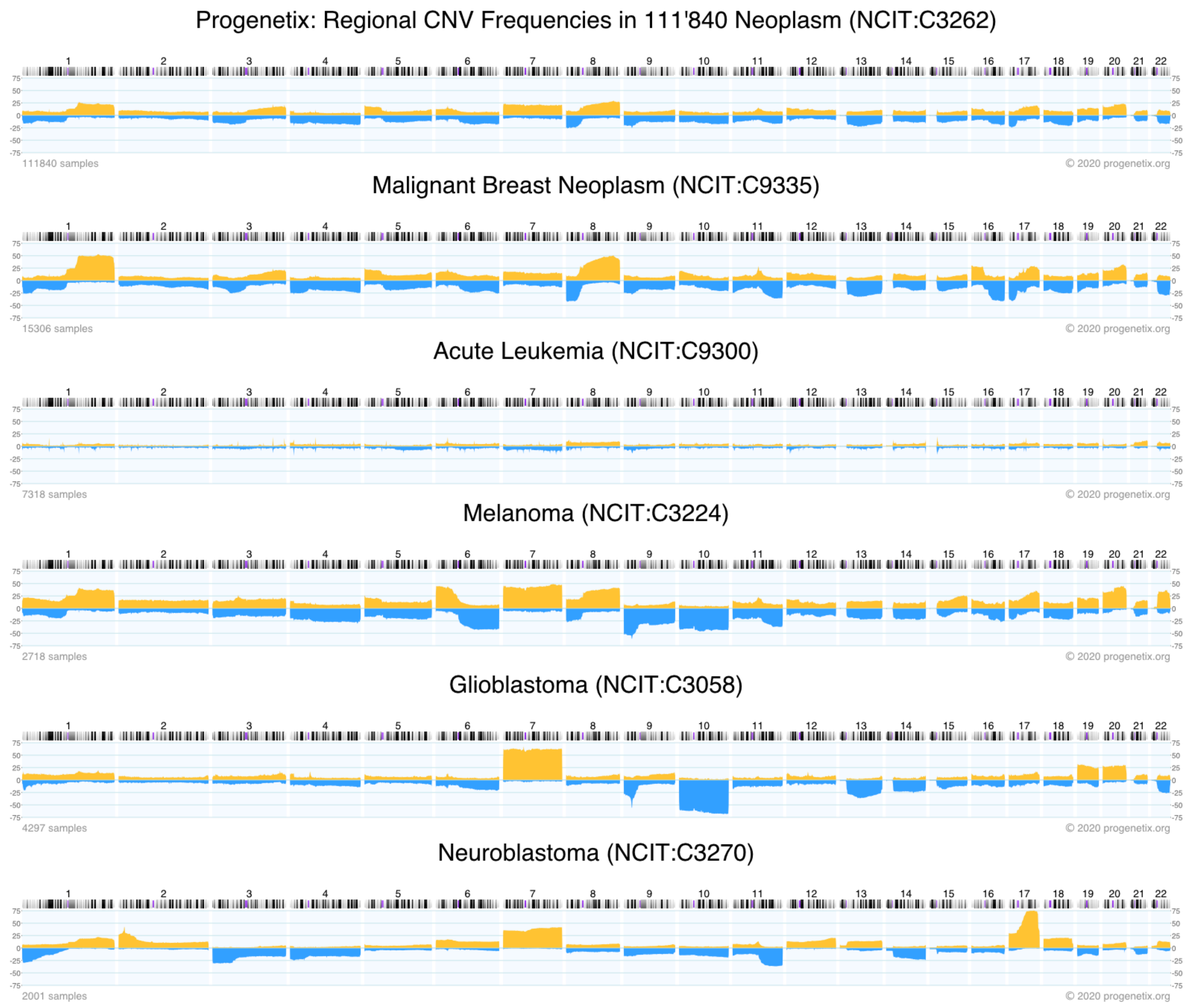
Progenetix provides a demonstration of how an open genomic reference resource can be built around GA4GH standards and how it can be used to support ongoing and future initiatives in GA4GH standard development as well as in driving ELIXIR implementation studies.

Beacon+ tests Beacon API extensions



*Beacon+ - built on top of the Progenetix infrastructure - has been instrumental in developing and testing Beacon extensions such as **structural variant** queries and **handover** data delivery (v1.n) or **filters** for querying biological and technical annotations (v2.n).*

Progenetix provides regional CNV frequency profiled for most cancer types

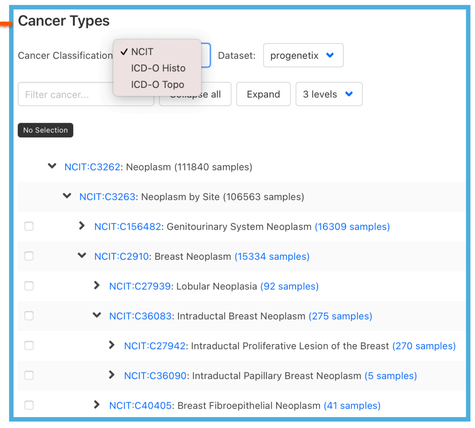


Genomic copy number frequency profiles in some tumor types, plotted from the Progenetix database API. The histograms detail the frequency of genomic duplications/amplifications (yellow; up) and deletions (blue, down) for the corresponding region in a given tumor type or all samples (top).

Modern Hierarchical Ontologies for Flexible Data Use

The use of hierarchical ontologies for biological classifications and parameters as well as for identifiers and technical metadata is imperative to make data accessible, reusable and amenable to computational mining and analysis methods.

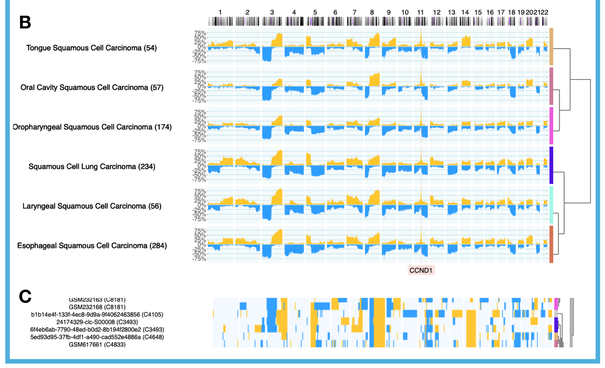
In Progenetix the systematic integration of "classical" property codes (e.g. International Classification of Diseases in Oncology; ICD-O 3) and their translation into hierarchical ontologies with registered identifiers (e.g. NCIt Neoplasm Core, MONDO, EFO...) empowers internal data structures as well as federated query implementations such as through Beacon v2 "filters".



Beacon *Handover* for Data Analytics

The Beacon protocol omits direct data delivery to avoid potential exposure of patient data. Within this paradigm, the Beacon v1.n *handover* extension adds options for retrieval or analysis of the matched data, outside the Beacon response itself. Beacon v2 will add other delivery methods to be used in conjunction with authentication and authorization technologies.

Standard Beacon queries together with dedicated handover objects in Progenetix enable a variety of services: variant display in the UCSC browser, sample data download (above, A); retrieval sample of specific genome profiles for subgroup visualization (below, B) or single sample similarity comparisons (C; all plots showing CNV data).



Conclusion

Development of protocols, schemas and standards in biomedical research and genomics applications happens in the continuous interplay between technological advances and epistemic advances.

The agile development principles and extensive data content of the Progenetix resource utilizes and promotes standards of the GA4GH ecosystem.

University of Zurich[UZH] · SPHN Swiss Personalized Health Network · SIB · elixir

For more information please visit progenetix.org and the code repositories at github.com/progenetix