Towards a pragmatic definition of cell type

This manuscript (permalink) was automatically generated from lubianat/technotype@e44735a on September 16, 2020.

Authors

- **Tiago Lubiana**
  0000-0003-2473-2313 · lubianat
  Computational Systems Biology Laboratory, University of São Paulo

Abstract

Introduction

One of the first classes in any undergraduate major in life sciences is the histology class. The students are tasked with identifying cell types across various tissues, looking for color and shape patterns in hematoxylin-eosin stains. The text-books, such as the one by Junqueira & Carneiro[**???**], work as "manuals" (in the Kuhnian sense)[1], perpetuating our paradigms of what we know about a couple of hundred cell types.

We advanced our tools, but our concept of "cell type" is still largely based on century-old histochemical techniques - such as the Golgi-stains of neurons, immortalized in the drawings of Ramon y Cajal[2]. This is noticeable even in the names given to cell types, which often harbour histological mentions: "erithrocytes", "eosinophil", "basophils", "oxyphilic cell of the thyroid" . That implies that the concepts we use are drawn from studies of microanatomy. The connection with anatomy leads to thinking about cell types as anatomical entities as if they were clearly dissectable and fixed in an organism. This may be the reason why attempts to quantifying cell types frequently mention the order of only "hundreds" of human cell types [3] [4] [5].

New techniques have challenged this anatomy based conceptualization. From multiplexed flow cytometry to patch clamping, single-cell techniques opened our eyes for an amazing diversity. With a burst of new categories, novel cell "subtypes" and "families" started to pop up in the literature. The rise of new cell types has become especially evident in the past few years, with the explosion of single-cell omics and the creation of the Human Cell Atlas [6] (which explictly aims to caracterize *all* cell types) and the HUBMAP [**???**] projects.

The advances in biology require us to find better answers for how to define a cell type. The concept might not have a "true" meaning, in a material-realistic sense. Nevertheless, we can strive to find nominal, pragmatic definitions for the pragmatic challenges of large scale biology. Otherwise, how can we precisely label single-cell data? How can we formalize the discovery of new cell types? How can we integrate the knowledge from the millions of scientific articles published every year?

The need for conceptual advance is perceived by the scientific community [7], and new perspectives are arising. One core line of thought on the definition of cell type is evolutionary: the cell type as an evolutionary unit defined by a Core Regulatory Complex (CoRC) of transcription factors. That definition enables the drawing of parallels between the evolution of other biological entities (such as genes, proteins and species) to the evolution of cell types. Models of how multicellular life works greatly benefit from concepts such as "sister types" (cell types that diverged from a single ancestor), "cell type homology," (cell types in different species that share a common evolutionary origin) and "cell type convergence" (cell types that execute similar functions, but which are not directly evolutionarily related). [8] [9]

However, in an analogous way to how different concepts of species coexist [10], our quest to define cell types may take various forms depending on the application. The challenge of representing cell types in the context of evolution is conceptually different from the challenge of representing cell types in biomedical experimentation. In that direction, it is notable the groundwork of the Cell Ontology [11] [12] [13] project and the International Workshop on Cells in Experimental Life Sciences[14] [15]. Their contributions base much of the views here and will be discussed in detail, througout the article.

The conceptual quest addressed by this work is one of the research synthesis. Which cell type definition can be crafted for rigorously describing biomedical experiments?

For that goal, the body of the article is divided in 4 parts. In Part 1, I will bring a proposal of a set of rules that are sufficient for defining cell types. In Part 2, I'll propose a small set of names for differentiating main classes of cell types. In Part 3, I will adress the logical consequences of the proposed definitions. And in Part 4, I will discuss the pragmatic challenges for employing such definitions.

-# A set of 3 + 1 rules for defining a cell type

In an opinion article published in Cell Systems in 2017, researchers presented their views on the "Conceptual Definition of 'Cell Type' in the Context of a Mature Organism?" [7]. The opinions vary, and do not converge to a consensus. Many of the scientists see a core role of cells' functions in defining cell types, a slippery road, as the meaning of "function" in biology is elusive [16].

In one recent attempt to define cell types for single cell RNA-Seq, Aevermann et al came up with a set of needs: "The minimum set of necessary and sufficient marker genes selectively expressed by the cell type", "A parent cell class in the CL", and "A specimen source description (anatomic structure þ species)." [**???**] They have great merit in defining clear guidelines for marking a cell type. The explicit requirement of markers is reasonable for the field of single-cell RNA-seq, where marker information is abundant. Notably, the Cell Ontology has even used markers for defining cell types across domains, an approach specially employed for immune cells [12] .

The use of markers, however, leaves us with a conceptual problem: definitions of cell type used by electrophysiologists, or even in manuals of histology classes, are not based on markers. Rigorously, this would leave aside a whole part of what we consider biomedical knowledge. Moreover, gene markers are not defined for cell types that span multiple species, a problem already discussed on the Cell Ontology report of 2011 [12]. Multispecies markers would require us to explicitly consider homology, adding an extra layer of confusion.

My pragmatic definition of cell type (for eukaryotic, multicellular organisms) consists of 3 + 1 simple rules. A cell type is any class of cells which *MUST* be:

- Rigorously defined
- Theoretically useful
- Identifiable for a defined taxon

And that *SHOULD* be:

- Logically related to other cell types

The meaning of *MUST* and *SHOULD* follow the the RFC guidelines for requirement levels, where *MUST* represents an "an absolute requirement of the specification" and *SHOULD* represents that there may exist valid reasons in particular circumstances to ignore a particular item, but the full implications must be understood and carefully weighed before choosing a different course". In the following paragraphs, I will clarify the meaning and motivation of the rules.

For rule 1, by "rigorously defined", I mean that a cell can be identified as belonging or not to the class given a set of criteria. An example of such a rigorous criteria is "expression of the CD3 protein, expression of the CD4 protein, lack of expression of the CD8 protein." This implies the need for using rigorous definitions of what is a "CD3 protein" , and what "expression" means. Similar definitions are used in practice immunology for defining cell types (even though it is not always simple [17]). In the case of markers, then, any combination of markers (or lack thereof) defines a different *cell type*. In this framework, however, *any* rigorous definition is acceptable, and small differences are enough for constituting new cell types. For example, a valid definition would be a "big cell" defined as a cell with a

length of more than 50 micrometers on any axis. Another valid definition is a “cardiocyte” as any cell in a heart.

The recognition of multiple valid types of rules is not new. The first Cell Ontology article, in 2005, explicitly acknowledged criteria based on “‘function’ (for example, electrically_excitable_cell, secretory_cell, photosynthetic_cell), histology (for example, epithelial_cell, mesenchyme_cell), lineage (for example, ectodermal_cell, endodermal_cell) and ploidy (for example, haploid_cell, polyploid_cell)”. [11]. The Cell Ontology, however, integrates cell types from different phyla. The integrated cell types are useful as points of convergence for databases, or for structured human learning. However, for annotation of scientific articles, a definition of the *scope* of taxons under a cell type may increase the clarity.

Rule number 2 is, thus, a specification of a rigorousness: one rigorousness criteria is to define the taxons in which a given cell type is expected to have manifestations. We should be able to identify a cell of the type in any individual of the taxon of interest, given the appropriate conditions (e.g. stage of life and biological sex). I will call this set of taxons the *scope* of the cell type. Note that, as cell types can be defined by function, and functions can converge, a global definition cannot restricted to monophyletic taxa (“clades”).

Knowing the scope is important to avoid the pitfalls of extrapolation. One recurrent extrapolation is assuming that theories corroborated with mice experiments are valid for other organisms. This extrapolation is an instance of the classic problem of induction, detailed thoroughly in the Logic of Scientific Discovery. A specification of the scope a researcher is referring to would make inductional claims explicit and enable proper evaluation of claims.

Currently, the largest authority on cell type definitions, the Cell Ontology, defines *“general cell types”* which spans multiple animal taxons.[13] These definitions are conceptually important, but their use for labeling studies plainly is dangerous. It is not safe to assume that a “mouse neutrophil” is simply a “neutrophil” that happens to be in a “mouse”. A definition of scope is essential to tease apart general claims from study-specific claims.

By providing a specific scope, we can make explicit what are currently implicit predictions. If we take a theory that “the scope of neutrophils is mammals,” it explicitly predicts that we expect whales to have neutrophils, for example. If we cannot identify neutrophils after severe tests for any single species, we might agree that the claim has been falsified. Then, we would have to consider this previous concept as unreal, and tailor our understanding of neutrophils to new scopes.

Rule number 3 is a rule of practical concern. There is a massive amount of “rigorous classes” that one scientist might come up with, due to the combinatorial nature of classes, far outnumbering the reported number of atoms in the observable universe. For that reason, a criterion of usefulness is necessary for deciding when a class of cells is considered a cell type. The simplest criteria of usefulness is one based on the individual: a valid cell type is whatever class any individual rationally finds useful.

Rule number 4 is one practical extension of the “usefulness” rule: a cell type has to be logically anchored to other cell types for it to be useful. Which means that a definition of a cell class is (for research synthesis concerns) of lower usefulness if it can’t be considered a “subclass” of other cell type. For our practical concerns, all imaginable cell types are subclasses of “cell of eukaryotes”. This is presented as a recommendation instead of a requirement as, in practice, it might be an overhead not strictly necessary for claims of discovery of new cell types and similar tasks.

The ontological organization is important for integrating knowledge across studies. A “transcriptomically-defined” cell type and a “electrophysiologically-defined” cell type cannot be

rigorously said to be the same, but they can be grouped in a “superclass” that contains cells that match either one or the other criteria. Practically, when describing a cell type, one should make an effort to insert it into the universe of interrelated cell types, even if that implies creating new “superclasses”.

The consequences of these set of criteria will be discussed further in the following sections.

Naming classes of cell types

Before analyzing the consequences of the criteria raised on part 1, it is important to make a set of naming conventions for different classes of cell types. The conventions are necessary to avoid confusion. Much of the literature mixes cell types in one species (e.g., when dealing with a cell type as an evolutionary unit) and multispecies (e.g., in the cell ontology). Current advances in the taxonomy of living beings are calling artificial the classifications of “genus,” “families,” “order,” and similar rankings. The level of a “species” is better defined, and useful in practice (with discernible theoretical divergences) [18] [19]. Given the importance of the concept species, I derive a “species-centric” view on the naming of classes of cell types. The three classes I propose are:

- Archetypes, for which the scope of the definition is beyond the level of species. For example, “mammal neutrophils.”
- *Stricto sensu* cell types, for which the scope of the definition is precisely one species. For example, “human neutrophils.”
- Infratypes, cell types for which the scope is below the level of species. For example, considering the mouse strain “C57BL/6J” [20], “C57BL/6J neutrophils”.

By adopting a more precise vocabulary, we can flesh out misunderstandings and communicate clearly. At the level of individual scientific experiments, we usually work at the “infratype” level: the samples come only from a subpopulation of the species of interest, and cannot be assumed to be “randomly sampled” from all individuals. This has important practical considerations for, once more, avoiding failing implicitly for the problem of induction.

Moreover, in individual experiments, we not only work with infratypes, we work with very specific infratypes, guided by non-random research setups and pragmatic choices. For example, we might call “CD4 T cells” what are actually CD3+, CD4+, CD8+ cells from the axillary lymph node of 2-month-old chow-fed female C57BL6/J mice from the mouse-house of the Institute of Biochemistry of the University of São Paulo collected in the mornings around 10 pm. Albeit really specific, all the mentioned facets (markers, anatomical location, age, biological sex, strain, housing conditions, circadian clock and diet) are known to alter what we know about cell types. Thus, we benefit from using a name for these very specific classes:

- Technotype: A specific, experimentally defined cell type, which harbors on its definitions the exact conditions from which the cell types were sampled.

Even if really specific, a technotype is still a class. Unless a study used only one single-cell, it likely contained some sampling method, which is the class for which hypotheses are actually tested, for example. This is the most "granular’ cell type in our pragmatic view for research synthesis. This is the type that can be strictly annotated in single-cell RNA-seq analysis, for example.

Single claims are made and tested for technotypes, and this claims can be logically combined in “upper” ontological levels for making claims with a higher degree of universality. This propagation of knowledge to upper levels, however, should not be implicit. (see Yarkoni 2020 for an analogous problem in the psychological sciences [21]). As defended by Popper, knowledge should travel “quasi-

inductionally" by fostering hypothesis with higher degrees of generality, which can then be tested for the more universal class [**???**].

Logical consequences of the definition

One notable logical consequence of the proposed set of criteria is that the definition of a "cell state" is left as a subclass for "cell type". For the pragmatic purpose adopted here, I avoid dissecting dissection of the differences between persistent classes of cells (which I refer to as "traditional cell types") or the transient, fugacious classes of cells (which I refer to as the "traditional cell state"). Even though such a distinction is an important topic for theoretical research, it is not a requirement for representing biomedical experiments.

One example of this entailment is that the class "human cells in metaphase of mitosis" can be considered a cell type, as they can be rigorously defined and are restricted to a taxon. Even though "metaphase" itself is still a biological process, we can describe all cells executing this process as from a single cell type.

However, does a dividing fibroblast stop being a fibroblast, even if temporarily? Again, I do not aim to answer this in a philosophical-ontological sense. Pragmatically, if the rigorous definition used (e.g., expression of a marker) still holds during duplication, this cell can be assigned to two disjunct classes: "fibroblasts" and "doubling cells"! It is, thus, essential to consider that cells can belong to at least two disjunct classes.

Often, cell types are described taxonomically, related in one single hierarchy, a tree [22] [23] [24] . Cells can be assigned to disjunct classes. Thus, it is not possible to annotate cell types with a single identifier using a taxonomic tree, in which each concept is represented by a single node with one (and only one) direct parent node. Cell types need to be represented ontologically (in the computational sense), which can be thought of multiple, intertwining taxonomies, taking into account different ways of classifying cells.

Another logical consequence of the definition is that concepts of "subtype" become redundant with "cell type." A "subtype," then, is a concept that only makes sense when talking about classes with different degrees of universality. Thus, claims to discover new cell "subtypes" or "types" differ only stylistically and can be considered indistinguishable in the perspective of research synthesis. s

Practical consequences of the definition

In the previous section, I discussed the logical entailments of accepting the proposed rules as valid. Here, I will extend the pragmatic considerations on using such a system for real-world applications.

By using the set of rules, we can better evaluate claims of discover of new cell types. With vast amounts of data and loose definition of cell types, it becomes uncannily easy to claim a new cell type. However, suppose one claims to discover a new "stricto sensu" cell type. In that case, one has to provide enough evidence that cells from this class are identifiable across all individuals of a species. A claim of an "archetype" would require evidence for existence in more than one species. Consequently, experiments that only use a specific strain of mice can only claim to discover an infratype.

An example of the discovery of a new "archetype" is the pair of articles published in Nature in 2018 [25] [26] about the newly found "ionocyte", a class of cells in the trachea enriched for expression of genes homologous to the *CFTR* gene. Both studies displayed evidence for such a class in both mouse

and human samples, corroborating the existence of an archetype. This discovery has been denominated by both articles as a discovery of a new cell type.

Another example of cell type discovery is a pioneer article by Villani et al [27]. It describes subclasses of monocytes and dendritic cells in humans, and pragmatically uses markers for their definition. The patients were recruited from “the Boston-based PhenoGenetic project (...) and the Newcastle community.” It is arguable that they did not have a random sample of humanity, and the observed results might not hold for different populations. This discovery of infratypes has also been described as a discovery of a new “cell type”.

Thus, by stating the scope of the cell type discovered and rigorously specifying its characteristics, a claim of discovery can be compared in the light of the evidence, very much like what we have done for centuries for claims of new animal species.

However, a new problem arises. How to name all these specific cell types? How to humanely understand so many “cell types” with such subtle differences?

For accurately, pragmatically classifying cell types from the perspective of research synthesis, we need rigorous definitions. These are very specific in nature, and every single empirical article might include several unique technotypes. This makes nomenclature a nightmare. Which proverbial names should we use to differ between B Cells that were selected by slightly different combinations of markers? I avoid this challenge, focusing on the identification of concepts that are computationally useful.

As described by Sabina Leonelli, the challenges brought up by big data in biology require an advance of our philosophical theories [28]. I argue that the inverse is also true: to advance the theoretical foundations of modern biology, we need to harness the computational tools. The quest for naming cell types becomes simpler when the goal doesn’t require human readability at every step. Instead, by harnessing the computer power to record complex definitions, we can focus on higher level abstractions (not unlike using contact names to access hundreds of phone numbers).

Classes in ontologies can have numeric identifiers. By assigning each technotype a Unique Resource Identifier, a URI, and by inserting the URI in a knowledge graph, such as the Cell Ontology [11] [12] [13] or Wikidata [29] [30], we can start dealing with the complexity of cell types definition across biology. For humans, each article and each dataset could explicitly declare the class it refers by a name (for example, “neutrophil”), avoiding natural language ambiguities while maintaining readability.

The Cell Ontology currently holds less than 4.000 cell types. The number of rigorous and useful cell types, however, is considerably larger. By the rules fo deductive logic, whenever you combine two classes (“neutrophil” +“human”) you give rise to a third class (“human neutrophil”). We need to consider several classes: rigorous descriptors, species, anatomical locations, stages of life, biological sexes, strains, stages in the circadian clock and more. The possible number of cell types is of many order of magnitudes higher than currently available. The scale denotes a logistical challenge, as we would require a more significant number of active collaborators in extension and maintenance of such a knowledge base. One way to progress is open systems such as Wikidata, where life scientists can add their ’cell types" with a low entry barrier. The development of such a system is a direction for future research to operationalize the descriptions here.

The idea of “technotype”, if coupled to the possibility for every researcher to craft their “cell type” of interest, can solve the problem of correctly labeling cell types in single-cell experiments. The non-existence of exact matches in CL (even when combined with other ontologies) renders it theoretically impossible to annotate articles and datasets without incurring in induction issues. As mentioned in the previous section, by having a knowledge connecting the concepts, we would still be able to

compare results from different researchers, but now explicitly stating the level of abstraction in which they can be compared.

A branch of computational single-cell development has dedicated itself to find tools for labeling single-cell experiments. While some approaches avoid using ontologies [22] [31], others aim at anchoring the classes to the Cell Ontology [32] or to MeSH IDs[33] [34] via manual match based on "expert comparison" of labels. These manual matches have been fed to algorithms such as BLAST2CO [34] to predict best "matches" for a single-cell cluster. Even though cell-type prediction is an exciting area of research, its current approach is inherently imprecise. Unless the cells were sampled in the same way across articles, drawn at random from the same population of individuals, they represent strictly different classes, even if very similar.

As mentioned by Sartivijai, Diehl and Yongqun, "Ontology classes should always represent certainty, and not probability."[15]. Considering the "technotype" level for single-study cell types improves that precision. We can change the task from finding a "match" to the cells in a given current experiment to finding a candidate class for a "point of insertion" of the described technotype, even directly (in a parent-child relation to previous class), or by creating a new super class (sibling relation to previous class). Albeit subtle, this extra layer of consideration allows more precise conceptualizations of labeling mechanisms, a theoretical advance that can pave the way for improvements in automatic classification algorithms.

Final remarks

In this article, I proposed a set of 3 rules (rigorous description, taxon scope restriction, and theoretical usefulness) and 1 recommendation (link to an ontology of cell types) to define cell types. I proposed 4 namings to clarify discussions on the topic: archetypes (a class with a scope above species level), *stricto sensu* cell types (a class with scope equal to one species), infratypes (a class with scope below the species level) and technotypes (the exact cell type defined for an experimental setup). The concept of the "technotype" can be harnessed as the unit for classifying cells, in an analogous way of how the "species" is the conventional unit for classifying organisms into higher-order taxa.

I dissected some logical entailments of such definition, which, admittedly, might conflict with current views on defining cell types. I do not aim at solving such conflicts but propose a different way of organizing our knowledge about cell types. Of note,

This article is intended to clarify some of the meanings and provide directions to the future development of the theoretical basis of cell type definition. The discussion on cell types' definition is still on its infancy, and we need human power to tackle the huge theoretical challenges. Biologists, philosophers, and computer scientists ought to distill the details of defining cell types, powering the Human Cell Atlas, and the life sciences research enterprise of this century.

References

1. **The structure of scientific revolutions**
   Thomas S. Kuhn, Ian Hacking
   The University of Chicago Press (2012)
   ISBN: 9780226458113

2. **Neuronal cell types**
   Richard H Masland
   Current Biology (2004-07) https://doi.org/dg84br
   DOI: 10.1016/j.cub.2004.06.035 · PMID: 15242626

3. **Search BioNumbers - The Database of Useful Biological Numbers**
   https://bionumbers.hms.harvard.edu/search.aspx

4. **How Many Types of Cells Are in the Human Body?**
   ibswit
   (2017-05-17) https://askabiologist.asu.edu/questions/human-cell-types

5. **What Is Your Conceptual Definition of "Cell Type" in the Context of a Mature Organism?**
   Cell Systems
   (2017-03) https://doi.org/d38b
   DOI: 10.1016/j.cels.2017.03.006 · PMID: 28334573

6. **The Human Cell Atlas**
   Aviv Regev, Sarah A Teichmann, Eric S Lander, Ido Amit, Christophe Benoist, Ewan Birney, Bernd Bodenmiller, Peter Campbell, Piero Carninci, Menna Clatworthy, … Human Cell Atlas Meeting Participants
   eLife (2017-12-05) https://doi.org/gcnzcv
   DOI: 10.7554/elife.27041 · PMID: 29206104 · PMCID: PMC5762154

7. **What Is Your Conceptual Definition of "Cell Type" in the Context of a Mature Organism?**
   Cell systems
   (2017-03-22) https://www.ncbi.nlm.nih.gov/pubmed/28334573
   DOI: 10.1016/j.cels.2017.03.006 · PMID: 28334573

8. **The evolution of cell types in animals: emerging principles from molecular studies.**
   Detlev Arendt
   Nature reviews. Genetics (2008-11) https://www.ncbi.nlm.nih.gov/pubmed/18927580
   DOI: 10.1038/nrg2416 · PMID: 18927580

9. **The origin and evolution of cell types**
   Detlev Arendt, Jacob M. Musser, Clare V. H. Baker, Aviv Bergman, Connie Cepko, Douglas H. Erwin, Mihaela Pavlicev, Gerhard Schlosser, Stefanie Widder, Manfred D. Laubichler, Günter P. Wagner
   Nature Reviews Genetics (2016-11-07) https://doi.org/f9b62x
   DOI: 10.1038/nrg.2016.127 · PMID: 27818507

10. **Species Concepts and Species Delimitation**
    Kevin De Queiroz
    Systematic Biology (2007-12) https://doi.org/c34kzf
    DOI: 10.1080/10635150701701083 · PMID: 18027281

11. **:{unav)**
Jonathan Bard, Seung Y Rhee, Michael Ashburner
Genome Biology (2005) https://doi.org/dfxc74
DOI: 10.1186/gb-2005-6-2-r21 · PMID: 15693950 · PMCID: PMC551541

12. **Logical Development of the Cell Ontology**
Terrence F Meehan, Anna Maria Masci, Amina Abdulla, Lindsay G Cowell, Judith A Blake, Christopher J Mungall, Alexander D Diehl
BMC Bioinformatics (2011-01-05) https://doi.org/c7kw6x
DOI: 10.1186/1471-2105-12-6 · PMID: 21208450 · PMCID: PMC3024222

13. **The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability**
Alexander D. Diehl, Terrence F. Meehan, Yvonne M. Bradford, Matthew H. Brush, Wasila M. Dahdul, David S. Dougall, Yongqun He, David Osumi-Sutherland, Alan Ruttenberg, Sirarat Sarntivijai, … Christopher J. Mungall
Journal of Biomedical Semantics (2016-07-04) https://doi.org/gg99b9
DOI: 10.1186/s13326-016-0088-7 · PMID: 27377652 · PMCID: PMC4932724

14. **Cells in experimental life sciences - challenges and solution to the rapid evolution of knowledge**
Sirarat Sarntivijai, Alexander D. Diehl, Yongqun He
BMC Bioinformatics (2017-12-21) https://doi.org/gg99b7
DOI: 10.1186/s12859-017-1976-2 · PMID: 29322916 · PMCID: PMC5763506

15. **Cells in ExperimentaL Life Sciences (CELLS-2018): capturing the knowledge of normal and diseased cells with ontologies**
Sirarat Sarntivijai, Yongqun He, Alexander D. Diehl
BMC Bioinformatics (2019-04-25) https://doi.org/gg99b8
DOI: 10.1186/s12859-019-2721-9 · PMID: 31272374 · PMCID: PMC6509796

16. **The meanings of “function” in biology and the problematic case of de novo gene emergence**
Diane Marie Keeling, Patricia Garza, Charisse Michelle Nartey, Anne-Ruxandra Carvunis
eLife (2019-11-01) https://doi.org/ggjnmv
DOI: 10.7554/elife.47014 · PMID: 31674305 · PMCID: PMC6824840

17. **Reporting and connecting cell type names and gating definitions through ontologies**
James A. Overton, Randi Vita, Patrick Dunn, Julie G. Burel, Syed Ahmad Chan Bukhari, Kei-Hoi Cheung, Steven H. Kleinstein, Alexander D. Diehl, Bjoern Peters
BMC Bioinformatics (2019-04-25) https://doi.org/ghbk9r
DOI: 10.1186/s12859-019-2725-5 · PMID: 31272390 · PMCID: PMC6509839

18. ***PhyloCode***
Wikipedia
(2020-07-10) https://en.wikipedia.org/w/index.php?title=PhyloCode&oldid=967070715

19. **PhyloCode: Division I. Principles** http://phylonames.org/code/divisions/1/

20. **000664 - C57BL/6J** https://www.jax.org/strain/000664

21. **The Generalizability Crisis**
Tal Yarkoni
(2019-11-22) https://doi.org/ggdf7h
DOI: 10.31234/osf.io/jqw35

22. **CHETAH: a selective, hierarchical cell type identification method for single-cell RNA sequencing**
Jurrian K de Kanter, Philip Lijnzaad, Tito Candelli, Thanasis Margaritis, Frank CP Holstege
Nucleic Acids Research (2019-09-19) https://doi.org/gg99dp
DOI: 10.1093/nar/gkz543 · PMID: 31226206 · PMCID: PMC6895264

23. **Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain**
Ken Sugino, Erin Clark, Anton Schulmann, Yasuyuki Shima, Lihua Wang, David L Hunt, Bryan M Hooks, Dimitri Tränkner, Jayaram Chandrashekar, Serge Picard, … Sacha B Nelson
eLife (2019-04-12) https://doi.org/ghbc3p
DOI: 10.7554/elife.38619 · PMID: 30977723 · PMCID: PMC6499542

24. **How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology**
John C. Marioni, Detlev Arendt
Annual Review of Cell and Developmental Biology (2017-10-06) https://doi.org/ggb632
DOI: 10.1146/annurev-cellbio-100616-060818 · PMID: 28813177

25. **A revised airway epithelial hierarchy includes CFTR-expressing ionocytes**
Daniel T. Montoro, Adam L. Haber, Moshe Biton, Vladimir Vinarsky, Brian Lin, Susan E. Birket, Feng Yuan, Sijia Chen, Hui Min Leung, Jorge Villoria, … Jayaraj Rajagopal
Nature (2018-08-01) https://doi.org/gdwskh
DOI: 10.1038/s41586-018-0393-7 · PMID: 30069044 · PMCID: PMC6295155

26. **A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte**
Lindsey W. Plasschaert, Rapolas Žilionis, Rayman Choo-Wing, Virginia Savova, Judith Knehr, Guglielmo Roma, Allon M. Klein, Aron B. Jaffe
Nature (2018-08-01) https://doi.org/gdwsj7
DOI: 10.1038/s41586-018-0394-6 · PMID: 30069046 · PMCID: PMC6108322

27. **Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors**
Alexandra-Chloé Villani, Rahul Satija, Gary Reynolds, Siranush Sarkizova, Karthik Shekhar, James Fletcher, Morgane Griesbeck, Andrew Butler, Shiwei Zheng, Suzan Lazo, … Nir Hacohen
Science (2017-04-20) https://doi.org/f94x5t
DOI: 10.1126/science.aah4573 · PMID: 28428369 · PMCID: PMC5775029

28. **The challenges of big data biology**
Sabina Leonelli
eLife (2019-04-05) https://doi.org/gfzw8q
DOI: 10.7554/elife.47381 · PMID: 30950793 · PMCID: PMC6450665

29. **Wikidata as a knowledge graph for the life sciences**
Andra Waagmeester, Gregory Stupp, Sebastian Burgstaller-Muehlbacher, Benjamin M Good, Malachi Griffith, Obi L Griffith, Kristina Hanspers, Henning Hermjakob, Toby S Hudson, Kevin Hybiske, … Andrew I Su
eLife (2020-03-17) https://doi.org/ggqqc6
DOI: 10.7554/elife.52614 · PMID: 32180547 · PMCID: PMC7077981

30. **Wikidata** https://www.wikidata.org/wiki/Wikidata:Main_Page

31. **Probabilistic gene expression signatures identify cell-types from single cell RNA-seq data**
Isabella N. Grabski, Rafael A. Irizarry

bioRxiv (2020-01-23) https://doi.org/gg99dq
DOI: 10.1101/2020.01.05.895441

32. **ontoProc: Ontology interfaces for Bioconductor, with focus on cell type identification**
https://www.bioconductor.org/packages/release/bioc/vignettes/ontoProc/inst/doc/ontoProc.html#conceptual-overview-of-ontology-with-cell-types

33. **CellMeSH: Probabilistic Cell-Type Identification Using Indexed Literature**
Shunfu Mao, Yue Zhang, Georg Seelig, Sreeram Kannan
Cold Spring Harbor Laboratory (2020-05-31) https://doi.org/gg99dr
DOI: 10.1101/2020.05.29.124743

34. **Searching large-scale scRNA-seq databases via unbiased cell embedding with Cell BLAST**
Zhi-Jie Cao, Lin Wei, Shen Lu, De-Chang Yang, Ge Gao
Nature Communications (2020-07-10) https://doi.org/gg4mm3
DOI: 10.1038/s41467-020-17281-7 · PMID: 32651388 · PMCID: PMC7351785