OncoCL-KB, a Knowledgebase for Integration of Clinical and Molecular Cancer Data

Mary E. Dolan

The Jackson Laboratory, Bar Harbor, ME 04609, USA
Mary.Dolan@jax.org

Abstract. OncoCL-KB is a knowledgebase built on OncoCL, our ontology for describing cancer cell types. OncoCL provides a semantic framework for integrating cancer-associated molecular data, and the properties imparted by the corresponding disrupted cellular pathways, with conventional pathology data associated with cancer progression.

OncoCL makes use of a number of other mature biomedical and clinical ontologies. In particular, OncoCL builds upon the cell type ontology (CL) to define a canonical cell that then undergoes oncogenic change and tumorigenesis with the acquisition of the cancer hallmarks as described by Hanahan and Weinberg. OncoCL-KB embeds annotated data sets – including cancer-associated genes and genomic variants, cancer-associated pathways, cancer stem cell markers, and cancer mouse models – in the OncoCL semantic framework. OncoCL-KB provides a resource that can contribute to a better understanding of cancer predisposition, earlier diagnosis, and better therapies.

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

Keywords: Ontology, cancer, knowledgebase

1 Introduction

The conceptual basis of OncoCL is that cancer cell phenotypes develop through the acquisition of cancer hallmarks described by Hanahan and Weinberg [1]. Hanahan and Weinberg suggest that most, if not all, cancer cell phenotypes derive from a succession of alterations in cell behavior, those being: 1) self-sufficiency in growth signals, 2) insensitivity to antigrowth signals, 3) evasion of apoptosis, 4) limitless replicative potential, 5) sustained angiogenesis, 6) tissue invasion and metastasis, 7) reprogramming of energy metabolism, 8) evading immune destruction, 9) genomic instability and mutation and 10) tumor-promoting inflammation. Our model represents oncogenic change in a cell as the acquisition of these functional hallmarks.

To serve as a framework for integration of diverse data, OncoCL constructs a composite 'mesophenotype' – finer in scale than an observable phenotype but coarser in scale than the full genotype – by reusing selected cancer-relevant branches of a number of previously developed, independent ontologies. OncoCL builds upon the cell

type ontology, CL [2] as a representation of a canonical cell. To capture cellular changes from conventional pathology, we use (among others) PATO, the Phenotypic Quality Ontology [3] with its description of 'morphology', 'potency', and so on. To capture the molecular changes that give rise to cancer hallmarks, we construct a molecular 'phenotype' described with terms from VariO, the variant ontology [4]. In accord with established ontology development principles, all relations are drawn from the Relation Ontology, RO [5], and all terms are embedded within the Basic Formal Ontology, BFO [6].

2 OncoCL-KB: Implementation and Availability

We have created a prototype knowledge base that embeds annotated data sets from a variety of public sources as instances in OncoCL. Our sources include: NCI Thesaurus, Reactome, UniProt, Sanger Cancer Gene Census. Our data include: cancer-associated genes and genomic variants, cancer-associated pathways, associations of abnormal cell with cancer types. We have implemented OncoCL in Protégé [7] with its expressive formal semantics and functionality to check for logical consistency and valid inferential structure. We have incorporated a user-developed plug-in [8] that enables integration of images of, for example cells and tissues, with appropriate classes. Our prototype OncoCL-KB is free and open to all users through the project website: https://www.jax.org/research-and-faculty/tools/oncocl.

References

1. Hanahan D., Weinberg R.A.: The hallmarks of cancer. Cell, 100, 57-70 (2000); Hanahan D., Weinberg R.A.: The hallmarks of cancer: The Next Generation. Cell, 144, 646-674 (2011)
2. Bard J., Rhee S.Y., Ashburner M.: An ontology for cell types. Genome Biology, 6, R21(2005)
3. Gkoutos G.V., Green E.C., Mallon A.M., Hancock J.M., Davidson D.: Using ontologies to describe mouse phenotypes. Genome Biol. 6(1):R8 (2005)
4. Vihinen, M.: Variation Ontology for annotation of variation effects and mechanisms. Genome Research, 24(2), 356–364 (2014)
5. Smith B., Ceusters W., Klagges B., Kohler J., Kumar A., Lomax J., Mungall C.J., Neuhaus F., Rector A., Rosse C.: Relations in Biomedical Ontologies. Genome Biology, 6:R46 (2005)
6. Smith B., Kumar A., Bittner T.: Basic Formal Ontology for Bioinformatics. Journal of Information Systems. 1-16 (2005)
7. Rubin D.L., Knublauch H., Fergerson R.W., Dameron O., Musen M.A.: Protégé-OWL: Creating Ontology-Driven Reasoning Applications with the Web Ontology Language. AMIA Annu Symp Proc. 2005: 1179 (2005)
8. Balhoff J.: Protégé plugin which displays image depictions of classes and individuals, https://github.com/balhoff/image-depictions-view