Samhita Mallavarapu
PI: Dr. Van Savage | Mentor: Paheli Desai-Choudhary
UCLA Computational and Systems Biology

Comparison of Neuron Radius and Length Scaling Ratios from Angicart++ Image Reconstructions and Existing Morphological Reconstruction Data

Abstract

Studying the structure and function of neurons, as the messengers of information from the brain and through the nervous system, is the key to understanding the composition and structure of the brain as a whole. In this study, we sought to validate the use of the novel software, Angicart++, originally developed for use on the blood vessels, as a method of studying the morphology of neurons that we are interested in. By doing this, we would be closer to finding a qualitative description, and eventually a computational model, for neurons and the neuromorphic structure of the brain. We validate this model by comparing the results obtained from data using two reconstruction methods: the SWC file method and the Angicart++ method. The data to conduct this analysis is sourced from the Allen Brain Atlas in the form of SWC files and image stacks. Once the analysis of each method was completed, we found significant variation between the outputs of each method. This could be due to the fact that we only conducted this analysis on one neuron's data so we should take a wider sample of neurons to conduct a more accurate analysis.

Introduction

Neurons are the way that information travels throughout the brain and nervous system. They are connected through an extensive network, made of axons and dendrites, that work in conjunction with each other to relay information and drive an organism to respond to the

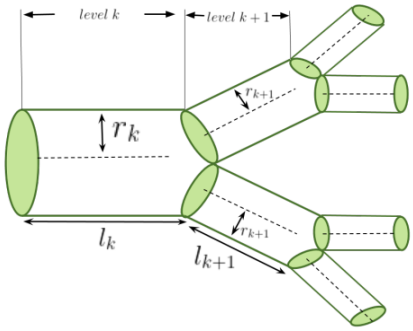
surrounding stimuli. It is possible to learn information about the structure and function of the neural network by taking a closer look at a neuronal level. Taking a look at the structure and function of a singular neuron will reveal not only information about that particular neuron but will also provide data to compare to that of other neuron types. Comparing these properties will allow for a better understanding of the association between the structure and function of the neuronal network as a whole.

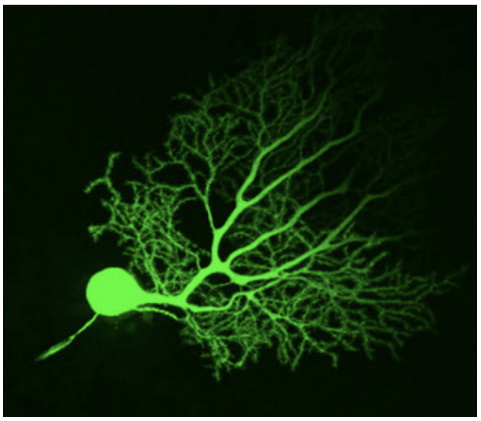
To achieve this goal, we employed a mathematical modeling approach that takes into account the theoretical biophysical properties of the network. We formulated cost functions and imposed constraints, and then solved for the theoretical predictions for the radius and length scaling ratios. We found the radius scaling ratio to be $\frac{r_k+1}{r_k}$ and the length scaling ratio to be $\frac{l_k+1}{l_k}$. These theoretical predictions were then validated using experimental data obtained from different neuron types.

To obtain the experimental data, we employed two different methods: the SWC method and the Angicart++ method. Both methods involve using data from the Allen Brain Atlas (ABA), a comprehensive database that provides information about various neurons. The SWC method involved downloading SWC files from the ABA and conducting an analysis of these datasets. The SWC files are structured as pixel-by-pixel data, so it is necessary to use the python script swc-to-ratios.py to organize the data by branch and extract the average radius and length of each branch.

The Angicart++ method involved downloading raw data images from the ABA and running an analysis to extract data from the images. Although the software was primarily intended for the analysis of blood vessels, we noticed the structural similarities between blood

vessels and branching neurons and decided to apply the program to neural networks to validate the model for this purpose. After running Angicart++, we looped through the outputs, calculated the scaling ratios using the python script getratios.py, and created histograms to show the distributions of these ratios. We then compared the means in the data to the means in the theoretical predictions calculated from the SWC files. If the outputs of Angicart++ were found to be consistent with the SWC calculations, it would validate the use of Angicart++ in the analysis of neuronal networks.





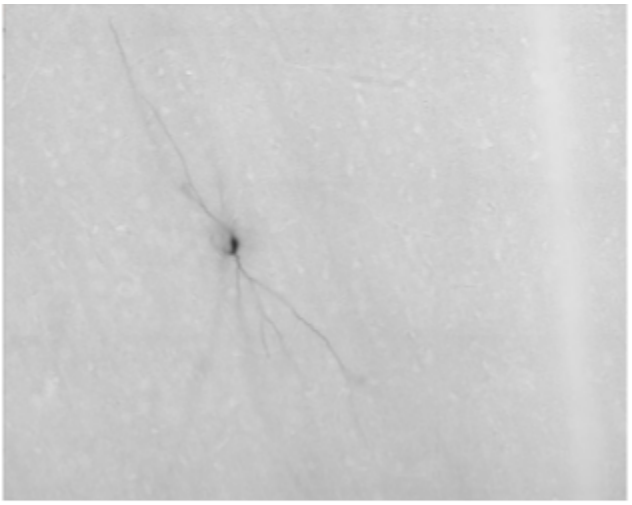
Data

The Allen Brain Atlas (ABA) is a comprehensive database that provides information about various neurons, including electrophysiological, morphological, and transcriptomic data. According to the ABA documentation, the staining methods for the cells were programmed using the intelliPATH FLX® protocol run plan and were rinsed on an autostainer using 1X TBS wash buffer (ABA, n.d.). In this study, we focused on human cells found within the ABA.

The ABA database is comprised of image stacks that can be utilized for 3D analysis as well as standard SWC files. We employed the Cell Feature Search function of the ABA to focus on the cells from the Middle Temporal Gyrus (MTG) layer 6. The MTG is a region of the brain that controls functions such as motion observation, deductive reasoning, language processing,

and producing facial expressions [Desai-Chowdhry et al., 2021]. We identified our primary subject, an eighteen-year-old male patient, by searching through the database and focusing on those cell images that displayed a clear branching pattern.

For the SWC method of analysis, we downloaded the SWC files of this cell, extracted the radius and length ratios using the python code swc-to-ratios.py, and then created histograms from these ratios using the python code histogram.py. For the Angicart++ method, we started by downloading the image stacks from the ABA and inverted the colors of the images using the python code invert.py. We found that each image in the stack was 7580×11,264 pixels. This image size was incompatible with Angicart++ however, we overcame this issue by downsampling the images until the dimensions were 470×701 pixels. This took four cycles of running the downsampling code. We also normalized the data by removing unnecessary images, resulting in a more accurate set of pictures for Angicart++ analysis.



This is a picture of a raw image from Allen Brain Atlas, before any of the pre-processing steps we took to get it ready for Angicart++.

Methods

Existing Reconstruction Data Analysis

The pixel-by-pixel data of the SWC files obtained from the Allen Brain Atlas necessitated the utilization of the python script, swc-to-ratios.py, in order to extract the average radius and length of each branch and organize this data by branch. The script swc-to-ratios.py operated by counting the number of pixels within the file, followed by identification of the pixel in which branching occurred. The script then proceeded to identify the pixels of the parent and

daughter branches, which were subsequently used to separate the branches. Subsequently, the parent branches were broken down into levels and an output file was generated, specifying the pixel ID labels for each point, the x, y, z spatial coordinates, the radius at each point, and the parent pixel IDs.

The average radius of each branch was calculated using the following formula, where k represented each branch, and the pixels i ranged from 1 to N_k, with N_k being the last pixel of each branch:

$$r_k = \sum_{i=1}^{N_k} \frac{r_i}{N_k}$$

The length of each branch was extracted by calculating the Euclidean distances between each point within the branch using the following formula:

$$l_k = \sum_{i=1}^{N_k} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2 + (z_i - z_{i-1})^2}$$

Once the radius and length of each branch had been calculated, the values were utilized to compute the scaling ratios. The radius scaling ratios were obtained by dividing the daughter radius by the parent radius and were subsequently listed and outputted into a .dat file. The same process was repeated with the intention of obtaining the length scaling ratios of the branches. The radius scaling ratios .dat file and the length scaling ratios .dat file were then used to create histograms using the python script histogram.py. These histograms contained information such as the mean and the Standard Error of the Mean (SEM), which were both listed and displayed by solid lines on the plots.

In order to create a more comprehensive distribution that took into account variations within samples, the SWC files of two additional samples of the same neuron type that also had

clear images were obtained and analyzed using the script swc-to-ratios.py. The python script bind.py was used to combine the radius ratios list .dat files of all three neuron samples and the python script histogram.py was used to create a histogram from the combined values. The same process was repeated for the length ratios list .dat files.

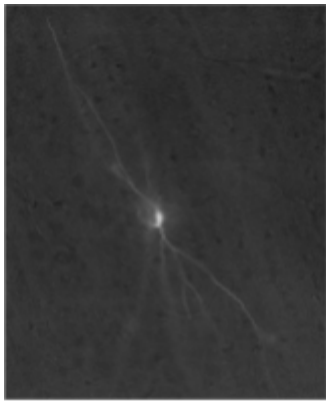
Image Processing with Angicart++ and Analysis

The initial step in utilizing the Angicart++ method for the analysis was obtaining the necessary data. Image stacks were downloaded from the Allen Brain Atlas in the form of .jpg files, which were subsequently numbered and prepared for use with the python script, invert.py. The script required the utilization of the OpenCV library, specifically the import name cv2. OpenCV is utilized for computer vision, machine learning, and image processing, and is an open-source library full of algorithms to accomplish these tasks. The invert.py script was used to invert each of the downloaded .jpg image files and convert them into .png files. It was found that the image files had dimensions of 7580×11,264 pixels. To optimize the performance of Angicart++, the images were downsampled four times to meet the requirement that both dimensions be less than 1000 pixels. Images that were deemed to lack valuable information were removed before running the images through Angicart++ at various threshold values.

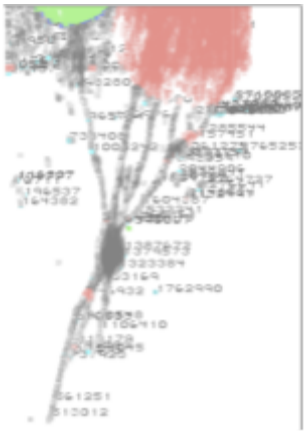
Angicart++ was used to analyze the image stacks as a 3D image, classifying the voxels in the image as part of the vessel (or neuron in this case) based on an intensity threshold. A threshold value was determined manually by trial-and-error relative to the visual of the vessels, resulting in a network mask on which spatial criteria were used to identify the endpoints, centerline, and branching points of the network. The software subsequently removed voxels that could be removed without disrupting the network, allowing for the quantification of the network's measurements. [Newberry et al, 2015]

Unlike the output obtained from the SWC file analysis, the output of Angicart++ was organized by branch and structured in two .tsv files per chosen threshold. One of the output files contained roots while the other output file did not contain roots. We focused on the file that presented the roots. A threshold analysis was conducted by running the analysis for seventeen different thresholds that fell within the range from 0.255 to 0.575. Once the .tsv files were obtained, four data columns were extracted and put into their own .dat files. The columns extracted were the parent names, the vessel names, the vessel radius (we chose to look at the observed radius), and the vessel length. Then, we use the python script replacena.py to replace any "N/A"s in the parentname.dat file with a 0 in order to normalize the data so that the next steps of the analysis could be done.

The data was then looped through and processed using the python code getratios.py, which imported the four .dat files and used them to locate the parent and daughter branches. The ratio of the daughter radius to the parent radius was taken to obtain the radius scaling ratio distribution and the ratio of the daughter length to parent length was taken to obtain the length scaling ratio distribution. Both of these lists were printed as separate .dat files. The getratios.py script subsequently created histogram distributions for each of the thresholds, resulting in the creation of thirty-four histograms during the threshold analysis. A table was hand-written, including a column for the threshold value, the mean radius, and the mean length from each histogram. These were converted into 3 .dat files that were imported into the python script threshold.py, which was used to create the threshold analysis plots from which the optimal threshold value was selected for comparison with the SWC file analysis. If the outputs of Angicart++ turned out to be relatively similar to the SWC calculations, it would validate the use of Angicart++ on neurons.



This is a picture of the neuron image after all the pre-processing steps we took to get it ready for Angicart++.



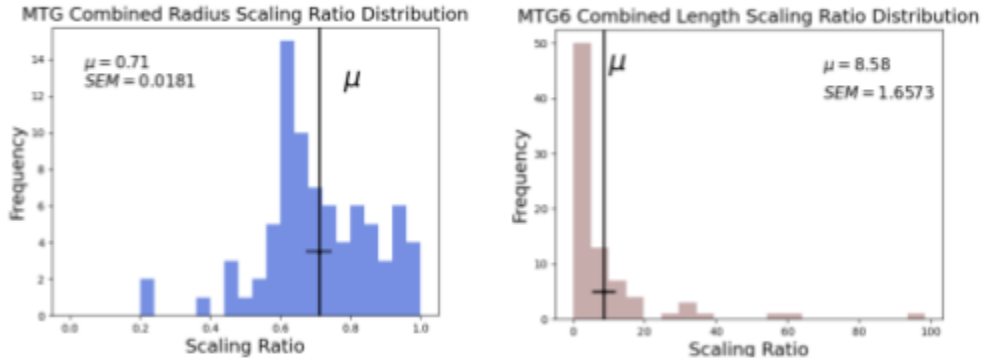
This is a picture of Angicart++ processing the image.

Results

The findings of this study are presented in four separate sections. The first section pertains to the analysis of the combined radius and length scaling ratios. The second section pertains to the threshold analysis. The third section pertains to the comparison of the radius scaling ratios between the SWC files and the Angicart++ data, while the fourth section pertains to the comparison of the length scaling ratios between the SWC files and the Angicart++ data. These results will be discussed in detail in the following sections of this paper.

Combined Radius and Length Scaling Ratios

To increase the representativeness of the data and decrease the standard error, we obtained the SWC files of three neurons and combined their data to create separate histograms of the combined radius and branch length. The combination of these three datasets resulted in more continuous histograms as compared to the histograms constructed from individual SWC files.

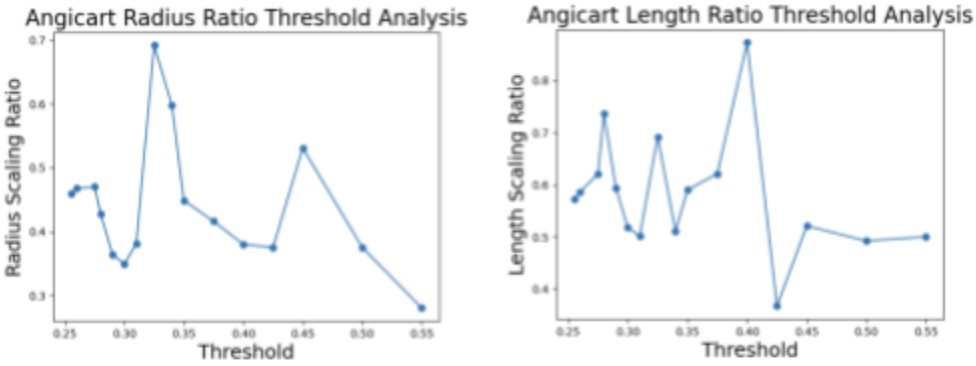


The left histogram is the MTG Combined Radius Scaling Ratio distribution. The right histogram is the MTG Combined Length Scaling Ratio distribution.

As can be observed, the distribution of the radius scaling ratio is uni-modal and exhibits a slight right skewness. The mean radius scaling ratio is 0.71, with a standard error of 0.0181 and a limited number of outliers. Additionally, the range of the radius scaling ratio falls between 0.2 and 1.0. In contrast, the distribution of the length scaling ratio is uni-modal and exhibits a pronounced right skewness, with a mean length scaling ratio of 8.58 and a standard error of 1.6573. The range of the length scaling ratio is broader, spanning from 0 to 100.

Threshold Analysis

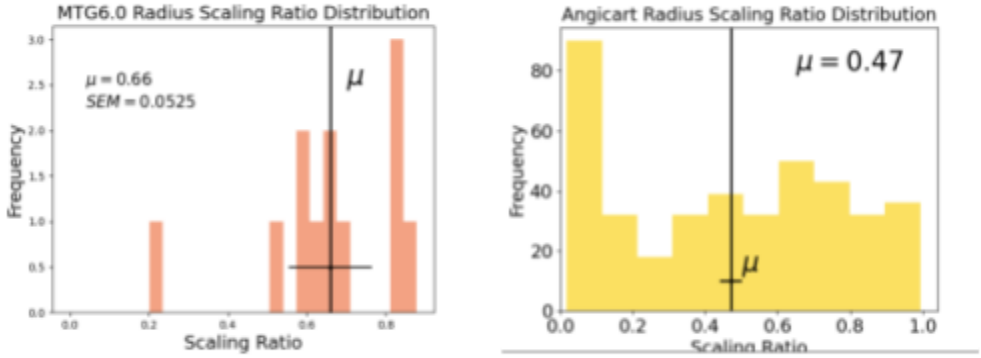
Prior to determining the most appropriate Angicart++ output for the comparisons, a threshold analysis was conducted. To achieve this, a range of seventeen different thresholds, ranging from 0.255 to 0.575, were utilized. The results were then plotted, as illustrated in the accompanying figures.



The left histogram is the Angicart++ Radius Ratio Threshold Analysis. The right histogram is the Angicart++ Length Ratio Threshold Analysis.

Based on the results of the threshold analysis, the optimal threshold value was determined to be 0.260, as the data plateaus at this threshold value in the radius plot, and a slight plateau is observed in the length plot. This value was selected as the threshold for further analysis.

Comparison of the Radius Scaling Ratios Between SWC Files and Angicart++ Data

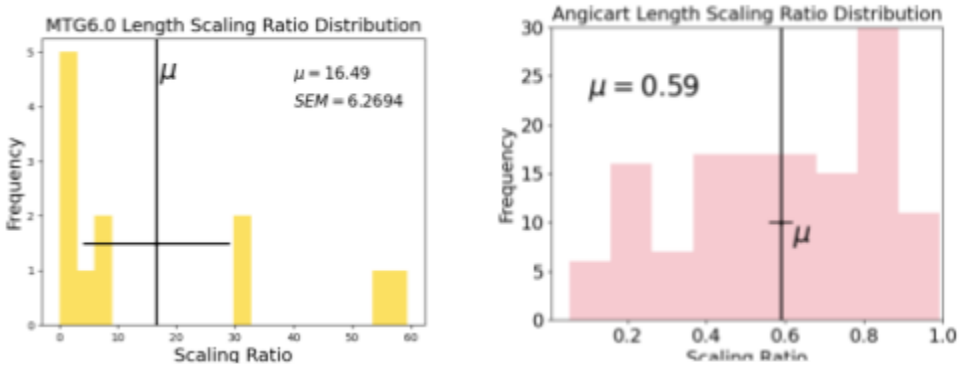


The left histogram is the SWC Radius Ratio Scaling Distribution. The right histogram is the Angicart++ Radius Ratio Distribution.

As illustrated in the accompanying figures, the distribution of the radius scaling ratio obtained from the SWC files is uni-modal and exhibits a left skewness. The mean radius scaling ratio is 0.66, with a standard error of 0.525 and a limited number of outliers. Additionally, the range of the radius scaling ratio falls between 0.2 and 1.0. Gaps in the distribution of the data are also observed. This results in a relatively long standard error (as represented by the horizontal

line). On the other hand, the distribution of the radius scaling ratio obtained from Angicart++ is uni-modal and exhibits a pronounced right skewness. The mean radius scaling ratio is 0.47. Although a precise standard error could not be calculated, the distribution is relatively continuous over its range between 0.0 and 1.0. This results in a significantly shorter standard error for the Angicart++ Radius Scaling Ratio distribution.

Comparison of the Length Scaling Ratios Between SWC Files and Angicart++ Data



The left histogram is the SWC Length Ratio Scaling Distribution. The right histogram is the Angicart++ Length Ratio Distribution.

As depicted in the accompanying figures, the distribution of the length scaling ratio obtained from the SWC files is uni-modal and exhibits a left skewness. The mean length scaling ratio is 16.49, with a standard error of 6.2694 and outliers. Additionally, the range of the length scaling ratio falls between 0 and 60. Gaps in the distribution of the data are also observed. This results in a relatively long standard error (as represented by the horizontal line). On the other hand, the distribution of the length scaling ratio obtained from Angicart++ is uni-modal and exhibits a left skewness. The mean length scaling ratio is 0.59. Although a precise standard error could not be calculated, the distribution is relatively continuous over its range between 0.0 and 1.0. This results in a significantly shorter standard error for the Angicart++ Length Scaling Ratio distribution.

Conclusion

In this study, we aimed to analyze the structural characteristics of neuronal networks using SWC files. Specifically, we sought to investigate the distribution of radius and length scaling ratios and compare them with Angicart++ data. To accomplish this, we employed a combination of three SWC files instead of using singular SWC files. The rationale behind this decision was to reduce the error associated with the analysis and obtain a more representative distribution.

Our analysis revealed that the combined histograms of the radius and length scaling ratios displayed smaller error bars compared to the singular SWC file histograms. The radius scaling ratio distribution was found to be uni-modal and slightly skewed to the right, with a limited number of outliers. Conversely, the distribution of the length scaling ratio was uni-modal and notably right-skewed, with a higher number of outliers. This discrepancy could be attributed to the larger variance within the length compared to the radius. Additionally, as more branching occurs, the radius of the daughter branches is typically smaller than the parent branch, which is not necessarily the case when examining the length of the branch.

To further improve the accuracy of our analysis, we performed a threshold analysis using 17 different threshold values. This approach enabled us to identify the threshold value that resulted in optimal output. Our analysis revealed that 0.26 was the optimal threshold value for comparison.

When comparing our results with Angicart++ data, we observed significant differences in the distribution of radius scaling ratios. Specifically, the SWC file distribution displayed gaps and was left-skewed, whereas the Angicart++ distribution was more continuous and right-skewed. Furthermore, the means of the distributions were dissimilar, with the SWC file

distribution displaying a higher mean. The error bars also varied in length, with the SWC file distribution displaying a longer error bar. These discrepancies could be attributed to the smaller number of data points used to derive the ratios from the SWC files, as compared to Angicart++ data.

Similar observations were made when comparing the length scaling ratios between the SWC files and Angicart++ data. The SWC file distribution displayed gaps and was right-skewed, whereas the Angicart++ distribution was more continuous and left-skewed. Additionally, the means of the distributions were dissimilar, due to the scaling ratios being on different scales altogether. The SWC file distribution also displayed a longer error bar compared to the Angicart++ distribution.

In conclusion, our study revealed that the combination of three SWC files resulted in a more representative distribution with a smaller error. The threshold analysis enabled us to identify an optimal threshold value for comparison. However, when comparing our results with Angicart++ data, we observed significant differences in the distribution of radius and length scaling ratios, which could be attributed to the smaller number of data points used for the SWC files. Further research is necessary to fully understand these discrepancies and improve the accuracy of our analysis.

Future Directions

This study was focused on analyzing the structural characteristics of a single neuron sourced from the Allen Brain Atlas, resulting in a limited understanding of the outputs. To gain a more comprehensive understanding of the accuracy of the Angicart++ method's outputs in comparison to the SWC file method's outputs, it is necessary to expand the sample size. As a next step, it is recommended to obtain additional samples from the Allen Brain Atlas and

conduct a similar analysis. This will enable us to validate the Angicart++ outputs and determine its applicability in neuronal network analysis.

If Angicart++ is found to be a valid tool for analyzing neuronal networks, it would provide us with a means to better understand the structural properties of neurons and eventually the neuromorphic structure of the brain. This could also open up new avenues of research, such as investigating why neurons with specific functions have certain structural characteristics or why different neuron types exhibit distinct branching networks.

References

[1] Allen Institute for Brain Science. Allen Human Brain Atlas. 2010 [cited 2021 May 9]. Available from: http://celltypes.brain-map.org/experiment/morphology/613396614.

[2] Desai-Chowdhry P, Brummer A, Savage V. How Axon and Dendrite Branching Are Governed by Time, Energy, and Spatial Constraints . bioRxiv 2021.07.15.452445; doi: https://doi.org/10.1101/2021.07.15.452445.

[3] Newberry MG, Ennis DB, Savage S, Cox VM. Testing Foundations of Biological Scaling

Theory Using Automated Measurements of Vascular Networks. PLoS Computational Biology. 2015; 11(8): e1004455.

[4] Savage VM, Deeds EJ, Fontana W. Sizing up Allometric Scaling Theory. PLoS Computational Biology. 2008; 4(9):e1000171.