Towards a pragmatic definition of cell type

This manuscript was generated from lubianat/technotype@6d8a10c on December 20, 2020.

Authors

- **Tiago Lubiana** 0000-0003-2473-2313 · lubianat Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil

- **Helder I Nakaya** 0000-0001-5297-9108 Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo, São Paulo, Brazil

Abstract

The concept of cell type is key for modeling biology. Recent technological advances are prompting us to rethink what we understand by cell type and how we classify them. There is currently no consensus for a definition of cell type, which makes it hard to integrate knowledge across life sciences. We propose here that a cell type should represent any class of cell that is (1) explicitly defined; (2) is identifiable within a taxon; and (3) is theoretically useful. We also specified four classes of cell types: *sensu stricto* cell types, archetypes, infratypes, and technotypes. They respectively specify cell type concepts applied to a single species, multiple species, populations below the species level, and particular experiments. The flexible and rigorous framework we propose can base annotation of single-cell omics datasets, and reconcile knowledge about cells across all different domains of science.

Introduction

One of the basic subjects in any undergraduate major in life sciences is histology. The students are required to identify cell types across various tissues and look for color and shape patterns in hematoxylin-eosin stains. Textbooks, like Junqueira’s Basic Histology [1] work as manuals that perpetuate the paradigms (in the Kuhnian sense) [2] of what we know about a few hundred cell types.

Our concept of “cell type” is still based on centuries-old histochemical techniques, such as the Golgi-stains of neurons immortalized by Ramon y Cajal [3]. The histological influence is noticeable even in the names given to cell types, such as "erythrocytes", "eosinophils", "basophils", and "oxyphilic cells of the thyroid". The concepts we use are drawn from studies of microanatomy. This connection with anatomy leads us to think about cell types as anatomical entities as if they are dissectible and fixed in an organism. The limits of resolution perpetuated by the histological-anatomical view may be why attempts to quantify cell types use the scale of "hundreds" of human cell types [4,5].

New techniques have challenged this anatomy based conceptualization. From flow cytometry to patch clamping, to single-cell RNA-seq, we saw a burst of new categories, and novel cell "subtypes" and "families" popped up in the literature. The bursting intensified in the past few years, with the rise of projects to characterize *all* human cell types, like the Human Cell Atlas and HUBMAP [4,5].

The advances in biology require us to find better answers for how to define a cell type. Such a concept might not even have a "true" meaning, in a philosophical-realistic sense. Nevertheless, we can strive to find nominal, pragmatic definitions for the real 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 millions of published scientific articles?

The need for a conceptual advance is being perceived by the community [6,7,8,9,10,11,12], and new perspectives are rising. In an opinion article published in Cell Systems in 2017, the researchers presented their views on the conceptual definition of 'cell type' in the context of a mature organism [6]. Many of the scientists believed that cell functions have a core role in defining cell types, which is a slippery road, as the very meaning of "function" in biology is elusive [13]. The opinions were varied, and no consensus was achieved.

One core line of thought is based on the cell type as an evolutionary unit defined by a Core Regulatory Complex (CoRC) of transcription factors. That definition enables the drawing of parallels, from 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) [14,15]

However, as much as different concepts of species coexist [16], our quest to define cell types may take various forms. 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 second direction, the groundwork of the Cell Ontology [17,18,19] and of the International Workshop on Cells in

Experimental Life Sciences [20,21] are notable. Their contributions base much of the views here and will be discussed in detail throughout the article.

We chose to use the term “cell type” to emphasize the focus on types as classes (or “kinds”) in contrast to real-world objects. The similar term “cell state” is used both to describe classes (e.g. activated T-cell) and real-world observations (e.g. the current state of a specific cell). Other similar notions, such as a “cell set”, “cell population” and “cell cluster” can also reminisce of a specific, countable group of cells, frequently from the same experiment.

The term “cell class” is also used in the literature, and would be a suitable synonym for our notion of cell type, as the main goal here is to refine the human-based theoretical classes. Classes that we can instantiate, i.e. assign to an observation of any real cell, in the same way we assign the class *Homo sapiens* to each and every human. The term “cell identity” has also been suggested for avoiding the cell type/cell state dilemma [22], but the notion of identity is slightly different from the idea of class. We opted to frame our work around the term “cell type” due to its historical usage and familiarity for the life sciences community.

The conceptual quest addressed by this work is one of research synthesis and is summarized in the following question: Which cell type definition can be crafted for rigorously describing biomedical experiments?

Towards that goal, the body of the article is divided into 4 parts. In Part 1, we propose a set of rules that are necessary and sufficient for defining cell types. Part 2 offers a small set of names for differentiating the main classes of cell types. In Part 3, we address the logical consequences of the proposed definitions, while Part 4 is a discussion of the pragmatic challenges envisaged in employing such definitions.

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

Our pragmatic definition of cell type (for eukaryotic, multicellular organisms) consists of 3 + 1 simple rules (Figure 1). A cell type is a class of cells that must be:

1. Explicitly defined
2. Identifiable for a defined taxon
3. Theoretically useful

And that should be:

4. Hierarchically related to other cell types

Figure 1: The set of 3 + 1 rules for defining a cell type.

Here, "must" represents an absolute requirement, whereas "should" suggests that "there may exist valid reasons in particular circumstances to ignore a particular item" (as per RFC 2119 [23]).

For rule 1, we mean that the cell type needs to be followed by a clear definition that would allow rational judgments of whether a singular cell belongs to the type or not. Such definitions should provide necessary and sufficient criteria for classification. An example is a cell type defined by "expression of the proteins CD3 and CD4, but lacking CD8." Even though there is still some ambiguity (see [24,25] for longer discussions), it already states clear and reasonable criteria. The degree of rigorousness cannot be decided a priori, as we still do not have a rigorous framework for representing biological knowledge, but we should strive to make definitions as rigorous as possible. Other examples of what could be explicit definitions are as follows:

- "Big cell" is a class of cells that have a length of more than 50 micrometers on any axis.
- "Human cortical neuron" is a class of cells in human cortex that are capable of producing an action potential.
- "Leukocyte" is a class of cells found in animal blood which are achromatic cells.

The recognition of multiple valid characteristics to define types is not new. The first Cell Ontology article, in 2005, explicitly acknowledged criteria based on function, histology, lineage, and ploidy.[17] These features were combined in the definitions of "species-neutral" cell types[19], arguably useful for integrating databases or for teaching biology. Gradually, we are acknowledging that we might need more specific classes to characterize experimental biology, leading to the definition of species-specific types defined by granular characteristics. [18,26].

Rule 2 is an explicit criterion that must be followed while discussing cell types scientifically; we need to define the taxa for which a given cell type is expected to manifest. The cell type then needs to be discoverable in any individual of the taxon (or taxa) of interest, given the appropriate conditions (e.g., stage of life and biological sex). The set of taxa covered by a cell type is called here a taxonomic scope (or just scope) of the cell type. Note that, as cell types can be defined by function and functions can converge, the taxonomic scope is not restricted to monophyletic taxa (clades). The definition of taxon used here is liberal and applies to any class of organisms that any researcher identifies explicitly as a unit.

Knowing the scope is important to avoid the pitfalls of extrapolation. A recurrent theme is that theories corroborated by mouse experiments are valid for human cell types. Such extrapolation is an instance of the classic problem of induction, which is discussed thoroughly in "The Logic of Scientific Discovery" [2]. The taxonomic scope allows us, researchers, to be clear regarding our claims, and better discern what we claim to be true for a strain, a species or any other class of organisms.

Rule 3 deals with a practical concern. Rigorously, there is an infinite number of explicit definitions that any scientist might come up with. One simple proof of this infinitude is that size-based cell definitions (as for "big cell" above) may alone consider any of the infinite real numbers. Thus, a cell type "bigger than 7.835 micrometers" might fit the first two rules, but will likely fail rule 3. If we, as a research community, want to characterize all human cell types, it is necessary to have a finite number of cell types. Rule 3 could be paraphrased as: a valid cell type is a class of cells that any researcher rationally finds useful for a theoretical perspective of reality. For example, a recent study used single-cell RNA-seq experiments to assign 275,000 Drosophila cells into 200 cell types [27]. Since these 200 cell types were useful for Özel and colleagues when describing the world, they automatically satisfied rule 3.

Rule 4 is a practical extension of the usefulness rule: a cell type has to be hierarchically-related to other cell types for increased usefulness. This means that a definition of a cell class is (for research synthesis concerns) less useful if it cannot be considered a "subclass" of another cell type. For practical concerns, all imaginable mammalian cell types are subclasses of a "eukaryotic cell" (defined as any cell of an eukaryotic organism) and likely can be subclasses of more specific cell types. The rule 4 is presented as a recommendation instead of a requirement as, in practice, it might be an overhead and not strictly necessary for tasks like claiming the discovery of a new cell type.

Ontological organization is important for integrating knowledge across studies. A cell type that is based on its transcriptome is not the same as one based on its electrophysiology. They can, nevertheless, be connected by a superclass that matches either one or the other criterion. For example, the green-OFF bipolar cells of the retina and the Syt2-/NK3R+ cells of the retina are considered to be the same cell type [28].

However, as these features are often measured separately, we have, in fact, two individual classes for which knowledge is produced. These classes, then, can be combined in the superclass “(green-OFF) OR (Syt2-/NK3R+) cells” for the integration of claims across domains. 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 this set of criteria will be discussed further in the following sections.

2. Naming classes of cell types

To facilitate communication among life scientists, we propose a set of naming conventions for different classes of cell types. Much of the literature mix cell types in one species (e.g., when dealing with a cell type as an evolutionary unit) or in multispecies (e.g., in the Cell Ontology). It is useful to distill these different concepts into names. Given the importance of the concept of species in biological classification [29], we derive a species-centric view on the naming of classes of cell types. The four classes (Figure 2) we propose are as follows:

- archetypes, for when the taxonomic scope of the type is beyond the level of species; for example, “mammal neutrophils.”
- *sensu stricto* cell types, for when the taxonomic scope of the type corresponds to a single species; for example, *Mus musculus* neutrophils.”
- infratypes, for when the taxonomic scope is below the level of species; for example, considering the mouse strain “C57BL/6J”, “neutrophils from C57BL/6J mice”.
- technotypes, for specific, experimentally defined cell types that harbor in their definition the precise conditions of the cells sampled; “2-month-old male C57BL/6J, Ly-6G+ CD11b+ M-CSF R- CD244- neutrophils”.

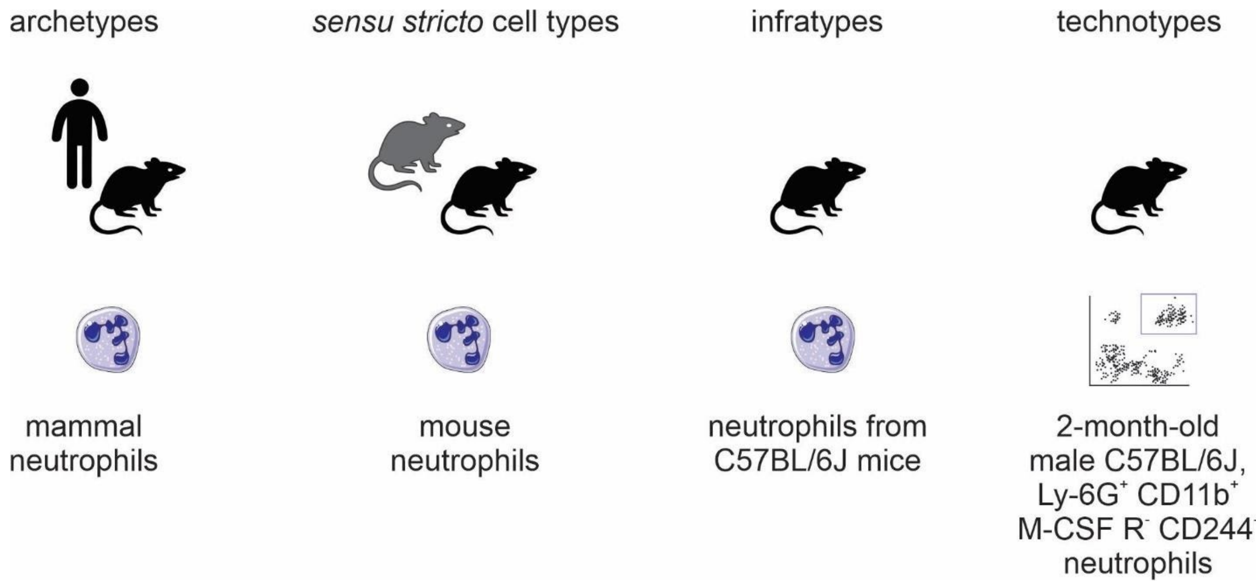

Figure 2: Names for classes of cell types.

By adopting a precise vocabulary, we can avoid misunderstandings and communicate more clearly. At the level of individual scientific experiments, scientists rarely reach the *sensu stricto* cell type level; the samples come only from a subpopulation of the species of interest and cannot be assumed to be randomly sampled from all individuals of the species. This has important practical considerations to, once again, avoid failing implicitly at the problem of induction.

Besides, in individual experiments, we work with cells of very specific classes. They are not only infratypes but very specific infratypes defined by non-random research setups and pragmatic choices. For example, we might call “CD4 T cells” what are 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 on several mornings around 10 pm. Although quite 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 the experimentally-constrained cell classes: technotypes.

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

Single claims are made and tested for technotypes, and the claims can be logically combined in “upper” ontological levels for reaching a higher degree of universality.

The propagation of knowledge to upper levels cannot be implicit (see Yarkoni 2020 for an analogous problem in the psychological sciences [30]). As Popper defends, knowledge should travel “quasi-inductionally” by fostering hypotheses with higher degrees of generality, which can then be tested for the more universal class [2].

3. Logical consequences of the definition of a cell type

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, we avoid the dissection of the differences between persistent classes of cells (which we refer to as traditional cell types) or the transient, fugacious classes of cells (which we refer to as traditional cell states). 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 explicitly defined and restricted to a taxon. Even though “metaphase” itself is a biological process, we can describe all cells executing this process as a singular cell class.

However, does a dividing fibroblast stop being a fibroblast, even if temporarily? Again, we do not aim to answer this in a philosophical-ontological sense. Pragmatically, if the explicit definition used for fibroblast (e.g., expression of a marker) still holds during duplication, this cell can be assigned to two classes that are not hierarchically related: “fibroblasts” and “doubling cells”. If cells can be assigned to multiple classes that are not hierarchically related, 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. This is in conflict with attempts to classify cell-types using single hierarchies in the form of a tree [31] [32] [33].
Cell types need to be represented ontologically with multiple inheritance, which can be thought of as multiple, intertwining trees that take into account different ways of classifying cells (Figure 3).

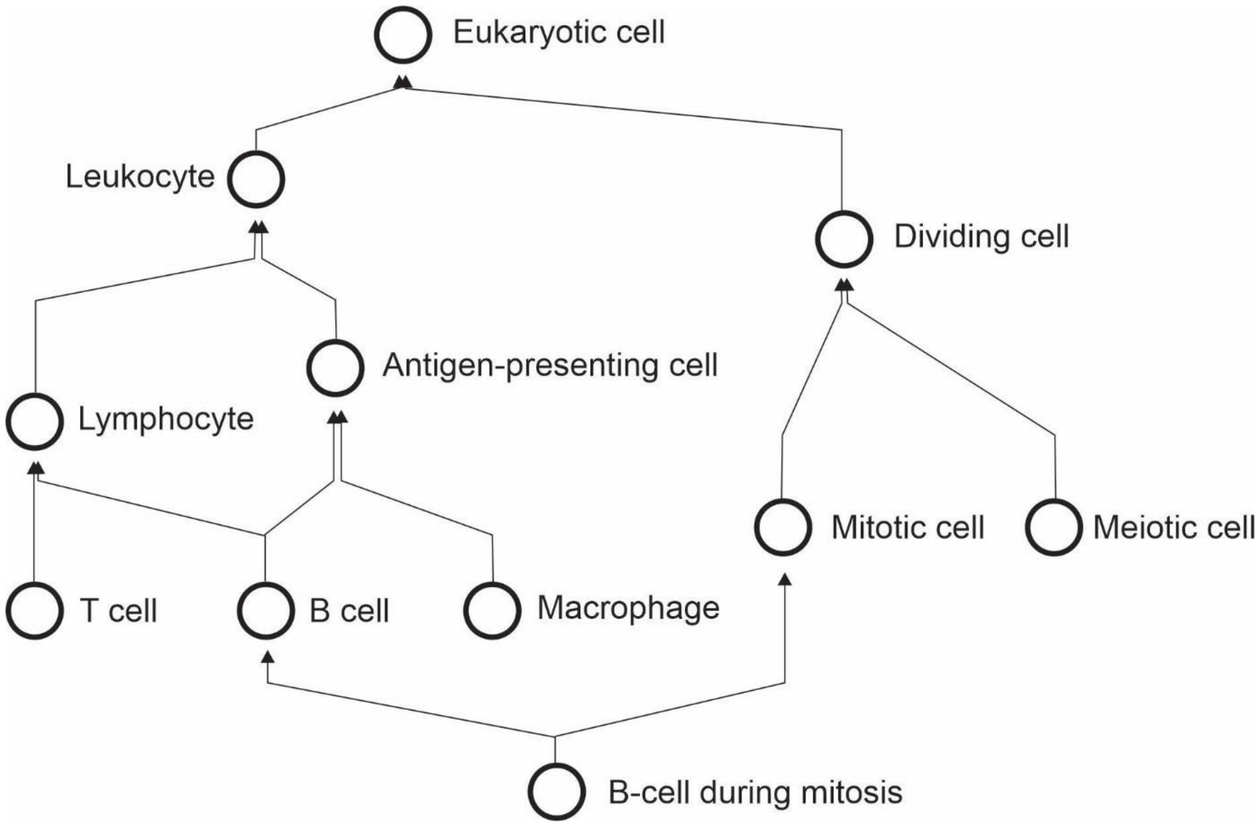

Figure 3: The cell type hierarchy is not a tree - it requires multiple inheritance for completeness.

Another logical consequence of the definition is that the concept of subtype becomes redundant with the concept of cell type. The notion of subtype, then, only makes sense when discussing classes with different degrees of universality. Thus, claims to discovery of new cell “subtypes” or “types” differ only stylistically and can be considered indistinguishable from the perspective of research synthesis.

4. Practical consequences of the definition of a cell type

In the previous section, we discussed the logical entailments of accepting the proposed rules as valid. Here, we extend the pragmatic considerations on using such a system for real-world applications. In a 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 (Cell Ontology)”, and “A specimen source description (anatomic structure þ species)”. [34] Their approach has great merit in defining clear guidelines for marking a cell type. The requirement of markers is reasonable for the field of single-cell RNA-seq, where marker information is abundant. The Cell Ontology has used markers for defining cell types, an approach employed in particular for immune cells [18,24,25].

The use of markers, however, leaves us with a conceptual problem – definitions of cell type used by electrophysiologists, or even in the manuals of histology classes, are not based on markers. Rigorously adopted, this requirement would leave aside an entire segment of what we consider biomedical knowledge. Moreover, gene markers are not defined for cell types that span multiple species, a problem already discussed in the Cell Ontology report of 2011 [18]. Thus, our set of rules was crafted to accommodate the different ways that people classify cells.

In fact, with so many different takes on the field, vast amounts of data, and loose definitions of cell type, it becomes uncannily easy to claim a new cell type. Our set of rules may contribute to formalizing cell type discovery.
If one explicitly claims to have discovered a new *sensu stricto* cell type, one should provide enough evidence that cells from this class are identifiable across all individuals of a species (given the constraints as age and biological sex). A claim of an archetype would require evidence of existence in more than one species. Consequently, experiments that only use a specific strain of mice have a more robust claim if the expectation is limited to the infratype.

An example of the discovery of a new archetype is the pair of articles published in Nature in 2018 [35,36] about the newly found “ionocyte”, a class of cells in the trachea enriched for the 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 of an archetype has been denominated by both articles as a discovery of a new cell type.

Another example of cell type discovery is found in a pioneering article by Villani et al [37]. The authors describe subclasses of monocytes and dendritic cells in humans and pragmatically use markers for their definition. The patients were recruited from “the Boston-based PhenoGenetic project (...) and the Newcastle community.” Arguably, 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 the discovery of a new cell type.

An example from the Villani et al article is the discovery of the “AXL+ SIGLEC6+ AS Dendritic cell”. This and other cell types are presented in the article as part of a “Human dendritic cell atlas”, generalizing the theory for the whole of humanity. The jump from technotype (which takes into consideration also descriptors like “healthy” and “age between 25 and 40 years”) to infratype (“all humans in this population scope”) to cell type *sensu stricto* (all humans) is depicted in Figure 4 and exemplifies the logical flow in our proposed framework.

Of note, “dendritic cells” are one of the cell types most thoroughly modeled by the Cell Ontology. [38,39] The current definition of the dendritic cell (CL_0000451) is coupled to the definition of leukocyte (CL_0000738), which defines it as " An achromatic cell of the myeloid or lymphoid lineages capable of ameboid movement“. This definition is not reconcilable with the “dendritic cells” studied by Villani et al. We have no way of knowing if the cells in their work are achromatic or capable of ameboid movement. That might sound pedantic and might, unfortunately, be so, but the logical requirements of computational systems lead to both biocurators and computers being seen as pedantic. This high level of precision is necessary to accurately depict not only the complexities of cell types but also of research settings.

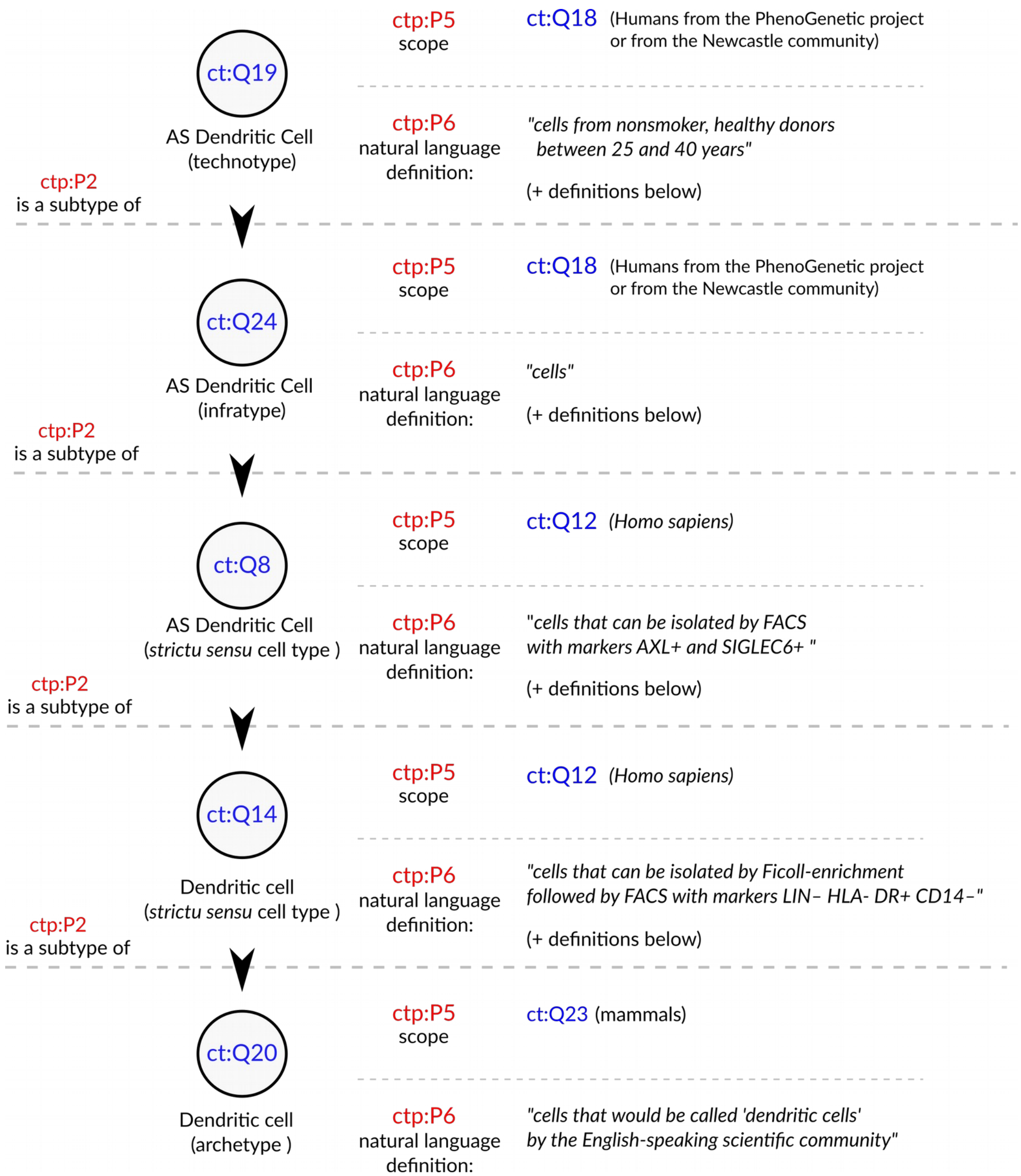

Figure 4: Conceptualization of a set of the cell types in Villani et al, 2017 [37]. The depicted cell types were manually curated from the article, where they are either implicitly or explicitly mentioned. Identifiers for cell types are written in pseudocode based on the Turtle serialization for RDF (https://www.w3.org/TR/turtle/) and represent valid URIs (described in the database https://celltypes.wiki.opencura.com/wiki/Main_Page). URI: Universal Resource Identifier; RDF: Resource Description Framework.

Even if we are not able to represent all the aspects that go into a cell type definition using ontologies, we can use an explicit "natural language definition" property to define cell types. As David Osumi-Sutherland puts in his 2017 article about cell type classification: there is a "*mismatch between quantified logic, which records assertions about all members of a class, and the messy, noisy reality of biology and the data we collect about it.*" [26]. Luckily, do not need to have all the biology formalized before we deal with cell types. Taking the example in Figure 4, all cell types treated as "dendritic cells" in the literature are valid subclasses of the dendritic cell archetype (ct:Q20). Such a subclassing system might lack the power to computationally check the validity of definitions. However, by the principle of minimal commitment [40], it could provide a coherent scaffold for representing experimental data (e.g., from single-cell transcriptomics) and allow logically robust data integration.

The commitment to logical coherence will require us to deal with many more types than we are used to. Given the variety of species on Earth, the complexity of multicellular life, and the diversity of research settings, a count of cell types may far exceed the mark of one million. Sabina Leonelli stated that the challenges thrown up by big data in biology require the advancement of our philosophical theories [41]. We agree and argue that the converse is also true: to advance the theoretical foundations of modern biology, we need to harness the power of computational tools. Computational ontologies provide a solution for dealing with complex concepts. Classes in ontologies can have alpha-numeric identifiers. We can, thus, assign each technotype a Unique Resource Identifier, a URI, similar to the Cell Ontology (CL)[17,18,19] or in the knowledge graph of Wikidata [42]. The power of using knowledge graphs for integrating knowledge about cell types is gaining momentum [11], and they rely heavily on the precise usage of unique identifiers.

The classification of cells into types and the naming of cell types are parallel tasks. While there has been progress on rules for naming cell types (particularly in neuroscience [43,44,45]), nomenclature is outside the scope of this article. Using identifiers/URIs without semantic sense already suffices for our purposes. Semantically void identifiers also help us to steer away from the Aristotelian essentialist view upon cell types, as discussed by Rowe and Stone in 1977 [46]. Identifiers can have labels that can be freely changed, while keeping a persistent URI. Our effort to refine the logical aspects of cell-type definitions can be combined with any commonly agreed naming/labeling system.

The URIs at the level of technotype allow precise labeling of cell types in real-world experiments. The technotype annotation empowers researchers to craft their cell type of interest, and connect this cell type to a common network of knowledge. Several single-cell transcriptomics tools try to assign labels to cells. While some approaches avoid ontologies [31,47], others utilize the Cell Ontology [48,49,50] or MeSH IDs[49,51] to identify the most likely cell type label for each cell or cell cluster. Different studies, however, almost always study different technotypes. Thus, the task of finding the exact type of cells in a given experiment, algorithms could try to find where the new technotypes should be inserted in an ontological network. For example, instead of claiming that cells from study A and study B are myeloid dendritic cells, we can claim that both cell types belong to the myeloid dendritic cell branch. By embracing these real differences between studies (and cells), the precise metadata of the study will enable a precise statement of the cell type. This will ultimately allow the coherent reuse of publicly available data.

Final remarks

In this article, we have proposed a set of three rules (rigorous description, taxon scope restriction, and theoretical usefulness) and one recommendation (hierarchical linking) to be followed when defining cell types. We have also proposed four types of naming to clarify discussions on the topic: archetypes (a class with a scope above species level), *sensu stricto* 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 a manner analogous to how the “species” is the conventional unit for classifying organisms into higher-order taxa. We have dissected some logical entailments of such definition, which admittedly might conflict with current views on defining cell types. We do not aim to solve such conflicts or negate the other perspectives but only to propose a unique way of organizing our knowledge on cell types. This article clarifies some of the meanings and provides directions for the future development of the theoretical basis of a cell type definition. The discussion on cell types’ definition is still in its infancy, and we need human power to tackle these 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.

Acknowledgements

We would like to express our gratitude and acknowledge the researches that dedicated time specifically to help us to discuss and refine the basis for developing concepts here presented. Nominally, we thank Kleber Neves, Gabriel Lovate, Cesar Prada, Diógenes Saulo Lima, Lucas Cardozo, Juliane Fernandes, Pedro Medeiros, Érika Molina, Antonio Pedro Vieira, João Vitor Cavalcante, Maria Fernanda Forni, Diorge Souza, Jean Bezerra, Gabriel Sato, Roberta Andrejew and Dimitrius Pramio. Part of this work was supported by grant #2019/26284-1, São Paulo Research Foundation (FAPESP).

References

1. **Junqueiras Basic Histology: Text And Atlas**
Luiz Carlos Uchôa Junqueira, Anthony L. Mescher
McGraw-Hill Education (2018-01-01) https://www.wikidata.org/wiki/Q102104590
ISBN: 9781260026184

2. **Logik der Forschung**
Karl Popper
(1934-01-01) https://www.wikidata.org/wiki/Q1868040
ISBN: 9783161484100

3. **Histologie du système nerveux de l'homme & des vertébrés.**
Santiago Ramón y Cajal
(1909-01-01) https://www.wikidata.org/wiki/Q51488921

4. **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

5. **The human body at cellular resolution: the NIH Human Biomolecular Atlas Program**
HuBMAP Consortium
Nature (2019-10-09) https://doi.org/gf92ht
DOI: 10.1038/s41586-019-1629-x · PMID: 31597973 · PMCID: PMC6800388

6. **What Is Your Conceptual Definition of "Cell Type" in the Context of a Mature Organism?**
Paul Blainey, Hans Clevers, Cole Trapnell, Ed Lein, Emma Lundberg, Alfonso Martinez Arias, Joshua R. Sanes, Jay Shendure, James Eberwine, Junhyong Kim, … Mathias Uhlén
Cell systems (2017-03-01) https://www.wikidata.org/wiki/Q87649649

7. **A periodic table of cell types**
Bo Xia, Itai Yanai
Development (2019-06-15) https://doi.org/ggctwf
DOI: 10.1242/dev.169854 · PMID: 31249003 · PMCID: PMC6602355

8. **Exciting times to study the identity and evolution of cell types**
Maria Sachkova, Pawel Burkhardt

Development (2019-09-19) https://doi.org/ghdb9v
DOI: 10.1242/dev.178996 · PMID: 31537583

9. **The Human Cell Atlas: from vision to reality.**
Orit Rozenblatt-Rosen, Michael J. T. Stubbington, Aviv Regev, Sarah Teichmann
Nature (2017-10-01) https://www.wikidata.org/wiki/Q47565008
DOI: 10.1038/550451a

10. **Human Cell Atlas and cell-type authentication for regenerative medicine**
Yulia Panina, Peter Karagiannis, Andreas Kurtz, Glyn N. Stacey, Wataru Fujibuchi
Experimental and Molecular Medicine (2020-09-15) https://www.wikidata.org/wiki/Q99418657
DOI: 10.1038/s12276-020-0421-1

11. **A community-based transcriptomics classification and nomenclature of neocortical cell types**
Rafael Yuste, Michael J. Hawrylycz, Nadia Aalling, Argel Aguilar-Valles, Detlev Arendt, Rubén Armañanzas, Giorgio A. Ascoli, Concha Bielza, Vahid Bokharaie, Tobias B. Bergmann, … Ed S. Lein
Nature Neuroscience (2020-08-24) https://www.wikidata.org/wiki/Q98665291
DOI: 10.1038/s41593-020-0685-8

12. **The evolving concept of cell identity in the single cell era**
Samantha A. Morris
Development (2019-06-27) https://www.wikidata.org/wiki/Q93086971
DOI: 10.1242/dev.169748

13. **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

14. **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

15. **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

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

17. **An ontology for cell types**
Jonathan Bard, Seung Y. Rhee, Michael Ashburner
Genome Biology (2005-01-01) https://www.wikidata.org/wiki/Q21184168
DOI: 10.1186/gb-2005-6-2-r21

18. **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

19. **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

20. **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

21. **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

22. **Current best practices in single-cell RNA-seq analysis: a tutorial**
Malte D. Luecken, Fabian J. Theis
Molecular Systems Biology (2019-06-19) https://www.wikidata.org/wiki/Q64974172
DOI: 10.15252/msb.20188746

23. **Key words for use in RFCs to Indicate Requirement Levels**
Scott Bradner
(1997-01-01) https://www.wikidata.org/wiki/Q104060055

24. **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

25. **flowCL: ontology-based cell population labelling in flow cytometry**
Mélanie Courtot, Justin Meskas, Alexander D. Diehl, Radina Droumeva, Raphael Gottardo, Adrin Jalali, Mohammad Jafar Taghiyar, Holden T. Maecker, J. Philip McCoy, Alan Ruttenberg, … Ryan R. Brinkman
Bioinformatics (2015-04-15) https://doi.org/f7cc46
DOI: 10.1093/bioinformatics/btu807 · PMID: 25481008 · PMCID: PMC4393520

26. **Cell ontology in an age of data-driven cell classification**
David Osumi-Sutherland
BMC Bioinformatics (2017-12-21) https://doi.org/ghcbdk
DOI: 10.1186/s12859-017-1980-6 · PMID: 29322914 · PMCID: PMC5763290

27. **Neuronal diversity and convergence in a visual system developmental atlas**
Mehmet Neset Özel, Félix Simon, Shadi Jafari, Isabel Holguera, Yen-Chung Chen, Najate Benhra, Rana Naja El-Danaf, Katarina Kapuralin, Jennifer Amy Malin, Nikolaos Konstantinides, Claude Desplan
Nature (2020-11-04) https://www.wikidata.org/wiki/Q101226729
DOI: 10.1038/s41586-020-2879-3

28. **The neuronal organization of the retina.**
Richard H. Masland
Neuron (2012-10-17) https://www.wikidata.org/wiki/Q34307217
DOI: 10.1016/j.neuron.2012.10.002

29. **International Code of Phylogenetic Nomenclature (PhyloCode)**
Philip D. Cantino, Kevin de Queiroz
(2020) http://phylonames.org/code/

30. **The generalizability crisis**
Tal Yarkoni
PsyArXiv (2019-11-22) https://psyarxiv.com/jqw35

31. **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

32. **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

33. **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

34. **Cell type discovery using single-cell transcriptomics: implications for ontological representation.**
Brian D Aevermann, Mark Novotny, Trygve Bakken, Jeremy A Miller, Alexander D Diehl, David Osumi-Sutherland, Roger S Lasken, Ed S Lein, Richard H Scheuermann
Human molecular genetics (2018-05-01) https://www.ncbi.nlm.nih.gov/pubmed/29590361
DOI: 10.1093/hmg/ddy100 · PMID: 29590361 · PMCID: PMC5946857

35. **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

36. **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

37. **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

38. **An improved ontological representation of dendritic cells as a paradigm for all cell types**
Anna Masci, Cecilia N Arighi, Alexander D Diehl, Anne E Lieberman, Chris Mungall, Richard H Scheuermann, Barry Smith, Lindsay G Cowell
BMC Bioinformatics (2009) https://doi.org/cpxdhs
DOI: 10.1186/1471-2105-10-70 · PMID: 19243617 · PMCID: PMC2662812

39. **Hematopoietic cell types: Prototype for a revised cell ontology**
Alexander D. Diehl, Alison Deckhut Augustine, Judith A. Blake, Lindsay G. Cowell, Elizabeth S. Gold, Timothy A. Gondré-Lewis, Anna Maria Masci, Terrence F. Meehan, Penelope A. Morel, Anastasia Nijnik, … Christopher J. Mungall
Journal of Biomedical Informatics (2011-02) https://doi.org/c6dmmh
DOI: 10.1016/j.jbi.2010.01.006 · PMID: 20123131 · PMCID: PMC2892030

40. **Toward principles for the design of ontologies used for knowledge sharing?**
Thomas R. Gruber
International Journal of Human-Computer Studies (1995-11-01)
https://www.wikidata.org/wiki/Q47303277
DOI: 10.1006/ijhc.1995.1081

41. **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

42. **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

43. **Common Cell type Nomenclature for the mammalian brain: A systematic, extensible convention**
Jeremy A. Miller, Nathan W. Gouwens, Bosiljka Tasic, Forrest Collman, Cindy T. J. van Velthoven, Trygve E. Bakken, Michael J. Hawrylycz, Hongkui Zeng, Ed S. Lein, Amy

Bernard
arXiv (2020-11-13) https://www.wikidata.org/wiki/Q104247451

44. **Neuron Names: A Gene- and Property-Based Name Format, With Special Reference to Cortical Neurons**
Gordon M. Shepherd, Luis Marenco, Michael L. Hines, Michele Migliore, Robert A. McDougal, Nicholas T. Carnevale, Adam J. H. Newton, Monique Surles-Zeigler, Giorgio A. Ascoli
Frontiers in Neuroanatomy (2019-01-01) https://www.wikidata.org/wiki/Q64065346
DOI: 10.3389/fnana.2019.00025

45. **Name-calling in the hippocampus (and beyond): coming to terms with neuron types and properties.**
D. J. Hamilton, D. W. Wheeler, C. M. White, C. L. Rees, A. O. Komendantov, M. Bergamino, Giorgio A. Ascoli
Brain informatics (2016-06-09) https://www.wikidata.org/wiki/Q37656899
DOI: 10.1007/s40708-016-0053-3

46. **Naming of neurones. Classification and naming of cat retinal ganglion cells.**
Rowe MH, Stone J
Brain, Behavior and Evolution (1977-01-01) https://www.wikidata.org/wiki/Q41052480
DOI: 10.1159/000125660

47. **Probabilistic gene expression signatures identify cell-types from single cell RNA-seq data**
Isabella N. Grabski, Rafael A. Irizarry
bioRxiv (2020-01-23) https://www.wikidata.org/wiki/Q104371272
DOI: 10.1101/2020.01.05.895441

48. **ontoProc: processing of ontologies of anatomy, cell lines, and so on**
Vince Carey
Bioconductor (2020)
https://www.bioconductor.org/packages/release/bioc/html/ontoProc.html

49. **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

50. **Unifying single-cell annotations based on the Cell Ontology**
Sheng Wang, Angela Oliveira Pisco, Aaron McGeever, Maria Brbic, Marinka Zitnik,

Spyros Darmanis, Jure Leskovec, Jim Karkanias, Russ B. Altman
(2019-10-20) https://www.wikidata.org/wiki/Q104057222

51. **CellMeSH: Probabilistic Cell-Type Identification Using Indexed Literature**
Shunfu Mao, Yue Zhang, Georg Seelig, Sreeram Kannan
(2020-05-31) https://www.wikidata.org/wiki/Q104371393