

INTRODUCTION

Open Access



Cells in Experimental Life Sciences (CELLS-2018): capturing the knowledge of normal and diseased cells with ontologies

Sirarat Sarntivijai¹, Yongqun He^{2*} and Alexander D. Diehl³

From Second International Workshop on Cells in Experimental Life Sciences (2018)
Corvallis, OR, USA. 7 August 2018

Abstract

Cell cultures and cell lines are widely used in life science experiments. In conjunction with the 2018 International Conference on Biomedical Ontology (ICBO-2018), the 2nd International Workshop on Cells in Experimental Life Science (CELLS-2018) focused on two themes of knowledge representation, for newly-discovered cell types and for cells in disease states. This workshop included five oral presentations and a general discussion session. Two new ontologies, including the Cancer Cell Ontology (CCL) and the Ontology for Stem Cell Investigations (OSCI), were reported in the workshop. In another representation, the Cell Line Ontology (CLO) framework was applied and extended to represent cell line cells used in China and their Chinese representation. Other presentations included a report on the application of ontologies to cross-compare cell types and marker patterns used in flow cytometry studies, and a presentation on new experimental findings about novel cell types based on single cell RNA sequencing assay and their corresponding ontological representation. The general discussion session focused on the ontology design patterns in representing newly-discovered cell types and cells in disease states.

Introduction

The rapid advancement of cell technologies has inevitably led to challenges in keeping up with the volume of the data being produced as well as the dynamic evolution of data formats and knowledge representation. Experimental cell cultures and cell lines are widely used and often generated de novo as part of an experimental protocol, while normalization of experimental cell data produced in different laboratory settings is difficult due to the non-synchronous nature of multiple laboratories working on similar questions on the same timeline. New knowledge obtained by high-resolution technologies (e.g., mass cytometry or single-cell RNA sequencing) adds more data volume that requires robust analysis and representation, especially regarding novel cell populations that do not correspond to existing classes of the community-based Cell Ontology (CL) [1] or Cell Line

Ontology (CLO) [2], two cell-related reference ontologies in the Open Biomedical Ontology (OBO) framework [3].

The first International Workshop on Cells in Experimental Life Science (CELLS-2017) was held in Newcastle, UK, on September 2017 on the theme of identifying challenges from the ontology semantic perspective, and the experimental laboratory perspective [4]. The CELLS-2017 workshop included 7 oral presentations, and eventually five full-length articles [5–9] were published in the special issue of the first CELLS workshop in *BMC Bioinformatics* (<https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-18-supplement-17>).

Following the success of the first CELLS workshop, the 2nd International Workshop on Cells in Experimental Life Science (CELLS-2018) was held on August 7, 2018, in conjunction with International Conference on Biomedical Ontology 2018 (ICBO-2018), August 7–10, 2018, Corvallis, Oregon, USA. CELLS-2018 was organized with a focus on the themes of addressing the

* Correspondence: yongqunh@med.umich.edu

²University of Michigan Medical School, Ann Arbor, MI, USA
Full list of author information is available at the end of the article



challenges of knowledge representation of newly-discovered cell types and of cells in disease states.

Summary of the talks and papers presented at this workshop

The half-day CELLS-2018 workshop included five oral presentations and a general discussion session. Four of the proceedings papers, corresponding to four oral presentations, have been extended, further reviewed, and are published in this special issue for the CELLS-2018 workshop in *BMC Bioinformatics*.

In this workshop, two new ontologies were reported. Serra et al. presented the development of the Cancer Cell Ontology (CCL) that represents cancer cell types related to a variety of hematologic malignancies using logical axioms that capture the expression of cellular surface markers using Protein Ontology (PRO) [10] classes [11]. The ontology includes human and computer-readable definitions for 300 classes of blood cancers. The logical axioms of the ontology can be used to classify patient cell surface marker data into appropriate diagnostic categories, to allow for integration of the ontology into existing tools for flow cytometry data analysis to facilitate the automated diagnosis of hematologic malignancies.

The other reported new ontology is the Ontology for Stem Cell Investigations (OSCI) [12]. Stem cells and stem cell lines have been widely used in biomedical research. CL and CLO represent *in vivo* stem cells and *in vitro* stem cell line cells, respectively. However, neither CL nor CLO has a focus on the representation of stem cell investigations, which is the main goal of the community-driven OSCI ontology. OSCI imports stem cell and stem cell line related classes from CL and CLO. OSCI also imports classes needed to describe stem cells and stem cell experimental methods from other ontologies such as PRO [10] and the Ontology for Biomedical Investigations (OBI) [13]. Many metadata types associated with various stem cell investigations have been identified and represented as classes in OSCI. Two use cases of OSCI include applying OSCI to systematically explore experimental variables in induced pluripotent stem cell line cell studies related to bipolar disorder, and applying OSCI to model and represent stem cell gene markers and relations identified using semi-automated literature mining.

The Chinese National Infrastructure of Cell Line Resource (NICR) stores and distributes over 2000 cell lines in China. To support worldwide interoperable cell line information representation and usage, Pan et al. reported the development of CLO-NICR ontology that represents these cell lines using the semantic framework of CLO [14]. As a subset of the master CLO release, CLO-NICR can also be used in a stand-alone way to support applications for biomedical research in China NICR. A new ontology design pattern specifically to CLO-NICR is proposed to bridge cell

line cells to disease studies using laboratory animal models. The authors also demonstrate the Chinese language representation of the cell line cell information. The CLO-NICR development was one of the research efforts from the OntoChina (<http://www.ontochina.org>) [15], a new initiative supporting an ontology developer community program in China to promote collaborative ontology development and applications around the world.

In flow cytometry, cells are gated into different populations based on common features like forward or side scatter or expression patterns of protein markers. The cell populations are typically identified via a general name of the cell type (e.g. 'T cells') and the specific marker pattern observed (e.g. CD14⁻/CD3⁺). In a CELLS-2018 presentation, Vita et al. applied ontologies to connect cell population descriptions and gating definitions [16]. In their study, ontologies were used to cross-compare cell types and marker patterns in the ImmPort Immunology Database and Analysis Portal [17]. The gating definitions were parsed using classes from the Cell Ontology (CL) [1] and the Protein Ontology (PRO) [10]. Logical axioms in the CL were used to detect discrepancies between submitted data about cell populations and cell type classes in the CL. New OBI classes were also generated to describe and capture data acquisition. This study resulted in suggestions for new logical axioms and cell types for addition to the CL.

One challenge is how to ontologically represent the cell types newly discovered using experimental technologies. In a CELLS-2018 podium presentation, Mohamed Keshk of Richard H. Scheuermann's group presented a provisional cell type ontology that represents newly identified novel cell types experimentally identified using single cell RNA sequencing (scRNAseq), an extension of the work Dr. Scheuermann's group presented in the CELLS-2017 workshop [5]. The data being gathered from the experimental work is represented in a standard semantic format that can be exchanged, retrieved, and inferred over using standard approaches and tools. This format supports the translation and representation of biological knowledge to become findable, accessible, interoperable and reproducible (FAIR), per the FAIR principles [18].

As can be seen, the presentations at the CELLS-2018 workshop covered a range of issues in the representation of cell types: cell types and experimental methodology related to them from under-represented domains including cancer cells and stem cells, the ontological representation of cell lines from a major national resource in China, and the challenge of representing experimental results about cells based on flow cytometry or single cell RNA sequencing. If there is a common theme to these efforts, it is the drive to present a coherent ontological representation of existing and new knowledge about cells that builds upon existing efforts in the CL, CLO,

and OBI, and is likewise extensible for the representation of new data and methodology related to cell types, within the OBO Foundry framework.

Challenges and opportunities

In the CELLS-2018 workshop, we also organized a general discussion session. Our discussion focused on the design pattern representation of normal vs diseased conditions in cells. For example, what is the difference between normal and diseased cells, and how to ontologically represent the transformation from a normal cell to diseased cell, and what are the common design patterns for normal and diseased cells? In addition, we discussed how to represent additional information (e.g., gene markers) in ontology? The five presentations described above fit in well with the discussion theme.

One challenge in cell ontology development is to how to specify the definition of cell types. The de facto convention used for the creation of many classes in the CL has resulted in the creation of cell type classes that are named and defined both textually and logically by the cell surface markers expressed by the cell type using a *post-composed* approach. This is may be less than ideal as the community discovers new cell types based on subdividing known cell populations with additional markers. As we move beyond marker definition of cell types developed via flow cytometry to defining cell types via single cell RNA sequencing or proteomics, a cell type could potentially be defined by a combinations of hundreds of marker genes or proteins. Should we add more markers to cell types and use these markers for cell type definitions? As reflected by the work of Dr. Scheurmann's group and others [19], we might therefore consider what constitutes the minimal combination set of markers that can uniquely define a new cell type? Under some circumstances, defining cell types by cell surface or internal markers may not be sufficient. For example, to define brain cells, we may also need to include other physiological and morphological properties such as cell location and function in order to uniquely define a specific brain cell type. However, in many older experimental protocols, all the cells of a given tissue are homogenized prior to RNA or protein extraction, resulting in the loss of specific location and function information in the process. Other types of experiment rely on, for instance, immunofluorescence or green fluorescence protein (GFP) tagging to examine the expression and localization of protein markers in cells. Given the diversity of the experimental approaches for studying cells, we need to be careful in how we represent experimental results in the form of ontological definitions of cell types. Some results may be true under specific experimental conditions, such as during cellular activation; however, under other conditions, the results may not be

true or may only be true for a particular state of a cell type. Ontology classes should always represent certainty, and not probability. Definitive identification and representation of cell types requires synthesizing knowledge gained from multiple experimental approaches for a cell type definition to be accepted as truth. Another question is how to infer cell function. How can we ontologically represent cell function and related pathways in normal or diseased cells? Can we infer the function by merging cell morphology and gene expression? Can we use function analysis via, for instance GO term enrichment of expressed genes to identify key biological processes a cell may be participating in, and thereby establish a functional signature that can be used to infer cell type? These questions may need to be addressed outside an ontology class level. Solutions under discussion include inference via instances of the main class, or using other object/annotation properties to build the inferred knowledge.

As the basic unit of life, the cell remains central to biological and biomedical research. Given that challenges presented by studies of cells and their ontological representation and applications, the attendees of the workshop welcome and expect more CELLS workshops in the future. And, of course, these challenges in themselves are also opportunities. For example, the Human Cell Atlas project, which aims to map all the cells in the human body, has recently been initiated from vision to reality [20], and is already generating a wealth of data about cell types. Many precision medicine studies, such as the Kidney Precision Medicine Project (KPMP: <http://kpmp.org>), aim to understand specific cell-disease associations and guide the control and treatment of diseases under different clinical conditions [21]. A variety of recent studies have discovered many new cell types, which sometimes show different characteristics in normal or disease states [22–26]. These clinical, experimental, and disease conditions, as well as new knowledge about cell types, need more in-depth ontological modeling, representation, and analysis. New tools are also needed to support more productive ontological research in the cell field, including the development of machine learning methods that work across many data sets to identify necessary and sufficient criteria for cell type definitions across the whole body. We anticipate promising research progress in the cell-related ontology field in the years to come.

Conclusion

The CELLS-2018 workshop provided a venue for researchers to report their updates and new studies, and to discuss challenges and innovative solutions in the development and application of biomedical ontologies to represent and analyze in vivo and in vitro cell- and cell line-related knowledge and data, including stem cell technologies.

Acknowledgements

We appreciate help provided by the CELLS-2018 Program Committee, which includes three of our meeting co-organizers (Sirarat Sarntivijai, Yongqun He, and Alexander Diehl), Asiyah Yu Lin, Parisa Nejad, Rebecca Racz, Nicole Vasilevsky, Danielle Welter, and Jie Zheng. We also appreciate the support by the organizers of the International Conference on Biomedical Ontology 2018. All the discussion and participation of the CELLS-2018 workshop attendees are also appreciated.

Funding

YH was supported by U54HL127624 (NHLBI). AD was supported by 5UL1TR001412 (NCATS). Publication costs of this article are not supported by sponsorship.

Availability of data and materials

Content and agenda of the CELLS workshop is available at <https://sites.google.com/site/cellsworkshop/>.

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 20 Supplement 5, 2019: Selected articles from the Second International Workshop on Cells in Experimental Life Sciences (CELLS-2018) at the 2018 International Conference on Biological Ontology (ICBO-2018)*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-20-supplement-5>.

Authors' contributions

All authors contributed to the writing and editing of the manuscript and have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹ELIXIR, Wellcome Genome Campus, Hinxton, Cambridge, UK. ²University of Michigan Medical School, Ann Arbor, MI, USA. ³Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, Buffalo, NY, USA.

Published: 25 April 2019

References

- Diehl AD, Meehan TF, Bradford YM, Brush MH, Dahdul WM, Dougall DS, He Y, Osumi-Sutherland D, Ruttenberg A, Sarntivijai S, et al. The cell ontology 2016: enhanced content, modularization, and ontology interoperability. *J Biomed Semantics*. 2016;7(1):44.
- Sarntivijai S, Lin Y, Xiang Z, Meehan TF, Diehl AD, Vempati UD, Schürer TC, Pang C, Malone J, Parkinson H, et al. CLO: the cell line ontology. *J Biomed Semantics*. 2014;5:37.
- Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, et al. The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol*. 2007;25(11):1251–5.
- Sarntivijai S, Diehl AD, He Y. Cells in experimental life sciences - challenges and solution to the rapid evolution of knowledge. *BMC Bioinformatics*. 2017;18(Suppl 17):560.
- Bakken T, Cowell L, Aevermann BD, Novotny M, Hodge R, Miller JA, Lee A, Chang I, McCarrison J, Pulendran B, et al. Cell type discovery and representation in the era of high-content single cell phenotyping. *BMC Bioinformatics*. 2017;18(Suppl 17):559.
- Kafkas S, Sarntivijai S, Hoehndorf R. Usage of cell nomenclature in biomedical literature. *BMC Bioinformatics*. 2017;18(Suppl 17):561.
- Osumi-Sutherland D. Cell ontology in an age of data-driven cell classification. *BMC Bioinformatics*. 2017;18(Suppl 17):558.
- Ong E, Xie J, Ni Z, Liu Q, Sarntivijai S, Lin Y, Cooper D, Terryn R, Stathias V, Chung C, et al. Ontological representation, integration, and analysis of LINC cell line cells and their cellular responses. *BMC Bioinformatics*. 2017; 18(Suppl 17):556.
- Ong E, Sarntivijai S, Jupp S, Parkinson H, He Y. Comparison, alignment, and synchronization of cell line information between CLO and EFO. *BMC Bioinformatics*. 2017;18(Suppl 17):557.
- Natale DA, Arighi CN, Blake JA, Bult CJ, Christie KR, Cowart J, D'Eustachio P, Diehl AD, Drabkin HJ, Helfer O, et al. Protein ontology: a controlled structured network of protein entities. *Nucleic Acids Res*. 2014;42(Database issue):D415–21.
- Serra L, Duncan W, Diehl A. The Cancer cell ontology. *BMC bioinformatics*. 2019;XX(this issue):XX-XX. <https://doi.org/10.1186/s12859-019-2722-8>
- He Y, Duncan WD, Cooper DJ, Hansen J, Iyengar R, Ong E, Walker K, Tibi O, Smith S, Serra L et al: OSCI: standardized stem cell ontology representation and use cases for stem cell investigation. *BMC Bioinformatics* 2019, XX(this issue):XX-XX. <https://doi.org/10.1186/s12859-019-2723-7>
- Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, Clancy K, Courtot M, Derom D, Dumontier M, et al. The ontology for biomedical investigations. *PLoS One*. 2016;11(4):e0154556.
- Pan H, Bian X, Yang S, He Y, Liu Y, Yang X. The cell line ontology-based representation, integration and analysis of cell lines used in China. *BMC Bioinformatics*. 2019;XX(this issue):XX-XX. <https://doi.org/10.1186/s12859-019-2724-6>
- He Y, Yu H, Yang X, Shao C, Zhou W, Zhu Y, Wang H, Liu Q, Xie J, Zheng J, et al. Ontology: foundation of biomedical big data and precision medicine research. *Chin J Bioinformatics*. 2018;16(1):7–14.
- Vita R, Overton JA, Diehl A, Cheung K, Dunn P, Kleinstein S, Peters B. Connecting cell population descriptions and gating definitions through ontologies. *BMC Bioinformatics*. 2019;XX(this issue):XX-XX. <https://doi.org/10.1186/s12859-019-2725-5>
- Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, Hu Z, Zalocusky KA, Shankar RD, Shen-Orr SS, et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data*. 2018;5:180015.
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018.
- Aevermann BD, Novotny M, Bakken T, Miller JA, Diehl AD, Osumi-Sutherland D, Lasken RS, Lein ES, Scheuermann RH. Cell type discovery using single-cell transcriptomics: implications for ontological representation. *Hum Mol Genet*. 2018;27(R1):R40–7.
- Rozenblatt-Rosen O, Stubbington MJ, Regev A, Teichmann SA. The human cell atlas: from vision to reality. *Nature News*. 2017;550(7677):451.
- He Y, Steck B, Ong E, Mariani L, Lienczewski C, Balis U, Kretzler M, Himmelfarb J, Bertram JF, Azeloglu E et al. KTAO: a kidney tissue atlas ontology to support community-based kidney knowledge base development and data integration. In: International conference on biomedical ontology 2018 (ICBO-2018): august 7–10, 2018 2018; Corvallis, Oregon, USA. 1–6.
- Jordao MJC, Sankowski R, Brendecke SM, Sagar LG, Tai YH, Tay TL, Schramm E, Armbruster S, Hagemeyer N, et al. Single-cell profiling identifies myeloid cell subsets with distinct fates during neuroinflammation. *Science* 2019;363(6425):eaat7554. <https://doi.org/10.1126/science.aat7554>.
- Gan Y, Li N, Zou G, Xin Y, Guan J. Identification of cancer subtypes from single-cell RNA-seq data using a consensus clustering method. *BMC Med Genet*. 2018;11(Suppl 6):117.
- Mrdjen D, Pavlovic A, Hartmann FJ, Schreiner B, Utz SG, Leung BP, Lelios I, Heppner FL, Kipnis J, Merkler D, et al. High-dimensional single-cell mapping of central nervous system immune cells reveals distinct myeloid subsets in health, aging, and disease. *Immunity*. 2018;48(2):380–95 e386.
- Der E, Ranabothu S, Suryawanshi H, Akat KM, Clancy R, Morozov P, Kustagi M, Czuppa M, Izmirly P, Belmont HM, et al. Single cell RNA sequencing to dissect the molecular heterogeneity in lupus nephritis. *JCI Insight*. 2017;2(9).
- Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. 2017; 356(6335).