Published in final edited form as:

J Biomed Inform. 2011 February ; 44(1): 75–79. doi:10.1016/j.jbi.2010.01.006.

Hematopoietic Cell Types: Prototype for a Revised Cell Ontology

Alexander D. Diehl^{1,†}, Alison Deckhut Augustine², Judith A. Blake¹, Lindsay G. Cowell³, Elizabeth S. Gold⁴, Timothy A. Gondré-Lewis², Anna Maria Masci³, Terrence F. Meehan¹, Penelope A. Morel⁵, Anastasia Nijnik⁶, Bjoern Peters⁷, Bali Pulendran⁸, Richard H. Scheuermann⁹, Q. Alison Yao², Martin S. Zand¹⁰, and Christopher J. Mungall¹¹

¹The Jackson Laboratory, Bar Harbor, ME

²National Institute of Allergy and Infectious Disease, Bethesda, MD

³Duke University Medical Center, Durham, NC

⁴Institute for Systems Biology, Seattle, WA

⁵University of Pittsburgh, Pittsburgh, PA

⁶University of British Columbia, Vancouver, BC

⁷La Jolla Institute for Allergy and Immunology, La Jolla, CA

⁸Emory University, Atlanta, GA

⁹University of Texas Southwestern Medical Center, Dallas, TX

¹⁰University of Rochester Medical Center, Rochester, NY

¹¹Lawrence Berkeley Laboratory, Berkeley, CA

Abstract

The Cell Ontology (CL) aims for the representation of *in vivo* and *in vitro* cell types from all of biology. The CL is a candidate reference ontology of the OBO Foundry and requires extensive revision to bring it up to current standards for biomedical ontologies, both in its structure and its coverage of various subfields of biology. We have now addressed the specific content of one area of the CL, the section of the ontology dealing with hematopoietic cells. This section has been extensively revised to improve its content and eliminate multiple inheritance in the asserted hierarchy, and the groundwork was laid for structuring the hematopoietic cell type terms as cross-products incorporating logical definitions built from relationships to external ontologies, such as the Protein Ontology and the Gene Ontology. The methods and improvements to the CL in this area represent a paradigm for improvement of the entire ontology over time.

Keywords

ontology; hematopoietic cells; immunology

© 2010 Elsevier Inc. All rights reserved.

[†]corresponding author: Alexander Diehl, Mouse Genome Informatics, The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, adiehl@informatics.jax.org, fax +1-207-288-6131.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

The Cell Ontology (CL) is a biomedical ontology originally built to represent *in vivo* and *in vitro* cell types, including those observed in specific developmental stages, of all the major model organisms.[1] The CL now aims to become a reference ontology within the Open Biomedical Ontology (OBO) Foundry (www.obofoundry.org).[2] The CL both serves the terminology needs of data annotation and provides a base ontology from which compound terms in other ontologies can be derived by means of cross-product term formation.[3] Within the Mouse Genome Informatics resource (www.informatics.jax.org), for instance, the CL is used in conjunction with Gene Ontology (GO) during annotation of mouse gene products to indicate the cell type in which a gene product is active, an approach that is now being adopted by other model organism databases within the GO Consortium. In ontology development for the GO CL terms are employed in the formation of new GO terms using the cross-product of core GO biological process terms and CL terms: for instance, the GO term “leukocyte differentiation” can be defined using the GO term “cell differentiation” and the CL term “leukocyte.”[4 (this issue)] The Immunology Database and Analysis Portal (www.immport.org) is using the CL as a reference of cell types for the mapping of results from the analysis of flow cytometry data. The CL is also frequently used to compose descriptions of phenotypes.[5]

The Cell Ontology is currently constructed using two relations, *is_a* and *develops_from*. The first relation is used to relate specific cell types to more general cell types (for example, between “T cell” and “lymphocyte”); the latter relation is used to indicate cell lineage relationships (for example, between “neuron” and “neuroblast”). The ontology, as it was initially developed, relies upon a number of artificial high level terms to capture types of cellular qualities, such as “cell in vivo,” “cell by organism,” and “cell by class,” a term which itself has the *is_a* child terms “cell by function,” “cell by histology,” “cell by lineage,” “cell by ploidy,” etc. These subclasses of cells have further *is_a* children denoting more specific qualities of cells. Depending on the qualities of a particular cell type it may have one or more of these terms as *is_a* ancestors. For instance, in the original form of the ontology the cell type “macrophage” is a direct subtype of “mononuclear phagocyte” and “professional antigen presenting cell,” and an indirect subtype of “cell by function,” “cell by histology,” “cell by nuclear number,” and “animal cell” (Figure 1A).

With its multiple inheritance structure, the original CL can be described as having separate ontologies of cell types delineated by particular cellular qualities overlaid upon each other, i.e. an ontology with multiple axes of differentiation that are variously and sometimes arbitrarily applied to individual cell types. Furthermore the high level terms themselves do not represent actual cell types, so the ontology is not a true *is_a* hierarchy. This unwieldy ontological construct is not ideal for developing proper inference about cell types, nor does it always provide obvious placement of new cell type terms.

Discussions among interested parties in the past few years have focused on how best to restructure the CL to eliminate the complexity of its multiple inheritance structure. One way to achieve this restructuring is to use an increasingly common methodology for developing ontologies: Assign logical definitions to classes based on their properties, and then let automated tools – called reasoners – infer the multiple inheritance hierarchy. This strategy exploits the work done in other ontologies. For example, neurons can be classified according to the types of chemical entities that are released, and the ontology of chemical entities can be used to determine the subsumption relationships between types of neuron.

The hematopoietic/immune cell types are of particular interest because of their roles in the immune response and consequent involvement in human health and disease. These cell types

in particular have been the focus of two rounds of intensive curation in recent years. A set of improvements for hematopoietic cells was done in 2006 in conjunction with the revision of the terms for immunological processes in the GO.[6,7] At that time 80 new hematopoietic cell type terms were introduced, many other terms were revised, and many improvements in ontology structure were made for these specific cell types.

A second, more extensive round of revisions to the hematopoietic cell type terms in CL is described herein. These revisions grew out of the proceedings of a National Institute of Allergy and Infectious Disease (NIAID) sponsored “Workshop on Immune Cell Representation in the Cell Ontology,” held in May 2008, where domain experts and biomedical ontologists worked together on two goals: 1) revising the existing terms and developing additional terms for T cells, B cells, natural killer cells, monocytes and macrophages, and dendritic cells, and 2) establishing a new paradigm for a comprehensive revision of the whole of the CL. These changes in the representation of hematopoietic cells were needed to represent these cell types in a more complete and accurate manner. The goals were to represent all major hematopoietic cell types identified in the literature in the ontology and to define these cell type in an in-depth manner that greatly increases the descriptiveness of the ontology for data annotation and logical inference.

Methods

The NIAID workshop attendees discussed how best to characterize hematopoietic/immune cell subsets, as well as how to improve the overall ontological structure of the corresponding portions of the CL and of the CL ontology in general. The consensus view was that the current multiple inheritance structure of the CL is unsustainable and that existing and new terms for hematopoietic cells should be logically defined via their structural parts and qualities as represented in other ontologies. Much discussion centered on what might be the optimal axis of differentia for these hematopoietic terms. It was recognized that, in many cases, these cell types are defined largely, but not solely, by the expression of particular marker proteins either at the cell surface (e.g. receptor proteins) or internally (e.g. transcription factors). The presence of these proteins as part of a cell is considered a structural feature of the cell, and we have chosen to use the relation *has_part* (or an appropriate sub-relation) from the OBO Relation Ontology (RO) to relate particular cell types to protein terms from the Protein Ontology (PRO) or protein complex terms from the cellular component ontology of the Gene Ontology (GO).[8,9]

However, for certain cell types such as macrophages, the full molecular characterization of different types is still not complete in the literature, and anatomical location effectively serves as a major differentia for these cells, which can be expressed via the relation *part_of*. For other cell types, functional or lineage criteria serve as differentia for the complete definition of the cells. Functional criteria include the ability to execute or participate in particular processes that relate to individual cell types, such as those referred to by the GO biological process terms “leukocyte mediated cytotoxicity” or “cytokine production,” or processes that involve coordination of multiple cell types, such as that referred to by the GO term “T-helper 1 type immune response.”

For this type of criteria we will use the *capable_of* type-level relation, defined in terms of the *bearer_of* and the *realized_by* type-level relations (these two relations will be incorporated in a future version of the RO). Formally, C *capable_of* P if there exists some D such that C *bearer_of* D and D *realized_by* P. Thus, we have focused on structural criteria where possible as the primary differentia, but have utilized other types of differentia when necessary. This flexibility is required to adhere to the commonly accepted biological definitions of individual cell types.

The Cell Ontology has been developed heretofore as an OBO-formatted ontology, and in developing the hematopoietic terms, we have relied on OBO-Edit 2.0 for editing the ontology, in conjunction with a text-editor for simpler modifications of the ontology file. [10] OBO-Edit provides textual and graphical interfaces that facilitate many aspects of ontology development, including the formation of cross-products, and the program worked well for our purposes.

Results

Reflecting the above considerations, we have taken a two-stage approach to further development of the hematopoietic cells in the Cell Ontology. In the first stage, which is now complete, we revised current terms and added new terms so that all hematopoietic cell type terms now have textual definitions that contain all the necessary details to define the cells logically. These terms have been directly incorporated into the existing ontology. Figure 2A shows a typical OBO term stanza for one of these new terms, “induced T-regulatory cell.”

We have also separated the hematopoietic terms from the complex hierarchy of the original CL as much as possible, so that the section of the ontology containing these terms represents a true *is_a* hierarchy. Figure 1B shows the simplified hierarchy for the cell type “macrophage.” In restructuring the ontology for the hematopoietic cells, we have eliminated the multiple inheritance via the artifactual high-level terms such as “cell by histology,” “cell by nuclear number,” or “cell by function.” Instead, information about cellular qualities is captured in the textual definitions where it is relevant, and is being used to build logical definitions for the cell types in the second stage of the work, as described below.

The version of the CL incorporating the changes in the hematopoietic terms accomplished in this first stage of revision has been given the working name “CL1.5” (Note this does not refer to the CVS revision number within the cell.obo file itself, but rather the data-version tag). Within CL1.5 there are many concrete improvements to CL content in the area of hematopoietic cells. We have created many new terms for individual cell types, including over 40 terms for T-lineage cells, over 40 terms for B-lineage cells, several natural killer cell terms, over 30 terms for monocytes and macrophages, and over 30 terms for dendritic cells. Other new terms have been introduced for various hematopoietic progenitor cell types. As discussed above, most of these new terms have been defined by structural criteria (protein expression) sometimes in conjunction with functional or anatomical relationships. An exception to this general rule is that most of the new macrophage terms are defined based on their anatomical location with protein expression criteria added where supported by the literature. All these new and revised terms are present in the publicly available version of the CL (www.obofoundry.org/cgi-bin/detail.cgi?id=cell). We have provided references within the ontology to published articles or textbooks that were used in developing the individual definitions for the majority of hematopoietic terms. For the rest, the term references are to the curators who developed the definitions based on their expert knowledge. All terms in the CL that were revised or developed during the NIAID workshop and its follow-up work have also been given the reference ID GO_REF:0000031, which refers to a brief description of the workshop in a list of references maintained at the GO Consortium web site (www.geneontology.org).

The ontology structure has been improved in important areas such as T cell and B cell development. Lineage relationships via the *develops_from* relation are now provided for many additional cell types. In general the hematopoietic terms are presented as species neutral, but species-specific information is incorporated in some definitions where necessary and comments have been added to provide clarity to data annotators, especially in cases where certain cell types have no close homologue in another species.

The second stage of development will be the extension of the hematopoietic term definitions into full cross-products as discussed above. The revised definitions provided in the first step will enable this extension in a fairly efficient manner depending upon the availability of the necessary terms in external ontologies. The initial step in this direction was taken by Masci and colleagues, who developed a dendritic cell ontology, DC-CL, which is based on cross-product principles and is the foundation of the revised dendritic cell terms in CL1.5.[11] DC-CL terms for types of dendritic cells are based on structural criteria (surface protein expression) with a few cell types also defined by relationships to functions or dispositions. DC-CL utilizes an expanded range of relation types based on those in the OBO Relation Type Ontology (www.obofoundry.org/cgi-bin/detail.cgi?id=relationship) in order to be more expressive about the cellular location and degree of protein expression (e.g. *has_plasma_membrane_part*, *has_high_membrane_amount*). We intend to use these relations together with the definitions provided by Masci et al. to provide logical definitions for the hematopoietic terms.

Recently, the Gene Ontology Consortium obtained an ARRA Competitive Revision grant to allow the cross-product/logical definition approach to be extended to the whole of the CL to create version “CL2.0.” As a first step we are developing the hematopoietic terms of CL 1.5 into an external mini-ontology, “Hemo-CL,” based on these cross products. A provisional version of Hemo-CL is available at obo.cvs.sourceforge.net/viewvc/obo/obo/ontology/anatomy/cell_type/hemo_CL.obo. Figure 2B shows the OBO term stanza for term “induced T-regulatory cell” as it is represented in Hemo-CL. This is illustrated graphically in Figure 2C. We are working with the curators of the Protein Ontology to ensure that the 600+ protein terms needed for Hemo-CL are found in the Protein Ontology. Completion of the Hemo-CL subontology is expected in 2010.

Discussion

The Cell Ontology is an essential core component of the OBO Foundry and has great potential for aiding in data annotation and analysis. With the improvements implemented for CL1.5 and planned for hemo-CL/CL2.0, we expect the CL to fulfill its promise in the area of hematopoietic cell representation. The ontology now has fairly complete coverage of these cell types in an improved hierarchy and with up-to-date molecular definitions. These changes will enable more robust inference across the ontology, provide greater utility for annotation of hematopoietic cell type data, and strengthen the use of the CL as a reference ontology.

Our two-stage approach has worked well in carrying out the needed additions and revisions in the ontology content in this area, and in outlining a clear plan for the full restructuring of the hematopoietic cell section of the ontology. We will now be able to carry out the full restructuring of the hematopoietic cells, and we will work closely with other communities of biologists to improve the CL in a variety of subfields based on our approach outlined here. We will also work with formal ontologists on the set of relations that will be used in the cross-product definitions. The relations defined by Masci et al such as *has_plasma_membrane_part*, as well as the *capable_of* relation introduced here could be considered undesirable as they potentially open the door to a slew of unprincipled relations. We maintain that these ‘macro’ style relations are both useful and principled, as they are formally defined in terms of existing relations and classes, and can be expanded out to a more verbose form if necessary. However, we will continue to work with the wider community of OBO Foundry developers to come up with a solution that is optimal for everyone.

The improvements in the representation of hematopoietic/immune cells in the Cell Ontology will have a great impact on the utility of the ontology in a variety of areas. For instance, classification of cell types based on flow cytometry data will now be possible from the molecular definitions of particular cell types in the CL. The hierarchical nature of the subsumption relationships will enable cell types to be classified according to varying levels of detail, depending on the number and type of flow cytometry parameters studied, corresponding to the specific cell surface molecules expressed. Thus a simple combination of markers used in a flow cytometry experiment designed to study T cells, such as using antibodies against the alpha-beta T cell receptor and the CD4 and CD8 coreceptors, will allow the identification of cell types corresponding to higher level terms in the T cell hierarchy of the CL. The addition of parameters, such as CD25 or intracellular staining for particular cytokines, will distinguish among more granular cell types in the ontology. Furthermore, the ontology itself provides a list of subtypes of particular high-level cell types, which can guide a researcher in the choice of additional parameters to study. Ideally, such an ontology-based flow cytometry system would provide for automated analyses and classifications of cell types in both basic research and clinical settings.

We expect also that cell type terms from the CL will be applied in annotations not only in the GO, but in combinations with other ontologies as well. For example, CL terms might be employed in conjunction with terms from the Infectious Disease Ontology (www.infectiousdiseaseontology.org) as part of a machine-readable description of the life history of an infectious disease within a host. The CL can be used to describe the cell types within the host for which a virus exhibits a specific tropism, and also to describe the cell types of the immune system that are most active in the immune response to the viral infection. In this way a canonical view of a typical infection can be described, and clinical findings could then be checked against this ontological based model to identify differences between individual patient immune responses, which may help guide the choice of therapies.

Furthermore, we expect the Cell Ontology to become a source of important metadata for high throughput gene expression data sets, which are often tied to particular cell types. Similarly, we see clear value in using the CL in the labeling of images and videos of cells, and for other cases where data is derived from defined cell types. Use of CL terms to specify elements of computational and mathematical models of immune responses will also facilitate more rigorous cross-model comparisons. Linking data identified with a particular cell type to the corresponding CL term will allow identification of GO terms that reference the CL term in their names or definitions. The Cell Ontology term "macrophage," for instance, is linked to the GO term, "macrophage differentiation" and 38 other GO terms related to biological processes in macrophages. Many of these GO terms have genes from particular species annotated to them. Thus, a person viewing a cell image annotated with a particular CL term will be able find related GO terms and their associated genes.

Summary

The work described herein represents a significant advance in the representation of hematopoietic cell types in the Cell Ontology, and will increase the utility of the ontology for data annotation, integration, and analysis in this domain in both research and clinical settings. Additionally, the workshop approach we utilized has provided a general framework for future development of these terms, and indeed the whole of the CL, based on the cross-product/logical definition approach.

Acknowledgments

We thank NIAID for the support of the workshop and follow-up teleconferences. ADD, TFM, and JAB are supported by NHGRI grant HG002273, RHS by NIAID contract N01AI40076, PAM by NIAID contract

N01AI50018, MZ by NIAID contract N01AI50020, LGC by NIAID contract R01 AI077706 and AI50019, AMM by NIAID contract AI50019, and BP by NIAID contract N01AI50019.

References

1. Bard J, Rhee SY, Ashburner M. An ontology for cell types. *Genome Biol* 2005;6:R21. [PubMed: 15693950]
2. Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 2007;25:1251–1255. [PubMed: 17989687]
3. Hill DP, Blake JA, Richardson JE, Ringwald M. Extension and integration of the gene ontology (GO): combining GO vocabularies with external vocabularies. *Genome Res* 2002;12:1982–1991. [PubMed: 12466303]
4. Mungall CJ, Bada M, Berardini TZ, et al. Cross-product extensions of the Gene Ontology. *J Biomed Inform.* this issue.
5. Mungall CJ, Gkoutos G, Smith C, et al. Intergrating phenotype ontologies across multiple species. *Genome Biology*. 2009 (accepted).
6. Diehl AD, Lee JA, Scheuermann RH, Blake JA. Ontology development for biological systems: immunology. *Bioinformatics* 2007;23:913–915. [PubMed: 17267433]
7. Diehl AD. unpublished.
8. Smith B, Ceusters W, Klagges B, et al. Relations in biomedical ontologies. *Genome Biol* 2005;6:R46. [PubMed: 15892874]
9. Natale DA, Arighi CN, Barker WC, et al. Framework for a protein ontology. *BMC Bioinformatics* 2007;8 Suppl 9:S1. [PubMed: 18047702]
10. Day-Richter J, Harris MA, Haendel M, et al. OBO-Edit--an ontology editor for biologists. *Bioinformatics* 2007;23:2198–2200. [PubMed: 17545183]
11. Masci AM, Arighi CN, Diehl AD, et al. An improved ontological representation of dendritic cells as a paradigm for all cell types. *BMC Bioinformatics* 2009;10:70. [PubMed: 19243617]

Figure 1.
Term placement for cell type term “macrophage” in the Cell Ontology.
A. In the original Cell Ontology, showing its multiple inheritance structure.
B. In the “CL1.5” incorporating the revisions to the hematopoietic cell types.

Figure 2.

Examples of improvement in the representation of hematopoietic cells.

A. OBO term stanza representative of CL1.5 term definitions for the term “induced T-regulatory cell.”

B. OBO term stanza representative of CL2.0 showing logical definition of the same term as in A.

C. Graphical view of the term relationships in B. (synonyms in A and B omitted for brevity)