Capturing cancer initiating events in OncoCL, a cancer cell ontology

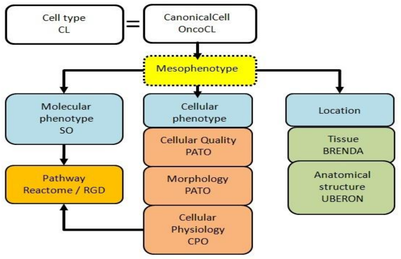
Mary E. Dolan
The Jackson Laboratory, Bar Harbor, Maine 04609 USA

The Jackson Laboratory
Leading the search for tomorrow's cures

HOW OncoCL WORKS

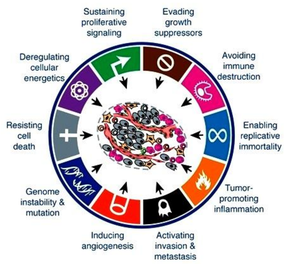
CANCER CELL MESOPHENOTYPE

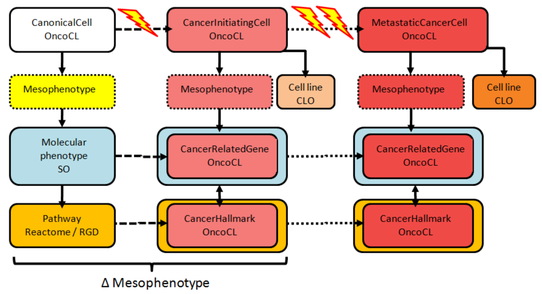
OncoCL provides a framework for consistent annotation of cancer-associated data allowing queries for how a particular cancer cell type is situated in a pathway of cancer progression. OncoCL constructs a 'mesophenotype' to describe the cancer cell reusing other ontologies as much as possible. The CanonicalCell is a normal cell type as represented in CL, the cell type ontology. The CanonicalCell has Mesophenotype comprised of canonical characteristics of that cell type: normal genotype, normal cellular phenotype (such as potency, proliferative capacity, morphology, function) and normal tissue and anatomical location.



REPRESENTING CANCER PROGRESSION

In OncoCL, cancer progression is represented as shown below: a CanonicalCell undergoes oncogenic change in a CancerRelatedGene thus acquiring a CancerHallmark to become a CancerCell with a change in Mesophenotype. The CancerCell can go on to acquire additional CancerHallmarks due to additional molecular changes in other CancerRelatedGenes to become a different type of CancerCell.

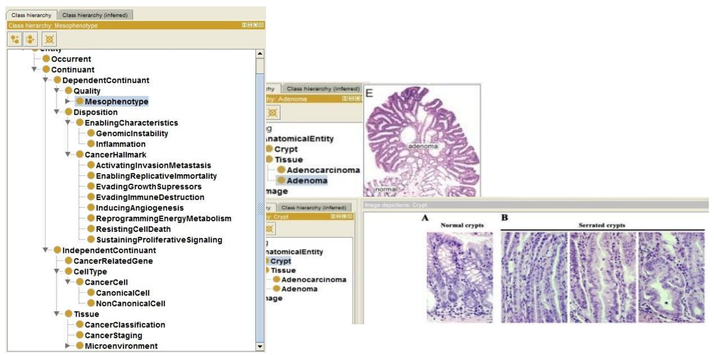




IMPLEMENTATION

OncoCL is implemented in Protégé-OWL. We construct five classes: CanonicalCell, CancerCell, Mesophenotype, CancerRelatedGene and CancerHallmark. These classes are embedded in BFO, the Basic Formal Ontology, which facilitates logical consistency checking. As described above, the classes are built on relevant parts of other existing biomedical ontologies: UBERON for anatomy; BRENDA, tissues; PATO, morphology and other qualities; Reactome, pathways; CPO, cellular phenotypes related to GO, the gene ontology; SO, sequence features.

Protégé allows browsing of the class hierarchy and definitions of all terms. We include a library of images where appropriate to facilitate annotation (Balhoff). Here we show images for the classes 'Adenoma' and 'Crypt.'



ABSTRACT

We have developed an ontology, OncoCL, to classify cancer cells and provide a framework for consistent annotation of cancer-associated data from conventional surgical pathology and cancer molecular biology for the purpose of access, comparison, and analysis. The cell type ontology, CL, describes normal cell types and was not designed to capture the pathology of cancer cells. OncoCL builds upon CL, as a canonical cell (represented in CL) undergoes oncogenic change and tumorigenesis with the acquisition of the cancer hallmarks described by Hanahan and Weinberg.

The characterization of cancer initiating cells and cancer initiation events present particular challenges – for example, the representation of the self-renewal and differentiation potential of cancer stem cells compared with those of canonical (normal) stem cells. But we know that the distinction of high-risk precursor lesions with a high likelihood of developing into cancer, compared with indolent disease, depends on the synthesis of complex, heterogeneous data related to cancer initiating cells. OncoCL is a flexible resource specifically developed to integrate these diverse data through the reuse of a number of other biomedical ontologies. This work will present the problems we encountered capturing cancer initiating events and the solutions we implemented to address them.

This work is supported by NIH with funding through NCI CA155825.

CHARACTERIZING CANCER INITIATING CELLS IS CHALLENGING

CSC-based therapies must distinguish CSCs not only from non-CSC tumor cells but also from normal stem cells. Various CSC-specific methods have been studied and seem promising, among them:

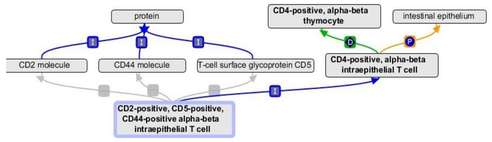
- The identification of distinctive CSC surface markers
- Overexpression, mutations, variant isotypes of CSC surface markers
- Evidence of glycosylation of CSC surface marker proteins
- microRNA regulation of 'stemness' genes in CSCs.
- The existence of abnormal but pre-cancerous stem cells.

We have extended the OncoCL mesophenotype to capture these data.

CAPTURING CELL SURFACE MARKER DATA

CL describes immune cells in terms of cell surface markers (Diehl et al). The graph shows, for example, that a 'CD2-positive, CD5-positive, CD44-positive alpha-beta intraepithelial T cell' is a type of 'CD4-positive, alpha-beta intraepithelial T cell' which develops from 'CD4-positive, alpha-beta thymocyte' and is part of 'intestinal epithelium.' The cell type under consideration is positive for cell surface markers CD2, CD44, and CD5, which is captured using the 'has_plasma_membrane_part' relation to those proteins.

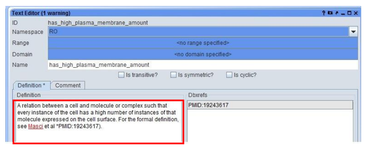
We use this procedure in OncoCL.




```
name: CD2-positive, CD5-positive, CD44-positive alpha-beta intraepithelial T cell
definition: "Intraepithelial T cells with a memory phenotype of CD2-positive, CD5-
positive, and CD44-positive." [PMID:16048545]
is_a: CD4-positive, alpha-beta intraepithelial T cell
has_plasma_membrane_part CD2 molecule
has_plasma_membrane_part CD44 molecule
has_plasma_membrane_part T-cell surface glycoprotein CD5
created_by: tmeehan
creation_date: 2009-10-20
```

RELATION ONTOLOGY IS KEY TO LINKING CANCER STEM CELLS TO CELL SURFACE MARKERS

According to good ontology design principles we use only relations in RO, the relation ontology. To describe cancer stem cells we relate them to cell surface markers by means of RO relations: 'has_plasma_membrane_part', 'has_high_plasma_membrane_amount' and 'has_low_plasma_membrane_amount'.

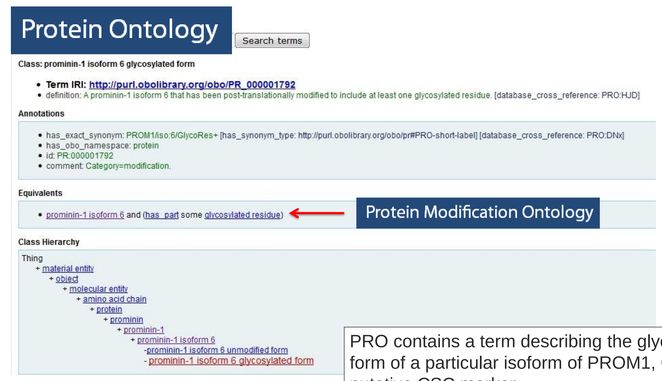


BIOMEDICAL ONTOLOGIES PROVIDE THE PIECES

Biomedical ontologies provide a shared understanding of a domain that is human intelligible and computer readable and, consequently, a representational system to support the integration and retrieval of the biological information. In addition, mappings among ontologies enables data integration.

PROTEIN ONTOLOGY, PRO

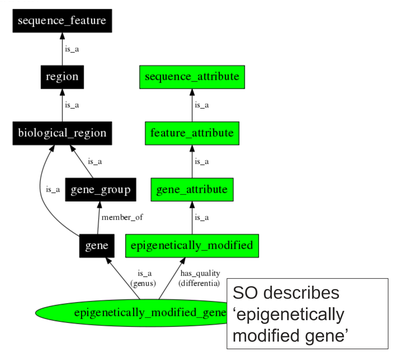
We use PRO (Natale et al) to describe CSC surface markers. PRO provides a formal classification of protein classes including isoforms, variants and modified forms with protein modifications, such as glycosylation, described by PSI-MOD, the protein modification ontology.



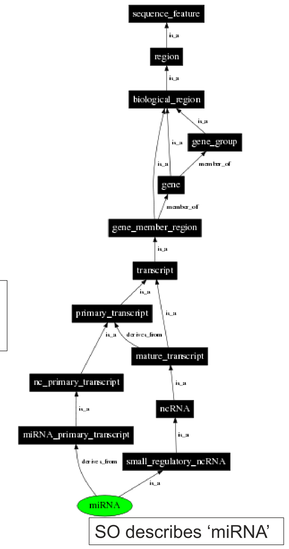
PRO contains a term describing the glycosolated form of a particular isoform of PROM1, CD133, a putative CSC marker.

SEQUENCE ONTOLOGY, SO

We use SO (Mungall et al) to describe CSC-specific microRNAs and CSC-specific epigenetic modifications.



SO describes 'epigenetically modified gene'



SO describes 'miRNA'

BioPortal
MicroRNA Ontology
Ontology for MicroRNA Target Prediction
Summary Classes Notes Mappings Widgets

SO is used by both the MicroRNA Ontology and the Ontology for MicroRNA Target Prediction, two resources that are under development and that can link to additional curated data.

CONCLUSIONS

We have augmented OncoCL to better characterize cancer stem cells by allowing the capture of data related to CSC surface markers including distinguishing CSC surface protein modifications, CSC-specific microRNAs and CSC-specific epigenetic modifications.

In future work we plan to include descriptions of the tumor microenvironment, characteristics to distinguish EMT state from CSC phenotype, and extend the mesophenotype to include aspects that would distinguish CSCs that arise from normal stem cells from those that arise from other progenitor or differentiated cells.

REFERENCES

Balhoff J. (2011) https://github.com/balhoff/image-depictions-view
Diehl, AD et al. (2011) Hematopoietic cell types: Prototype for a revised cell ontology. *Journal of Biomedical Informatics*. 44(1),
Hanahan D, Weinberg RA. (2000) The hallmarks of cancer. *Cell*, 100, 57; Hanahan D, Weinberg RA. (2011) The hallmarks of cancer: The Next Generation. *Cell*, 144, 646.
Mungall, CJ et al. (2011) Evolution of the Sequence Ontology terms and relationships. *Journal of Biomedical Informatics* 44(1),87.
Natale DA et al. (2011) The Protein Ontology: a structured representation of protein forms and complexes. *Nucleic Acids Res*. 39:D539.

www.jax.org