

# SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

## Classification of GABAergic interneurons by leading neuroscientists

Bojan Mihaljević<sup>1\*</sup>, Ruth Benavides-Piccione<sup>2</sup>, Concha Bielza<sup>1</sup>, Pedro Larrañaga<sup>1</sup> & Javier DeFelipe<sup>2</sup>

There is currently no unique catalog of cortical GABAergic interneuron types. In 2013, we asked 48 leading neuroscientists to classify 320 interneurons by inspecting images of their morphology. That study was the first to quantify the degree of agreement among neuroscientists in morphology-based interneuron classification, showing high agreement for the chandelier and Martinotti types, yet low agreement for most of the remaining types considered. Here we present the dataset containing the classification choices by the neuroscientists according to interneuron type as well as to five prominent morphological features. These data can be used as crisp or soft training labels for learning supervised machine learning interneuron classifiers, while further analyses can try to pinpoint anatomical characteristics that make an interneuron especially difficult or especially easy to classify.

### Background & Summary

There is currently no unique catalog of cortical GABAergic interneuron types<sup>1</sup>. Forming such a catalog is a major goal in neuroscience and is currently pursued by, among others, the Human Brain Project, the Allen Institute and the BRAIN initiative<sup>2,3</sup>. While high-throughput data generation may enable a fully data-driven classification of interneurons in near future, by clustering<sup>4,5</sup> molecular, morphological, and electrophysiological features, researchers currently use established morphological types such as chandelier, Martinotti, neurogliaform, and basket<sup>6–10</sup>.

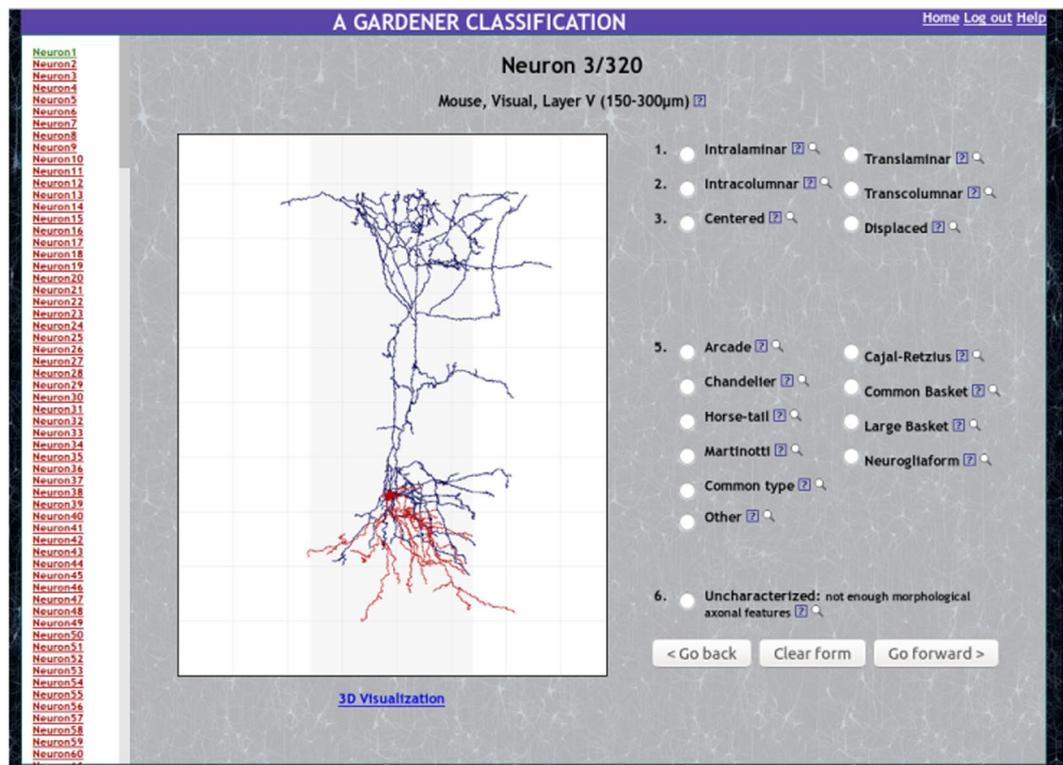
In 2013, we asked 48 leading neuroscientists to classify 320 interneurons by inspecting 2D and 3D images of their morphology (ref.<sup>7</sup> see Fig. 1). This landmark study was the first to quantify the degree of agreement among neuroscientists in morphology-based interneuron classification, showing high agreement for the chandelier and Martinotti types, yet low for most of the remaining types considered such as, for example, the large basket type. In addition to interneuron type, the neuroscientists also classified the cells according to prominent morphological features, such as whether an axon was intra- or trans-laminar.

In this report, we present the data collected by<sup>7</sup>, namely the labeling choices made by the 48 neuroscientists. We also provide the input that the neuroscientists had when classifying the cells: the 2D morphology images that they looked at, cell metadata they were shown, and the definitions of interneuron types and morphological features of interest. For 241 of the cells we provide their morphology reconstructions as well as their Neuromorpho.org<sup>11</sup> ids, so that one can obtain additional metadata. We report a posteriori data curation, such as identifying ten cells that were shown to the annotators rotated upside-down. We also provide an R package with utility functions for analyzing the data.

Besides enabling one to reproduce the study by<sup>7</sup>, these data allow for further analyses. They have been used to assign class labels for supervised<sup>7,12,13</sup> and semi-supervised<sup>14</sup> classification of interneurons, to cluster neuroscientists according to their classification choices<sup>15</sup>, to quantify neuroscientists' accuracy when identifying Martinotti cells from morphology images and use it as a baseline for assessing supervised classifiers<sup>16</sup>, as well as to contrast the classification choices of these 48 neuroscientists to those from a particular research group<sup>16</sup>.

Combining the here provided classification choices can give a crisp or soft (i.e., probabilistic) estimate of the type of these 320 neurons, insofar as the type can be accurately determined from an image of the morphology along with basic metadata. Since the classification choices come from many leading neuroscientists, these

<sup>1</sup>Computational Intelligence Group, Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid, Boadilla del Monte, 28660, Spain. <sup>2</sup>Laboratorio Cajal de Circuitos Corticales, Universidad Politécnica de Madrid and Instituto Cajal (CSIC), Pozuelo de Alarcón, 28223, Spain. \*email: [bmihaljevic@fi.upm.es](mailto:bmihaljevic@fi.upm.es)



**Fig. 1** The web application used to gather the neuroscientists' classification choices for cortical GABAergic interneurons.

combined estimates are objective, i.e., they represent a consensus among experts from different laboratories. Assessing and accounting for the accuracy of the annotators (e.g.<sup>17–19</sup>) might give better estimates of the type than if assuming that they are equally accurate, as we have done in our previous work. Further analyses may consider per-species or laminar differences in inter-neuroscientist agreement, or can try to pinpoint characteristics that distinguish interneurons that are especially difficult to classify from those that clearly belong to a given type. For example, while there was little inter-neuroscientist agreement on the large basket type, some cells were clearly members of this type as they were labeled as large basket cells by a majority of the neuroscientists.

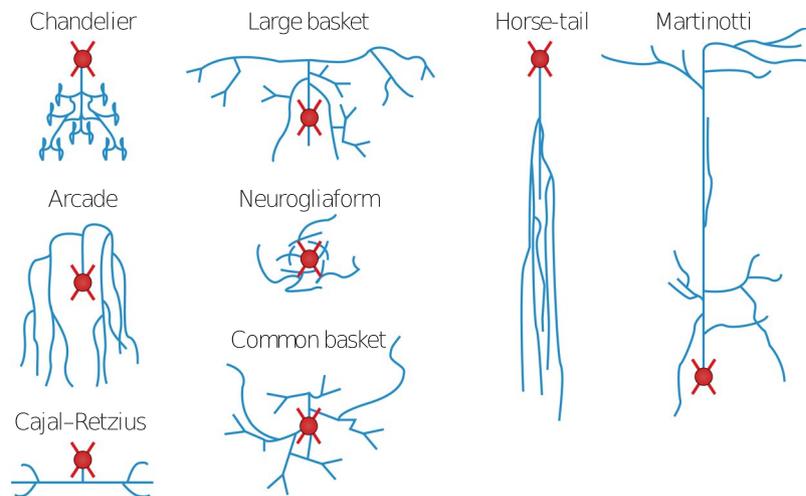
## Methods

All data were collected by<sup>7</sup> and data acquisition is described in their paper. Here we provide a self-contained description of the neurons, the classification experiment, and curation, so that the data can be used by referring to this publication only.

**Classification web application.** Each neuroscientist used the web application shown in Fig. 1 to classify interneurons. In addition to 2D images, which were available for all interneurons, 3D visualization was provided for 241 of the cells, allowing the neuroscientists to rotate and zoom the morphologies. The cell's brain area, cortical layer and estimated layer thickness were stated when available, as well as the species of the animal. A help page provided definitions of neuronal types and categories. The web application that the neuroscientists used to classify the cells can be accessed at <http://cajalbbp.es/gardenerclassification/>. Throughout the paper, we use the term 'annotators' to refer to the 48 neuroscientists that participated in the study, *annotating* (i.e., classifying) the selected cells.

**Interneuron selection.** Reference<sup>7</sup> asked the neuroscientists to classify 320 cortical GABAergic interneurons. The authors downloaded 241 of these cells from Neuromorpho.org<sup>11</sup>, and obtained pictures of the remaining 79 interneurons by scanning images from scientific publications (our dataset includes the 2D images of all 320 cells; see below). The cells come from different cortical areas and layers of the mouse, rat, rabbit, cat, monkey and human. The authors obtained cell metadata from Neuromorpho.org and the scanned papers. The layer of the soma was unknown for 30 cells from Neuromorpho.org.

**Classification scheme.** Reference<sup>7</sup> proposed a classification scheme based mainly on patterns of axonal arborization. The scheme contemplates ten interneuron types (see Fig. 2): arcade, Cajal-Retzius, chandelier, common basket, common type, horse-tail, large basket, Martinotti, neurogliaform, and other. Other is meant to be



**Fig. 2** Interneuron types in the gardener's scheme. Figure from<sup>7</sup>. Reprinted by permission from Nature Reviews Neuroscience.

chosen when the neuroscientist finds none of the remaining nine types adequate and prefers to use an alternative name. Full definitions of the types are provided in the data (see below).

In addition to interneuron type, the classification scheme contemplates five high-level morphological features, such as whether or not the axon is restricted to the layer that contains the soma. These features, termed F1, F2, F3, F4, and F6 (F5 is the previously discussed interneuron type) have the following categories: (F1) intralaminar and translaminar; (F2) intracolumnar and transcolumar; (F3) centered and displaced; (F4) ascending, descending, and both; (F6) characterized and uncharacterized. The uncharacterized category of F6 means that a cell's reconstruction is not good enough to reliably classify it. When labeling a cell as uncharacterized in feature F6, the neuroscientist cannot annotate it according to any of the remaining five features, F1-F5. F4 is only applicable for cells that are labeled as translaminar and displaced in F1 and F3, respectively. Full definitions of features F1-F6 are provided in the data.

**Annotation.** A total of 48 neuroscientists participated in the study. 42 of them fully classified all 320 neurons, thus providing  $42 \times 320 \times 6 = 80,640$  labels. Six neuroscientists classified a subset of the 320 cells (150, on average), with four of them failing to assign labels to all of F1-F6 for some cells, thus providing a total of another 4,452 labels.

**A posteriori curation.** We found, by comparing 97 of our Neuromorpho.org reconstructions to reconstructions that we got directly from the original laboratory, that ten cells were rotated upside-down at Neuromorpho.org and were thus displayed as such to the neuroscientists. We report which cells were shown with their morphologies upside-down (see below).

We provide the additional metadata that we obtained from Neuromorpho.org: the original cell type (the 'Secondary Cell Class' attribute at Neuromorpho.org), the reconstructing laboratory ('Archive'), and the name of a reference article. We also provide morphology reconstruction files that we downloaded from Neuromorpho.org for 241 cells.

### Data Records

All data are hosted at figshare<sup>20</sup>. The neuroscientists' annotations are stored in `annotations.csv` (see example in Table 1). The numbering for annotators 1 to 42 follows that in<sup>7</sup>—e.g., the ids can be matched to those used, for example, in their Fig. 19 of Ref.<sup>7</sup>—while ids larger than 42 correspond to neuroscientists that labeled less than 320 interneurons and were thus not considered in<sup>7</sup>. The neurons also maintain the ids, ranging from 1 to 320, used in the original classification web application and by<sup>7</sup>. The 'complete' column in `annotations.csv` indicates where the neuron was completely or partially labeled by a particular neuroscientist. In all files, 'None' means that an entry was not applicable to a given interneuron (e.g., F4 for annotator 1, cell 1, in Table 1), while an empty entry (e.g., F2 for annotator 45, cell 1, in Table 1) denotes a missing value.

Definitions for alternative interneuron type names provided by the neuroscientists, which they used in column 'other' of Table 1, are stored in `alternative-types.csv` (see Table 2). There are a total of 269 alternative type names, of which 251 are unique (241 are unique if we remove the '?' symbol from type names). 163 types have a name but no definition (e.g., see annotator 14 in Table 2) while one type, assigned to 38 neurons by a single neuroscientist, also lacks a name (last row in Table 2). The `annotators.csv` file lists the names and affiliations in 2013 of the 48 neuroscientists that participated in the study. The names are given in an alphabetic order, unrelated to the annotator ids used in `annotations.csv`.

Basic cell metadata, with corrected brain area information, is in file `metadata.csv` (see Table 3). Column 'neuromorpho.name' corresponds to the 'Neuron name' attribute at Neuromorpho.org. Cells that were rotated upside-down are marked with TRUE in the 'rotated' column, while the original type, as reported at Neuromorpho.org

Annotator	Neuron	F1	F2	F3	F4	F5	F6	Other
1	1	intralaminar	intracolumnar	centered	None	neurogliaform	characterized	None
16	1	intralaminar	intracolumnar	centered	None	neurogliaform	characterized	None
43	203	translaminar	intracolumnar	centered	None	common basket	characterized	None
1	79	translaminar	intracolumnar	displaced	both	other	characterized	columnar basket
16	79	translaminar	intracolumnar	displaced	ascending	Martinotti	characterized	None
43	281	translaminar	intracolumnar	displaced	ascending	horse-tail	characterized	None
1	80	None	None	None	None	None	uncharacterized	None
16	80	intralaminar	intracolumnar	centered	None	common type	characterized	None
43	282	translaminar	intracolumnar	displaced	descending	horse-tail	characterized	None

**Table 1.** Annotation of interneurons 1, 79 and 80 by neuroscientists 1, 16, and 45.

Annotator	Type	Definition
1	columnar basket	This term is not new, I believe. In my view, these cells have a pattern of...
4	bitufted	I made a mistake. The bitufted cell should be classified as bipolar horsetail...
4	ascending horsetail	Cells with horsetail-shaped ascending axons...
4	bipolar horsetail	Cells with horsetail-shaped ascending and descending axons...
7	deep layer inhibitor	neuron with an axonal domain that targets preferentially deep cortical layer...
7	double bouquet	neuron with an axonal domain that targets both deep and superficial layers...
14	narrow arbor cell	
18	bitufted?	see above...
23	bitufted	
43		

**Table 2.** Examples of alternative type names and definitions provided by the neuroscientists. An alternative type is uniquely defined by the annotator id along with its name.

Neuron	Neuromorpho.name	Species	Area	Layer	Rotated	Original.type
1	None	Monkey	Visual	IV	FALSE	
2	2001-11-09-B-L23-dendax	Rat	Somatosensory	II/III	FALSE	Not reported
3	020801-2-ST	Mouse	Visual	V	FALSE	Somatostatin
						containing cell
6	None	Monkey	Visual	IV	FALSE	
29	C170998D-I4	Rat	Somatosensory	II/III	TRUE	Basket cell
35	C170897A-I1	Rat	Somatosensory	IV	TRUE	Basket cell

**Table 3.** Partial (eight out of nine columns in metadata.csv) metadata for six cells.

org, is provided in column 'original.type'. We provide the name of the original paper when the paper is reported at Neuromorpho.org (not shown in Table 3).

The `neuromorpho-swcs` folder contains morphology reconstructions for 241 interneurons that we downloaded from Neuromorpho.org in August, 2019. The reconstructions correspond to standardized morphology files from Neuromorpho.org, encoded in the SWC format. Each reconstruction filename is given by the 'neuromorpho.name' of the neuron followed by 'CNG.swc' (e.g., C170998D-I4.CNG.swc for neuron with 'neuromorpho.name' C170998D-I4). The `urls` file contains the Neuromorpho.org URLs that we downloaded the reconstructions from, so users can easily re-download them. While unlikely, some reconstructions may be affected by future curation efforts at Neuromorpho.org. We thus recommend users to consider re-downloading the reconstructions from Neuromorpho.org instead of using the reconstructions provided in `neuromorpho-swcs`.

The `annotation-input` folder contains input given to the neuroscientists during the classification experiment: the 2D images of neurons (in the `images` subfolder), exact definitions of the types and features (`instructions.pdf`), cells' metadata as it was shown to them in `metadata.csv`, and estimates of cortical layer thickness per species and brain area (`areamap.txt`).

### Technical Validation

Most of the neuronal reconstructions were thorough and were therefore considered as characterizable by the neuroscientists. Shortcomings of the study include the upside-down display of ten cells, unknown or unreported layer of the soma for 30 cells, and a mistakenly reported brain area for 43 cells.

The quality of the classification labels stems from the fact that they come from a diverse set of 48 leading neuroscientists. We did not control for labeling mistakes, as one could do by, for example, showing an interneuron twice to an annotator and comparing the two assigned labels. As a surrogate measure of consistency, we compared the interneuron type labels by four neuroscientists to their own classification of the same cells in their previous papers. The four neuroscientists were annotators that were co-authors of papers associated with the neurons that we obtained from Neuromorpho.org. We restricted our attention to cells whose original label (i.e., the ‘Secondary Cell Class’ attribute at Neuromorpho.org) was either chandelier, basket, Martinotti, or neurogliaform. We disregarded cells that were shown rotated upside down, thus obtaining 104 label pairs to compare. When matching the provided labels with the original ones, we considered arcade, common basket and large basket as matching the Neuromorpho.org basket type; we include the arcade type here because it is often regarded as an alternative name for the nest basket type (see, e.g., Table 1 in<sup>8</sup>).

The labels that the four neuroscientists assigned in our study matched their original labels 86% of the time (in 91 out of 104 cells). We consider that this number indicates high labeling consistency by these four annotators, given the differences between our experimental setting and the setting in which they provided the original labels. For example, a major difference is that, when performing the original classification, authors are likely to have had access to additional data such as physiological or molecular characteristics of the cell, or low-level morphological features such as the distribution of boutons.

Note that even imperfect classification by many annotators can provide very accurate results when combining their output. Namely, statistics tells us that that combining the output of many diverse predictors that are individually better than random guessing tends to produce very accurate predictions<sup>21</sup>. Thus, as long as our group of annotators is diverse enough, and they are individually better than random guessing—which is to be expected given their expertise—combining their labels by simple majority will tend to provide good labels. More sophisticated combining methods<sup>17–19</sup> might identify and account for any systematic mistakes among annotators (e.g., some might be more familiar with certain cell types than with others).

### Usage Notes

The data can be downloaded from figshare<sup>20</sup>. Since the data consist of comma-separated values and plain text files, they can be easily handled with standard data analysis software. We also provide the `gardenr` R package with a utility functions to access the data and examples of analyses. The package can be installed from Github with `devtools::install_github('ComputationalIntelligenceGroup/gardenr')`. It loads the data into R and shows how to combine data from three csv files—namely, on annotations, metadata and alternative types data—to perform meaningful analyses. Some utility functions for combining data are provided, while advanced filtering and summarization are easy to perform with the `tidyverse` packages.

### Code availability

The code of the web application (Fig. 1) that<sup>7</sup> used to collect the neuroscientists’ inputs is not publicly available.

Received: 16 July 2019; Accepted: 20 September 2019;

Published online: 22 October 2019

### References

1. Ascoli, G. A. *et al.* Petilla terminology: Nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Reviews Neuroscience* **9**, 557–568 (2008).
2. Huang, Z. J. & Luo, L. It takes the world to understand the brain. *Science* **350**, 42–44 (2015).
3. Grillner, S. *et al.* Worldwide initiatives to advance brain research. *Nature Neuroscience* **19**, 1118–1122 (2016).
4. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience* **19**, 335–346 (2016).
5. Gouwens, N. W. *et al.* Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nature Neuroscience* **22**, 1182–1195 (2019).
6. Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience* **5**, 793–807 (2004).
7. DeFelipe, J. *et al.* New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nature Reviews Neuroscience* **14**, 202–216 (2013).
8. Markram, H. *et al.* Reconstruction and simulation of neocortical microcircuitry. *Cell* **163**, 456–492 (2015).
9. Tremblay, R., Lee, S. & Rudy, B. GABAergic interneurons in the neocortex: From cellular properties to circuits. *Neuron* **91**, 260–292 (2016).
10. Feldmeyer, D., Qi, G., Emmenegger, V. & Staiger, J. F. Inhibitory interneurons and their circuit motifs in the many layers of the barrel cortex. *Neuroscience* **368**, 132–151 (2018).
11. Ascoli, G. A., Donohue, D. E. & Halavi, M. Neuromorpho.org: A central resource for neuronal morphologies. *The Journal of Neuroscience* **27**, 9247–9251 (2007).
12. Mihaljević, B., Benavides-Piccione, R., Bielza, C., DeFelipe, J. & Larrañaga, P. Bayesian network classifiers for categorizing cortical GABAergic interneurons. *Neuroinformatics* **13**, 192–208 (2015).
13. Mihaljević, B., Bielza, C., Benavides-Piccione, R., DeFelipe, J. & Larrañaga, P. Multi-dimensional classification of GABAergic interneurons with Bayesian network-modeled label uncertainty. *Frontiers in Computational Neuroscience* **8**, 150 (2014).
14. Mihaljević, B. *et al.* Classifying GABAergic interneurons with semi-supervised projected model-based clustering. *Artificial Intelligence in Medicine* **65**, 49–59 (2015).
15. López-Cruz, P. L., Larrañaga, P., DeFelipe, J. & Bielza, C. Bayesian network modeling of the consensus between experts: An application to neuron classification. *International Journal of Approximate Reasoning* **55**, 3–22 (2014).
16. Mihaljević, B. *et al.* Towards a supervised classification of neocortical interneuron morphologies. *BMC Bioinformatics* **19**, 511 (2018).
17. Dawid, A. P. & Skene, A. M. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **28**, 20–28 (1979).
18. Welinder, P., Branson, S., Belongie, S. & Perona, P. The multidimensional wisdom of crowds. In *Advances in Neural Information Processing Systems* **23**, 2424–2432 (2010).

19. Raykar, V. C. & Yu, S. Eliminating spammers and ranking annotators for crowdsourced labeling tasks. *Journal of Machine Learning Research* **13**, 491–518 (2012).
20. Mihaljević, B., Benavides-Piccione, R., Bielza, C., Larrañaga, P. & DeFelipe, J. Classification of GABAergic interneurons by leading neuroscientists. *figshare*, <https://doi.org/10.6084/m9.figshare.c.4574186> (2019).
21. Kuncheva, L. I. *Combining pattern classifiers: Methods and algorithms*. (John Wiley & Sons, 2014).

### Acknowledgements

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness through the TIN2016-79684-P project. This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2).

### Author contributions

B.M. wrote the paper and curated the data. R.B.P., C.B., P.L. and J.D. gathered the data and carried out the initial study and reviewed this manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to B.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019