progenet x oncogenomic resource

Qingyao Huang¹², Paula Carrio-Cordo¹², Bo Gao¹², Rahel Paloots¹², Ziying Yang¹², Michael Baudis¹²

¹ Swiss Institute of Bioinformatics, 8057 Zurich ² University of Zurich, Winterthurerstrasse 190, 8057 Zurich



Introduction

Progenetix resource

The Progenetix oncogenomics resource provides sample-specific cancer genome profiling data and biomedical annotations as well as provenance data for cancer studies. Especially through more than 100k genomic copy number number (CNV) profiles from over 500 cancer types, Progenetix is the most comprehensive reference resources for copy number aberration in cancer, empowers comparative analyses vastly exceeding individual studies and diagnostic concepts as well as supports development of data standards and exchange protocols through Global Alliance for Genomics and Health (GA4GH).

Database URL: progenetix.org

02

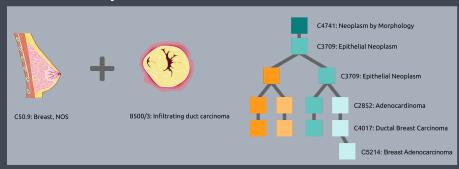
New meta-data

Domain-specific mapping

ICD-O system

v.s.

NCI thesaurus



ICD topography

Clinical and diagnostic aspects



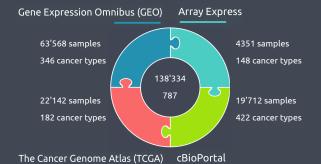
Uberon

Cross-species anatomy Developmental lineages Cross-database reference Spatial relations

Geographic provenance



New data sources Sample expansion



04

Data standards

CURIE, GA4GH, Phenopackets schema

FAIR

Findability

Accessibility

Interoperability

 ${\bf R}$ eusability

- Compact URI (CURIE) syntax, e.g. pgx:pgxbs-kftvgk8h
- GA4GH specification for variation data, i.e. individual -> biosample -> callset -> variant
- Phenopackets for phynotype data, i.e. id, phenotypes, EHR, provenance, biosample, biosample...



Beacon protocol

Features and prospects



- A Beacon answers a query for a specific variant against an individual or aggregated collection
- · Support single or range query
- Filter by disease, phenotype, ...
- Authentication (to come...)



New web interface

Many features...

Cancer Types		
The cancer samples in Progenetix are mapped to several classification systems. For each of the classes, agg of the corresponding samples can be initiated by clicking the sample number or selecting one or more classe		
Sample selection follows a hierarchical system in which samples matching the child terms of a selected class		
Cancer Classification: NCIT Cancer Core		
melanomal Hierarchy Depth: 2 levels ▼ No Selection		
▼ NCIT:C3262: Neoplasm (116232 samples)		
▼ NCIT:C3263: Neoplasm by Site (109317 samples)		
□ NCIT:C156482: Genitourinary System Neoplasm (16410 samples)		
□ VCIT:C156484: Malignant Genitourinary System Neoplasm (15938 samples)		
□ VCIT:C36076: Malignant Reproductive System Neoplasm (10383 samples)		



arrayMap Samples	
Gene Symbol ①	
Select	
Reference name ①	(Structural) Variant Type ①
9 v	DEL (Deletion)
Start or Position ()	End (Range or Structural Var.) ()
21500001-21975098	21967753-22500000
Minimum Variant Longth ①	Maximal Variant Length
	8
Cancer Classification(s) (I)	
NCIT-C3058: Glioblastoma (4375) X	x ~
Genotypic Sex (I)	Biosample Type ()
Select	Select V
Filters ①	Filter Logic Multiple term selected!
	AND V
City ①	
Select V	
21500001 24150000 21561753 52500000 Overy Database	



Just scan and visit ...