

# SCIENTIFIC DATA

**OPEN**

## An open repository for single-cell reconstructions of the brain forest

Masood A. Akram<sup>1</sup>, Sumit Nanda<sup>1</sup>, Patricia Maraver<sup>1</sup>, Rubén Armañanzas<sup>1</sup> & Giorgio A. Ascoli<sup>1</sup>

Received: 2 October 2017

Accepted: 8 January 2018

Published: 27 February 2018

NeuroMorpho.Org was launched in 2006 to provide unhindered access to any and all digital tracings of neuronal morphology that researchers were willing to share freely upon request. Today this database is the largest public inventory of cellular reconstructions in neuroscience with a content of over 80,000 neurons and glia from a representative diversity of animal species, anatomical regions, and experimental methods. Datasets continuously contributed by hundreds of laboratories worldwide are centrally curated, converted into a common non-proprietary format, morphometrically quantified, and annotated with comprehensive metadata. Users download digital reconstructions for a variety of scientific applications including visualization, classification, analysis, and simulations. With more than 1,000 peer-reviewed publications describing data stored in or utilizing data retrieved from NeuroMorpho.Org, this ever-growing repository can already be considered a mature resource for neuroscience.

<sup>1</sup>Center for Neural Informatics, Structures & Plasticity, Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA 22030, USA. Correspondence and requests for materials should be addressed to R.A. (rarmanan@gmu.edu) or to G.A.A. (email: ascoli@gmu.edu).

## Introduction

At the turn of the millennium the field of neuroinformatics was still in its infancy, but many researchers had already embraced computer simulations as an important approach to ground the understanding of the nervous system into quantitative models<sup>1–3</sup>. Pioneering work in anatomy and physiology had long established that the tree-like morphologies of neuronal axons and dendrites play fundamental roles in network connectivity and signal processing. Progress in hardware and software had gradually enabled the practice of digitally reconstructing neuronal morphology three-dimensionally from live microscopic feeds or image stacks<sup>4</sup>. These data were especially useful to carry out comprehensive statistical analyses and to build realistic computational models. However, interactions between experimentalists and theoreticians were the exception rather than the norm, and sporadic data exchange relied on serendipitous peer-to-peer contacts.

In that context, NeuroMorpho.Org was envisioned as an open, public-access, online portal where digital reconstructions of neurons from different animal species, brain regions, cell types, experimental conditions, and acquisition methods could be stored and freely retrieved for re-use in any scientific, educational, or artistic application<sup>5</sup>. The first version of the database was released in 2006 with an initial set of less than 1,000 cells<sup>6</sup> and was immediately adopted with broad enthusiasm in the neuroscience community<sup>7</sup>. Since then the repository continuously evolved, adding both new data and improved functionality in each of 30 releases over 11 years. Nevertheless, the key operational principles have remained stable: identifying suitable data in peer-reviewed publications, requesting the source files from authors, and making them available to the community after processing into ready-to-reuse form.

Here we describe the current status and future prospects of this mature resource from multiple perspectives, including those of the users, contributors, and developers. We start with an overview of the content and web-based accessibility, then we describe the community usage of these data, and lastly we provide a backstage view of the curation process enabling dense coverage of this data type<sup>8</sup>.

## Results

### Database content

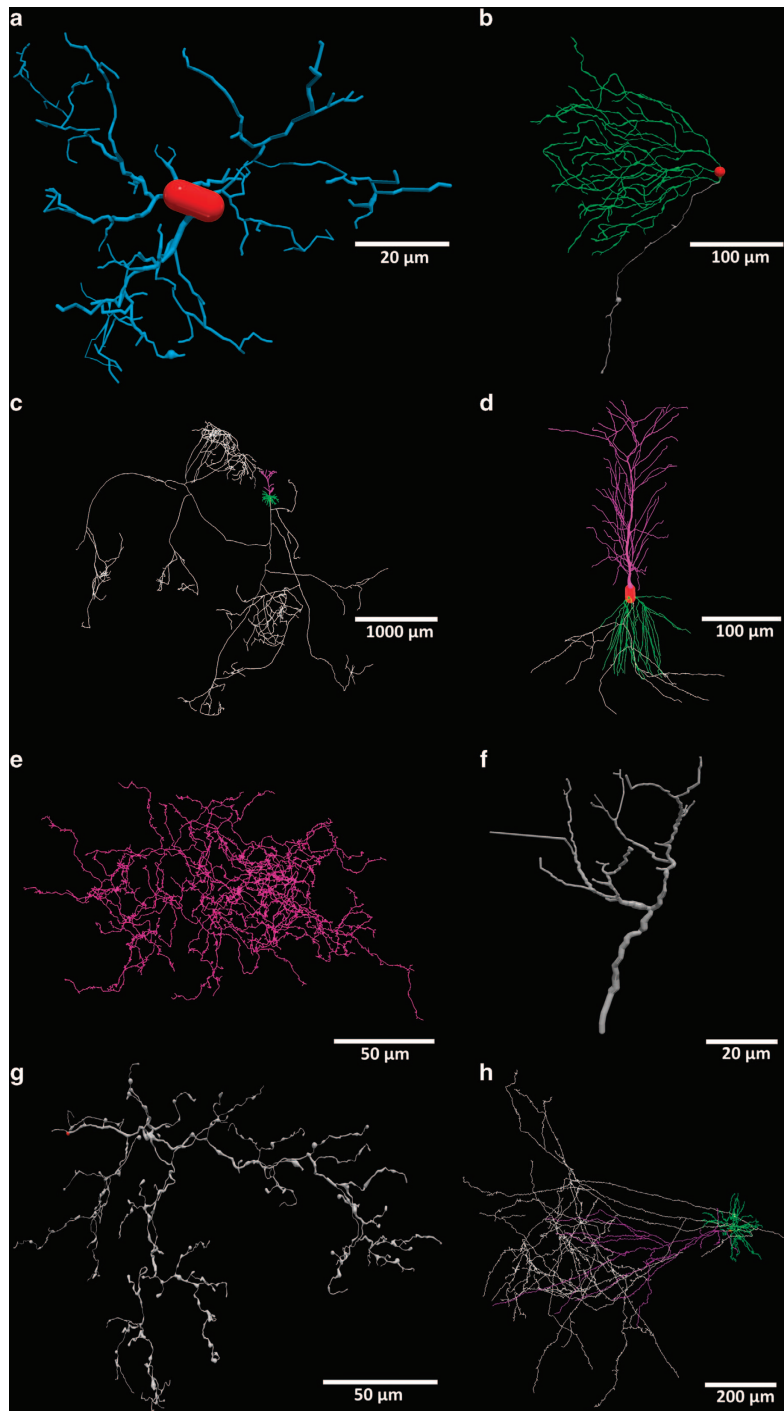
Version 7.3 of NeuroMorpho.Org (fall 2017 release) contains 80,012 cells contributed by 416 laboratories and described in 714 peer-reviewed articles. These data come from ~50 anatomical regions of more than 40 animal species ranging from nematodes to humans and notably encompassing both common experimental model (mouse, rat, fruit fly, etc.) and less usual sources (e.g., manatee, leopard, and giraffe). Most pertinently, all major cell types of the nervous system are represented across a number of dimensions (Fig. 1), including sensory receptors, glia, local interneurons and projection neurons releasing excitatory, inhibitory or modulatory neurotransmitters.

For every cell, the database includes the original morphological tracing file as provided by the contributors, the standardized version, and a detailed log of all changes. The basic data format describes the represented trees shape as a set of 3D coordinates (in micrometers), each associated with a tag indicating the type of structure (soma, axon, dendrite, apical, unspecified neurite, glial process or custom-defined), thickness, and the identity of the connected point in the path to the root. In addition to this digital reconstruction file, every cell is displayed with an image, an animation, and a user-interactive in-browser display. For certain users, this latter functionality may require minor JAVA setting adjustments (detailed in the Frequently Asked Questions) due to recent security updates on several operating systems. Moreover, all cells are comprehensively annotated with metadata<sup>9</sup> regarding the animal subject (species, strain or genotype, sex, age, weight), the cell studied (anatomical region, molecular expression, physiological features), the methodological procedure (experimental condition, tissue sectioning, specimen staining, imaging resolution, tracing system), and linked identifiers (PMID and DOI) of the corresponding referenced publication. Last but not least, a battery of geometric features (length, angles, branch topology, fractal dimension, etc.) is extracted and provided with each individual cell<sup>10</sup>.

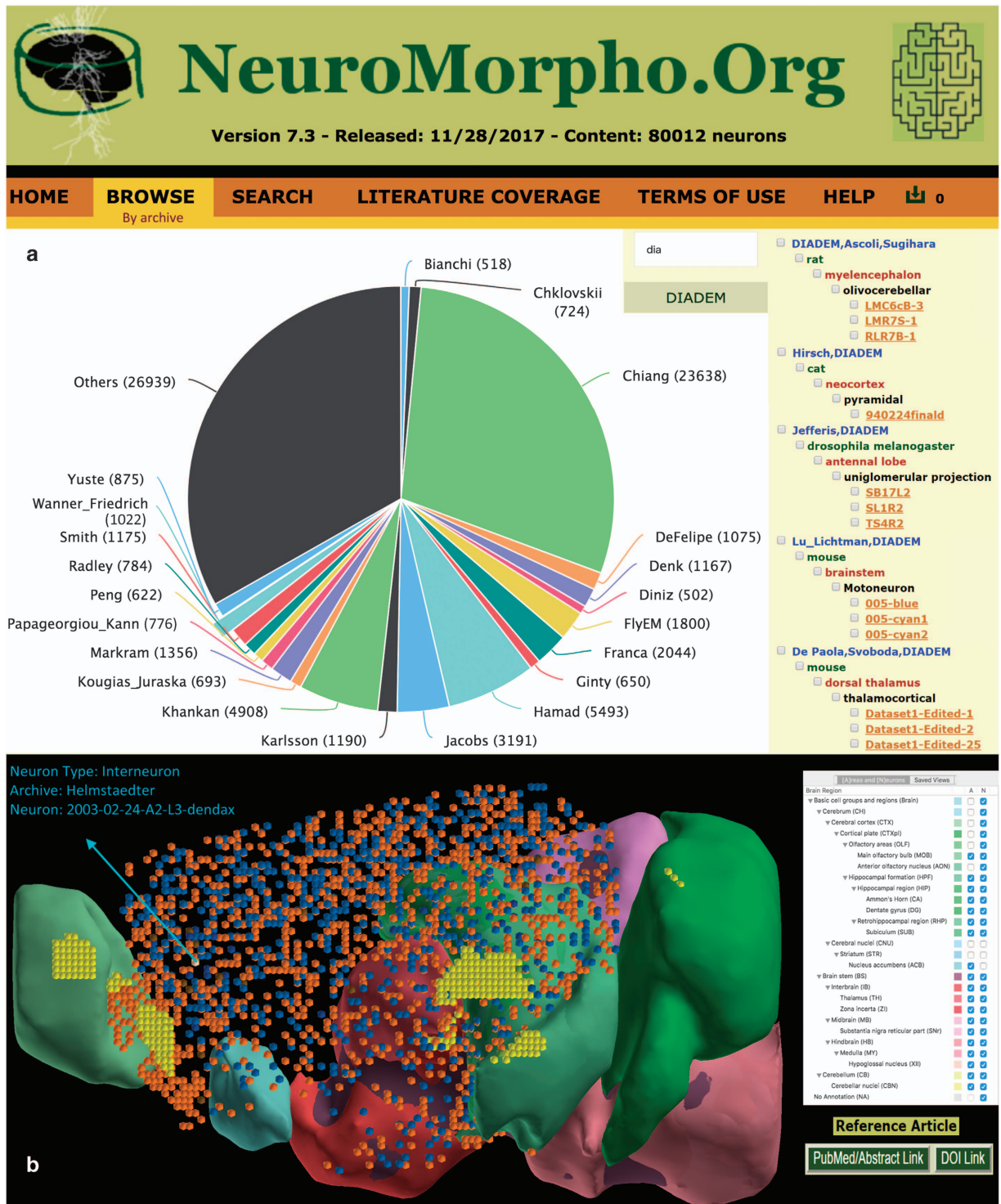
### Web functionality

NeuroMorpho.Org provides a user-friendly graphical interface to access the data through any modern web browser. Visitors can sample a random set of neurons or browse the entire repository by cell type, brain region, animal species, or contributing laboratory, corresponding to the intuitive elements most immediately associated with every study (what, where, which, who). In all cases, data can be selected and downloaded or simply dynamically visualized in rapid sequence with simple cursor movements. Furthermore, the separate Neuron Atlas application, which can be freely downloaded from the Browse menu, repurposes an earlier implementation of the Allen Brain Atlas software to map the spatial distribution of neurons across the major anatomical subdivisions of the rodent brain. The interactive 3D display links to the individual neuron pages and also enables direct visualization of the metadata and the single neuronal morphologies (Fig. 2).

The web front-end also offers multiple search functions. The metadata search option enables one to quickly identify content by conveniently combining filters through logically organized drop-down menus. For example, selecting mouse or drosophila as species populates the strain menu with completely different content. Choosing 'C57BL/6' from the former while setting experimental protocol to '*in vitro*' pulls (as of version 7.3, released in November 2017) 4,803 cells, which can be displayed as a summary or downloaded in bulk. The morphometry search is similarly organized allowing users to select the



**Figure 1.** Digital reconstructions of cellular morphologies in NeuroMorpho.Org. (a) Microglia cell from mouse spinal cord<sup>52</sup>, with processes in blue and cell body in red (Data Citation 1). (b) Ganglion cell from mouse retina<sup>53</sup>, with dendrites in green, axon in silver, and soma in red (Data Citation 2). (c) Pyramidal neuron from mouse neocortex<sup>54</sup>, with apical dendrites in magenta, basal dendrites in green, and long projection axon in silver (Data Citation 3). (d) Pyramidal neuron from rat hippocampus<sup>55</sup>, with apical dendrites in magenta, basal dendrites in green, and (incomplete) axon in silver (Data Citation 4). (e) Interneuron from mouse retina<sup>56</sup>, with unspecified neurites in pink (Data Citation 5). (f) Direction sensitive mechanoreceptor from cricket peripheral nervous system, with unspecified neurites in pink<sup>57</sup> (Data Citation 6). (g) Olivocerebellar neuron from rat myelencephalon, with axon in silver<sup>58</sup> (Data Citation 7). (h) Pyramidal neuron from cat neocortex<sup>59</sup>, with apical dendrites in magenta, basal dendrites in green, and long projection axon in silver (Data Citation 8).



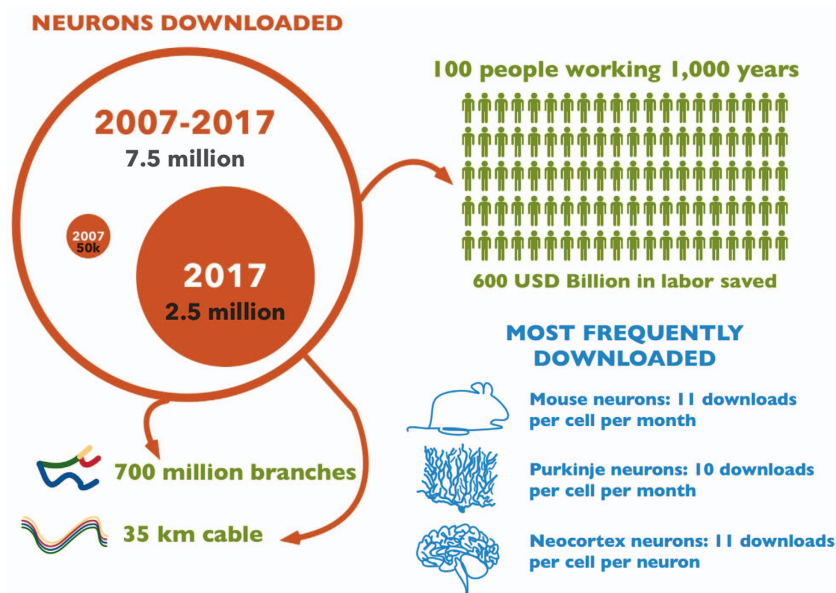
**Figure 2. Browse functionality.** (a) Summary view of selected neurons from the Browse-by-Archive pane, with links to individual entries and options for download. (b) NeuronAtlas, a free downloadable application for exploring digital reconstructions of neuronal morphologies from rat and mouse brains.

quantitative ranges of geometric parameters of interest. For instance, querying for neurons with at least 100 branches and a maximum path distance from the soma greater than 1 mm retrieves 5,706 cells (v.7.3).

A distinct search mechanism adopts a familiar 'Google-like' keyword bar with Boolean logic and wildcards. Alternatively, the advanced OntoSearch function leverages logical reasoning by generalization and specification of hierarchically organized knowledge<sup>11</sup>. Moreover, users can query the database

<p><b>a</b></p> <pre> 0:  "neuron_id" 1:  "neuron_name" 2:  "archive" 3:  "reference_pmids" 4:  "reference_doi" 5:  "note" 6:  "age_classification" 7:  "brain_region" 8:  "cell_type" 9:  "species" 10: "strain" 11: "scientific_name" 12: "stain" 13: "experiment_condition" 14: "protocol" 15: "slicing_direction" 16: "reconstruction_software" 17: "objective_type" 18: "original_format" 19: "domain" 20: "attributes" </pre>	<p><b>b</b></p> <pre> 0:  "principal cell" 1:  "interneuron" 2:  "projection" 3:  "dLGN-projecting" 4:  "pyramidal" 5:  "neurogliaform" 6:  "granule" 7:  "Embryo-born" 8:  "day4-born" 9:  "sensory receptor" 10: "mechanosensory" 11: "glutamatergic" 12: "GABAergic" 13: "cholinergic" 14: "nitroergic" 15: "serotonergic" 16: "fast-spiking" 17: "irregular spiking" 18: "Calretinin (CR)-positive" 19: "Somatostatin (SOM)-positive" 20: "Parvalbumin (PV)-negative" </pre>
<p><b>c</b></p> <pre> neuron_name:  "AM48-2" neuron_id:    2245 depth:       36.99 height:      583.85 bif_ampl_local: 73.53 branch_Order: 19 contraction: 0.77 eucDistance: 550.73 fractal_Dim: 1.05 fragmentation: 7096 n_bifs:      142 n_branch:    285 n_stems:     1 partition_asymmetry: 0.56 pathDistance: 1015.12 pk_classic:  2 surface:     9475.06 volume:      568.5 width:       396.67 diameter:    0.24 length:      12566.7 </pre>	<p><b>d</b></p> <pre> pmid: 25838038 doi: "10.1093/cercor/bhv053" journal: "Cerebral Cortex" species: "Rat" paper_ID: 1592 brain_Region: "Neocortex" cell_type: "Projection neuron" data_Status: "Positive response" first_Author: "Narayanan R.T." last_Author: "Oberlaender M." #neurons: 153 ocdate: "2015-06-30" paper_Title: "Beyond Columnar Organization: Cell Type- and Target Layer-Specific Principles of Horizontal Axon Projection Patterns in Rat Vibrissal Cortex" Details: "Positive response" system_Used: "NeuroLucida" year: 2015 </pre>

**Figure 3.** NeuroMorpho.Org's API gives developers and data miners access to database content and information allowing integration with other neuroinformatics tools. Complete usage documentation and definitions are available at [neuromorpho.org/apiReference.html](http://neuromorpho.org/apiReference.html). (a) Sample of available JSON objects characterizing experimental methods. (b) Selected entries specifying the dimension of cell types. (c) Subset of a given cell's morphometric characteristics defined as attribute-value pairs. (d) Bibliographic details for a data set from an individual publication.



**Figure 4.** Infographics summary of data download activity from the NeuroMorpho.Org portal.

content using the Literature Search, which is further elaborated on in the Methods section. The most recently added search functionality makes information machine-accessible through a public Application Programming Interface (API) using JavaScript Object Notation (JSON), a lightweight data-interchange format. The NeuroMorpho.Org API provides large-scale data access through anchors to neuron metadata, morphometric measurements, and literature references, in addition to brain regions, animal species, and more (Fig. 3).

### Usage statistics

Digital reconstructions shared through NeuroMorpho.Org are re-used for various purposes including training and dissemination, implementation of morphologically realistic computational simulations, assessments of potential synaptic contacts, validation of developmental growth models, and comparative statistical analyses across different neuron types, anatomical locations or experimental conditions. The usage of the repository can be quantified in terms of unique website visits, number of downloads, and derived publications. As of December 2017, the web page has been visited above 270,000 times from more than 75,000 unique internet addresses in 166 countries. The number of hits increased from less than 3,000 in the first year to over 15,000 in the June-August 2017 trimester. Notably, two-thirds of this traffic comes from the United States and China.

A total number of 7.5 million neurons were downloaded in standardized format along with 4.5 million ancillary files (original ‘as received’ morphologies and standardization logs) before the 7.3 release. Approximately 50,000 files were downloaded in the first year of the project up to a record 2.5 million in June-August 2017 (Fig. 4). The download volume can also be tallied in number of neurite branches (700 million) or total cable length (35 km). Based on an average of three working days per neuron (one for histology, one for imaging, one for tracing), it would take nearly 100,000 people working for a year to produce the amount of data downloaded in the course of this project. The most popular species and brain region are respectively mouse and neocortex (11 downloads per cell per month). The most frequently downloaded neuron types are Purkinje, newborn granule, and extraverterd pyramidal cells (10 downloads per cell per month). Complete statistics are updated online at every release.

The standard way to quantify impact in scientific research is by tracking citations. At least 543 publications cite NeuroMorpho.Org, 353 of which describing results that were directly based on data obtained from this resource. This ‘secondary’ usage substantially adds to the primary service of providing a data repository and standardization process. In this regard, before describing representative examples of re-usage application in the next section, it is worth mentioning that leading publishers, including Springer-Nature, Elsevier, and the Public Library of Science, recommend NeuroMorpho.Org as a trusted database for deposition of digitized morphological reconstructions. This is particularly important given the growing opportunity for ‘data publication’<sup>12</sup>.

### Representative applications

One of the most typical usages of digital reconstructions of neuronal morphology consists of constraining and validating data-driven models of electrophysiological activity or of developmental structural dynamics. For example, pyramidal cells taken from NeuroMorpho.Org are frequently studied in

simulations of activity<sup>13</sup> or structure<sup>14</sup>. Often these studies are extended to obtain novel computational or theoretical conclusions. For instance, the comparative analysis of dendritic complexity across animal taxonomy was recently related to the discriminative capacity of signal integration<sup>15</sup>, and the virtual generation of neuronal trees revealed a balance between the minimization of total wiring and signal conduction time<sup>16</sup>. An open problem increasingly addressed with morphological quantification is the unbiased classification of neuron types<sup>17</sup>. Application of co-clustering techniques on a large set of neocortical cells from many different labs and experimental preparations revealed a clear separation between principal neurons located in different cortical areas and depths<sup>18</sup>. Leveraging modern machine learning methods also yielded methods for content-based retrieval of morphologies, where similar neurons can be efficiently identified from a large collection based on sparse examples<sup>19</sup>.

Earlier examples of secondary discoveries have been extensively reviewed elsewhere<sup>20</sup>. Evidence of broader impact has also emerged beyond basic scientific research. The open availability of digital data is encouraging uptake in divulgation efforts, from undergraduate education<sup>21</sup> to 3D printable models<sup>22</sup>. By their own nature, many expository or outreach applications are not associated with peer-reviewed references, but are nonetheless beneficial to global societal progress. As representative examples we mention the selection as ‘Site of the Month’ in Neuroscience For Kids ([faculty.washington.edu/chudler/neurok.html](http://faculty.washington.edu/chudler/neurok.html)), the creation of an online teaching module in BrainU ([brainu.org](http://brainu.org)), and the use as testing dataset in a past competition of the Chinese Applied Math Olympiads ([neuromorpho.org/china\\_contest.jsp](http://neuromorpho.org/china_contest.jsp)). Similarly, data from NeuroMorpho.Org are used to train commercial expert systems ([drvtechnologies.com/aivia5](http://drvtechnologies.com/aivia5)) and to produce multimedia art installations ([vimeo.com/191338612](http://vimeo.com/191338612)). Only a fraction of these outcomes could be predicted at the outset, reinforcing the notion that open data availability may generate genuinely new opportunities for discovery and creativity that would be missed in a more restrictive ‘sharing upon request’ model.

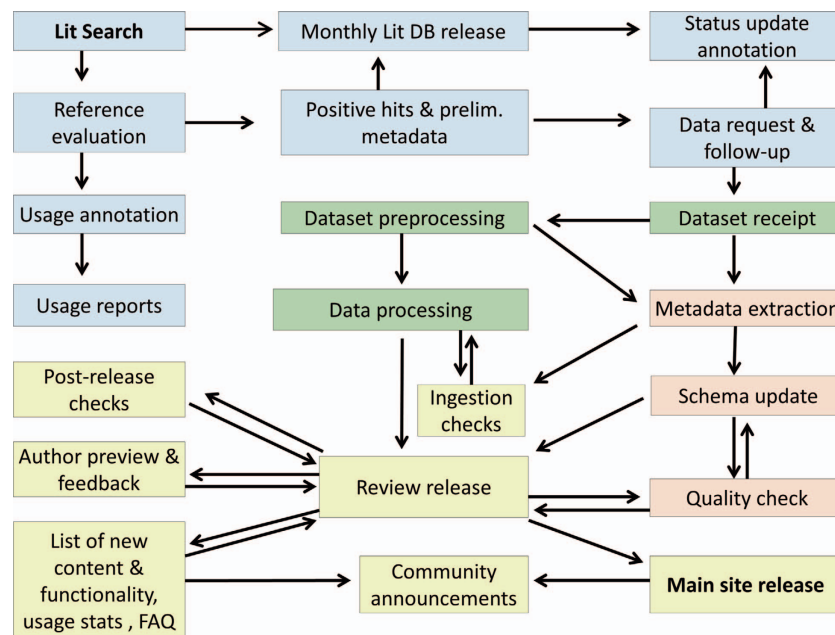
## Discussion

After more than a decade of continuous operation, NeuroMorpho.Org is now considered a stable and mature resource in the neuroscience community. Strikingly, we curators receive as much positive feedback and expressions of gratitude from data contributors as from data users. On the one hand this reflects the value added by the systematic standardization and annotation processes that are detailed in the Methods section. On the other, the above observation strongly refutes the entrenched view of a unidirectional cost-benefit flow from the ‘heroic’ experimental data producers to the ‘parasitic’ computational data modelers<sup>23</sup>. On the contrary, it corroborates the alternative idea of synergistic cooperation even in the absence of direct collaboration, since re-usage demonstrably augments the impact of the original dataset<sup>24</sup>. At the same time, we treasure constructive criticisms from the community, and the vast majority of new and enhanced functionalities were implemented over time to address outside suggestions and requests.

The pace of growth in database content continued to rise substantially over the lifespan of the project. While reconstructions initially accumulated at an average rate of less than 500 per year, most recently the yearly increase of data received passed the figure of 25,000. This rather dramatic acceleration is due to several compounded causes. Most prominently, the field of computational neuroanatomy became increasingly ‘hot’ and the number of publications describing digital reconstruction of neuronal and glial morphology has grown by an order of magnitude over the past 11 years (from ~7 to ~70 articles per month), even faster than the general upward trend in biomedical research in general and neuroscience in particular. Secondly, the progressive automation of the tracing process is boosting the typical dataset size<sup>25</sup>, from less than 20 neurons per study in the first year of operation to over 100 in the most recent one, with a notable single contribution of over 16,000 reconstructions<sup>26</sup>. Third, the attitude towards data sharing is gradually improving, from a hesitant 25% of positive responses in 2006 to a more encouraging 55% in 2017. In order to keep up with the expanding volume of incoming datasets, the information technology and data management infrastructure had to be progressively modernized and improved, as described in the Methods.

The content is also adapting to the parallel evolution of neuroscience research. For instance, glia (astrocytes, oligodendrocytes, microglia, and other types), once considered the ‘dark-matter’ of the nervous system or the ‘forgotten’ brain cells, are now recognized to contribute crucially to fundamental physiological processes, such as neural development<sup>27</sup>, signal processing<sup>28</sup>, and synaptic plasticity<sup>29</sup>. Furthermore, glial cells are involved in major neuropathologies, including Alzheimer’s<sup>30</sup> and Parkinson’s<sup>31</sup> diseases, stroke<sup>32</sup> and epilepsy<sup>33</sup>, as well as traumatic brain<sup>34</sup> and spinal cord injuries<sup>35</sup>. Specifically, glia morphology represents one of the most useful biomarkers of brain function and dysfunction, as exemplified by enlargement of activated microglia upon rising neuronal death<sup>36</sup>, loss of myelination related to retraction of oligodendrocytes<sup>37</sup>, and altered astrocyte architecture in response to pharmacological treatment<sup>38</sup> and neurotoxicity<sup>39</sup>. In line with this mounting emphasis and awareness, we started receiving spontaneous requests from researchers to include digital tracings of glia arbor morphology into the repository. We thus appropriately modified the metadata schema, search engine, and ingestion scripts, and release 7.3 now contains 3,069 glial reconstructions from 8 initial datasets.

In the foreseeable future, NeuroMorpho.Org is similarly expected to keep growing both quantitatively (content amount) and qualitatively (content type). Continuous integration with community resources will emphasize neuronal classification both through comprehensive programs such as the National



**Figure 5.** Flowchart of the operational procedures of NeuroMorpho.Org from start (Literature search) to end (Main site release), highlighting literature mining (blue), morphological standardization (green), metadata annotation (pink), and data release (yellow).

Institutes of Health-fostered Brain Initiative Cell Census Network (BICCN) and domain-specific projects like Hippocampome.Org<sup>40</sup>. This growth is going to necessitate the design and implementation of ergonomic, smart, and robust tools for richer metadata annotation benefiting authors and data curators alike. Moreover, it will be essential to incorporate the anatomical embedding of the reconstructed neurons in a representation of the surrounding tissue. Technically, this only requires the specification of a triad of non-aligned three-dimensional neuronal tracing points relative to a common coordinate framework. This augmented representation is within reach at least for fly, mouse, and human datasets, and would synergistically complement ongoing efforts to capture correlated pre- and post-synaptic circuit connectivity along with the digital morphology<sup>41</sup>. Last but not least, progress is advancing towards the extension of the arbor reconstruction format to accommodate temporal and multi-channel information<sup>42</sup>. These additions are especially important in light of recent experimental breakthrough in live time-lapse super-resolution imaging which could reveal critical biochemical details of neuronal dynamics.

## Methods

This section describes the current operational pipeline of NeuroMorpho.Org. The overall process can be broadly divided into four main task aggregates: literature mining, data processing, metadata annotation, and public release (Fig. 5). The execution of most steps is logically serial for each dataset, though the team organization is largely parallel: while releases are lumped into a few versions per year, new datasets are continuously identified, requested, received, processed, and annotated.

Given NeuroMorpho.Org's mission of freely distributing digital reconstructions from peer-reviewed publications, it is natural for the data identification process to begin with literature mining<sup>43</sup>. Using a comprehensive battery of keywords in appropriate combinations and several full-text search engines, we collect a set of newly published articles every month which possibly describe tracing datasets. Since the monthly queries were optimized to minimize the number of missed articles, only approximately a third of the shortlisted potential hits actually contain relevant data. Furthermore, users can also directly suggest articles to mine by checking the reference status using the Literature Search function. Team curators carefully evaluate each article and extract preliminary metadata information for every positive instance, including number of reconstructions, brain region, cell type, reconstruction system, bibliographic reference, and corresponding author contact. These minimal details are necessary and sufficient to invite data submission. The outcome of the request (after multiple reminders and follow-ups as needed) determines whether each given dataset is available (received and in processing pipeline) or unavailable (unanswered: ~70%, declined: ~15%, or declared lost: ~15%). A small but increasing proportion of data (5% in 2017) is sent spontaneously by the authors before publication. In addition to new data, the literature mining staff also tracks published usage of data downloaded from the repository. The entirety of the above information is recorded in a public literature database<sup>44</sup> that is updated monthly (Fig. 6).



a		b	
Availability Status		Summarized by Reconstruction Information	
Available	Unavailable	Human	Monkey
Determining availability		Cat	Rat
		Mouse	Drosophila
		Other_Mammals	Other_Vertebrates
		Other_Invertebrates	
PMID/DOI	Contacted authors		
28060701	Perisomatic GABAergic synapses of basket cell effectively control principal neuron activity in amygdala networks Requested on 02/17/2017 I. M. Veres, G. A. Nagy, N. Hajos eLife in processing pipeline	17924570	Cerebral cortex 111 Pyramidal neurons using NeuroLucida
27353294	Early-onset Purkinje cell dysfunction underlies Cerebellar ataxia in proximal multifunctional protein-2 deficiency Requested on 08/10/2016 Stephanie De Munter, Simon Verheijden, Esther Vanderstuyft, Ana Rita Malheiro, Pedro Brites, David Gall, Serge N. Schiffmann, Myriam Baes Neurobiology of Disease in repository	24453325	Hippocampus 15 GABAergic inhibitory interneurons using Neuromantic
2768598X	Comparative morphological analysis of chandelier cells in the auditory cortex of rodents and vertebrates Requested on 12/31/2015 Jaemin Lee, Jinhoi Song, Seongyea Jo, Min Kyung Kang, Seokho Kim, Hosung Bae, Mariel Chang, Suhwan Knox, Tae Kang, Sang Koh, Eliezer Seok, Hee Jun Cho Nature Communications No response	22002739	Barrel cortex 291 Subplate neurons using NeuroLucida
2825626X	Critical periods in gustatory reception for the emergence of taste-aversive associative conditioning Requested on 04/07/2014 Dalkiran Akif, Ruhi Yavuz, D. Deyman, Murat Salihoglu, Gang Peng, E. Shen The Journal of Physiology Request declined	20187150	Anterior olfactory nucleus 282 Type I, Type II and Pyramidal neurons using NeuroLucida
27057368	Nutritional Omega-3 Deficiency Alters Glucocorticoid Receptor-Signaling Pathway and Neuronal Morphology in Regionally Distinct Brain Structures Associated with Emotional Deficits Requested on 02/24/2017 T. Larrivee, M. L. Hilal, Smedt-Peyrusse, De, N. Sans, S. Laye Neural Plasticity Positive response	2522047X	Dentate gyrus 54 Interneurons using Simple Neurite Tracer
25546402	Type 2 wide-field amacrine cells in TH::GFP mice show a homogenous Synapse distribution and contact small ganglion cells Requested on 08/01/2017 B. Bruggen, A Meyer, F Boven, R Weiler, K Dedek European Journal of Neuroscience Invited	1682782X	Hippocampus 113 Pyramidal neurons using NeuroLucida
		2637722X	Nucleus accumbens 405 Medium spiny neurons using Imaris
		2287296X	Medial prefrontal cortex 768 Pyramidal and Medium spiny neurons using NeuroLucida
		27173096	Hippocampus 50 Pyramidal Neurons and Interneurons using Neurite Tracer
		22162008	Entorhinal cortex 119 Principal neurons using NeuroLucida
		25838038	Neocortex 153 Projection neurons using NeuroLucida

**Figure 6. Literature database and search.** (a) Retrievable information by availability status. To avoid singling out individual authors, the identifiers for declined or unanswered queries are fictional in this illustration; actual unabridged data are publicly posted online in the ‘Literature Coverage’ section of NeuroMorpho.Org. (b) Preliminary metadata annotation for all references deemed to describe digital reconstructions of neuronal morphology.

If the authors choose to share their traced cells, the next steps are metadata annotation and morphological standardization. The authors are invited to contribute to the metadata annotation; their input is complemented with information extracted from the corresponding publication and team curators formalize the resulting description using control vocabularies and formal ontologies prior to ingestion<sup>11</sup>. New terms are judiciously added as needed to the corresponding schema. Meanwhile, every neuron is assigned a unique machine-readable identifier while also maintaining its original human-assigned name after minor modifications if necessary to avoid occasional duplications or special characters known to interfere with smooth web interactions.

The raw neuronal reconstructions, which can be acquired in more than 20 different tracing systems and file types, are then translated into a common non-proprietary format<sup>45</sup> and undergo a series of consistency checks such as detection of non-positive diameters, overlapping points, and disconnected branches among many others. A comprehensive description of the complete standardization process is also posted on NeuroMorpho.Org along with the open-source checking software. Inconsistencies are edited using a variety of programs, including CVapp<sup>45</sup>, Neuromantic<sup>46</sup>, neuTube<sup>47</sup>, Vaa3D<sup>48</sup>, TREES Toolbox<sup>49</sup>, and several more as needed<sup>50</sup>. Importantly, all edits are tracked in accompanying log reports.

After completing morphological editing, data processing is finalized by extraction of morphometric parameter and generation of static images<sup>51</sup> and rotating animations for each neuron. All data, metadata, and ancillary files are then ingested into the previous version of the database to produce a password-protected Review site. This step is essential to enable all contributing authors to preview their data and request changes prior to public release. Furthermore, at this stage we can implement a number of cross-checks, which include: detecting duplicated cells, broken web links, and missing data files; ensuring the accuracy and consistency of new metadata entries and proper integration of enhanced functionalities; and updating usage statistics and frequently asked questions. Upon resolving any ensuing issues if any, the new version is finally openly released and publicly announced through mailing lists and social media.

As with any large-scale scientific database, occasional errors in data or metadata are detected and reported after public release. We try and correct all mistakes immediately if possible, without waiting for the next release. In case of simple typos, the correction is implemented silently. Any change in metadata, in contrast, is disclosed in the following release with a link to detailed documentation (see e.g., last bullet of v6.3 in the ‘What’s New’ page). When correcting the morphological files themselves, a link to the previous version is provided in the Notes of the corresponding neuron page (cf. sixth bullet of v6.1 in the ‘What’s New’ page).

## References

- Sejnowski, T. J., Koch, C. & Churchland, P. S. Computational neuroscience. *Science (New York, N.Y.)* **241**, 1299–1306 (1988).
- Dayan, P. Computational modelling. *Current opinion in neurobiology* **4**, 212–217 (1994).
- Hines, M. L. & Carnevale, N. T. NEURON: a tool for neuroscientists. *The Neuroscientist: a review journal bringing neurobiology, neurology and psychiatry* **7**, 123–135 (2001).
- Senft, S. L. A brief history of neuronal reconstruction. *Neuroinformatics* **9**, 119–128 (2011).
- Ascoli, G. A. Mobilizing the base of neuroscience data: the case of neuronal morphologies. *Nature reviews. Neuroscience* **7**, 318–324 (2006).
- Ascoli, G. A., Donohue, D. E. & Halavi, M. NeuroMorpho.Org: a central resource for neuronal morphologies. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **27**, 9247–9251 (2007).
- Ascoli, G. A. Successes and rewards in sharing digital reconstructions of neuronal morphology. *Neuroinformatics* **5**, 154–160 (2007).
- Halavi, M. *et al.* NeuroMorpho.Org implementation of digital neuroscience: dense coverage and integration with the NIF. *Neuroinformatics* **6**, 241–252 (2008).
- Parekh, R., Armananzas, R. & Ascoli, G. A. The importance of metadata to assess information content in digital reconstructions of neuronal morphology. *Cell and tissue research* **360**, 121–127 (2015).
- Scorcioni, R., Polavaram, S. & Ascoli, G. A. L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature protocols* **3**, 866–876 (2008).
- Polavaram, S. & Ascoli, G. A. An ontology-based search engine for digital reconstructions of neuronal morphology. *Brain informatics* **4**, 123–134 (2017).
- Akins, V. T., Weragalaarachchi, K., Picardo, M. C. D., Revill, A. L. & Del Negro, C. A. Morphology of Dbx1 respiratory neurons in the preBotzinger complex and reticular formation of neonatal mice. *Scientific data* **4**, 170097 (2017).
- Popovic, M. A., Carnevale, N., Rozsa, B. & Zecevic, D. Electrical behaviour of dendritic spines as revealed by voltage imaging. *Nature communications* **6**, 8436 (2015).
- Lopez-Cruz, P. L., Bielza, C., Larranaga, P., Benavides-Piccione, R. & DeFelipe, J. Models and simulation of 3D neuronal dendritic trees using Bayesian networks. *Neuroinformatics* **9**, 347–369 (2011).
- Zippo, A. G. & Biella, G. E. Quantifying the Number of Discriminable Coincident Dendritic Input Patterns through Dendritic Tree Morphology. *Scientific reports* **5**, 11543 (2015).
- Cuntz, H., Forstner, F., Borst, A. & Hausser, M. One rule to grow them all: a general theory of neuronal branching and its practical application. *PLoS computational biology* **6** (2010).
- Armananzas, R. & Ascoli, G. A. Towards the automatic classification of neurons. *Trends in neurosciences* **38**, 307–318 (2015).
- Lu, Y., Carin, L., Coifman, R., Shain, W. & Roysam, B. Quantitative arbor analytics: unsupervised harmonic co-clustering of populations of brain cell arbors based on L-measure. *Neuroinformatics* **13**, 47–63 (2015).
- Conjeti, S. *et al.* Neuron-Miner: An Advanced Tool for Morphological Search and Retrieval in Neuroscientific Image Databases. *Neuroinformatics* **14**, 369–385 (2016).
- Parekh, R. & Ascoli, G. A. Quantitative investigations of axonal and dendritic arbors: development, structure, function, and pathology. *The Neuroscientist: a review journal bringing neurobiology, neurology and psychiatry* **21**, 241–254 (2015).
- Chu, P., Peck, J. & Brumberg, J. C. Exercises in Anatomy, Connectivity, and Morphology using Neuromorpho.org and the Allen Brain Atlas. *Journal of undergraduate neuroscience education: JUNE: a publication of FUN, Faculty for Undergraduate Neuroscience* **13**, A95–a100 (2015).
- McDougal, R. A. & Shepherd, G. M. 3D-printer visualization of neuron models. *Frontiers in neuroinformatics* **9**, 18 (2015).
- Longo, D. L. & Drazen, J. M. Data Sharing. *The New England journal of medicine* **374**, 276–277 (2016).
- Ascoli, G. A., Maraver, P., Nanda, S., Polavaram, S. & Armananzas, R. Win-win data sharing in neuroscience. *Nature methods* **14**, 112–116 (2017).
- Nanda, S. *et al.* Doubling up on the fly: NeuroMorpho.Org Meets Big Data. *Neuroinformatics* **13**, 127–129 (2015).
- Chiang, A. S. *et al.* Three-dimensional reconstruction of brain-wide wiring networks in *Drosophila* at single-cell resolution. *Current biology: CB* **21**, 1–11 (2011).
- Zuchero, J. B. & Barres, B. A. Glia in mammalian development and disease. *Development (Cambridge, England)* **142**, 3805–3809 (2015).
- Song, I. & Dityatev, A. Crosstalk between glia, extracellular matrix and neurons. *Brain research bulletin* **136**, 101–108 (2017).
- Dzyubenko, E., Gottschling, C. & Faisner, A. Neuron-Glia Interactions in Neural Plasticity: Contributions of Neural Extracellular Matrix and Perineuronal Nets. *Neural plasticity* **2016**, 5214961 (2016).
- Rodriguez, J. J., Butt, A. M., Gardenal, E., Parpura, V. & Verkhratsky, A. Complex and differential glial responses in Alzheimer's disease and ageing. *Current Alzheimer research* **13**, 343–358 (2016).
- Greggio, E., Bubacco, L. & Russo, I. Cross-talk between LRRK2 and PKA: implication for Parkinson's disease? *Biochemical Society transactions* **45**, 261–267 (2017).
- Song, F. E. *et al.* Roles of NG2-glia in ischemic stroke. *CNS neuroscience & therapeutics* **23**, 547–553 (2017).
- Vezzani, A., Pascente, R. & Ravizza, T. Biomarkers of Epileptogenesis: The Focus on Glia and Cognitive Dysfunctions. *Neurochemical research* **42**, 2089–2098 (2017).
- Sajja, V. S., Hlavac, N. & VandeVord, P. J. Role of Glia in Memory Deficits Following Traumatic Brain Injury: Biomarkers of Glia Dysfunction. *Frontiers in integrative neuroscience* **10**, 7 (2016).
- Ahuja, C. S. *et al.* Traumatic spinal cord injury. *Nature reviews. Disease primers* **3**, 17018 (2017).
- Plaza-Zabala, A., Sierra-Torre, V. & Sierra, A. Autophagy and Microglia: Novel Partners in Neurodegeneration and Aging. *International journal of molecular sciences* **18** (2017).
- Dulamea, A. O. Role of Oligodendrocyte Dysfunction in Demyelination, Remyelination and Neurodegeneration in Multiple Sclerosis. *Advances in experimental medicine and biology* **958**, 91–127 (2017).
- Ahmed, S., Gull, A., Khuroo, T., Aqil, M. & Sultana, Y. Glial Cell: A Potential Target for Cellular and Drug Based Therapy in Various CNS Diseases. *Current pharmaceutical design* **23**, 2389–2399 (2017).

39. Blanco-Suarez, E., Caldwell, A. L. & Allen, N. J. Role of astrocyte-synapse interactions in CNS disorders. *The Journal of physiology* **595**, 1903–1916 (2017).
40. Wheeler, D. W. *et al.* Hippocampome.org: a knowledge base of neuron types in the rodent hippocampus. *eLife* **4** (2015).
41. Takemura, S. Y. *et al.* A connectome of a learning and memory center in the adult *Drosophila* brain. *eLife* **6** (2017).
42. Nanda, S. *et al.* Design and implementation of multi-signal and time-varying neural reconstructions. *Sci. Data* **5**, 170207 doi:10.1038/sdata.2017.207 (2018).
43. Halavi, M., Hamilton, K. A., Parekh, R. & Ascoli, G. A. Digital reconstructions of neuronal morphology: three decades of research trends. *Frontiers in neuroscience* **6**, 49 (2012).
44. Ascoli, G. A. Sharing Neuron Data: Carrots, Sticks, and Digital Records. *PLoS biology* **13**, e1002275 (2015).
45. Cannon, R. C., Turner, D. A., Pyapali, G. K. & Wheal, H. V. An on-line archive of reconstructed hippocampal neurons. *Journal of neuroscience methods* **84**, 49–54 (1998).
46. Myatt, D. R., Hadlington, T., Ascoli, G. A. & Nasuto, S. J. Neuromantic—from semi-manual to semi-automatic reconstruction of neuron morphology. *Frontiers in neuroinformatics* **6**, 4 (2012).
47. Feng, L., Zhao, T. & Kim, J. neuTube 1.0: A New Design for Efficient Neuron Reconstruction Software Based on the SWC Format. *eNeuro* **2** (2015).
48. Peng, H., Long, F., Zhao, T. & Myers, E. Proof-editing is the bottleneck of 3D neuron reconstruction: the problem and solutions. *Neuroinformatics* **9**, 103–105 (2011).
49. Cuntz, H., Forstner, F., Borst, A. & Hausser, M. The TREES toolbox—probing the basis of axonal and dendritic branching. *Neuroinformatics* **9**, 91–96 (2011).
50. Parekh, R. & Ascoli, G. A. Neuronal morphology goes digital: a research hub for cellular and system neuroscience. *Neuron* **77**, 1017–1038 (2013).
51. Torben-Nielsen, B. An efficient and extendable python library to analyze neuronal morphologies. *Neuroinformatics* **12**, 619–622 (2014).
52. Ohgomori, T., Yamada, J., Takeuchi, H., Kadomatsu, K. & Jinno, S. Comparative morphometric analysis of microglia in the spinal cord of SOD1(G93A) transgenic mouse model of amyotrophic lateral sclerosis. *The European journal of neuroscience* **43**, 1340–1351 (2016).
53. Badea, T. C. & Nathans, J. Morphologies of mouse retinal ganglion cells expressing transcription factors Brn3a, Brn3b, and Brn3c: analysis of wild type and mutant cells using genetically-directed sparse labeling. *Vision research* **51**, 269–279 (2011).
54. Economo, M. N. *et al.* A platform for brain-wide imaging and reconstruction of individual neurons. *eLife* **5**, e10566 (2016).
55. Ishizuka, N., Cowan, W. M. & Amaral, D. G. A quantitative analysis of the dendritic organization of pyramidal cells in the rat hippocampus. *The Journal of comparative neurology* **362**, 17–45 (1995).
56. Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168–174 (2013).
57. Jacobs, G. A. & Theunissen, F. E. Extraction of sensory parameters from a neural map by primary sensory interneurons. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **20**, 2934–2943 (2000).
58. Brown, K. M., Sugihara, I., Shinoda, Y. & Ascoli, G. A. Digital morphometry of rat cerebellar climbing fibers reveals distinct branch and bouton types. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **32**, 14670–14684 (2012).
59. Martinez, L. M. *et al.* Receptive field structure varies with layer in the primary visual cortex. *Nature neuroscience* **8**, 372–379 (2005).

## Data Citations

1. Ohgomori, T. *NeuroMorpho.Org* NMO\_66876 (2017).
2. Badea, T. C. *NeuroMorpho.Org* NMO\_10752 (2014).
3. Chandrashekar, J. *NeuroMorpho.Org* NMO\_51118 (2017).
4. Amaral, D. G. *NeuroMorpho.Org* NMO\_00940 (2006).
5. Denk, W. *NeuroMorpho.Org* NMO\_63366 (2017).
6. Jacobs, G. A. *NeuroMorpho.Org* NMO\_04590 (2009).
7. DIADEM, Ascoli, G. A., Sugihara, I. *NeuroMorpho.Org* NMO\_08839 (2013).
8. Hirsch, J. A., *NeuroMorpho.Org* NMO\_07293 (2012).

## Acknowledgements

The authors thank Zulfa Ghannouchi for assistance in literature mining, Kayvan Bijari for useful discussion, Usama Saifi for help with the illustrations, Pruthvik Narayanaswamy and Suyameendra Wadki for software programming and data ingestion, and the undergraduate data processing team ([neuromorpho.org/acknowl.jsp#NMO](http://neuromorpho.org/acknowl.jsp#NMO)) for careful morphological proofing. This work was supported in part by grants R01NS039600 (NIH), R01NS086082 (NIH), U01MH114829 (NIH), and DBI-1546335 (NSF). Publication of this article was funded in part by the George Mason University Libraries Open Access Publishing Fund.

## Author Contributions

All authors are core members of the NeuroMorpho.Org team. M.A.A. is responsible for literature evaluation and metadata annotation; S.N. is responsible for data conversion and external tool integration; P.M. is responsible for literature management and access statistics; R.A. is the technical lead, supervises the software programmers, and is responsible for data ingestion/release as well as database maintenance/development; G.A.A. is the project founder and overall director, interacts with data owners, and is responsible for data requests, management, processing, and quality-control. M.A.A. and G.A.A. wrote the manuscript and created the illustrations with the remaining co-authors providing edits and crucial feedbacks.

## Additional Information

**Competing interests:** The authors declare no competing financial interests.

**How to cite this article:** Akram, M. A. *et al.* An open repository for single-cell reconstructions of the brain forest. *Sci. Data* **5**:180006 doi: 10.1038/sdata.2018.6 (2018).

**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 Author(s) 2018