OncoMX tutorial

1.Background

Biomedical ontology is getting important. It integrates the information about various model organisms, obtains under different conditions and stores in heterogeneous database. A controlled vocabulary of biomedical terms (which organized in order to allow expression of complex relationships in machine readable form) is need for biomedical ontology. Biomarkers have demonstrated effectiveness in quantifying biological molecule that can be found, which not only reveal normal or abnormal process but also a condition of cancer.

2.OncoMX

According to the National cancer institute 'OncoMX is a cancer mutation and expression knowledgebase developed to enable exploration of cancer biomarkers alongside relevant experimental evidence and functional annotations.' OncoMX IS licensed under a Creative Commons Attribution 4.0 International License

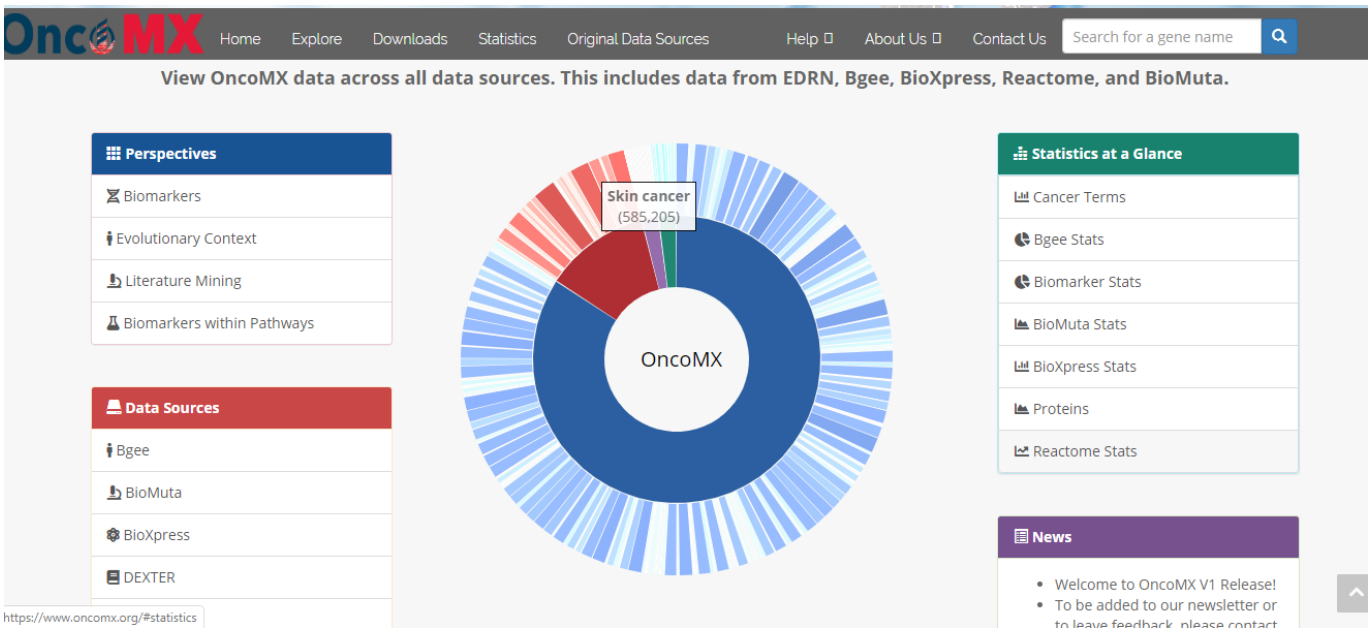
Ontologies used in OncoMX and its description:

Ontologies		Description
GO	Gene ontology	It provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data from GO Consortium members, as well as tools to access and process these data
NCI-t	NCI Thesaurus	It provides reference terminology for many NCI and other systems. It covers vocabulary for clinical care, translational and basic research, and public information and administrative activities
UMLS	Unified Medical Language System	It integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information systems and services, including electronic health records.
MeSH	Medical Subject Headings	It organizes content of Medline by imposing labels
HPO	Human Phenotype Ontology	It aims to provide a standardized vocabulary of phenotypic abnormalities encountered in human disease
Uberon	Uberon	It Integrates cross-species ontology covering anatomical structures in animals
DO	Disease Ontology	It integrates disease and medical vocabularies through extensive cross mapping of DO terms to MeSH, ICD, NCI's thesaurus, SNOMED and OMIM."

Contributing resources and its description

Contributing resources	Description
BioMuta	Mutation in cancer
BioXpress	Differential expression in cancer
Bgee	Gene expression data in animals
EDRN	accelerate the translation of biomarker information into clinical applications
DiMeX	A Text Mining System for Mutation-Disease Association Extraction

Main page



Landing page—access and explore gene specific cancer data

Menu bar—Home, Explore, Downloads, Statistics, Original Data Sources

Dashboard-- OncoMX data across all data sources. This includes data from EDRN, Bgee, BioXpress, Reactome, and BioMuta.

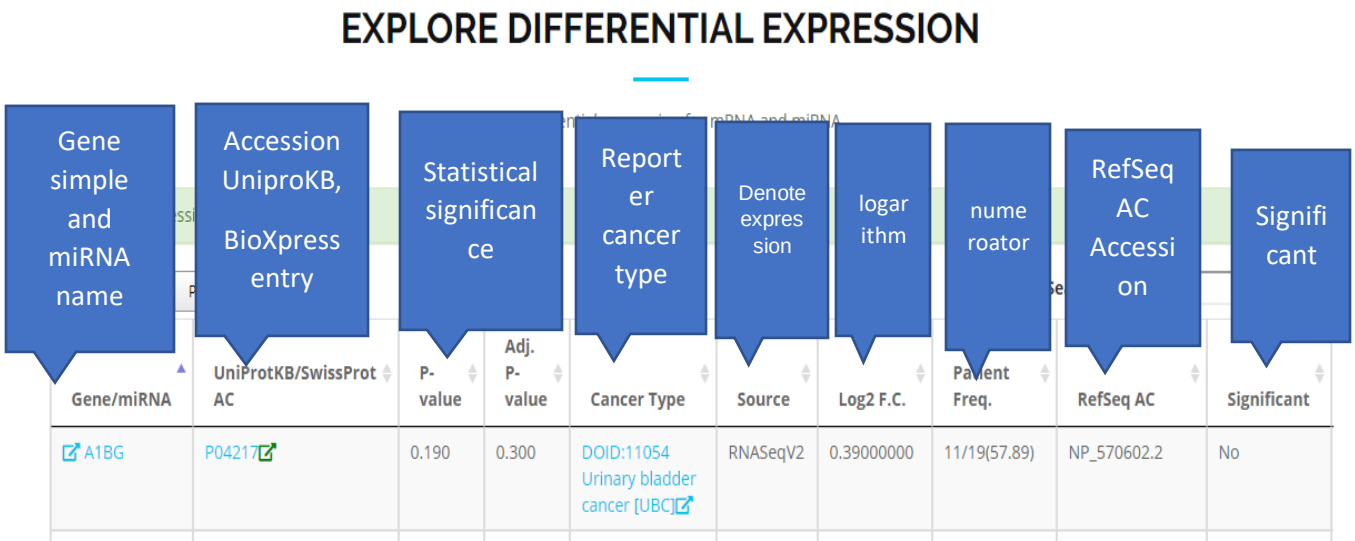
Explore tables—Perspectives, Data Sources, Statistics at a Glance, News

External downloads—Download Excel files

About team—Team member

Contact

Explore Differential expression table:



Differential expressions were analyzed at the study level and calculated as differences between groups of tumors with a specific cancer type and matched adjacent normal samples. Whether the available results can be screened based on cancer type or the p value corresponding to the differential expression analysis

3.How to search

Biological question:

What are the top biomarkers can be used to detect liver cancer and its detail information?

Parameters: biomarkers, liver cancer

Liver cancer: Liver cancer happens when liver cells develop changes (mutations) in their DNA. A cell's DNA is the material that provides instructions for every chemical process in your body. DNA mutations cause changes in these instructions. One result is that cells may begin to grow out of control and eventually form a tumor — a mass of cancerous cells.

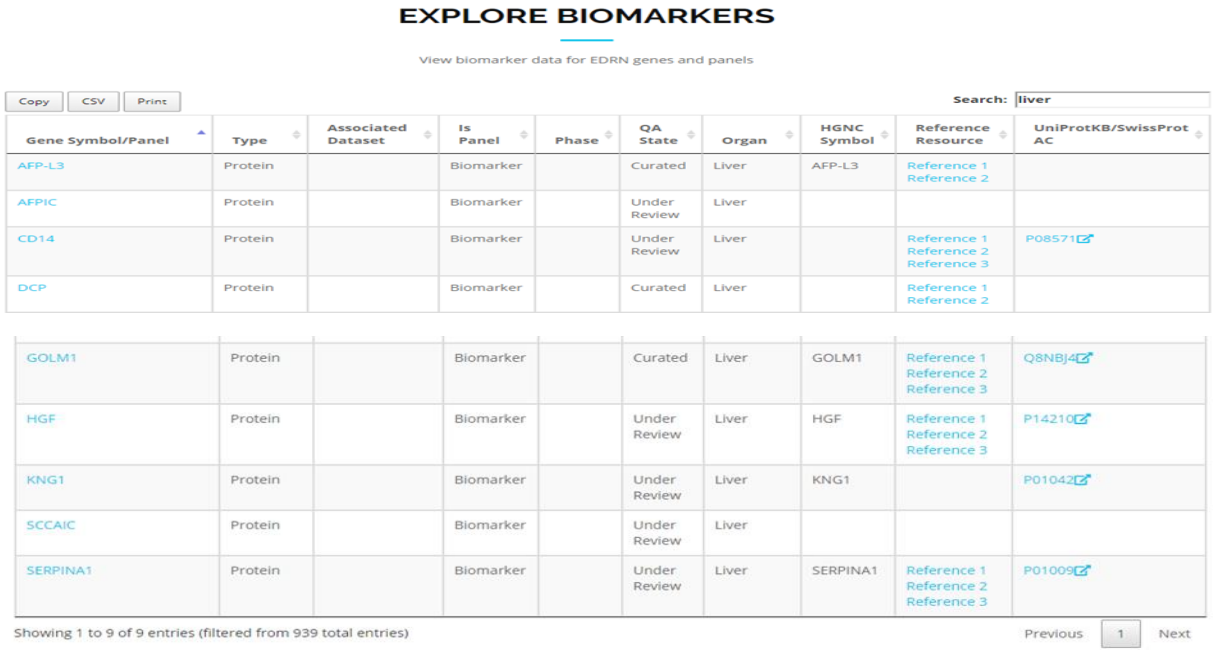
Biomarkers: ①Biological molecule that can be found

②Indicates normal or abnormal (tumor) process

③Indicates a condition of cancer.

General results:

From OncoMx-→use biomarker section→find 9 biomarker can be used to detect liver cancer which include(AFP-L3, AFPIC, CD14, DC, GOLM1,HGF, KNG1, SCCAIC, SERPINA1)



EXPLORE BIOMARKERS

View biomarker data for EDRN genes and panels

Copy CSV Print Search: liver

Gene Symbol/Panel	Type	Associated Dataset	Is Panel	Phase	QA State	Organ	HGNC Symbol	Reference Resource	UniProtKB/SwissProt AC
AFP-L3	Protein		Biomarker		Curated	Liver	AFP-L3	Reference 1 Reference 2	
AFPIC	Protein		Biomarker		Under Review	Liver			
CD14	Protein		Biomarker		Under Review	Liver		Reference 1 Reference 2 Reference 3	P08571
DCP	Protein		Biomarker		Curated	Liver		Reference 1 Reference 2	
GOLM1	Protein		Biomarker		Curated	Liver	GOLM1	Reference 1 Reference 2 Reference 3	Q8NBJ4
HGF	Protein		Biomarker		Under Review	Liver	HGF	Reference 1 Reference 2 Reference 3	P14210
KNG1	Protein		Biomarker		Under Review	Liver	KNG1		P01042
SCCAIC	Protein		Biomarker		Under Review	Liver			
SERPINA1	Protein		Biomarker		Under Review	Liver	SERPINA1	Reference 1 Reference 2 Reference 3	P01009

Showing 1 to 9 of 9 entries (filtered from 939 total entries) Previous 1 Next

Detailed for biological question:

Step-by-step instructions:

1. Visit the OncoMX using your browser and login to query the public datasets.

2. Once logged in, access the Biomarkers from the landing page. Click on the “Biomarkers” tab. (Figure 1)

Figure 1 Access the BIOMARKERS Browser

1. In order to query for the biomarkers in liver cancer. For this
   a. Click the BIOMARKERS Browser-→navigate to EXPLORE BIOMARKERS page(View biomarker data for EDRN genes and panels).
   b. Type in ‘liver’ in the search bar-> Click Search as shown the right (Figure 2).

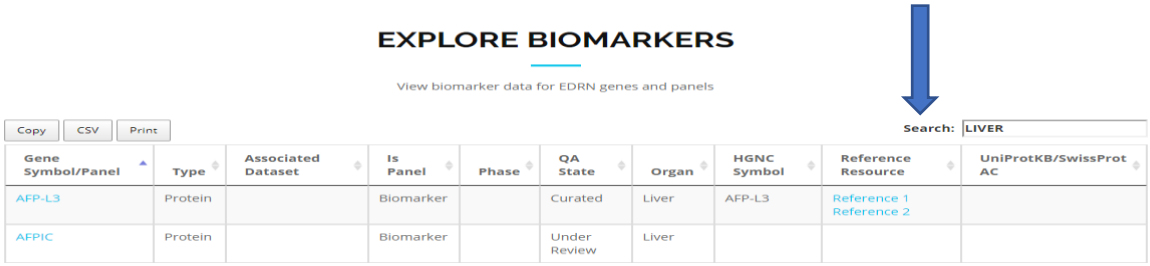


Figure 2 Access the EXPLORE BIOMARKERS page

   c. To find detail information about the biomarker in liver cancer. Scroll down the EXPLORE BIOMARKERS page, under the bottom of this page, there is a table (Figure 2) listed the biomarkers , click on one of the top biomarker ‘AFP-L3’-→Navigate to a details page(Figure 3), which includes the EDRN Title, Allases, Description, EDRN Publications et al. (answer to the proposed question).

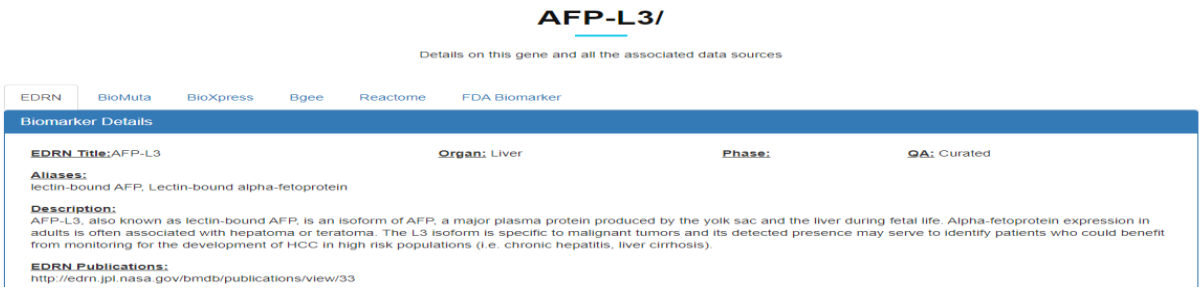


Figure 3 Access the biomarkers details information

4.How to use data portal

1. data portal Homepage https://data.oncomx.org/. Its navigation bar contains

Home, Explore, Downloads, Statistics (View OncoMX data across all data sources. This includes data from EDRN, Bgee, BioXpress, Reactome, and BioMuta.), Original Data Sources (displays a collective list of all the data sources that will be merged in this portal). (Figure 4)

Figure 4 Access the data site page

2. Under navigation bar, there is dataset collection. In the search bar, you can search by BCO ID, name or contributor.
3. Under dataset collection, there is a list of sub-menus (which can be check or uncheck), you can filter by categories, species, status, file type.
4. Scroll down the data site page, under the bottom of this page, there are tables (Figure 4) listed the total of 18 datasets, click on one of the datasets “FDA breast cancer biomarkers” Navigate to a details page (Figure 5). Each row represents one gene linked to its respective test. Genes are labeled by relevant identifiers/accessions from UniProtKB, HGNC, and EDRN. Tests are distinguished by manufacturer, FDA submission ID(s), clinical trial ID(s), and PubMed ID(s)
5. On the right top of the table, there are menus (which include BCO JSON/README/DOWNLOAD).
   You can click BCO JSON to get joson scirp.
   You can click README to get every step information.
   You can click DOWNLOAD to get Excel file about data information.
6. All OncoMX data are available for download at the parallel data site, www.data.oncomx.org. This repository contains all the data in OncoMX as well as custom datasets generated for and/or by collaborators. All data are uniformly packaged and described following the BioCompute Object (BCO) model to enable users access to relevant provenance details.

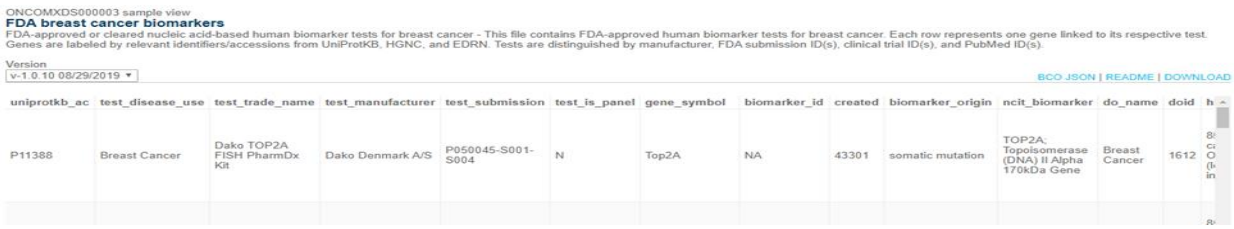
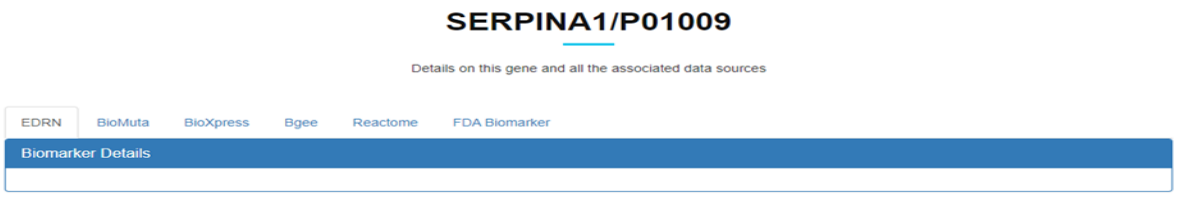


Figure 5 FDA breast cancer biomarkers table

5.Critiques

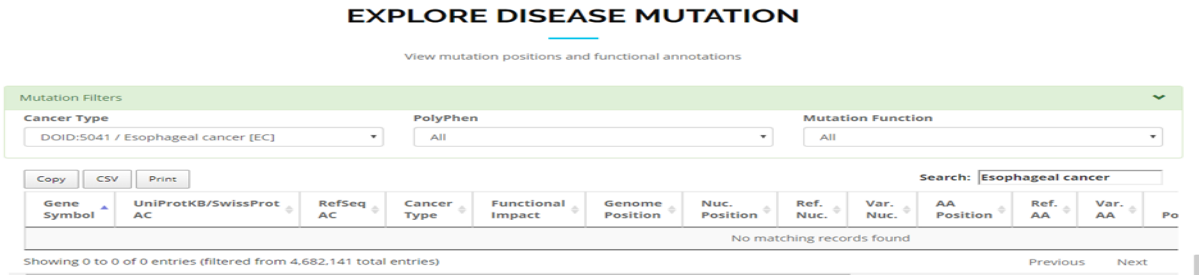
① Lack of details information about biomarkers

For example: SERPINA1/P01009 is one of the biomarkers in liver cancer, however when click it, there is not details information about it.



② Lack of information about explore disease mutation

In order to view mutation positions and functional annotations, type cancer type DOID:5041/ Esophageal cancer, there are no information about Esophageal cancer.



In order to view mutation positions and functional annotations, may use a more comprehensive data integration platform to handle a variety of different types of data.

Reference

Faria, D., Pesquita, C., Mott, I. *et al.* Tackling the challenges of matching biomedical ontologies. *J Biomed Semant* **9,** 4 (2018) doi:10.1186/s13326-017-0170-9