On the Formal Representation and Annotation of Cellular Genealogies.

Patryk Burek[1] and Nico Scherf [2,3] and Heinrich Herre[4]

[1] Institute of Computer Science, Faculty of Mathematics, Physics and Computer Science, Marii Curie-Sklodowskiej University, Lublin, Poland,
[2] Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, School of Medicine, TU Dresden, Dresden, Germany,
[3] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany,
[4] Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany
patryk.burek@poczta.umcs.lublin.pl
nico.scherf@tu-dresden.de
heinrich.herre@imise.uni-leipzig.de

Abstract. Time-lapse microscopy is a primary experimental tool for biologists to study development: the dynamic process by which an entire organism forms from an individual cell. The domain of these cellular dynamics is quite complex, and thus, demands a conceptual and computational architecture to support the integration of knowledge obtained across experiments and theories. In previous work, we have addressed the conceptual level and developed an axiomatic theory of cellular genealogies. In this work, we will address the other fundamental part of theory formation: the experimental level, where we have to deal with actual observations and discoveries. In the case of experiments from time-lapse microscopy, we need to go from the individual images taken at discrete time points to a full conceptual description of the underlying continuous cellular processes. In this work, we take a first step to bridge the general theory T(CO) and the experimental level by investigating individual cases. Any time-lapse experiment is linked to a real spatiotemporal genealogy, and we assume that these entities are particular instances of the general theory. We will investigate how this individual experimental information can be organised and represented.

Keywords: Knowledge management, Ontology of biological reality, Theories of Developmental Biology, Microscopy, Time-lapse imaging, Cell tracking.

1 Introduction

Cellular dynamics and interactions shape multicellular life as it develops from a single fertilised egg into a complex organism. These (inter-)cellular processes also maintain the structure and function of the organism during its lifetime. To fully understand the principles underlying this self-organising process, we have to be able to observe and analyse the cellular dynamics and cellular states from experiments [1]. One milestone of time-lapse experiments has undoubtedly been the reconstruction of the embryonic lineage tree of the nematode *Caenorhabditis elegans* [2]. From these roots, modern fluorescence microscopy has turned into a powerful tool to resolve the dynamics of thousands of cells together with readouts of cellular states by fluorescent labels [1, 3] across a wide range of biological questions from developmental biology to stem cell biology and oncology. But imaging and visualizing cellular dynamics is only one part of the problem, we also have to extract and quantify the resulting cellular dynamics. Beyond simple experiments with only a few cells, manual analysis of cell tracking experiments is mostly infeasible. Consequently, a variety of methods have been developed to computationally track individual cells in time-lapse movies [4, 5]. Beyond computational tracking of cells there waits another challenge, however: How can we formalise and extract knowledge from the automated (or manual) tracking results? Here, we need to develop and refine concepts and theories to make sense of the patterns we observe [6]. The first step is to establish standard data formats that serve as the core to annotate and share the tracking results [7]. We should base these annotations on a solid theoretical foundation and carefully develop the underlying terminology and formal concepts themselves as theories about the biological world [8]. We have recently [9] made a first step into this direction and developed the essential parts of a conceptual architecture that supports integration and interoperability of cell tracking experiments. This framework is based on the *Cellular Genealogy* as a fundamental notion for the development of a *Cell Tracking Ontology*. Some core components and patterns of which have already been presented in [10, 11]. In this work, we will now explore the experimental level of theory formation, where we have to deal with actual observations. Both aspects need to be addressed when developing an empirical theory about an area of reality. In the case of time-lapse experiments, we need to go from individual images taken at discrete time points by a microscope to a full, conceptual description of the underlying continuous cellular processes. Here, we take a step to bridge the general theory T(CO) and the experimental level by investigating individual cases. We will examine how different experimental information can be organised and represented.

2 Towards a formal theory of cellular genealogies

Developmental biology is the science that investigates how a variety of interacting processes (at the molecular, cellular and tissue level) generate the various shapes, size, and structural features that arise throughout the life cycles of multicellular organisms. This field also encompasses the biology of regeneration, metamorphosis, and the growth and the differentiation of stem cells in the adult organism and is thus intimately linked with stem cell biology and basic research in regenerative medicine and oncology.

We would like to note one fundamental problem here: there is no clear consensus on how to define the boundary between the animate and inanimate. Typical defining properties of life are, among others, metabolism, adaptivity and interaction with the environment, self-organisation, reproduction, heredity, and growth. These conditions define a system which must satisfy at least the following basic properties. It should have a boundary, demarcating the system from the environment, and it should have inner parts.

It should further be able to sense and interact with the environment[1]. In biology, the cell is the simplest system satisfying these assumptions. Thus, in our view, the self-organised development of a cellular genealogy, starting from a zygote, seems to be an essential feature of the animate.

Hence, the ontology of biology should consider the existence of cellular genealogies as one of the essential features demarcating biology from other fields of natural science, as physics or chemistry. The cellular genealogy of an animal is determined by the whole developmental process of this animal from the initial zygote to the multicellular organism that is focused on the cell level. At any time-point of an animal's life, a collection of cells are present. During development, these cell collectives permanently change, by e.g. cell division, cell differentiation and cell death. Hence, a cellular genealogy is a process which is determined by the development of an animal's cell collectives. Cellular genealogies possess a certain structure which can be specified by using the notion of a cell collective and cell situation, and the process connecting them. There are various important structural parameters of a cellular genealogy. How many cells exist in a complete genealogy of an animal? What can be said about the sequence of the maximal time-intervals during which there is no change of the corresponding cell-collectives? As a next step, these cell collectives can be extended by adding relations, such as the morphology of the single cells, or cell groups, or their localisation in space. Cell collectives extended by additional relations are called cell situations. Since cells may divide and eventually die the number of cells within a region under consideration (e.g. a developing organism) changes through time. Let us consider a time-segment (time-interval) *I*, such that during *I* no cell-division and no cell death occurs. Then, the cells existing during *I* form a collective Cells(*I*) that can be considered as a continuant through *I*.

In [9], we introduced the notion of Cell-Collective-Genealogy (denoted by CollGen), and Cell-Situation-Genealogy (SitGen). The lifetime of an organism is assumed to be a closed time interval. We assume that the time is presented by time-points and time-intervals, whereas the time-points have the order-type of the real numbers. Let us consider a time-segment (time-interval) I such that during I no cell-division and no cell death occurs; then we call the set of cells associated with this interval a cell collective. During times when the number of cells changes, new cells may occur, and cells may disappear (i.e. die). We consider the life of an organism Org from fertilisation to death. Org starts as a single cell, the zygote, develops into a multicellular structure through time collectives of cells, lives in a dynamic equilibrium and finally dies, i.e. the dynamic, functional structures dissolve. We divide the lifetime T of Org into a sequence of non-overlapping time-intervals I(1), ... I(n) such that the following conditions are satisfied:

(1) The intervals I(m) have a first point (they are left-closed), but no last point (right open). More precisely, they have the form [a(m), a(m+1)) specifying the

[1] Cf. *Autopoiesis* as an attempt to define living matter using concepts from general systems theory such as self-organisation.

set {c : a(m) ≤ c < a(m+1)}, where 0 ≤ m ≤ n. Further, LifeT(Org) = ⋃{ I(m) : 0 ≤ m ≤ n}.

(2) Let Cells(I(k)) be the set of cells existing during I(k), then no cell death or division occurs during the interval I(k), k≺ n. Further, we assume that Cells(I(k)) ≠ Cells (I(k+1)).

These conditions imply further properties: From Cells(I(k)) to Cells(I(k+1)) the number of existing cells changes. We consider two types: cell division and cell death. If a division of a cell c ∈ I(k) occurs then this process ends up with two daughter cells starting their existence at the left boundary of the interval I(k+1). Analogously, if a cell undergoes cell death during I(k) then this ends at the left-boundary of I(k+1). The final definition of CollGen(start) then must specify which cells from Cells(k) are related to which cells in Cells(k+1). To this end, we introduce two relations: div(x,y,z): a cell x of Cells(k) undergoes a cell division during I(k) resulting in two daughter cells y and z starting their existence at the left-boundary of I(k+1). We also introduce the relation id(x,y) stating that x belongs to Cells(k) and y belongs to Cells(k+1) and both cells are identical. We further say that a cell x in Cells(k) has a successor cell y in Cells(k+1), if y is either a daughter cell of z or if y is identical with x, denoted by succ(x,y). The cell collective genealogy CellGen(c(0)), is then specified by the following system CollGen(c(0)) = ({Cells(k) | 0 ≤ k ≤ n}, div(x,y,z), id(x,y)). We call the intervals I(k) invariance intervals because during these intervals no cell-change occurs. The structure of such a cell-collective genealogy is an important, uniquely determined feature of the organism. The following theorem we postulate without proof.[2]

Theorem. For any organism Org there exists a uniquely determined cell-collective genealogy associated with Org.

As outlined, for every cell collective x there is a uniquely determined time-interval I such that no changes occur during I. This time-interval is called the *invariance interval* of the cell-collective; it has a left-boundary and no right-boundary.

A cell situation genealogy is an extension of the cell collection genealogy: We start with the system CollGen(c(0)) and extend any collective of cells(k) into an object-situation Sit(k). Sit(k) contains exactly the cells of Cells(k) as objects, and it is embedded into an object-situation with the timeframe I(k) and a specified spaceframe. The collective Cells(k) then spans a certain space, which contains at least the spatial convex closure of the objects in Cells(k). We must specify the situation type determined by a signature Σ. A situation-genealogy is based on signature-extension of a cell collective genealogy, i.e., to the signature Σ(0) further symbols from a signature Σ(1) are added; we assume that Σ (0) ∩ Σ(1) = ∅. For every signature Σ(1) we may introduce a model-structure that models the corresponding cell-situation genealogy, called SitGen = (CollGen, int(Σ(1)).

[2] Because of limitation of space for the current paper, the proof is presented [9].

Based on the signature Σ(0) we created an initial theory about cell-collective genealogies, denoted by T(0)(CG), and presented in [9]. The further development and refinement of this theory is a future research field of its own. Here, we would like to concentrate on the experimental level of theory formation.

3 The experimental framework

Humans access the independent reality by various components and levels of their cognitive systems. The immediate interaction between the subject and reality, subject and object, is realised through the senses by the process of perception. These sense-data are organised, clustered, and then concepts and relations between the data are formed. We call this basic information, acquired by the subject, phenomena, and data. A theory about a domain should formulate certain conditions that explain the domain's phenomena. The higher levels of cognition use principles of causation to establish a theory about a part of reality.

A theory T consists of propositions which are postulated to be true in D. An experiment is a mediator between a theory and the real domain under consideration. We want to get data about the CG which are not captured by the given theory T(CG). What can be said about the types of the involved cells, and about the structure of the cellular genealogies of concrete species? The Gene Ontology (GO) provides many features about the cells which are not yet considered in the current theory TG. However, all this information is needed to extend the initial theory T(CG)(0) so we can get a complete picture of the behavioural dynamics of cells. Time-lapse experiments are one important source of such information.

These real-world genealogies are analysed by cell tracking experiments. Such experiments yield snapshots by a microscope M of a continuously developing cell situation genealogy SitGen. Related to SitGen the microscope M generates a finite sequence of images that correspond to presentic situations, determined by SitGen. These images are called frames, and the resulting finite ordered set of frames is called the frame sequence of the experiment. An experiment of this type establishes a relation between SitGen, the microscope M, and the frame sequence FSeq, denoted by Exp(SitGen, M, FSeq), whereas M serves as a mediator[3] between the original entity SitGen and the output FSeq in the form of a finite sequence of images. The frame-sequence provides important information about the evolving cell situation genealogy. The snapshot of a situation is an independent ontological entity, which is classified in the framework of GFO as a material presentic object, or simply as a material presential. With this assumed preconditions, a formal description of the relation Exp(SitGen,FSeq) is useful, because it provides a deeper understanding of the relation between the reality of SitGen and the data, generated by M, and provides a frame sequence FSeq, briefly denoted by FSeq(SitGen). The interaction between cellular genealogies and frame sequences are described by axioms. Here, we only sketch the basic ideas; further details are presented in [9].

[3] The development of such mediators (imaging techniques) play an important role in the advancement of science and its applications in general. A significant example is magnetic resonance imaging (MRI).

FSeq(x) ≔ x is a frame-sequence, and its components are called frames. Every frame is a snapshot of a situation, denoted by PSit. We introduce a linear ordering between the components of a frame-sequence, hence such a sequence can be presented by the structure FSeq = ({F(1), …, F(n)}, <), where F(1) < … < F(n). Let FSeq be a frame sequence, we say that a component G is a successor of the component F, if F < G and there is no frame between F and G; we say that G is subsequent to F.
We assume that in any frame there occur cells, that these cells are presentials, and any such presentic cell is a snapshot of a uniquely determined cell (with lifetime> 0). We introduce relations such as assoc(F,t): „the frame F is associated with the time point t" (F is a snapshot at time-point t) and component(x,y): x is a component of the frame y, distance(a,b,r): the presentic cell a and the presentic cell b have distance r.
For every frame sequence FSeq there exists a cell collective genealogy CollGen such that any component of FSeq is a snapshot of a cell collective in CollGen.

4 From Frames to the Representation of Cellular Genealogies

The data acquired by the experiment are taken from snapshots which are presented in the frame sequences. Hence, we use the frame-sequence and some basic knowledge about the sequences' structure. For the tracking of single cells (as individual instances) we must introduce constants c(1),..., c(n), denoting these (presentic) cells. These constants are associated with the different frames, F(1),…, F(n) being snapshots at certain time-points, say t(1),...,t(n). Since we are not sure, whether a cell **a** in frame F(i) is the same as the cell **b** in F(i+1) (the same for daughter cells and divisions etc.,), we are forced - in the first step – to consider the presentic cells for any frame separately. For this purpose, we may associate to any constant c a timestamp, say expressed by c@t (the presentic cell occurs in the frame F(t)). For the construction of a representation of the genealogy, we need to know whether some of the following conditions hold: id(a@i, b@(i+1)), or div(a@i, b@(i+1), c@(i+1)), Death(c@i) (and other relations according to the situation). To answer these questions background knowledge using existing ontologies, the concepts of which can be applied to annotate the frames and their parts. Another important method could be machine learning, and – of course – other methods of artificial intelligence. Symbolic artificial intelligence can be used to abstract temporal patterns from temporalised data (i.e. data of the form c@t).

We further distinguish atomic from complex data. Atomic data have the form of atomic sentences, for example, id(a@i, b@(i+1)) (with the meaning: a@i and b@(i+1) are snapshots of a cell c). Complex data are particular combinations of atomic data, for example we may consider id(a@i, b@(i+1)) /and id(b@(i+1), c@(i+2)) which says that the cells a@i, b@(i+1), c@(i+2) are equivalent, hence present the same cell.
Individual data can be annotated by additional information, taken from existing bio-ontologies. For example, the complex datum [id(a@i, b@(i+1)) and id(b@(i+1), c@(i+2))] can be annotated by a cell-type T.

Let us consider a frame F(i) = (PSit, c(1),...,c(m), r(1),...r(n)), and F(i+1) a successor frame of the sequence. F(i), F(i+1) reflect snapshots of certain situations S, S' of the genealogy. Are F(i) and F(i+1) snapshots of the same situations, or from different situations? One may either assume that both are from the same situation, or that F(i) and F(i+1) are taken from succeeding situations:

(1) F(i) and F(i+1) are from the same situation Si(k), i.e. no cell division or cell death occurs, and the cells in F(i) and F(i+1) are snapshots of the same cell. We then need to know for which c' in F(i) and c'' in F(i+1) there is a cell c in Sit (being an object), such that c' and c'' are snapshots of the same cell c. Furthermore, a cell division might occur during Sit. We need to know whether a certain cell in F(i) and F(i+1), identified as the same cell, is in the process of cell division, as might be deduced from the shape of the cell (e.g. its nuclear structures) or by some additional signal such as fluorescent markers of specific proteins [12]. The process of identifying cells across snapshots is called *cell tracking* and typically uses either engineered features or Machine Learning to establish a set of rules when the identified c' and c'' are "similar enough" to be considered snapshots of the same cell [4, 13, 14].

(2) F(i) is from situation Sit(j) and F(i+1) is from situation Sit(j), hence S(j+1) is the successor situation of S(j). In this case, there is a change of cells from S(j) to S(j+1). Then, we need to know how the cells in F(i) relate to the cells in F(i+1): Which cells in F(i) have no successor in F(i+1)? Which cells c in F(i) give rise to daughter cells c', c'' in F(i+1)? Which cells c' in F(i) and c'' in F(i+1) are snapshots of the same cell?

As mentioned above, we assume those assignments between observed cells in F(i) and F(i+1) have been estimated either using computational or manual cell tracking. The information from (1) and (2) can then be used to construct a formal representation out of the experiment. In constructing a representation of an individual genealogy, we must introduce constants in the language; every cell that we detect in a frame is denoted by a constant. The number of constants may change as new cells may occur (after cell division). When using FOL as a representation language, we thus add atomic sentences to the specification. For example, if c' and c'' are constants and we know that c' and c'' are daughter cells of c, we add to following sentences to the representation: daughter(c, c'), daughter(c, c''), div(a, b, c) etc. Analogously, we may add id(c, d), or (not exists x such that successor(c, x)), or Dead(c). We can also represent this information about the constants as a knowledge graph using a graph-theoretical representation. We summarise some of the representational formalisms using an example. FOL is the most expressive formalism. We distinguish (in a generalisation of similar notions of DL (description logic)), the FO-TBox, FO-Abox, and FO-extABox. An FO-Tbox (first-order TBox) contains those formulas with variables and quantifiers. FO-Abox (first-order ABox) contains only atomic sentences (i.e. no variables, no quantifiers), FO-extABox (extended first-order ABox) contains variable-free propositions composed of atomic sentences/propositions and propositional connectives. Let us consider a specific example to demonstrate FO-TBox, FO-Abox, and FO-extABox: $T = \{\forall x \forall y \forall z\, (\mathrm{div}(x,y,z) \rightarrow \mathrm{daughter}(y,x) \wedge \mathrm{daughter}\ (z,x) \wedge y \neq z)\}$ belongs to FO-TBox.

The atomic sentences {div(a,b,c), daughter(b,a), daughter(c,a)} belong to FO-ABox. The FO-ABox must be consistent with the FO-TBox. Hence, the FO-ABox must not contain the atomic sentence b = c. The FO-extABox could also contain sentences like c ≠ b. These propositions can then be formalised within DL, OWL using FOL/CL. As a result of the analysis of the frame-sequences, we get a set of atomic sentences. Our strategy is to develop a system FO-TBox and FO-ABox and transform this representation into DL and OWL.

We ultimately want to provide a solution that enables interoperability among cell tracking experiments. There is not yet a widely-accepted standard for storing, annotating and exchanging cell tracking results and the tools used in the domain usually come with their own ad hoc formats. However important first attempts have been made to define a standard data format [7]. Furthermore, there are already several ontologies available and organised within the Open Biological and Biomedical Ontology (OBO) Foundry [15]. Among those there are many ontologies that are relevant to cell tracking experiments. We are particularly interested in ontologies that describe (1) experiments, such as the Ontology for Biomedical Investigations (OBI) [16], (2) cells, e.g. the Cell Ontology [17] or (3) cell characteristics and behaviours, such as the Phenotype and Trait Ontology [18] or the Cell Behavior Ontology (CBO) [19]. Therefore, a straightforward approach (illustrated in Fig.1) is the annotation of raw data using one or several of these relevant ontologies. Raw data usually contains only presentic information obtained from FSeq(x) such as frames, presential cells and presential situations and therefore the proposed solution serves well for querying raw data on presentic entities, e.g. return all frames containing cells of a given type or shape. However, the existing frameworks would not support more advanced queries which go beyond presentic entities, e.g. one cannot query for cellular genealogies in which all cells of a certain sub-lineage died, or for a subpopulation of stem cells which gave rise to a certain pattern of differentiated cells. This is the consequence of the missing formalisation in the basic data structures that are typically used. Our work aims to close this gap. Therefore, on top of the solution presented so far, we also need to facilitate cross-experiment and cross-systems queries as well as data exploration that is not limited to presentic entities. In Figure 1, we propose an architecture built on the core formalism of cellular genealogies. It consists of n cell tracking systems, each supporting their own format for representing the results of the experiments. The raw data from each system is then translated into the OWL-Abox by means of the Cell Tracking Annotation (CTA) Tool, e.g. [20]. The CTA will provide interfaces for specific raw data formats and will support the automatic translation of raw data into an interoperable format based on OWL-TBox of the Cell Tracking Ontology (CTO). That way, the information about the presentic entities detected in the images, such as presentic cells, their characteristics or presentic cell collections can be correctly represented in the ontology. Next, the CTA automatically augments the presentic information by means of frame-sequences axioms. Basing on an OWL-ABox containing information on presentic entities and their sequences, CTA allows reconstruction of (1) time extended entities such as cells (objects) together with their characteristics, as well as (2) intercellular processes, such as cell divisions and finally, (3) complex structures such as cellular genealogies. Similarly, these reconstructed entities can themselves be further annotated with the help of the biomedical ontologies integrated into CTO. The whole knowledge graph extracted from the reconstructed entities is then added to the original OWL-ABox, which then can be used as a source for cross-system services such as, e.g. cross-system querying.

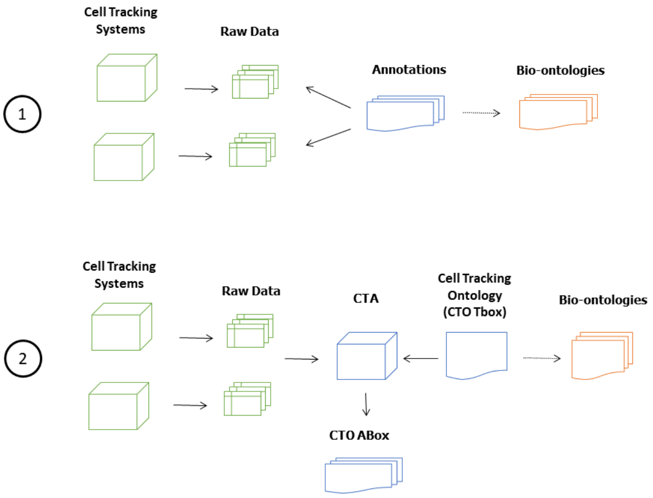

Fig. 1. (1) presents a straightforward architecture for introducing interoperability in cell tracking domain by annotation with existing bio-ontologies. (2) depicts the architecture based on the Cell Tracking Ontology and the Cell Tracking Annotation Tool which supports the transformation of raw data into CTO ABox which increases the possibilities of information retrieval on cellular genealogies.

5 Conclusions

As a continuation of the work presented in [9], we present here a generic framework for specifying a basic relation between empirical theories and the corresponding experiments as mediators between the theory and the world of individual entities. An essential component is the symbolic presentation of the data, acquired by experiments from real-world entities. We applied this framework to the domain D(CG) of cellular genealogies. The symbolic reconstruction and representation of cellular genealogies from data, acquired by experiments, uses techniques of information technology, including various forms of data representation as formal logics, description logics, and implemented languages like OWL. We argue that such a broad framework is needed as it provides the components and modules to achieve the overall aim that can be summarised by the following conditions:

1. Extraction and interpretation of biological data from systems-level experiments, and

support of the interoperability between and across different types of observations at the single-cell level (e.g. time-lapse microscopy and single-cell sequencing).

2. Integration of data and knowledge that should lead to new forms of organisation of biological knowledge.
3. Supporting and augmenting the scientific progress by the use of techniques of machine learning and symbolic artificial intelligence.

We hope that our approach and framework paves the way for further research topics in these directions.

References

1. Wallingford, J.B.: The 200-year effort to see the embryo. Science. 365, 758–759 (2019).
2. Schnabel, R., Hutter, H., Moerman, D., Schnabel, H.: Assessing Normal Embryogenesis in Caenorhabditis elegans Using a 4D Microscope: Variability of Development and Regional Specification. Dev. Biol. 184, 234–265 (1997).
3. Megason, S.G., Fraser, S.E.: Imaging in systems biology. Cell. 130, 784–795 (2007).
4. Ulman, V., Maška, M., Magnusson, K.E.G., Ronneberger, O., Haubold, C., Harder, N., Matula, P., Matula, P., Svoboda, D., Radojevic, M., Smal, I., Rohr, K., Jaldén, J., Blau, H.M., Dzyubachyk, O., Lelieveldt, B., Xiao, P., Li, Y., Cho, S.-Y., Dufour, A.C., Olivo-Marin, J.-C., Reyes-Aldasoro, C.C., Solis-Lemus, J.A., Bensch, R., Brox, T., Stegmaier, J., Mikut, R., Wolf, S., Hamprecht, F.A., Esteves, T., Quelhas, P., Demirel, Ö., Malmström, L., Jug, F., Tomancak, P., Meijering, E., Muñoz-Barrutia, A., Kozubek, M., Ortiz-de-Solorzano, C.: An objective comparison of cell-tracking algorithms. Nat. Methods. 14, 1141–1152 (2017).
5. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., Van Valen, D.: Deep learning for cellular image analysis. Nat. Methods. (2019). https://doi.org/10.1038/s41592-019-0403-1.
6. Wellmann, J.: Model and movement: studying cell movement in early morphogenesis, 1900 to the present. Hist. Philos. Life Sci. 40, 59 (2018).
7. Gonzalez-Beltran, A.N., Masuzzo, P., Ampe, C., Bakker, G.-J., Besson, S., Eibl, R.H., Friedl, P., Gunzer, M., Kittisopikul, M., Le Dévédec, S.E., Leo, S., Moore, J., Paran, Y., Prilusky, J., Rocca-Serra, P., Roudot, P., Schuster, M., Sergeant, G., Strömblad, S., Swedlow, J.R., van Erp, M., Van Troys, M., Zaritsky, A., Sansone, S.-A., Martens, L.: Community Standards for Open Cell Migration Data, https://www.biorxiv.org/content/10.1101/803064v1, (2019). https://doi.org/10.1101/803064.
8. Leonelli, S.: The challenges of big data biology. Elife. 8, (2019). https://doi.org/10.7554/eLife.47381.
9. Burek, P., Scherf, N., Herre, H.: On the Ontological Foundations of Cellular Development, https://www.biorxiv.org/content/10.1101/2020.05.30.124875v1, (2020). https://doi.org/10.1101/2020.05.30.124875.
10. Burek, P., Scherf, N., Herre, H.: A pattern-based approach to a cell tracking ontology. Procedia Comput. Sci. 159, 784–793 (2019).
11. Burek, P., Scherf, N., Herre, H.: Ontology patterns for the representation of quality changes of cells in time. J. Biomed. Semantics. 10, 16 (2019).
12. Zerjatke, T., Gak, I.A., Kirova, D., Fuhrmann, M., Daniel, K., Gonciarz, M., Müller, D., Glauche, I., Mansfeld, J.: Quantitative Cell Cycle Analysis Based on an Endogenous All-in-One Reporter for Cell Tracking and Classification. Cell Rep. 19, 1953–1966 (2017).

13. Moen, E., Borba, E., Miller, G., Schwartz, M., Bannon, D., Koe, N., Camplisson, I., Kyme, D., Pavelchek, C., Price, T., Kudo, T., Pao, E., Graf, W., Van Valen, D.: Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning, https://www.biorxiv.org/content/10.1101/803205v2, (2019). https://doi.org/10.1101/803205.
14. Kwok, R.: Deep learning powers a motion-tracking revolution. Nature. 574, 137–138 (2019).
15. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat. Biotechnol. 25, 1251–1255 (2007).
16. Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M.H., Bug, B., Chibucos, M.C., Clancy, K., Courtot, M., Derom, D., Dumontier, M., Fan, L., Fostel, J., Fragoso, G., Gibson, F., Gonzalez-Beltran, A., Haendel, M.A., He, Y., Heiskanen, M., Hernandez-Boussard, T., Jensen, M., Lin, Y., Lister, A.L., Lord, P., Malone, J., Manduchi, E., McGee, M., Morrison, N., Overton, J.A., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Schober, D., Smith, B., Soldatova, L.N., Stoeckert, C.J., Jr, Taylor, C.F., Torniai, C., Turner, J.A., Vita, R., Whetzel, P.L., Zheng, J.: The Ontology for Biomedical Investigations. PLoS One. 11, e0154556 (2016).
17. Diehl, A.D., Meehan, T.F., Bradford, Y.M., Brush, M.H., Dahdul, W.M., Dougall, D.S., He, Y., Osumi-Sutherland, D., Ruttenberg, A., Sarntivijai, S., Van Slyke, C.E., Vasilevsky, N.A., Haendel, M.A., Blake, J.A., Mungall, C.J.: The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. J. Biomed. Semantics. 7, 44 (2016).
18. Gkoutos, G.V., Schofield, P.N., Hoehndorf, R.: The anatomy of phenotype ontologies: principles, properties and applications. Brief. Bioinform. 19, 1008–1021 (2018).
19. Sluka, J.P., Shirinifard, A., Swat, M., Cosmanescu, A., Heiland, R.W., Glazier, J.A.: The cell behavior ontology: describing the intrinsic biological behaviors of real and model cells seen as active agents. Bioinformatics. 30, 2367–2374 (2014).
20. Wagner, S., Thierbach, K., Zerjatke, T., Glauche, I., Roeder, I., Scherf, N.: TraCurate: efficiently curating cell tracks, https://www.biorxiv.org/content/10.1101/2020.02.14.936740v1, (2020). https://doi.org/10.1101/2020.02.14.936740.