Hematopoietic Cell Types: Prototype for a Revised Cell Ontology

Alexander D. Diehl[1‡], Alison Deckhut Augustine[2], Judith A. Blake[1], Lindsay G. Cowell[3], Elizabeth S. Gold[4], Timothy A. Gondré-Lewis[2], Anna Maria Masci[3], Terrence F. Meehan[1], Penelope A. Morel[5], NIAID Cell Ontology Working Group[§], Anastasia Nijnik[6], Bjoern Peters[7], Bali Pulendran[8], Richard H. Scheuermann[9], Q. Alison Yao[2], Martin S. Zand[10], Christopher J. Mungall[11]

[1]The Jackson Laboratory, Bar Harbor, Maine; [2]National Institute of Allergy and Infectious Disease, Bethesda, MD, [3]Duke University Medical Center, Durham, NC, [4]Institute for Systems Biology, Seattle, WA, [5]University of Pittsburgh, Pittsburgh, PA, [6]University of British Columbia, Vancouver, BC, [7]La Jolla Institute for Allergy and Immunology, La Jolla, CA, [8]Emory University, Atlanta, GA, [9]U.T. Southwestern Medical Center, Dallas, TX, [10]University of Rochester Medical Center, Rochester, NY, [11]Lawrence Berkeley Laboratory, Berkeley, CA

Nature Precedings : hdl:10101/npre.2009.3635.1 : Posted 17 Aug 2009

Abstract

The Cell Ontology (CL) aims for the representation of in vivo and in vitro cell types from all of biology. Although the CL is a reference ontology of the OBO Foundry, it 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. A recent workshop sponsored by NIAID on hematopoietic cell types in the CL addressed both issues. The section of the ontology dealing with hematopoietic cells was extensively revised, and plans were set for structuring these cell type terms as cross-products with logical definitions built from relationships to external ontologies, such as the Protein Ontology and the Gene Ontology. The methods and improvement to the CL in this area represent a paradigm for improvement of the whole of the ontology over time.

Overview

The Cell Ontology (CL) is an OBO Foundry candidate ontology originally built to represent *in vivo* and *in vitro* cell types, including developmental stages, of all the major model organisms.[1] The CL now aims to become a reference ontology within the OBO Foundry.[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] At Mouse Genome Informatics, the CL is used in conjunction with Gene Ontology (GO) annotation of mouse gene products to indicate the cell type in which a gene product is active. The GO itself uses CL terms in the formation of new GO terms: for instance, the GO term "leukocyte differentiation" is a cross-product of the CL term "leukocyte" with the GO term "cell differentiation."

The Cell Ontology is constructed using two relationships, *is_a* and *develops_from*. The first relationship is used to build an ontology of cellular subtypes; the latter relationship is used to indicate cell lineage relationships. The ontology as it was initially developed relied 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, the well-defined cell type "erythrocyte" is a type of "erythroid lineage cell," "oxygen accumulating cell," "transporting cell," and "blood cell." It also has a *develops_from* relationship with "reticulocyte."

With its multiple inheritance structure, the original CL could be described as having separate ontologies of cell types delineated by particular cellular qualities overlaid upon each other, i.e. an ontology with

‡ corresponding author, adiehl@informatics.jax.org

§ Other members of the NIAID Cell Ontology Working Group: Christopher C. Cavnor, Patrick Dunn, Thomas B. Kepler, Jingming Ma, Yuri N. Naumow, Elena N. Naumova, Jeremy Seto, and Alessandro Sette.

multiple axes of differentia that are variously and sometimes arbitrarily applied to individual cell types. Furthermore the high level terms themselves are not 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.

Informal 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 with the aim of finding a single axis of differentia upon which to base the ontology. Participants in these general discussions about the CL gradually recognized that no consistent differentia such as cellular structure or lineage can adequately describe all cell types, and that the best solution for biologists is to represent the differences and relations between cell types as scientists working in various subfields of biology do, depending on their specific criteria for differentiating cell types.

Other criticisms about the CL include the fact that many terms do not have definitions or a complete set of synonyms. Also, cell types in many subfields of biology are poorly represented within the CL. A compounding issue has been the lack of a full-time curator for the ontology as a whole. Efforts at improvement have been made in certain areas of the ontology, and hematopoietic cell types in particular have been the focus of two rounds of intensive curation in recent years. Here we report on these revisions and examine the process as an example for the future development of the Cell Ontology.

Hematopoietic Cell Type Revisions

The first set of improvements for hematopoietic cells was done in 2006 in conjunction with the revision of the terms for immunological processes in the GO.[4,5] 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 cell types.

A second, larger round of revisions to the hematopoietic cell type terms in CL is described herein. These revisions are the product 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 and developing additional specific terms for T cells, B cells, natural killer cells, monocytes and macrophages, and dendritic cells, and 2) establishing a new paradigm for development of the CL. These changes in the representation of hematopoietic cells were needed to represent these cell types in a more complete manner so that all major cell types identified in the literature are found in the ontology and so that these cell types are defined 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 both specific groups of cell types of interest to immunologists as well as how to improve the overall ontological structure of these groups and 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 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 in many cases that 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 participants agreed that the relationship *has_part* from the OBO relationship ontology would be used to relate particular cell types to protein terms from the Protein Ontology.[6,7]

However, for certain cell types, such as macrophages, it was seen that the full molecular characterization of different types of macrophages is still not complete in the literature, and that anatomical location serves as a major differentia for these cells. 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 GO processes that relate to individual cells, such as "cytotoxicity" or "cytokine production," or GO processes that involve coordination of multiple cell types, such as "T-helper 1 type immune response." Thus, the participants at the workshop agreed to focus on structural criteria where possible as the primary differentia, but to accept other types of differentia when necessary. This flexibility should make it possible to stick to the commonly accepted biological definitions of individual cell types and to organize the ontology according to sound ontological principles while still reflecting organization of hematopoietic cell types seen in the literature.

The primary goal in revising the hematopoietic cell terms is to define all the terms according to logical

Nature Precedings : hdl:10101/npre.2009.3635.1 : Posted 17 Aug 2009

Nature Precedings : hdl:10101/npre.2009.3635.1 : Posted 17 Aug 2009

definitions based on relationships to external ontologies. The workshop participants recognized that reaching the full development of these terms as cross-products would be difficult at this time due to the lack of a full-time curator for the CL. Also, external ontologies, such as the Protein Ontology, are not yet complete in all the required terms. Yet at the same time, the new hematopoietic cell terms are needed for data annotation and development of cross-products in the GO and other ontologies.

Results: A Two-Stage Process

Reflecting the above considerations, the participants at the NIAID workshop agreed upon a two-stage approach to further development of the hematopoietic cells in the Cell Ontology. In the first stage, which is now complete, current terms were revised and new terms added by the experts at the workshop. The textual definitions for these terms contain all the necessary details to define the cells logically. These terms have been directly incorporated into the existing ontology. It was also decided to separate the hematopoietic terms from the multiple inheritance hierarchy of the original CL as much as possible, so that the section of the ontology containing these terms represents a true ontology hierarchy. This first-stage ontology has been given the working name "CL1.5." Figure 1A shows a typical OBO term stanza for one of these new terms, "induced T-regulatory cell."

The second stage will then be the development of the hematopoietic terms into full cross-products as discussed above. The extended definitions provided in the first step will hopefully enable this to be done in a fairly efficient manner depending upon the availability of the necessary terms in external ontologies. Ideally, this approach will be extended to the whole of the CL to create version "CL2.0." For the moment we plan to develop the hematopoietic terms of the CL into an external mini-ontology based on these cross products, "hemo-CL." Figure 1B shows the OBO term stanza for term "induced T-regulatory cell" as it will be represented in hemo-CL and CL2.0, illustrated graphically in Figure 1C. We have already been working with the curators of the Protein Ontology to ensure that protein terms needed

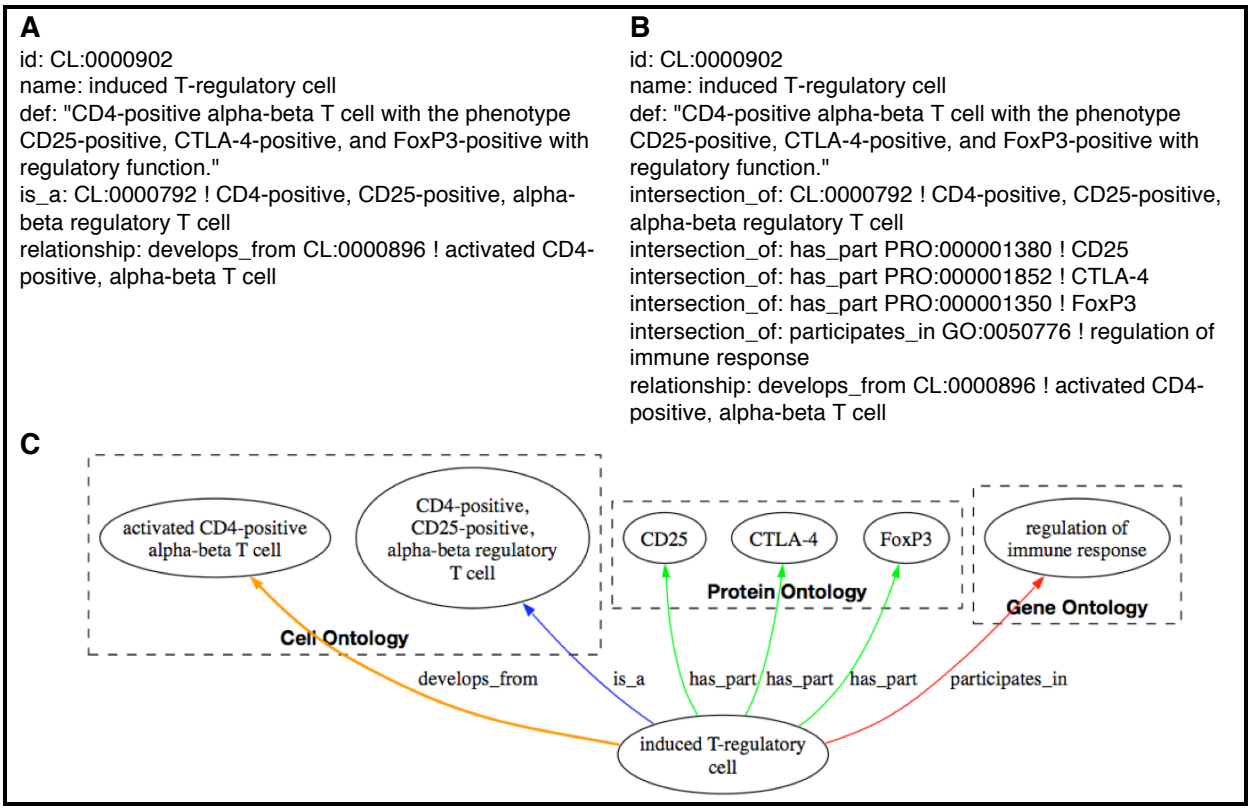

Figure 1. Examples of improvement in the representation of hematopoietic cells.
A. OBO term stanza representative of CL1.5 term definitions for the term "induced T-regulatory cell."
B. OBO term stanza representative of CL2.0 showing logical definition of the same term as in A.
C. Graphical view of the term relationships in B.

for hemo-CL are found in the Protein Ontology.

The initial step towards hemo-CL and CL2.0 has been taken by Masci and colleagues, who have 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.[8] DC-CL terms for types of dendritic cells are primarily 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 relationship types based on those in the relationship ontology in order to be more expressive about the cellular location and degree of protein expression (*has_membrane_part*, *has_high_membrane_amount*). It is likely that similar relationships will be employed in the construction of hemo-CL and CL2.0.

Specific Improvements in the Representation of Hematopoietic Cell Types

With the work initiated at the NIAID workshop and carried on afterwards, many concrete improvements to CL content in the area of hematopoietic cells have been achieved. Many new terms for individual cell types have been created, 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 were 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. The 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.

The ontology structure has been improved as well in important areas such as T cell and B cell development. Lineage relationships via the *develops_from* relationship have been provided for many additional cell types. In general the hematopoietic terms are intended to be species neutral, but species-specific information has been incorporated in some definitions where necessary and comments added to provide clarity to data annotators.

Discussion

The Cell Ontology is an essential core component of the OBO Foundry and has great potential for aiding data annotation and analysis. With the improvements described herein, implemented for CL1.5, and planned for hemo-CL/CL2.0, we expect the CL to fulfill much more of 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 using up-to-date molecular definitions. These changes will provide for more robust inference across the ontology and greater utility for annotation of hematopoietic cell type data, and will strengthen the use of the CL as a reference ontology for cross-product development.

The workshop approach, aided by an acting editor for this section of the ontology, has worked reasonably well in carrying out the needed additions and revisions in the ontology content in this area, and in outlining a clear plan for the future of the ontology. The section-by-section approach for improvement of defined parts of the Cell Ontology represents a paradigm for continued development of the CL and should prove even more useful once dedicated funding is achieved.

Acknowledgements

We thank NIAID for the support of the workshop and follow-up teleconferences. ADD and JAB are supported by NHGRI grant HG002273, RHS by NIAID grant N01AI40076, PAM by NIAID grant N01AI50018, and BP by NIAID grant N01AI50019.

References

1. Bard J, Rhee SY, Ashburner M. An ontology for cell types. Genome Biol. 2005;6:R21.
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.
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;12:1982-91.
4. Diehl AD, Lee JA, Scheuermann RH, Blake JA. Ontology development for biological systems: immunology. Bioinformatics. 2007;23:913-5.
5. Diehl AD, unpublished.
6. Smith B, Ceusters W, Klagges B, et al. Relations in biomedical ontologies. Genome Biol. 2005;6:R46.
7. Natale DA, Arighi CN, Barker WC, et al. Framework for a protein ontology. BMC Bioinformatics. 2007;8 Suppl 9:S1.
8. 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.

Nature Precedings : hdl:10101/npre.2009.3635.1 : Posted 17 Aug 2009