QUANTIFICATION OF NEURON TYPES IN THE RODENT HIPPOCAMPAL FORMATION USING COMPUTATIONAL METHODS

by

Sarojini Manjusha Attili

A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Neuroscience

Committee:

_____	Dr. Giorgio Ascoli, Committee Chair
_____	Dr. Diek Wheeler, Committee Member
_____	Dr. Padmanabhan Seshaiyer, Committee Member
_____	Dr. Saleet Jafri, Committee Member
_____	Dr. Saleet Jafri, Department Chairperson
_____	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
_____	Dr. Fernando R. Miralles-Wilhelm, Dean, College of Science

Date: _____

Summer Semester 2021
George Mason University
Fairfax, VA

Quantification of Neuron Types in the Rodent Hippocampal Formation Using Computational Methods

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Sarojini Manjusha Attili
Master of Science
George Mason University, 2016
Bachelor of Individualized Study
George Mason University, 2005

Director: Giorgio A. Ascoli, Professor
Department of Molecular Neuroscience

Summer Semester 2021
George Mason University
Fairfax, VA

DEDICATION

This dissertation is dedicated to my loving husband Srinivas Attili.

ACKNOWLEDGEMENTS

I would like to thank my husband, Srini for his constant encouragement and motivation throughout the last eight years of my academic journey; my boys Abhi and Sid for their patience and understanding on many evenings and weekends when I was occupied with working towards my thesis and could not spend time with them.

To my advisor, Dr. Giorgio Ascoli: I feel very fortunate to have had this opportunity to work with you. I sincerely believe that I wouldn't have been able to come this far in accomplishing my academic goals if it was not for you. You have been a wonderful mentor; I learnt so much from you in these last five years. Being a mother of young kids can be challenging especially while trying to balance family and a PhD. Thank you for understanding my constraints and granting me the much-needed flexibility of location and hours; you have been very kind in acknowledging my situation.

I would like to thank my committee members Drs Wheeler, Padmanabhan and Jafri for agreeing to serve on my committee and for their valuable input; my co-author Dr. Moradi for his contribution and support. Finally, I would like to thank my fellow lab members for support and discussions; and the Krasnow Institute for Advanced Study for providing this opportunity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

QUANTIFICATION OF NEURON TYPES IN THE RODENT HIPPOCAMPAL FORMATION USING COMPUTATIONAL METHODS

Sarojini Manjusha Attili, Ph.D.

George Mason University, 2021

Dissertation Director: Dr. Giorgio A. Ascoli

Computational biology is a powerful means to achieve a deeper understanding of the brain through modeling and simulations. To accomplish such a task, it is crucial to have the necessary data from all relevant components. Hippocampome.org is a mature knowledge repository that is aimed at understanding the hippocampus which is largely implicated in learning and memory. It has classified neurons in the hippocampal formation – Dentate Gyrus, CA3, CA2, CA1, Subiculum and Entorhinal Cortex into 122 types based on morphological patterns and neurotransmitters. Hippocampome.org houses a wealth of information on neuronal properties such as morphology, electrophysiology, biomarkers, circuitry, and axonal and dendritic densities for 122 neuron types. The ultimate goal of Hippocampome.org is to create biologically plausible computational models of the hippocampus. A key component needed to achieve this objective is the quantification of the population size of the classified neuron types. The addition of these data will enrich the knowledgebase and allow for modeling the hippocampal formation. The goal of my doctoral project is to provide the quantification data for the 122 neuron types classified by Hippocampome.org. Work includes three distinct studies: (1) Cell

numbers, distribution, shape, and regional variation throughout the murine hippocampal formation from the adult brain Allen Reference Atlas, (2) Operations research methods for estimating the population size of neuron types, and (3) Quantification of neuron types in the rodent hippocampal formation by data mining and numerical optimization.

INTRODUCTION

The brain has been an enigma for centuries; the Egyptians first coined a word for brain in hieroglyphic script (Kandel et al., 2013) in the sixteenth century BC. Ever since, there have been several theories and concepts about the mechanism and functionality of the brain. It is during the twentieth century AD that Neuroscience was recognized as a distinct unified academic discipline (GM Shepherd, 2010). Today, we have progressed in leaps and bounds in understanding the brain. Yet, we do not have a clear understanding of how the interaction of neurons produce specific behaviors such as thinking, acting, perceiving, or learning (Kandel et al., 2013). Connectivity of these cells into elaborate circuits underlies cognitive function but also susceptibility to dysfunction. Elucidating the properties of nerve cells and their circuits is critical in understanding disease and in the development of drug targets (Ecker et al., 2017). This requires a comprehensive characterization of the components or cell types of the brain which is the primary goal of the Brain Initiative Cell Census Network (BICCN). The BICCN is one of the seven high priority areas of the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative which was Launched on April 2, 2013. The key goal of the BRAIN Initiative is to develop innovative technologies to interrogate how the brain's cells and circuits interact at the speed of thought and, ultimately, to reveal the complex links between brain function and behavior (Mott et al., 2018).

The present work aims to fulfill a small part of that goal – estimating population size of neuron types for the rodent hippocampal formation. This dissertation is a collection of three projects with that common goal. Each study has been conducted and documented in the form of a chapter with a specific aim as follows.

Chapter 1 describes the first aim which is the quantification of hippocampal cells from the Allen mouse brain reference atlas. The goal of this work is to report the cell-by-cell soma segmentation in every sub-region and layer of the left hippocampal formation through the full rostral-caudal extent, except for the (already well characterized) principal layers of Cornu Ammonis (CA) and Dentate Gyrus (DG). These numbers, although not *sufficient*, are *necessary* for the quantification of distinct neuronal types classified on Hippocampome.org. In Chapter 1, we solve the problem using image analysis and computational techniques. The project involves extraction of the data (Nissl stained images) from the Allen mouse brain reference atlas, transformation of these images into readable data, applying the necessary corrections on the data to obtain accurate numbers, comparing our numbers to stereological counts, conducting analysis on the resultant data and reporting results/trends. The overall cell numbers (~600k cells in entorhinal cortex, ~200k in DG, ~430k in CA1-3, and ~300k in subiculum) and other details have been reported as results. Our results were further analyzed to understand trends related to cell numbers, spatial distributions, shape, size, and bimodality. Bimodal distributions were observed in the 30 different parcels of the hippocampal formation with respect to the cell size. We speculated that this bimodality in cell size could be attributed to cell type mainly

corresponding to neurons and glia. We list the proportion of small cells and misclassified proportions for all 30 parcels in table 3. Large cell proportions have been used to calculate neuronal estimates from total cell counts.

Having obtained the neuronal population estimates using total cell counts and bimodality proportions, we went ahead and used these data in the second chapter which aims at devising a methodology to estimate neuronal types using numerical optimization. This study was conducted to develop a generic solution to obtain population estimates for specific neuron types. Acknowledging the fact that our methodology primarily relied on the research data available, we studied the ways in which we could efficiently put together a solution to the problem. At a high level, a three-step process was followed that consisted of literature search, equation generation, and numerical optimization. We assessed six different numerical optimization algorithms and chose two which resulted in lowest residual errors. This study was crucial as it led to the origination of the method needed to be able to achieve the ultimate goal of this thesis. Chapter 2 provides an in-depth detail of the process for the sample area, dentate gyrus.

We next moved on to using our methodology on the entire hippocampal formation to find the counts for 122 neuronal types. Chapter 3 is the central objective of the thesis. The number of neurons and corresponding ranges for every type identified by Hippocampome.org have been computationally quantified using data from the existing scientific literature and applying the methodology devised in the previous study.

Improvements were made on the previously devised methodology during the research phase; the use of biomarker data was applied to the numerical optimization process which included the addition of inequalities. We could successfully extend our method to the entire hippocampal formation. We estimated the counts and ranges for all 122 neuron types as classified by Hippocampome.org. We examined the proportions of neuronal counts in the six subregions and twenty-six layers, distribution of interneurons expressing 19 types of biomarkers and neuron types with 23 firing phenotypes. Analysis was conducted to investigate the variation of population distributions and numerical densities with respect to volumes. Significant decline was observed in the numerical density of dendritic targeting neurons with increasing volume. Estimated counts of dendritic targeting neuron types were constant across subregions with respect to corresponding volumes indicating that these numbers are independent of the size of the subregion they belong to.

The results from Chapter 3 were consolidated into the updated version of Hippocampome.org. Figure 14 shows a prototype with the quantification matrix. This is a new feature that provides population estimates for the neuron types along with the related evidence and ranges for both rat and mouse. In conclusion, the methodology has been an effective solution to this problem and can be applied across several areas to obtain estimated populations of different cell types given the availability of credible scientific data.

CHAPTER ONE: CELL NUMBERS, DISTRIBUTION, SHAPE, AND REGIONAL VARIATION THROUGHOUT THE MURINE HIPPOCAMPAL FORMATION FROM THE ADULT BRAIN ALLEN REFERENCE ATLAS (SAROJINI M. ATTILI, ..., GIORGIO A. ASCOLI*)

Quantifying the distribution of cells in every brain region is fundamental to attaining a comprehensive census of distinct neuronal and glial types. Until recently, estimating neuron numbers involved time-consuming procedures that were practically limited to stereological sampling. Progress in open-source image recognition software, growth in computing power, and unprecedented neuroinformatics developments now offer the potentially paradigm-shifting alternative of comprehensive cell-by-cell analysis in an entire brain region. The Allen Brain Atlas provides free digital access to complete series of raw Nissl-stained histological section images along with regional delineations. Automated cell segmentation of these data enables reliable and reproducible high-throughput quantification of regional variations in cell count, density, size, and shape at whole-system scale. While this strategy is directly applicable to any regions of the mouse brain, we first deploy it here on the closed-loop circuit of the hippocampal formation: the medial and lateral entorhinal cortices; dentate gyrus (DG); areas Cornu Ammonis 3 (CA3), CA2, and CA1; and dorsal and ventral subiculum. Using two independent image processing pipelines and the adult mouse reference atlas, we report the first cellular-level soma segmentation in every sub-region and non-principal layer of the left hippocampal formation through the full rostral-caudal extent. It is important to note that our techniques excluded the layers with the largest number of cells, DG granular and CA pyramidal, due to dense packing. The numerical estimates for the remaining layers are corroborated by

traditional stereological sampling on a data subset and well match sparse published reports.

Introduction

Neuronal and glial numbers are an important attribute in the characterization of distinct functional regions of the nervous system. Cell counts and densities vary considerably between and within brain areas as well as across species and life-span development (Bayer et al. 1982; Herculano-Houzel et al. 2011; Long et al. 1998; Meyer et al. 2010). These numbers are moreover susceptible to pathologies, pharmacological treatment, and genetic alterations (Fitting et al. 2009; A.M. Insausti et al. 1998; Malberg et al. 2000; Rajkowska 2000). While the ratio between neurons and glial cells is still actively debated (Bahney and Bartheld 2017; Herculano-Houzel et al. 2013; Sherwood et al. 2006), the numbers of neurons in two inter-connected areas relate to circuit convergence and divergence, which are essential design elements of computational processes. Thus, determining the total number of cells in every brain region is fundamental to attaining a comprehensive census of distinct neuronal and glial types (Insel et al. 2013; Kandel et al. 2013; Kim et al. 2017).

The vast majority of reports of cell counts in the neuroscience literature are based on stereological methods (Grady et al. 2003; Schmitz and Hof 2005; West et al. 1991). In these approaches, the number of cells is accurately measured in a small but unbiased proportion of the volume of interest. The numbers for the whole target region can then be extrapolated under the assumption that the sample be representative. Since stereological

neuron counting requires the involvement of a human operator to identify cells in the stained tissue (Schmitdz et al. 2014), the procedure is inherently labor-intensive, time-consuming, and, in particular for cell-sparse regions, impractically inefficient (Boyce and Gundersen 2018). Until recently, the only alternative method that could afford routine comprehensive cellular counting in an entire brain region relied on nuclear identification from homogeneous suspensions (Herculano-Houzel and Lent 2005). While this relatively newer technique distinguishes neurons and glia, it makes it necessary to physically dissect each area of interest and cannot measure geometrical features such as cellular size, shape, or spatial distribution within the tissue.

Continuous growth in computing power and parallel progress in image recognition have substantially altered this status quo (Bhanu and Peng 2000; Peng et al. 2013). Specifically, several algorithms were recently designed to enable high-throughput soma detection and analysis (Hu et al. 2017; Kayasandik and Labate 2016; Luengo-Sanchez et al. 2015; Tapias and Greenamyre 2014; Zhang et al. 2018; Quan et al. 2013). Fully automatic cell segmentation modules are also implemented in freely available mainstream software programs such as ImageJ (Schindelin et al. 2015; Schneider et al. 2012) and CellProfiler (Bray et al. 2015; Lamprecht et al. 2007), allowing robust quantification of soma count, location, and geometry in large-scale applications. In parallel with these computational advances, the public availability of the Allen Brain Reference Atlas (Jones et al. 2009; Lau et al. 2008; Sunkin et al. 2012) provided unprecedented digital access to complete image series of Nissl-stained sections along with regional and laminar delineations.

These new neuroinformatics developments offer a potentially paradigm-shifting alternative to stereological sampling. Taken together, automated cell segmentation and comprehensive online sharing of raw histological datasets provide the opportunity for reliable and reproducible quantification of regional variations in cellular number, size, shape, and spatial distribution at whole-system scale. While this strategy is directly applicable to all regions of the mouse brain, we first deploy it here on the closed-loop of the hippocampal formation, consisting of the six layers of medial and lateral entorhinal cortices and the (three-to-five-layered) dentate gyrus (DG), Ammon's Horn areas CA3, CA2, and CA1, and dorsal and ventral subiculum.

On the one hand, the initial focus on the hippocampal formation reflects the great interest in this structure by the broad neuroscience community (Hasselmo and Stern 2015; Kandel 2004; Moser et al. 2017). On the other, most unbiased stereology techniques have historically been tested on the hippocampus (Boss et al. 1985; Miki et al. 2005; West et al. 1991), and the relative wealth of published information provides a useful validation benchmark. At the same time, even relatively basic questions on the number and density of cell in the mouse hippocampal formation, such as their medial-lateral, anterior-posterior, and laminar distribution in the entorhinal cortex or in the subiculum, remain unanswered.

Using publicly available software and the full images series of Nissl-stained coronal sections from the standard adult mouse brain Allen Reference Atlas, we present

the first complete cell-by-cell soma segmentation in all sub-regions and non-principal layers through the full rostral-caudal extent of the left hippocampal formation. Note that our estimates exclude the bulk of neurons in DG and CA, which are in the densely packed granular and pyramidal layers. A recent estimate for these regions in the mouse placed the total cell numbers (bilateral values) at about 1,250,000 for dentate gyrus, 870,000 for CA1, and 570,000 for CA3 (Murakami et al. 2018; cf. Amaral et al. 1990 for rat). We validate our computational counting approach in all other sub-regions and layers in three ways: by reproducing the analysis with two independent image segmentation pipelines; by confirming the results with traditional stereological estimates on a substantial tissue sample (~10% of total); and by comparing our findings to available data in the published literature. In addition to comprehensive cell counts, the reported quantitative analysis reveals definitive regional variation of soma geometry and spatial occupancy. With this study, we also release open-source all segmented images, the entire database of raw measurements, and our analysis scripts for further community mining.

Materials and methods

Image acquisition. Nissl-stained coronal section images of the adult (8-week old C57Bl/6J male) mouse brain containing any part of the hippocampal formation (sections 64 through 104) were downloaded from the Allen Institute for Brain Science's Mouse Reference Atlas (RRID:SCR_002978) at the highest available resolution (90 dpi). The hippocampal formation is composed of the hippocampus proper, which consists of the dentate gyrus and Cornu Ammonis areas 1-3 (CA1, CA2, CA3), the medial and lateral entorhinal cortices, and the dorsal and ventral portion of the subiculum. The Allen Brain

Atlas further delineates these regions into sub-areas and layers, giving rise to 45 hippocampal structures: strata lacunosum-moleculare, radiatum, pyramidale, and oriens of Cornu ammonis 1 and Cornu ammonis 2 (CA1slm, CA1sr, CA1sp, CA1so, CA2slm, CA2sr, CA2sp, CA2so); strata lacunosum-moleculare, radiatum, lucidum, pyramidale, and oriens of Cornu ammonis 3 (CA3slm, CA3sr, CA3slu, CA3sp, CA3so); the molecular, granular, and polymorphic layers of the dentate gyrus (DGmo, DGsg, DGpo); layers 1, 2, 2a, 2b, 2/3, 3, 4, 4/5, 5, 6a, and 6b of the lateral entorhinal cortex (ENTl1, ENTl2, ENTl2a, ENTl2b, ENTl2-3, ENTl3, ENTl4, ENTl4-5, ENTl5, ENTl6a, ENTl6b); layers 1, 2, 2a, 2b, 3, 4, 5, and 6 of the dorsal zone of the medial entorhinal cortex (ENTm1, ENTm2, ENTm2a, ENTm2b, ENTm3, ENTm4, ENTm5, ENTm6); layers 1, 2, 3, and 5/6 of the ventral zone of the medial entorhinal cortex (ENTmv1, ENTmv2, ENTmv3, ENTmv5-6); and strata moleculare, pyramidale, and radiatum of both the dorsal and ventral parts of the subiculum (SUBd-m, SUBd-sp, SUBd-sr, SUBv-m, SUBv-sp, SUBv-sr). We omitted the pyramidal layers of Cornu ammonis and the granular layer of dentate gyrus from analysis since their cells are too densely packed for effective segmentation with our automated pipeline, leaving 41 structures of the hippocampal formation.

Next, the scalable vector graphic (svg) plates corresponding to each structure and section (654 in total) were retrieved through the Allen Brain Atlas Application Programming Interface (RRID:SCR_005984) to use as masks for cropping individual structures from the entire coronal brain images. Using the freeware program Inkscape.org

(RRID:SCR_014479), each svg mask was sequentially pasted directly on top of the corresponding Nissl-stained image. The two were then clipped together so that the Nissl image was cropped in the shape of the svg mask and exported. This operation was repeated for every downloaded coronal section and svg plate, so that each of the resulting 654 cropped images represented one cross-section of an individual hippocampal formation structure (Fig. 1).

Image processing. We utilized two distinct image segmentation pipelines in parallel for independent processing and analysis of the acquired images: ImageJ (RRID:SCR_003070) and CellProfiler (RRID:SCR_007358). Both software tools read each of the 654 images and followed a series of steps that resulted in object-by-object segmentation of each image (Fig. 1). The ImageJ pipeline involved increasing contrast by 0.3% (ImageJ-suggested value) and enhancing sharpness (unspecified default parameter), conversion to binary image, and setting the minimum size threshold for object detection to the recommended value of 3 pixels. The CellProfiler pipeline involved using four modules: 'UnmixColors' to convert into grey scale and 'IdentifyPrimaryObjects' for optimal object identification (Otsu 1979; Sankur 2004) with the default thresholding factor of 3, 'MeasureObjectSizeShape' for specifying the objects to be measured, and 'ExportToSpreadsheet' for saving the results in the desired format. Altogether, this process generated a segmentation file and a .csv spreadsheet for each of the 654 Nissl stained images. The data processing scripts, related calibration files, and segmented

images are all released in open source

(hippocampome.org/php/data/ABA_Counts_Database.zip).

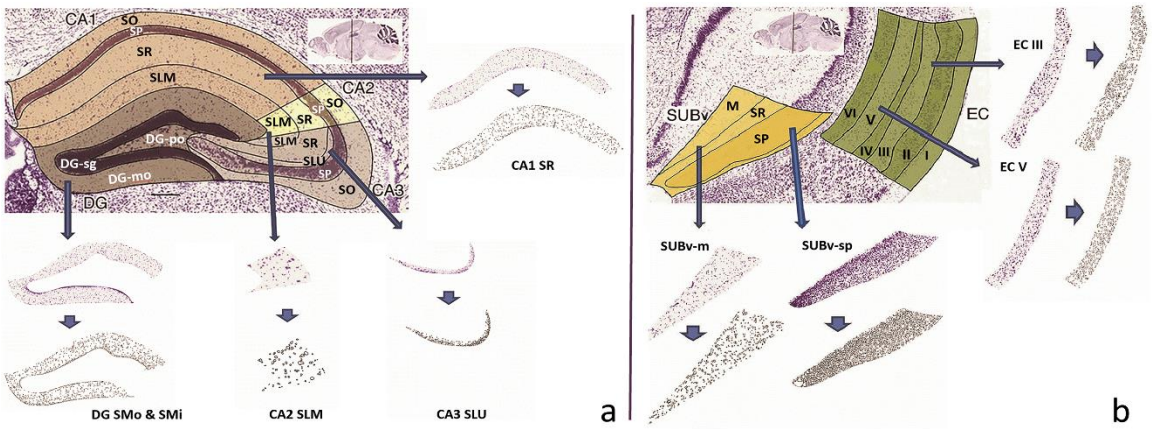

Figure 1 Image source and processing

a. Coronal section of hippocampus areas CA1, CA2, CA3, and Dentate Gyrus (sections illustrated: 74_CA1SR, 74_CA3SLU, 74_CA2SLM and 74_DGmo). b. Coronal section of the entorhinal cortex and the ventral subiculum (sections illustrated: 85_SUBvm and 85_SUBvsp). In both panels, arrows point to the corresponding sample Nissl stained sections as labeled and the shorter arrows point to the segmented images obtained from ImageJ software. Figure was created using ImageJ, Microsoft PowerPoint, and Adobe Photoshop.

Although ImageJ and CellProfiler were remarkably accurate in identifying and segmenting objects from the 2D images as evaluated by visual inspection, the counts could not be taken to reflect cell numbers directly due to four well-known deviations. *First*, several sections exhibited substantial cell clumping, which results in overestimating cell size and underestimating cell count, as each multi-cell clump is generally segmented as a single object. To solve this issue, we applied the corrective watershed algorithm (LaTorre et al. 2013) for separating the clumped cells (Fig. 2a). *Second*, the process of delineating the hippocampal structures within each coronal section splits every border-crossing cell into two, which yields overestimated cell counts and underestimated cell sizes. To alleviate this problem, bordering objects for each section were sorted by area and only the top half were included in the total count for that section, excluding the bottom half (Fig. 2b). *Third*, physical slicing of histological sections inevitably cuts all cells that intersect the surfaces. Since fragments of the dissected cells end up in an adjacent section, cell counts tend to be overestimated and cell sizes underestimated. To address this so-called 'lost caps' scenario (Hedreen 1998a), we applied the Abercrombie (Abercrombie 1946) correction (Fig. 2c). *Fourth*, the use of maximum intensity 2D projections through the slice depth risks missing the occluded cells located under those visible from the top view. To adjust the final counts accordingly, we derived a formula by extending Schellart's theory on count estimates in case of coinciding particles (Hader et al. 2001). Specifically, the tissue can be considered as if divided into "layers" each as thick as the average cell diameter; the number of observed cells (O) then equals the sum of the visible cells in all layers. In the top layer, where there is no occlusion, the number

of visible cells equals the real total number of cells (N) divided by the number of layers. In the second layer, the expected fraction of occluded cells corresponds to the areal occupancy of the top layer (the sum of the areas of all cells in the top layer, A, divided by the area of the section (S); thus the number of visible cells equals the real total number of cells (N) divided by the number of layers discounted by that areal occupancy. In the case of two layers, the above reasoning can be summarized in the following equations: $O = \frac{N}{2} + \frac{N}{2}\left(1 - \frac{A}{S}\right) = \frac{N}{2}\left(2 - \frac{A}{S}\right)$, which can be easily solved as: $N = 2O * \frac{S}{2S-A}$. In the case of three layers (Fig. 2d), the expected fraction of occluded cells in the third layer corresponds to the summed areal occupancy of the first and second layers. In this scenario, the formula relating the number of observed cells O to the real number of cells N based on the areal occupancy A/S is:

$O = \frac{N}{3} + \frac{N}{3}\left(1 - \frac{A}{S}\right) + \frac{N}{3}\left(1 - \frac{2A}{S} + \frac{A^2}{S^2}\right)$, which can be solved as: $N = 3O * S^2/(3S^2 - 3A * S + A^2)$.

Shape analysis, bimodality, and spatial distributions. Total cell counts were quadrupled since every fourth section from the coronal brain series of Allen's brain map was Nissl stained and all extracted measurements were converted into metric units multiplying by the reported pixel size of 1.047 µm on each side (Allen Data Production 2011). For each processed image, we extracted two measurements with ImageJ and CellProfiler from every segmented cell: the section area in squared micrometers; and the circularity, defined as $4\pi * \frac{area}{perimeter^2}$, where perimeter is the length of the cell

segmentation. While the area quantifies the cell size, circularity characterizes its shape, with a value of 1.0 corresponding to a perfect circle and values closer to 0 indicating increasingly elongated or tortuous shapes.

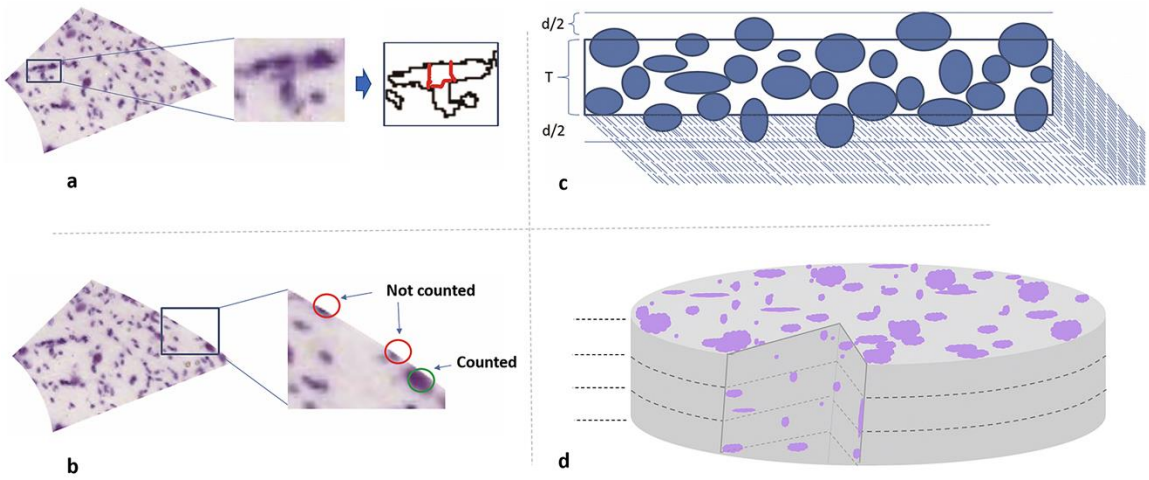

Figure 2 Computing cell counts from segmented object

a. Segmentations were de-clumped using the watershed algorithm (illustrated section: 70_CA2SLM). b. Bordering cells were sorted based on cell area and only upper half was counted for that section while lower half was considered to belong to neighboring areas. c. Cut cells due to sectioning were accounted for using Abercrombie formula: $N = n * [\frac{T}{T+d}]$, where N is the number of cells after correction, n is the number of all detected objects before correction, T is the section thickness, and d is the mean diameter. d. Section thickness is divided into equal layers where height of each layer equals mean cell diameter; occluded cells in the depth of the tissue were accounted for using the formula

for count estimates of aligned particles (see Methods). Figure was created using ImageJ, Microsoft PowerPoint, and Adobe Photoshop.

Analysis was conducted on all cells counted to understand the presence of multimodal distributions in the cell populations based on the size attribute. Kernel density estimation (KDE) plots were generated for each of the 30 parcels of the hippocampus. The KDE plots were fitted with a mixture of two Gaussian distributions, yielding for every parcel a mean and a standard deviation for each of the Gaussians and the relative weight between the two. The proportions of small and large cells per parcel as well as their Gaussian overlap were calculated from these parameters. Hartigan's Dip test (Maechler 2016) was run on all parcels to test for bimodality/multimodality.

Furthermore, for each processed image we computed three parameters capturing the overall spatial distribution of the cells: the first was volumetric density, defined as the total number of cells in the section divided by the section volume (the product of the mask area by the nominal thickness). The second parameter was the real occupancy (utilized in the occlusion correction described above), defined as the summed area of cells divided by the mask area. The third parameter defines the tiling or clustering tendency of the cells using two-tailed t-test statistics of the nearest neighbor distance (NND) distribution against the null hypothesis (Andrey et al. 2010). Specifically, we

randomly distributed within each mask area a number of points identical to that measured in the corresponding section. We then extracted the NND for every point using both the real and random locations. Lastly, we t-tested the real NND distribution against the random NND distribution. A significantly greater NND than random after Bonferroni correction indicates spatial tiling, while a significantly smaller NND than random indicates spatial clustering.

We computed the average and coefficient of variation of the individual cell-level measurements (size and shape) within each section and used the section average in subsequent analyses. We then analyzed all measurements section-by-section rostro-caudally within each hippocampal structure as well as compared the combined sections across structures. The 23 Allen Brain Atlas subdivisions of the medial and lateral regions of the entorhinal cortex were collated into 12 parcels to match the standard nomenclature of Hippocampome.org (Wheeler et al. 2015) as it follows:

- ENTl1 was re-termed LEC I
- ENTl2, ENTl2a, ENTl2b and half the number of cells of ENTl2-3 were combined into LEC II
- ENTl3 and half the number of cells from ENTl2-3 were combined into LEC III
- ENTl4 and half the number of cells from ENTl4-5 were combined into LEC IV
- ENTl5 and half the number of cells from ENTl4-5 were combined into LEC V
- ENTl6a and ENTl6b were combined into LEC VI

- ENTm1 and ENTmv1 were combined into MEC I
- ENTm2, ENTm2a, ENTm2b and ENTmv2 were combined into MEC II
- ENTm3 and ENTmv3 were combined into MEC III
- ENTm4 was re-termed MEC IV
- ENTm5 and half the number of cells from ENTmv5-6 were combined into MEC V
- ENTm6 and half the number of cells from ENTmv5-6 were combined into MEC VI

Stereological sampling. In order to compare our computational pipeline to traditional stereological measurements, we counted an unbiased sample of cells in 65 sections (10% of total) using the 2D probe 'MBF Fractionator' of the standard commercial software MicroBrightField StereoInvestigator (RRID:SCR_004314). The 65 images were selected randomly by picking 1-2 from each hippocampal structure. For each image, we traced the contour of interest and adjusted the size of the counting frame to allow reliable object identification and accurate marking (this varied for each image depending on the size and distribution of cells). We then specified the size of the Systematic Random Sampling grid based on the section size to ensure a representatively large sample, which is necessary for accurate statistical estimates. Once the grid was placed on the section, we marked each visible object inside the frame or on the inclusion line; after marking all objects in the grid, the process was moved to the next sampling site where the grid was placed. Once all the sampling sites were marked, the process was completed, and the probe run list

displayed the estimated population for the section. The data files and results were exported and are included in the shared database (hippocampome.org/php/data/ABA_Counts_Database.zip).

Results

Quantitative validation of overall approach. The cell identification process described in the "Methods" was critically assessed in three distinct ways: by cross-examining the segmentation results from the two independent software frameworks (ImageJ and CellProfiler); by comparing the count results to available data published in the peer-reviewed literature; and by repeating the analysis on a representative subset of the images with traditional unbiased stereology.

Visual inspection of the ImageJ and CellProfiler segmentations against the original images revealed remarkable consistency between the two programs as well as with intuitive evaluation. Specifically, the majority of the objects we would have manually classified with high confidence as cells upon visual inspection were identified as such by both automated pipelines, which also accurately and similarly delineated the body perimeters. When we reviewed a sample of the cells segmented by only one of the two programs and 'missed' by the other, most were difficult calls that we could classify as cells only with low confidence. We found no cases of objects identified and segmented by both ImageJ and CellProfiler, which we did not consider to be likely cells. On quantitative analysis, the discrepancy in overall cells count between the two programs in

the bulk of individual images was within 10%, with an overall difference of less than 103,000 out of a total over 1.5 million. Table 1 reports a summary of these comparisons by anatomical area along with data from published studies (see below). At the level of whole hippocampal formation (excluding the granule cell layer of the dentate gyrus and the pyramidal cell layers of CA1-3), the overall count difference between the two software programs was minimal, with an average absolute difference of 7% at the sub-region level. The exact image-by-image counts for both ImageJ and CellProfiler are included in the shared database.

Table 1 Validation of two independent image segmentation methods with previous studies

(Fitting et al. 2009; Mulders et al. 1997; Ramsden et al. 2003; Lister et al. 2006; Sousa et al. 1998; Rasmussen et al. 1996; Kim et al. 2017; Long et al. 1998; Kaae et al. 2012; Murakami et al. 2018; Erö et al. 2018; Bezaire et al. 2016; Andrade et al. 2000; Herculano-Houzel et al. 2013; Grady et al. 2003)

Region	Source	Counts	% Difference from ABA (ImageJ)
DG (except granule layer)	ABA (ImageJ)	196,067	
	ABA (CellProfiler)	176,748	-9.85
	Literature	253,286	29.18
	ABA (ImageJ)	173,200	

CA2+CA3 (except pyramidal layers)	ABA (CellProfiler)	154,733	-10.66
	Literature	176,728	2.04
CA1 (except pyramidal layer)	ABA (ImageJ)	252,533	
	ABA (CellProfiler)	247,169	-2.12
	Literature	227,388	-9.96
SUB	ABA (ImageJ)	294,294	
	ABA (CellProfiler)	279,333	-5.08
	Literature	289,171	-1.74
EC (lateral & medial)	ABA (ImageJ)	602,400	
	ABA (CellProfiler)	557,569	-7.44
	Literature	684,990	13.71
Total	ABA (ImageJ)	1,518,495	
	ABA (CellProfiler)	1,415,552	-6.78
	Literature	1,631,563	7.45

Comparing the counts obtained with our approach to existing data in earlier scientific studies (mostly using unbiased stereology or nuclear counts) requires a series of assumptions to pool results from a variety of diverse experimental procedures. First, the vast majority of reports on cell counts in the rodent hippocampal formation to date investigated rats rather than mice. In order to relate numerical values between species, we applied the linear scaling parameters reported for cortical structures (Herculano-Houzel et al. 2006): namely, the number of mouse cells in the cerebral cortex (including the hippocampus) equals approximately 34% of the number of rat cells; the number of mouse neurons equals approximately 41% of the number of rat neurons; and the number of mouse glial cells equals approximately 29% of the number of rat glial cells. A summary

of the derivation of these scaling rules from the source data is included in the Supplementary Material (Table S1). We utilized the same scaling factors across all sub-regions of the hippocampal formation.

We compiled all other additional assumptions, each specific to the interpretation of individual reports or handling of missing data, in the Supplementary Material as well (Table S2). Lastly, the Supplementary Material (Table S3) also summarizes the step-by-step computations to derive the literature values utilized in Table 1 for each separate sub-region (DG: Table S3a; CA2/3: Table S3b; CA1: Table S3c; subiculum: Table S3d; and entorhinal cortex: Table S3e).

The cell counts obtained with the above-described workflow from both image processing software programs were compared against literature-reported values (Table 1) sourced from 15 distinct studies (Fitting et al. 2009;
Mulders et al. 1997; Ramsden et al. 2003; Lister et al. 2006; Sousa et al. 1998; Rasmussen et al. 1996; Kim et al. 2017; Long et al. 1998; Kaae et al. 2012; Murakami et al. 2018; Erö et al. 2018; Bezaire et al. 2016; Andrade et al. 2000; Herculano-Houzel et al. 2013; Grady et al. 2003). The overall differences between our approaches and the average literature data for the whole hippocampal formation overall is remarkably contained (~8% for ImageJ and ~15% for CellProfiler). The average absolute difference at the sub-region level (<15%) was considerably smaller (less than half) than the typical variability between experimental studies (Fitting et al. 2009; Mulders et al. 1997;

Ramsden et al. 2003; Lister et al. 2006; Sousa et al. 1998; Herculano-Houzel et al. 2013; Grady et al. 2003).

As additional validation of our approach, we subjected a representative subset of the data (~10% of the images) to unbiased stereology (Fig. 3). This dataset spanned every parcel and ranged from very sparse (<20 cells per image) to considerably dense (>1000 cells per image). Stereological estimates very strongly correlated with our comprehensive counts on the same images ($R > 0.99$, $p < 10^{-6}$). The overall counts between the two methods differed by less than 5%, and the absolute deviation on an image-by-image case averaged less than 10%. The section-by-section stereological counts are also included in the shared database.

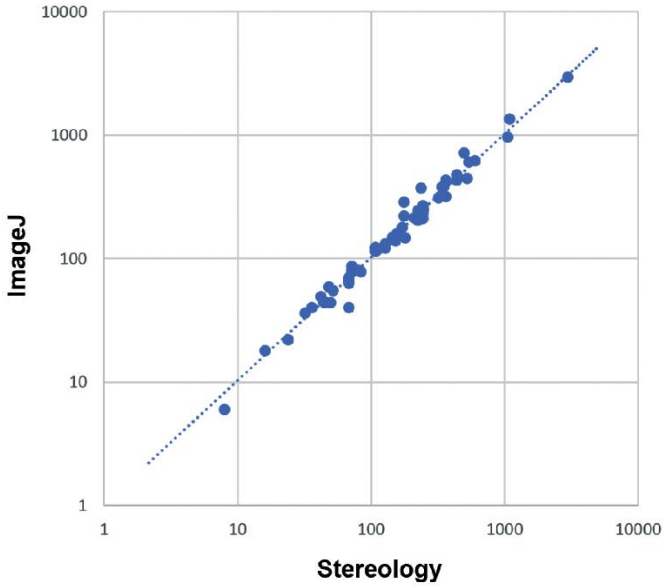

Figure 3 Comparing image processing methods to stereological methods

Figure was created using Microsoft Excel and Adobe Photoshop

Cell numbers, spatial distributions, shape, and size. After validating the analysis pipeline, we collated the measurements of cell count, spatial distribution, shape, and size for each of 30 anatomical parcels (Table 2) to reveal differences and similarities across hippocampal sub-regions and layers.

Table 2 Numbers, density, spatial occupancy, size, and shape.

Area and circularity are reported as averages (values in parenthesis are the coefficients of variation).

Parcel	Counts	Density (mm^{-3})	Tiling Ratio	Circularity Avg (CV)	Area (μm^2) Avg (CV)
DGmo	155,548	75,610	3/30	0.70 (0.28)	69.52 (1.16)
DGpo	40,519	111,556	5/23	0.65 (0.28)	122.08 (0.97)
CA3slm	2,665	72,650	0/13	0.71 (0.25)	49.65 (0.72)
CA3sr	53,065	67,964	0/24	0.73 (0.26)	61.65 (0.98)
CA3slu	22,348	114,525	0/22	0.59 (0.36)	82.95 (1.16)
CA3so	75,530	84,499	0/24	0.68 (0.28)	67.63 (0.93)
CA2slm	5,656	75,103	0/15	0.74 (0.24)	50.04 (0.72)
CA2sr	7,136	71,148	0/16	0.72 (0.26)	64.33 (0.97)
CA2so	6,800	81,953	0/16	0.73 (0.26)	62.14 (0.95)
CA1slm	94,865	80,619	2/24	0.74 (0.23)	56.64 (0.77)
CA1sr	82,744	58,478	0/24	0.67 (0.29)	84.79 (0.93)
CA1so	74,924	84,020	0/24	0.69 (0.26)	67.08 (0.84)
SUBdm	11,283	93,639	0/13	0.73 (0.26)	59.11 (0.85)
SUBdsr	10,203	73,840	1/13	0.72 (0.26)	63.61 (0.84)
SUBdsp	53,216	108,550	13/13	0.66 (0.25)	133.30 (0.69)
SUBvm	33,832	84,815	0/11	0.74 (0.25)	51.15 (0.92)
SUBvsr	21,273	81,799	0/10	0.70 (0.27)	62.96 (0.95)
SUBvsp	164,487	119,233	14/14	0.66 (0.25)	140.13 (0.70)
LEC I	77,508	100,409	0/35	0.63 (0.38)	55.59 (1.12)
LEC II	98,750	87,268	17/53	0.65 (0.26)	154.50 (0.82)
LEC III	62,936	77,518	9/32	0.68 (0.24)	131.62 (0.77)
LEC IV	28,113	63,647	7/28	0.69 (0.26)	109.69 (0.75)
LEC V	59,488	73,429	13/29	0.70 (0.23)	118.16 (0.77)
LEC VI	77,690	109,712	22/54	0.69 (0.23)	122.04 (0.66)
MEC I	37,217	99,948	0/19	0.69 (0.29)	47.21 (1.11)
MEC II	66,365	126,570	13/21	0.66 (0.25)	134.05 (0.90)
MEC III	45,397	121,965	13/17	0.67 (0.25)	134.78 (0.98)
MEC IV	7,251	132,969	2/10	0.66 (0.28)	91.63 (0.67)
MEC V	15,585	139,513	4/14	0.67 (0.26)	104.85 (0.69)
MEC VI	26,101	140,819	10/13	0.67 (0.24)	111.90 (0.68)

The number of cells in each parcel varies widely from less than 2700 in stratum lacunosum-molecular of CA3 to more than 160,000 in stratum pyramidale of ventral subiculum. This broad range primarily reflects the known overall volumes of each parcel (for rat hippocampus proper, see e.g. Ropireddy et al. 2012). In addition, we analyzed the results at the individual section-by-section level to investigate any trends in the data within each parcel along the rostro-caudal direction. The relative distribution of cells along the rostro-caudal extent was considerably non-uniform, but tightly corresponded to each relative section area, with a tendency for relatively higher values towards the caudal end for most parcels (Fig. 4). In quantitative terms, the number of cells within each section significantly correlated with that section's mask area ($R=0.9$, $p<10^{-6}$). This indicates that volume accounts for more than 80% of the within-parcel rostro-caudal variance in cell numbers. In other words, the volumetric cell density is essentially constant within each parcel throughout the rostro-caudal extent. Between parcels, in contrast, the volumetric cell density ranged broadly from 58,478 mm^{-3} in CA1 stratum radiatum to 140,000 mm^{-3} in layer 6 of medial entorhinal cortex.

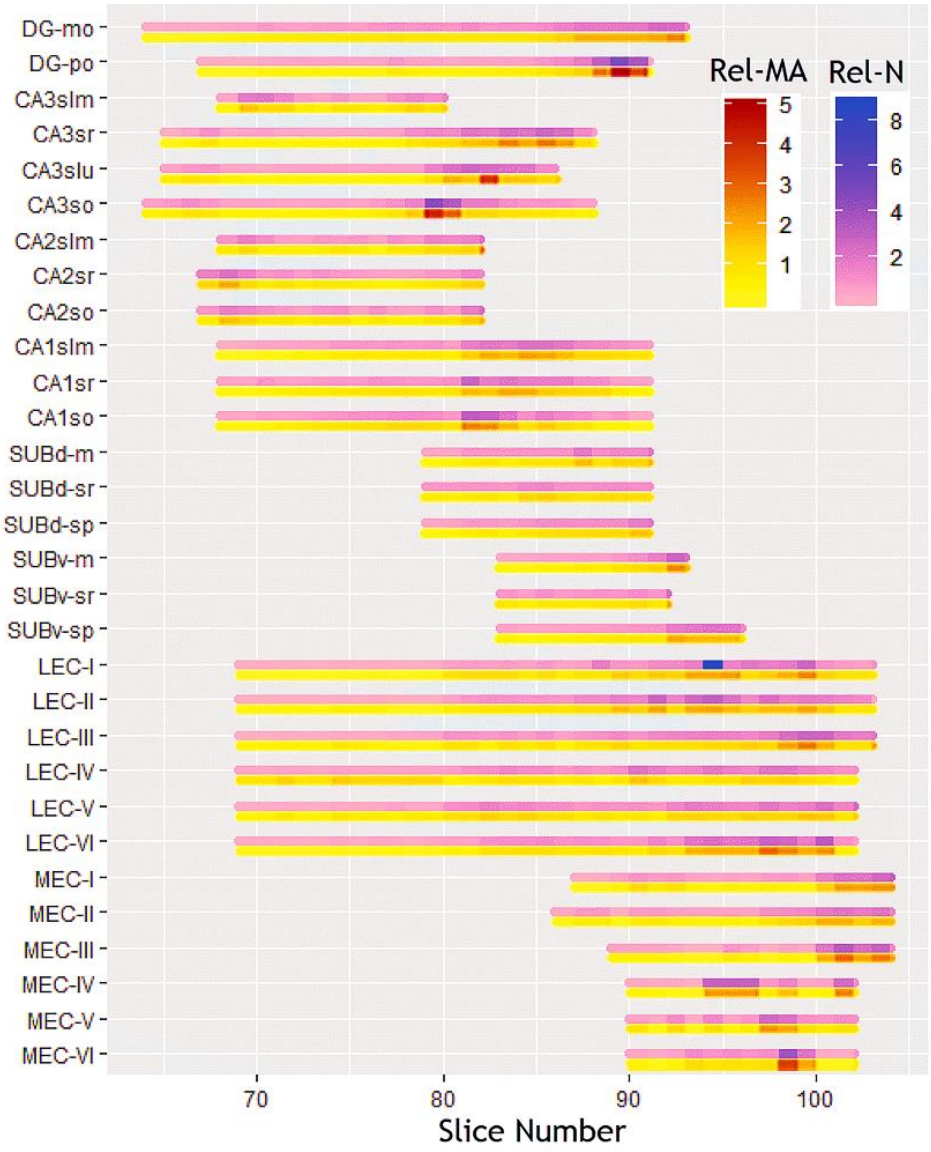

Figure 4 Rostro-caudal gradients of relative cell counts (N) and mask areas (MA)

Figure was created using Microsoft Excel, R, and Adobe Photoshop

Our statistical analysis of the spatial distributions of cell locations within each parcel was designed to identify two opposite types of deviations from the null hypothesis of random distribution: cell clustering (whereas distances from the nearest neighbors are typically smaller than random) and cell tiling (whereas distances from the nearest neighbors are typically larger than random). None of the 654 sections in our study displayed evidence of clustering (Table 2): 152 (23%) revealed significant tiling (Fig. 5)

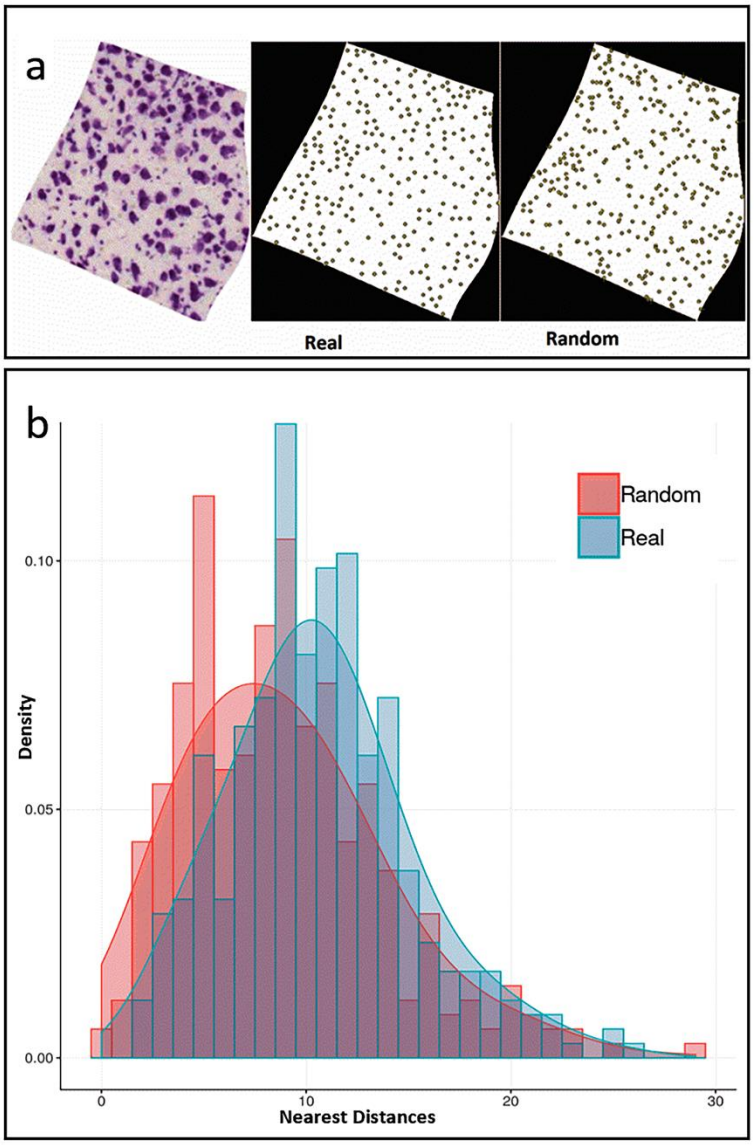

Figure 5 Spatial distributions

a. Nissl stained image (left) for a representative image (79_ENTl4-5), segmentation centroids (center), and randomized points (right); b. Frequency histograms and corresponding probability density functions for real and randomized nearest distances. Figure was created using Microsoft Excel, Microsoft PowerPoint, R, and Adobe Photoshop.

after Bonferroni correction and 502 (77%) could not be distinguished from random. Only the pyramidal layer in dorsal and ventral subiculum consistently featured cell tiling in all sections. The other subicular layers as well as layer 1 in medial and lateral entorhinal cortex and all CA2 and CA3 parcels had nearly no tiling in any section. The other entorhinal layers and dentate gyrus had an intermediate proportion of tiling sections, ranging from 10% in the dentate molecular layer to 76% in medial entorhinal layer 3, with no observable rostro-caudal trend. Interesting, the proportion of tiling section was significantly correlated to density across the 30 parcels ($R=0.72$, $p<10^{-5}$), suggesting a greater pressure for optimal space occupancy in cell-denser areas.

Cell shape was remarkably uniform throughout the hippocampal formation, with the average circularity only ranging between 0.59 (CA3 lucidum) and 0.74 (stratum lacunosum-moleculare CA1 and CA2, and molecular layer of ventral subiculum) and contained within-parcel coefficients of variations between 0.23 (CA1 lacunosum-moleculare, lateral entorhinal cortex layers V and VI) and 0.38 (lateral entorhinal cortex layer I). Circularity values were very consistent between image analysis pipelines (ImageJ vs CellProfiler) and did not vary along the rostro-caudal axis.

In contrast, cell size varied considerably from parcel to parcel, ranging in section area from less than 50 μm^2 in medial entorhinal layer 1 to more than 150 μm^2 in lateral entorhinal layer 2. As a general observation, cells in all layers of areas CA1-3 except CA3 lucidum and CA1 radiatum were twice as small as cells in all layers of the entorhinal cortex with the stark exceptions of both medial and lateral layers 1. The situation was more varied within the dentate gyrus, with cells in the polymorphic layer on average twice as large as in the molecular layer, and in the subiculum, with cells in the pyramidal layers two-and-a-half the size of those in the molecular layer. Cell size was moderately variable within parcel, with coefficient of variations contained between 0.66 and 1.16. No rostro-caudal gradients in cell size were observed in any parcel possibly with the sole exception of lateral entorhinal layer 3, in which rostral cells were marginally larger than caudal ones (similar to micrographs published in R. Insausti et al. 1998), though the trend was not statistically significant.

Parallel lines of evidence point to functional (Brun et al. 2008), physiological (Giocomo and Hasselmo 2008), and developmental (Ray and Brecht 2016) differences between the dorsal and ventral regions of the medial entorhinal cortex. Thus, we analyzed the cell size in the 10 sections of the Allen Reference Atlas that separately delineate dorsal and ventral MEC. We found that layer 2 cells (a large portion of which are spiny stellate neurons) were 40% larger and twice more variable in ventral than in dorsal MEC ($181.71 \pm 229.20 \mu m^2$, N=544 vs. $129.41 \pm 97.40 \mu m^2$, N=605; $p<10^{-8}$). Interestingly, the opposite relation was observed in layers 3 (dorsal: $136.0 \pm 89.59 \mu m^2$, N=225; ventral:

106.67±58.98μm^2, N=70; p<0.0006) and 5-6 (dorsal: 117.03±71.65μm^2, N=292; ventral: 84.98±57.10μm^2, N=45; p<10^{-7}), which are all dominated by pyramidal neurons. In contrast, we observed no dorsal-ventral gradients in the rest of the hippocampal formation.

Visual inspection of the Nissl images and corresponding processed segmentations suggested that the cell size distribution within a given parcel and section was typically not regular or uniform, but often skewed, with a substantial number of smaller cells within a more restricted size range and a broader right-tail distributed population of larger cells. Dip test multimodality analysis (Maechler 2016) revealed statistically significant bimodal distributions in a majority of parcels (23/30: Table 3).

Table 3 Small cell proportion, misclassified proportion, average areas of small and large cells.

(values in parenthesis are the coefficients of variation). * denotes significant dip test results.

Parcel	Proportion of small cells %	Misclassified proportion	Small Cells Area (μm^2) Avg (CV)	Large Cells Area (μm^2) Avg (CV)
DGmo*	61.24	0.14	37.46 (0.33)	85.81 (0.50)
DGpo*	55.18	0.11	49.78 (0.54)	187.17 (0.57)
CA3slm	65.45	0.14	35.54 (0.30)	74.46 (0.55)
CA3sr*	73.56	0.07	37.20 (0.39)	110.13 (0.51)
CA3slu*	60.00	0.13	37.15 (0.70)	139.15 (0.64)
CA3so*	60.70	0.11	35.78 (0.34)	105.19 (0.67)
CA2slm	69.00	0.10	33.96 (0.32)	76.26 (0.46)
CA2sr	68.18	0.11	37.27 (0.37)	95.82 (0.52)
CA2so	58.26	0.15	32.96 (0.34)	73.50 (0.48)
CA1slm*	62.00	0.14	37.77 (0.24)	72.43 (0.50)
CA1sr*	56.00	0.15	42.83 (0.42)	120.97 (0.54)
CA1so*	57.74	0.08	34.84 (0.31)	102.76 (0.53)
SUBdm*	59.03	0.14	35.60 (0.25)	71.70 (0.46)
SUBdsr*	67.06	0.10	37.33 (0.34)	100.78 (0.51)
SUBdsp*	21.61	0.16	47.19 (0.44)	153.32 (0.64)
SUBvm*	58.19	0.15	33.03 (0.32)	75.07 (0.71)
SUBvsr*	52.00	0.15	35.47 (0.40)	100.46 (0.75)
SUBvsp*	19.73	0.17	54.02 (0.46)	147.51 (0.58)
LEC1*	62.30	0.14	28.26 (0.44)	72.26 (0.54)
LEC2*	20.00	0.14	46.37 (0.42)	174.73 (0.63)
LEC3*	23.99	0.14	42.98 (0.41)	148.94 (0.59)
LEC4*	39.58	0.12	41.77 (0.46)	148.37 (0.51)
LEC5*	41.42	0.11	47.54 (0.44)	151.35 (0.46)
LEC6*	75.53	0.14	92.07 (0.57)	202.30 (0.48)
MEC1*	67.18	0.13	27.21 (0.50)	73.16 (0.68)
MEC2*	62.91	0.17	75.51 (0.62)	181.78 (0.43)
MEC3	76.01	0.12	91.36 (0.58)	223.74 (0.47)
MEC4	19.02	0.16	33.53 (0.40)	100.05 (0.55)
MEC5	53.00	0.20	63.78 (0.51)	138.33 (0.46)
MEC6*	76.43	0.14	77.19 (0.58)	164.28 (0.42)

The kernel density estimate (KDE) of all distributions could be well fitted with a Gaussian mixture model (Fig. 6), allowing the determination for every anatomical parcel of the mean and variance of smaller and larger cells, their relative proportion, and the fraction of cells that would be assigned to the incorrect group based on a sharp size threshold ("misclassified" area under the cross-over of the two Gaussians). Small cell areas varied considerably between parcels, from less than 30 μm^2 in medial entorhinal layer 1 to more than 90 μm^2 in lateral entorhinal layer 6. Similarly, large cell areas also varied over 3-fold between ~72 μm^2 in the molecular layer of the dorsal subiculum to ~224 μm^2 in layer 3 of the medial entorhinal cortex. The fraction of small cells was balanced overall (average 55%) but ranged widely even across adjacent parcels from more than three-quarters in medial entorhinal cortex layer 3 to less than one-fifth in layer 4. The proportion of misclassified cells was relatively modest, ranging from 7% in CA3 radiatum to 20% in medial entorhinal cortex layer 5 (average over all parcels: 13.4%).

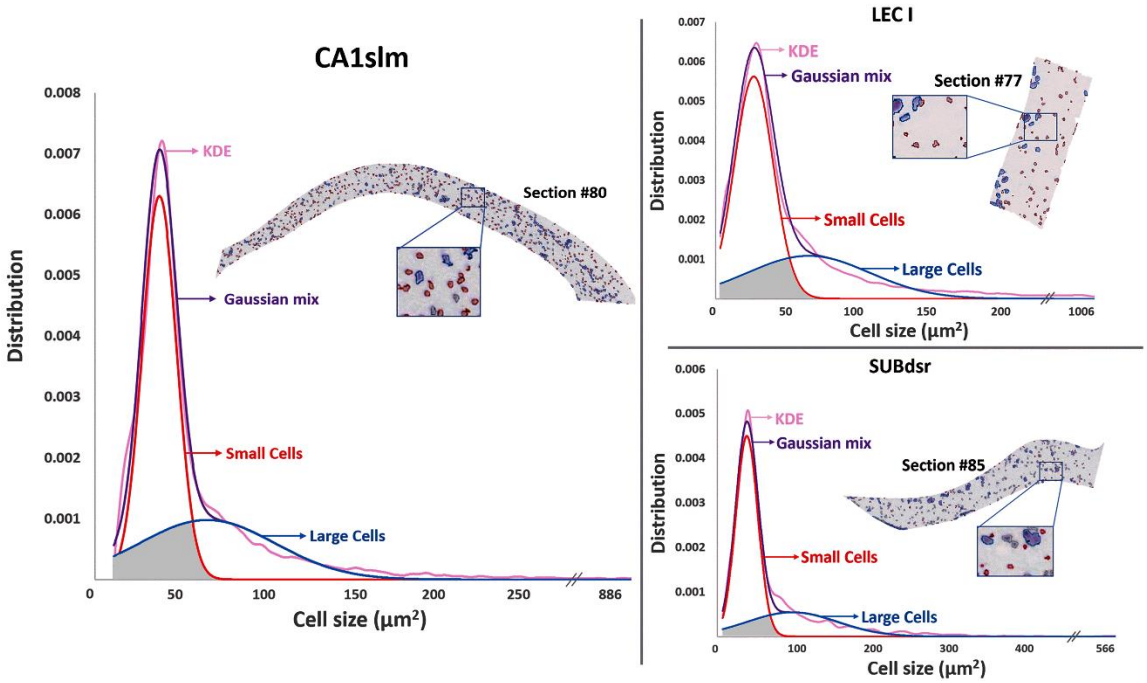

Figure 6 Bimodal cell size distributions of CA1slm (left), LEC I (top right), and SUBdsr (bottom right) with kernel density estimates (KDE) fitted by a Gaussian mixture model. The representative Nissl sections in the insets highlight small and large cells in blue and red, respectively. The gray area under the Gaussian curves represents the proportion of 'small' and 'large' cells that would be misclassified if selected based on a hard size threshold. Figure was created using Microsoft Excel, Microsoft PowerPoint, and Adobe Photoshop.

Data availability. The hippocampome.org/php/data/ABA_Counts_Database.zip archive contains all project files and scripts associated with this report, namely: masked segmented images (output files from ImageJ and CellProfiler) for each individual parcel and slice; scripts for corrections and spatial analysis in R; ImageJ scripts for processing, corrections and mask area calculations; CellProfiler calibration files; and stereology data and results files. The accompanying 'readme' document describes each folder and file in detail with instructions on usage. This content is freely available for further research or re-analysis.

DISCUSSION

The advent of high-throughput histology, accelerating progress in image processing, continuous increase in computing power, and the pervasive accessibility of informatics tools have ushered the era of whole brain cell-by-cell comprehensive analyses into the changing world of neuroscience research. Although there are valid reasons to still favor stereological sampling for certain applications (Schmitz et al., 2014), several recent applications have demonstrated the feasibility of systematically identifying all cell bodies in a mouse brain (Kim et al. 2017; Murakami et al. 2018). Most criticisms of "biased" sampling estimates such as in the Abercrombie formula (Hedreen 1998b) do not apply when considering comprehensive counts across the entire cellular population in a system. Here we show that one of the most popular and widely adopted resources in modern neuroscience, the adult mouse brain Allen Reference Atlas, can be used to obtain cellular-level segmentations throughout entire anatomical formations from the original Nissl stained images. After appropriate corrections to account for sectioning distortions,

cell agglomerates, border crossings, and tissue occlusion, we demonstrate that this approach is reproducible across independent image processing pipelines, accurate as matched against unbiased stereology, and reliable in comparison to available published values in the scientific literature. As a notable exception, dentate granule layer and Cornu Ammons pyramidal layers were excluded from this process as individual cells cannot be reliably identified in densely packed Nissl stained images. However, the total cell counts for these areas can be estimated by various other methods and have been reported for both rat and mouse in seminal studies over the past three and half decades (e.g. Bayer et al. 1982; Amaral et al. 1990; West et al. 1991; Rapp and Gallagher 1996; Rasmussen et al. 1996; Baldwin et al. 1997; Mulders et al. 1997; Calhoun et al. 1998; A.M. Insausti et al. 1998; Sousa et al. 1998; Lister et al. 2006; Hosseini-Sharifabad and Nyengaard 2007; Fitting et al. 2009; Kaae et al. 2012; Murakami et al. 2018).

A key advantage of comprehensive cell segmentation is that, in addition to providing complete counts in each anatomical parcel, it also enables quantitative analysis of cell shape, spatial occupancy, and size. We discovered that, while cross-section circularity (a simple shape measure) is relatively invariant throughout the hippocampal formation, spatial occupancy differs considerably across different parcels, but remains essentially constant within parcels along the entire rostro-caudal extent. Layers 2, 3, 5 and 6 of the medial entorhinal cortex exhibit the highest numerical densities (as also noted in Canto et al. 2008) along with the pyramidal layer of ventral subiculum. This is not surprising since these parcels are especially rich in principal cells, though not in the extreme packing

fashion of the dentate granule layer and CA1-3 pyramidal layers, which were excluded for that reason from analysis. Interestingly, higher occupancy correlated significantly with the fraction of images that demonstrated cell tiling, possibly indicative of pressure towards optimal placement. In contrast, we never observed statistically significant cell clustering. At first, this fact may seem at odd with reports describing cell islands or studies distinguishing island and ocean cells, especially in layer 2 of medial entorhinal cortex. At closer inspections, however, the existing experimental evidence is limited to primates (Goldenberg et al. 1995) or molecularly segregated (Sun et al. 2015) or projection defined (Kitamura et al., 2014) cell populations. Seminal studies in rats mention cell clusters, but only in a subset of sections (R. Insausti et al. 1998). It will be interesting for future studies to ascertain the extent of species specificity for this phenomenon.

Cell size also changed substantially among parcels but was additionally non-uniform within parcels, though again without a discernible rostro-caudal trend. Neuroscientists have maintained a long-standing interest in attempting to distinguish major cell classes on the basis of their somatic sizes. Unfortunately, however, GABAergic cortical interneurons span almost an order of magnitude of cell body areas from approximately 80 μm^2 to more than 700 μm^2 (Ascoli et al. 2008), while hippocampal pyramidal neurons fall well within that range with typical somatic areas of the order of 200 μm^2 (Nakatomi et al. 2002). Thus, it remains untenable to distinguish projection excitatory and local inhibitory neurons in the rodent hippocampal formation solely based on their soma

dimension. However, the main classes of glial cells are generally characterized by smaller cell bodies, from microglia in the 20-40 μm^2 range (Long et al. 1998) to oligodendrocyte and astrocytes in the 40-80 μm^2 range, respectively (Fitting et al. 2009). Thus, it may be possible at least in principle to separate neurons and glia on the basis of cell size.

The size bimodality analysis presented in this study revealed average areas of 45 μm^2 and 122 μm^2 for the "small" and "large" cells, respectively, across the 30 anatomical subdivisions of the hippocampal formation. These values are approximately consistent with the previously reported sizes of glia and neurons; moreover, the overall fraction of small cells (~55%) is also in line with the expected proportion of glial cells (Herculano-Houzel et al. 2006). Despite these convergent indications, extreme caution needs to be exercised in this interpretation, and the assignment of neurons and glia based on cell size should be deemed tentative at best. With these caveats in mind, it is still worth noting that, based on our segmentation results, layer 1 in medial entorhinal cortex has the smallest cell area across the hippocampal formation. Interestingly, independent experimental evidence suggests that this anatomical parcel is nearly devoid of neurons, whereas the glial density is approximately uniform across layers (Wu et al. 2005; Witter 2011). It is thus tempting to speculate that the number of cells we counted from those sections might correspond mostly or entirely to glia. Glial distribution is also known to vary across the hippocampus: in particular, stratum lacunosum-moleculare displays a higher density of both oligodendrocytes (Vinet et al. 2010) and microglia (Jinno et al. 2007). Intriguingly, this layer also displayed the smallest cell area in all three CA areas.

Direct quantification of distinct cell types (neurons and the three main classes of glia) from Nissl stained images is possible in principle but requires a higher resolution than currently available in public repositories to enable the examination of physical features of individual cells. At high resolution, specific rules can be applied to identify different cell types such as neurons, astrocytes, and microglia based on size, intensity, or texture (Garcia-Cabezas et al. 2016; Rajkowska et al. 2016). Appropriate thresholds for these characteristics can thus be incorporated in existing image processing software tools to efficiently classify objects directly upon segmentation. Similar methodologies can be adapted to quantify any cell type in biological sections as long as appropriate labels are available to mark them. Nonetheless, it is also important to consider that post-mortem handling may alter neuron numbers, stain absorption and selectivity, and other characteristics relevant to the above approach (Gonzalez-Riano et al. 2017). Notwithstanding these caveats, the general method presented here of cell-by-cell segmentation from comprehensive series of stained sections based on readily available image processing software is potentially paradigm-shifting and should be extensible to most of the remaining 700 distinct parcels of the Allen Brain Atlas.

Compliance with ethical standards

Conflict of interest

There is no conflict of interest.

CHAPTER TWO: OPERATIONS RESEARCH METHODS FOR ESTIMATING THE POPULATION SIZE OF NEURON TYPES (SAROJINI M. ATTILI … GIORGIO A. ASCOLI1)

Understanding brain computation requires assembling a complete catalog of its architectural components. Although the brain is organized into several anatomical and functional regions, it is ultimately the neurons in every region that are responsible for cognition and behavior. Thus, classifying neuron types throughout the brain and quantifying the population sizes of distinct classes in different regions is a key subject of research in the neuroscience community. The total number of neurons in the brain has been estimated for multiple species, but the definition and population size of each neuron type are still open questions even in common model organisms: the so called “cell census” problem. We propose a methodology that uses operations research principles to estimate the number of neurons in each type based on available information on their distinguishing properties. Thus, assuming a set of neuron type definitions, we provide a solution to the issue of assessing their relative proportions. Specifically, we present a three-step approach that includes literature search, equation generation, and numerical optimization. Solving computationally the set of equations generated by literature mining yields best estimates or most likely ranges for the number of neurons in each type. While this strategy can be applied towards any neural system, we illustrate its usage on the rodent hippocampus.

Introduction

A quantitative description of the brain's machinery is essential to understand the mechanisms of nervous system functions. The brain encompasses an extraordinary quantity and diversity of cells. The human brain contains nearly 100 billion neurons (Herculano-Houzel 2009) and the rodent brain contains around 100 million neurons (Herculano-Houzel et al. 2011). Neurons can be grouped into many distinct types based on their structural, physiological, and molecular features (Bota and Swanson 2007; Shepherd et al. 2019). The composition of balanced proportions of neuron types into elaborate networks enables the brain's many specific computations. Estimated counts of neuronal types, i.e. a "neuronal census," would enable more accurate and complete models of brain circuits. Towards this goal, the National Institutes of Health launched the BRAIN Initiative Cell Census Network (BICCN), a consortium of research projects tasked with generating a comprehensive molecular and anatomical cellular "parts list" within a three-dimensional reference mouse whole-brain atlas (Ecker et al. 2017).

Counting the neurons of each type in a region requires establishing the identity of millions of individual neurons. Rapid progress in genetic phenotyping is on the verge of enabling a comprehensive cell-level classification of neurons throughout the mouse cortex (Tasic et al. 2018). However, linking these growing molecular data to anatomical connectivity requires the analysis of the neuronal input and output elements, namely dendritic and axonal arbors. Full morphological characterization of axons and dendrites involves physical or optical tissue sectioning to follow the complex branching structures

in the dense three-dimensional space. This is a labor-intensive and error-prone procedure for a human to perform manually, underscoring the need for increasingly automated machine-learning approaches (Peng et al. 2015; Januszewski et al. 2018). Experimentally, the problem is exacerbated by the large disproportion between the total length of an individual axon (hundreds of millimeters) and its branch thickness (tens of nanometers), resulting in a very small ratio ($\sim 10{-}7$) between the volume of a neuronal projection and the territory it spans. This major obstacle will likely keep the acquisition of comprehensive structural data at single-neuron resolution below full-brain scale for many years. Therefore, indirect estimation of neuron type population counts is an important and useful endeavor.

The neuroscience literature contains a great deal of data relevant to the census problem. These include stereological sampling of neuronal densities in specific anatomical areas, morphological characterizations of collections of neurons from the same brain region, slice imaging of neurons stained for a particular molecular marker, and more (Hamilton et al. 2012; White et al. 2019). Each of these data types expresses facts about absolute or relative neuronal population sizes. Integrating such diverse sources of information for a neuronal census poses two primary challenges: formatting all relevant observations in terms of a common neuronal classification scheme; and inferring population sizes from the properly formatted evidence.

Solving the first challenge will ultimately require a broad consensus in the neuroscience community on how to define neuron types objectively and reproducibly (Armañanzas and Ascoli 2015). For the purpose of illustration, in this study we tentatively adopt a recent circuit-based classification proposal (Ascoli and Wheeler 2016) for which relatively abundant data are available for parts of the rodent brain such as the hippocampus. Solving the second challenge entails a workflow for integrating contrasting measurements and interpolating through missing data points. Operations research offers many techniques for leveraging inconsistent and/or incomplete information to achieve an optimal estimate for a set of target parameters. These techniques fall under the broad umbrella of mathematical optimization. Here we describe the use of mathematical optimization to obtain an estimated neuronal census. The neuronal population counts to be estimated are represented as free parameters. Data relating neuron types to their properties (e.g. from literature search or experiment) are formatted as equations in terms of these parameters. These equations are composed into an objective function, which can be optimized by a variety of algorithms. Thus, the novelty of this work consists of applying, for the first time, established operations research strategies to the open neuroscience problem of the brain cell census. The present study illustrates the proposed approach with a concrete application to a subregion of the hippocampal formation of the mammalian brain.

Methods

In order to describe the operations research aspects of our approach, we first explain how it is possible to derive a system of equations encoding constraints for a neuronal census. In the most general sense, every neuron type is associated with a distinct collection of properties (e.g. morphological, physiological or molecular) through a many-to-many relationship. In other words, no single property uniquely identifies a neuron type, and any property is typically associated with multiple neuron types. However, the full set of properties of a given neuron type is indeed different from that of all other neuron types. Useful constraints consist of measurements, observations or reports on neuronal properties that can link combinations of neuron types to specific numerical values.

Consider for instance a brain region with only two neuron types, A and B, and corresponding counts n_A and n_b. If a stereology experiment determines the total number of neurons in that region to be 1000, this provides a useful constraint (and corresponding equation) by setting the sum of the two target counts to the measured value ($n_A + n_B = 1000$). Now suppose that only neuron type A expresses a particular protein and an article reports that, out of 20 cells tested in that region, 15 were found to be positive for that protein while 5 were negative. This provides another useful constraint (and a second equation) by indicating a ratio (3:1) between the two target values ($n_A/n_B = 3$). In this simple case the number of equations equals the number of unknown counts, yielding a well-constrained system with a single exact solution ($n_A = 750$, $n_B = 250$).

In a more general sense, a system of equations is overdetermined if there are more constraints (equations) than parameters and underdetermined if there are fewer constraints than parameters. In the census problem, overdetermined systems may arise from multiple experiments measuring the same variable (e.g. the total number of neurons in a region) and yielding inconsistent results. Unless the equations are trivially redundant, overdetermined systems are inconsistent and thus do not have exact solutions. In this case numerical optimization may find a best estimate that minimizes the discrepancy from all available constraints. Underdetermined systems arise when insufficient constraints are available for one or more of the target unknowns. If none of the constraints are mutually inconsistent, an underdetermined system typically has an infinite number of solutions. In this case numerical optimization may find the range of values defining the possible solutions.

Classification and analysis framework

As a pilot study, we applied our methodology to estimate the population size of each known neuron type in the hippocampal subregion of the dentate gyrus (DG). This required a neuronal classification scheme for the dentate gyrus. We defined DG cell types based on the knowledge base Hippocampome.org, an online repository containing morphological, molecular, and physiological information on neurons of the rodent hippocampal formation (Wheeler et al. 2015). Hippocampome.org classifies neurons primarily by neurotransmitter released and the presence of axons and dendrites in the distinct subregions and layers of the rodent hippocampus (Fig. 7). It also includes

molecular biomarker (Hamilton et al. 2017) and electrophysiological properties (Komendantov et al., 2019) for each type. Hippocampome.org identifies 18 distinct neuron types in DG: 14 with cell bodies exclusively present in a single layer and 4 with cell bodies distributed across two layers. Thus, the target unknowns or decision variables for this neuronal census consist of the population counts for 22 layer-wise types, which we represent here with parameters $x1$, $x2$, … $x22$ (Table 4).

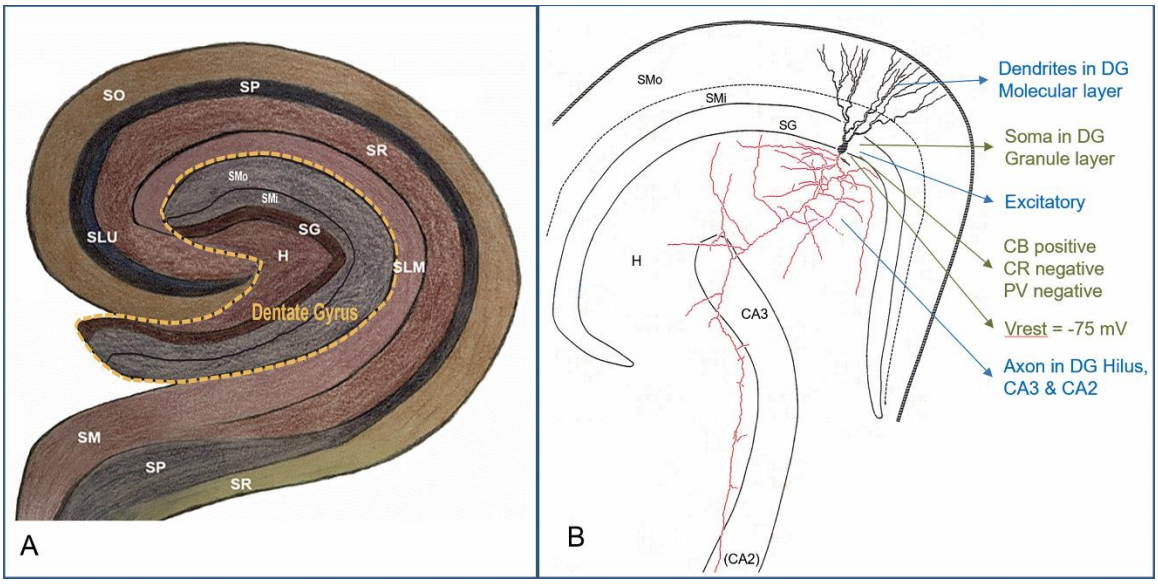

Figure 7 Hippocampome.org neuron type classification

A Layer organization of the rodent hippocampus, highlighting the dentate gyrus and surrounding regions. b A dentate gyrus granule cell (cell body and dendrites: black, axon: red) with color-coded properties (neurotransmitter and axonal-dendritic distributions: blue; molecular expression and electrophysiology: green). Label abbreviations: CB

calbindin, CR calretinin, H Hilus, PV parvalbumin, SG stratum granulosum, SLM stratum lacunosum-moleculare, SLU stratum lucidum, SM/SMi/SMo stratum moleculare (inner/outer), SP stratum pyramidale, SR stratum radiatum, Vrest resting voltage potential. (Color figure online).

Table 4 Layer-specific neuron types (left) and decision variables representing corresponding counts (right)

Neuron type	Decision variable
Granule	x_1
Hilar ectopic granule	x_2
Semilunar granule	x_3
Mossy	x_4
Mossy MOLDEN	x_5
AIPRIM	x_6
Axo-axonic	x_7
Basket	x_8
SG: Basket CCK+	x_9
SMi: Basket CCK+	x_{10}
H: HICAP	x_{11}
SG: HICAP	x_{12}
H: HIPP	x_{13}
SG: HIPP	x_{14}
HIPROM	x_{15}
MOCAP	x_{16}
MOLAX	x_{17}
SMi: MOPP	x_{18}
SMo: MOPP	x_{19}
Neurogliaform	x_{20}
Outer molecular layer	x_{21}
Total molecular layer	x_{22}

Our methodology in estimating the neuron type counts consists of three steps: (1) searching for actionable information regarding neuronal counts from the peer reviewed literature, (2) assembling a set of equations by mapping the extracted information to the

chosen neuron type classification scheme, and (3) numerically optimizing these equations by minimizing an objective function (discrepancy from empirical evidence) to derive type-specific counts (Fig. 8).

Literature Search

Our literature mining protocol began with an analysis of the bibliography of Hippocampome.org v.1.7 (hippocampome.org/php/Help_Bibliography.php). Hippocampome.org lists 496 publications used as evidence for the definition of neuronal types. Each of the dentate gyrus neuron types that are the subject of this study is associated with at least one, but typically several, such publication(s). The full text of each DG publication was evaluated for relevance and selected for further mining if it contained at least one of four kinds of data: stereology-based measurements of cell counts or densities; counts or densities derived from image processing techniques; morphological ratios obtained from studies that reconstructed small samples of neurons for electrophysiological analysis; and inferences based on volumetric estimates and indirect evidence.

Stereology aims to obtain unbiased estimates of cell numbers by inferring population sizes in three dimensions from two-dimensional slice images. A traditional stereological technique is the “optical dissector” (Russ and Deho 2001), which uses a varying focal plane to obtain many optical “slices” of an intact piece of tissue. A relatively newer method, the optical fractionator, transforms the highly anisotropic tissue into a

homogeneous suspension of free-floating nuclei which can then be counted microscopically or by flow cytometry and identified morphologically or immunocytochemically (Herculano-Houzel et al. 2015).

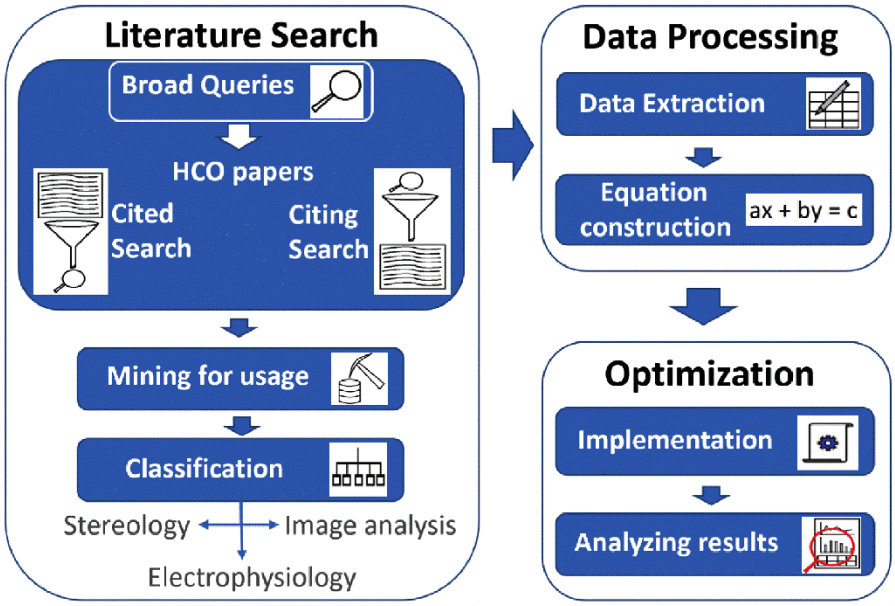

Figure 8 Methodological pipeline to estimate neuron type counts through three sequential phases

Literature mining, data processing, and numerical optimization. HCO: Hippocampome.org.

In general, stereology requires specific training, equipment, and histological processing, as well as appropriate sampling strategies, careful calibration, and rigorous statistical analysis (Bartheld 2001). Recently, newer *image processing* techniques for automated object segmentation have enabled cell counting from an entire brain region of interest without need of sampling (Bhanu and Peng 2000; Peng et al. 2013; Attili et al. 2019). *Morphological ratios* are derived from electrophysiological experiments, such as patch clamp recordings, designed to understand the cell properties in a specific neural system. Unlike stereology, electrophysiological cell sampling is not optimized for counting. Papers from the Hippocampome.org bibliography containing any of the above information were further mined for any *reference cited* in the context of the above information; moreover, all *references citing* the selected papers were also mined if their title explicitly referred to relevant neuron type data. Stereology-based measurements and image-processing calculations were considered more reliable than morphological ratios and indirect inferences. This is because stereology and image processing are designed to obtain accurate population counts, while morphological ratios typically come from experiments that use unclear sampling methodologies and inferences are based on uncertain assumptions. Therefore, stereology/image-processing constraints were weighted 10:1 against morphological ratios and indirect inference constraints. Rodent scaling rules (Herculano-Houzel et al. 2006) were used to integrate mouse data with most of the available information that is specific to rats.

Data processing

The sets of neurons described by the authors of the identified articles typically did not directly align with the Hippocampome.org classification scheme. Thus, constraint formation required mapping the literature-defined neuron types (literature types) to Hippocampome.org types. Since each Hippocampome.org type has a unique set of morphological, electrophysiological, and biochemical properties, we translated the description of each literature type into a similarly formalized set of properties, which we then used to match one or more Hippocampome.org types. The literature type could next be assigned the variable(s) *xi* associated with the matching type(s). When a literature type had properties matching multiple Hippocampome.org types, the sum of the variables representing the corresponding Hippocampome.org types was used. As an example of an equation generated from an electrophysiological experiment, one of the mined articles (Ceranik et al. 1997) states:

"Neurons from dentate gyrus outer molecular layer were recorded and filled with biocytin for video microscopy. 40 neurons were adequately stained. Out of these, 6 neurons were identified as displaced granule cells, 14 neurons had a local axonal arborization that was confined mainly to the OML, 3 projected to the stratum lacunosum moleculare of the CA1 region, and 17 neurons projected to the subiculum via the hippocampal fissure."

In this description, the author defines four different groups of neurons having somata in the DG outer molecular layer (SMo in Fig. 7 and Table 4 above). Based on the descriptions of their axons, the groups of 14 and 17 neurons were matched to unique

Hippocampome.org types MOPP and DG Neurogliaform, respectively. This allowed us to construct the equation $\frac{x_{19}}{x_{20}} = \frac{14}{17}$, where x_{19} and x_{20} are the parameters representing the respective outer molecular layer counts of MOPP and Neurogliaform cells. Equations were converted to a percent-error format for optimization ($\frac{17x_{19}}{14x_{20}} - 1 = \mathcal{E}$), where $\mathcal{E}$ is the error term (*residual*). The "displaced granule cells" and "3 projecting to the stratum lacunosum moleculare" represented groups with no corresponding Hippocampome.org types, so similar equations could not be constructed for these groups.

Optimization

Let a vector containing population counts $x_1, x_2, \ldots x_n$ be a "count vector" $\mathbf{x}$. Then, given a set of constraints $c_1, c_2, \ldots c_m$ and corresponding weights $w_1, w_2, \ldots w_m$, our goal is to find a count vector $\hat{x}$ that best satisfies the weighted constraints. There are multiple plausible ways to formulate this optimization problem. One approach is to express our constraints as a linear system of equations. This presents a linear least squares problem that can be easily solved with popular methods. Unfortunately, constraints are implicitly weighted in this formulation by the magnitudes of the known parameters they contain. The resultant massive and arbitrary weight disparities undesirably bias optimization results. Below we demonstrate this fact and present an alternative formulation that minimizes the weighted percent errors of each constraint.

Linear formulation

To formulate the problem as linear least squares, we need to derive from each constraint *ci* a linear equation in ***x***. Our constraint set consists of two kinds of constraints: sums and ratios:

$$x_6 + x_{11} + x_{13} + x_{15} = 16{,}801 \textit{ (sum)}$$

$$\frac{x_{19}}{x_{20}} = \frac{6}{11} \textit{ (ratio)}$$

Sum constraints are already linear equations in ***x***. Ratio constraints can be converted into linear equations in ***x***:

$$\frac{x_{19}}{x_{20}} = \frac{6}{11} \rightarrow 11x_{19} - 6x_{20} = 0$$

Thus, all constraints have a corresponding linear equation. This allows us to formulate the matrix equation $\boldsymbol{Ax} = \boldsymbol{b}$, with $\boldsymbol{A}$ an $m \times n$ matrix with rows corresponding to the left-hand side coefficients for each equation, and **b** a vector containing the right-hand-side constants or known parameters. The equation has no solution because our linear system is overdetermined and inconsistent, but there exists a best fit $\hat{x}$ that minimizes the sum of squared errors. We need to place boundary conditions on $\hat{x}$, since neuron counts cannot be too small or large. We also wish to differentially weight the constraints arising from different source experiments. Now let *Li* and *Ui* be lower and upper bounds for *xi*, and let $\boldsymbol{W}$ be an $m \times m$ diagonal matrix with weights $w_1, w_2, \ldots w_m$ the diagonal. Then we can define the problem as follows:

$$\mathcal{E}_i^L(x) = \sum_{j=1}^{n} \left(A_{ij} x_j\right) - b_i$$

$$\sum_{i=1}^{m} w_i \left(\mathcal{E}_i^L(x)\right)^2 = ||W^{\frac{1}{2}}(Ax - b)||^2$$

$$\hat{x} = arg_x min||W^{\frac{1}{2}}(Ax - b)||^2$$

subject to Li ≤ xi ≤ Ui for xi ∈ ***x***

This is a *constrained, weighted, linear least squares optimization problem*. Most numerical programming environments provide off-the-shelf routines that can efficiently solve this type of problem (e.g. scipy.optimize.lsq_linear in Python, bvls in R). Unfortunately, this formulation has an undesirable property for count estimation. Our constraints contain known parameters of widely varying magnitudes. Consider the error terms $\mathcal{E}_a^L(x)$ and $\mathcal{E}_b^L(x)$for two of our sum constraints:

$$\mathcal{E}_a^L(x) = (x_1 + x_7 + x_8 + x_9 + x_{12} + x_{14} + x_{16} + x_{17} + x_{21} + x_{22}) - 1{,}200{,}000$$

$$\mathcal{E}_b^L(x) = (x_6 + x_{11} + x_{13} + x_{15}) - 16{,}801$$

Now suppose we have a count vector ***x**** such that both constraints *a* and *b* are violated by some common factor *F*. Then $\mathcal{E}_a^L(x^*)$ and $\mathcal{E}_b^L(x^*)$ are:

$$\mathcal{E}_a^L(x^*) = 1{,}200{,}000F - 1{,}200{,}000 = 1{,}200{,}000(F - 1)$$

$$\mathcal{E}_b^L(x^*) = 16{,}801\,F - 16{,}801 = 16{,}801(F - 1)$$

The ratio of the errors is:

$$\frac{\mathcal{E}_b^L(x)}{\mathcal{E}_b^L(x)} = \frac{1{,}200{,}000(F - 1)}{16{,}801(F - 1)} = \frac{1{,}200{,}000}{16{,}801} \approx 71$$

Thus, when constraints c_a and c_b are equally violated in percentage terms, our squared error objective function penalizes the deviation from c_a approximately $71^2 = 5041$ times more than the deviation from c_b. Thus, constraints are implicitly weighted according to

the size of the known parameters (constants) they contain. This property is undesirable, since there is no reason to expect the precision of the source measurements to increase with their magnitude. One could eliminate the implicit weighting for sum constraints by normalizing with respect to the measured count. However, no equivalent operation is available for ratio constraints. Therefore, the linear formulation is a poor choice for application to counts estimation.

Percent error formulation

We can avoid the implicit weighting in the linear formulation by directly optimizing the percent errors of each constraint. Let $LHS_i(\boldsymbol{x})$ and RHS_i be the left and right sides of the measurement form of constraint c_i (all x_i on left, measured constant on right; see 'Equations' in supplementary materials). Incorporating weights and boundary constraints, minimization of the squared percent errors gives the optimization problem:

$$\mathcal{E}_i^P(x) = \frac{LHS_i(x) - RHS_i}{RHS_i} = \frac{LHS_i(x)}{RHS_i} - 1$$

$$\hat{x} = arg_x \min \sum_{i=1}^{m} wi(\varepsilon_i^P(x))^2$$

$$\text{Subject to } L_i \leq x_i \leq U_i \; for \; x_i \in x$$

This formulation does not suffer from the implicit constraint weighting of the linear approach. Equations constructed in this manner have been listed in Table 5.

Algorithms and implementations

A diverse array of algorithms exists for numerical optimization. The relative performance of different algorithms depends on the characteristics of the objective function and the tuning of algorithm hyperparameters, e.g. learning rates and boundary/initial conditions. A comprehensive review of available algorithms is beyond the scope of the present article. Instead we only describe the algorithms we applied to the DG neuronal census

Table 5 Representative equations and scientific meaning with original source, type of experimental evidence, and corresponding weights (wt.).

Equation	Interpretation	Source	Experimental evidence	Wt.
$\frac{x_6}{16801} + \frac{x_{11}}{16801} + \frac{x_{13}}{16801} + \frac{x_{15}}{16801} - 1 = \varepsilon_1$	Count of inhibitory hilar neurons is 16,801	Buckmaster & Jongen-Relo 1999, Table 1	Stereology	10
$\frac{x_1}{1080000} + \frac{x_7}{1080000} + \frac{x_8}{1080000} + \frac{x_9}{1080000} + \frac{x_{12}}{1080000} + \frac{x_{14}}{1080000} + \frac{x_{16}}{1080000} + \frac{x_{17}}{1080000} + \frac{x_{21}}{1080000} + \frac{x_{22}}{1080000} - 1 = \varepsilon_2$	Count of granule layer neurons is 1,080,000	Hosseini-Sharifabad and Nyengaard 2007, p209	Stereology	10
$\frac{x_2}{65420} + \frac{x_4}{65420} + \frac{x_5}{65420} + \frac{x_6}{65420} + \frac{x_{11}}{65420} + \frac{x_{13}}{65420} + \frac{x_{15}}{65420} - 1 = \varepsilon_3$	Count of all neurons in hilus is 65,420	Grady et al. 2003, Table 1	Stereology	10
$\frac{x_1}{1258848} + \frac{x_2}{1258848} + \frac{x_3}{1258848} + \frac{x_4}{1258848} + \frac{x_5}{1258848} + \frac{x_6}{1258848} + \frac{x_7}{1258848} + \frac{x_8}{1258848} + \frac{x_9}{1258848} + \frac{x_{10}}{1258848} + \frac{x_{11}}{1258848} + \frac{x_{12}}{1258848} + \frac{x_{13}}{1258848} + \frac{x_{14}}{1258848} + \frac{x_{15}}{1258848} + \frac{x_{16}}{1258848} + \frac{x_{17}}{1258848} + \frac{x_{18}}{1258848} + \frac{x_{19}}{1258848} + \frac{x_{20}}{1258848} + \frac{x_{21}}{1258848} + \frac{x_{22}}{1258848} - 1 = \varepsilon_4$	Total count of dentate gyrus neurons is 1,258,848	Erö et al. 2018, Supplementary data (Mouse bilateral values, halved and converted to rat)	Image processing	10
$\frac{x_3}{189693} + \frac{x_{10}}{189693} + \frac{x_{18}}{189693} + \frac{x_{19}}{189693} + \frac{x_{20}}{189693} - 1 = \varepsilon_5$	Count of all neurons in the molecular layer is 189,693	Attili et al. 2019, Table 2 (halved and scaled to rat)	Image processing	10
$\frac{x_2}{127550} + \frac{x_4}{127550} + \frac{x_5}{127550} + \frac{x_6}{127550} + \frac{x_{11}}{127550} + \frac{x_{13}}{127550} + \frac{x_{15}}{127550} - 1 = \varepsilon_6$	Count of all neurons in the hilus is 127,550	Murakami et al. 2018, supplementary data (halved and scaled to rat neurons)	Image processing	10

Equation	Interpretation	Source	Experimental evidence	Wt.
$\frac{11*x_{19}}{6*x_{20}} - 1 = \varepsilon_7$	The proportion of outer molecular layer MOPP cells to neurogliaform cells is 6:11	Armstrong et al. 2011, p1480, middle-right	Morphological ratio	1
$5 * x_7/(x_6 + x_{11} + x_{12} + x_{13} + x_{14} + x_{15} + x_{16} + x_{17} + x_{21} + x_{22}) - 1 = \varepsilon_8$	In granular layer and hilus, dendrite-targeting interneurons are 5 times more abundant than axon-targeting	Han 1994, p103, bottom-right	Morphological ratio	1
$\frac{(7*(x_8+x_9))}{(6*(x_{13}+x_{15}))} - 1 = \varepsilon_9$	For every 6 basket cells in granular layer, 7 hilar cells are found that project to outer molecular layer	Lubke et al. 1998, p1521–6	Morphological ratio	1
$\frac{(3*(x_6+x_{11}+x_{12}+x_{16}))}{(4*(x_{13}+x_{14}+x_{21}))} - 1 = \varepsilon_{10}$	The hilar and granular cells projecting to outer molecular layer are 3/4 of those projecting to inner molecular layer	Mott et al. 1997, p3992–3	Morphological ratio	1
$\frac{(64*(x_{10}+x_{18}))}{(169*x_3)} - 1 = \varepsilon_{11}$	Of 233 sampled cells in inner molecular layer, 64 are excitatory	Williams et al. 2007, p13757, top-right	Morphological ratio	1
$\frac{(x_{19}+x_{20})}{(2*x_{10}+2*x_{18})} - 1 = \varepsilon_{12}$	Interneurons are twice as abundant in outer as in inner molecular layer	Woodson et al. 1989, Fig. 2–7	Indirect inference	1

problem. Some of these algorithms require boundary and/or initial conditions for $\boldsymbol{x}$. Where required, we chose a lower bound of 0 for all neuron types, an upper bound of 1.2 million for granule cells (the principal cells of the DG), and an upper bound of 1 million for all other neuron types. We chose initial conditions to be consistent with estimates from an early modeling proposal (Morgan et al. 2007): 800,000 for granule cells, 90 for hilar ectopic granule cells, 15,000 for mossy cells, 5000 for mossy Molden cells, and 1000 for all other neuron types.

The *non-negative least squares* algorithm solves the linear least squares problem $arg_x min||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||^2$ with the constraint $\boldsymbol{x} \geq 0$ (Lawson and Hanson 1995). This algorithm does not take any hyperparameters. The *bounded-variable least squares* variant (Stark and Parker 1993) minimizes the same objective function, but subject to explicit boundary conditions. We used the respective R implementations **nnls** (Mullen and van Stokkum 2015) and **bvls** (Mullen 2015). Boundary conditions for **bvls** were set as described above. The *interior point* algorithm is popular for solving large nonlinear programming problems (Byrd et al. 2000). This algorithm requires boundary and initial conditions in addition to other hyperparameters. We used the MATLAB implementation **fmincon** (Mathworks, Natwick, MA, USA). Boundary and initial conditions were set as described above. The following hyperparameters were set to their default values: maximum number of iterations (1000), variable tolerance (10^{-6}), function tolerance (10^{-6}), constraint tolerance (10^{-6}), and unboundedness threshold (10^{-20}). The maximum number of function

evaluations was set to 100,000 and the derivative approximation was calculated by forward differences with LDL factorization and an initial barrier of 0.1.

Simulated annealing is a popular optimization method for complex non-linear objective functions with multiple local minima (Van Laarhoven and Aarts 1987; Xiang et al. 2013). This algorithm requires boundary and initial conditions in addition to other hyperparameters. We used two different implementations: MATLAB's **simulannealbnd** (fast annealing option) and R's **optim_sa** (Husmann et al. 2017). Boundary and initial conditions were set as described above. All other hyperparameters were set to their default values: function tolerance (10−6), maximum number of inner loop iterations (100,000), reannealing interval (100), outer loop temperature reduction (0.99), maximum function evaluations per variable (3000), stall iterations per variable (500), and initial temperature (100).

The *pattern search* method finds a sequence of points that approach an optimal point based on an adaptive mesh (Audet and Dennis 2003; Conn et al. 1997). The value of the objective function either decreases or remains the same from each point in the sequence to the next. This algorithm requires boundary and initial conditions in addition to other hyperparameters. We used the MATLAB implementation **patternsearch**. Boundary and initial conditions were set as described above. The polling method was set by two parameters: poll order algorithm ("Consecutive") and search function ("GPSPositiveBasis2N"). All other hyperparameters were set to their default values:

initial mesh size (1), expansion factor (2), contraction factor (0.5), initial penalty (10), penalty factor (100), bind tolerance (10^{-3}), size tolerance (10,000), mesh tolerance (10^{-6}), maximum number of iterations per variable (100),maximum function evaluations per variable (2000), variable tolerance (10^{-6}), function tolerance (10^{-6}), and constraint tolerance (10^{-6}).

Each of the algorithms returns an estimated count vector $\hat{x}$. The population counts for the 18 DG neuron types of Hippocampome.org can then be obtained simply by summing the layer-specific vector elements for each type (e.g. the count of Basket CCK+ cells equals $\hat{x}_9 + \hat{x}_{10}$) and rounding to the closest integer.

Results

The literature search and data processing procedures described above yielded 50 independent pieces of information (constraints) extracted from 32 distinct peer-reviewed scientific sources pertinent to the cell census of the unilateral (i.e. one hemisphere only) dentate gyrus of the adult rat. The experimental evidence for 22 constraints was based on stereology, for 6 on image processing, for 13 on morphological ratios, and for 9 on indirect inferences, resulting in a total sum of weights of 302. These constraints were formulated as mathematical equations (Table 5) and assigned weights based on the reliability of each source. Here we only present representative examples for each type of evidence for illustrative purposes. The full set of 50 equations used in the optimization,

their scientific interpretations, and source quotations are included in the ‘Equations’ tab of the supplementary materials at hippocampome.org/php/data/ANOR_suppl_mat.xlsx.

We tested three algorithms (interior point, pattern search, and simulated annealing) on the percent-error optimization problem and two (non-negative least squares and bounded variable least squares) on the linear least squares problem. Though non-negative and bounded variable least squares optimized the linear least squares objective, we scored their solution vectors using the percent-error objective for comparison with the other algorithms (Table 6). Interior point and pattern search had equivalent performance, superior to simulated annealing. Simulated annealing performance depended on the implementation (MATLAB was superior to R). Non-negative and bounded-variable least squares performed equivalently, indicating a lack of sensitivity to upper bounds for this problem.

Table 6 Optimization algorithms and their percent-error objective function values

Algorithm (implementation)	Least square sum
Non-negative least squares (R)	161.94
Bounded-variable least squares (R)	161.94
Interior point (MATLAB)	23.47
Pattern Search (MATLAB)	23.47
Simulated annealing (R)	48.21
Simulated annealing (MATLAB)	42.05

Although interior point and pattern search converged to the same objective value, their solution vectors were different. Individual vector elements (i.e. population counts) varied substantially (>30%) between the two vectors for 8 out of 18 neuron types (Table 7).

Table 7 Estimated dentate gyrus cell counts by neuron type for the two best-performing algorithms

Neuron type	Interior point	Pattern search
SG Granule	971,495	971,284
Hilar ectopic granule	90	90
Semilunar granule	112,075	112,094
Mossy	21,742	21,742
Mossy molden	6052	6052
AIPRIM	3658	6308
Axo-axonic	1086	1086
Basket	2698	4396
Basket CCK+	10,680	15,960
HICAP	3983	1662
HIPP	4523	4760
HIPROM	7605	7554
MOCAP	816	488
MOLAX	8275	8272
MOPP	19,634	12,912
Neurogliaform	19,135	18,900
Outer molecular layer	725	488
Total molecular layer	32	80

Interestingly, the midpoint of the vectors yielded the same objective value. This suggests that the two vectors exist within a continuous solution space. Non-negative and bounded-variable least squares yielded qualitatively different results, with many low population neuron types set to their lower bound. This is explained by the above analysis of implicit weighting in the linear least squares problem formulation. Solution vectors for all algorithms are provided in the 'Results' tab of the supplementary materials at hippocampome.org/php/data/ANOR_suppl_mat.xlsx.

These results cannot be validated directly due to the unavailability of independent neuron type count data, with the exception of granule cells and hilar ectopic granule cells. Nevertheless, the objective value for the two best-performing algorithms corresponded to a total residual error of 7.8% of the weight sum (302). This number may be interpreted as a measure of the residual deviation from published estimates and the existing disagreement within the literature. On the one hand, the results are most strongly determined by the stereological and image analysis data, since we weighted those sources ten times more than the evidence from morphological ratios and indirect inferences. On the other hand, in the absence of alternative approaches to determine neuron type counts, stereology and image analysis counts provide the most reliable indirect evidence to evaluate the plausibility of these estimates on a layer-by-layer basis. Specifically, we

summed the averages of the neuronal counts obtained from the interior point algorithm and pattern search across each of the three dentate gyrus layers: granular (SG), hilar (H), and molecular (SM). We then compared these values against the distributions of the corresponding estimates available in the literature (Fig. 9). Our results fall within one standard deviation of the mean in all three cases (SG: 988,759 vs. 1,115,971 ± 260,218, N = 13; H: 46,610 vs. 69,484 ± 39,079, N = 10; SM: 158,846 vs. 289,031 ± 124,334, N = 3). The values of all data points for this analysis are included in the 'stereology visualization' tab of the supplementary data at hippocampome.org/php/data/ANOR_suppl_mat.xlsx.

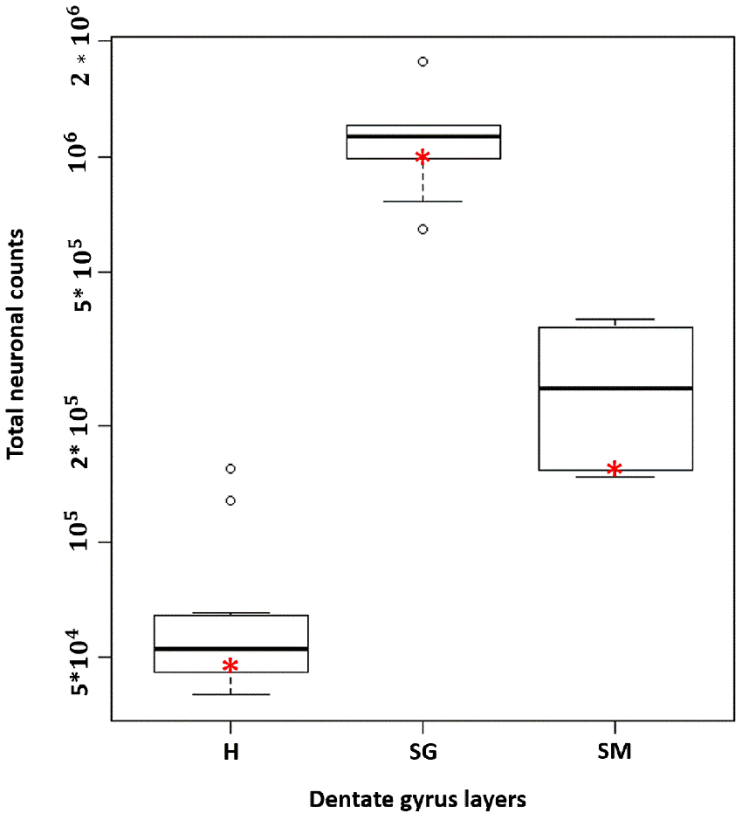

Figure 9 Layer-by-layer comparison of the average neuron counts:

From interior point and pattern search (red asterisks: layer totals from Table 7) against literature-based stereological and image analysis estimates in hilus (Grady et al. 2003; Fitting et al. 2009;Mulders et al. 1998; Ramsden et al. 2003; Lister et al. 2006; Sousa et al. 1998; Rasmussen et al. 1996; Erö et al. 2018; Murakami et al. 2018; Attili et al. 2019), granule layer (West et al. 1991; Hosseini-Sharifabad and Nyengaard 2007; Bayer et al. 1982;Mulders et al. 1998; Rasmussen et al. 1996; Kaae et al. 2012; Rapp and Gallagher 1996; Fitting et al. 2009; Sousa et al. 1998; Calhoun et al. 1998; Insausti et al. 1998; Erö et al. 2018; Murakami et al. 2018), and molecular layer (Erö et al. 2018; Murakami et al. 2018; Attili et al. 2019). The bottom and top of the boxes represent first and third quartiles respectively, the bold midline is the median, the whiskers indicate the span of

data within one and a half inter-quartile ranges from the box, and the circles are the remaining points. H hilus, SG Stratum granulare (granule layer), SM stratum moleculare (molecular layer).

Conclusions

This work demonstrates that the proposed pipeline of annotation, conversion into equations, and optimization can yield a viable solution to the neuronal census problem. Previous approaches to this open problem in neuroscience in another hippocampal region (area CA1) entailed deriving estimates for interneuron type populations from the literature chiefly based on expression of neurochemical markers (Bezaire et al. 2016). However, due to the extant sparsity of neurochemical marker data, that effort relied on a very large number of forced assumptions that are still awaiting empirical validation (Bezaire and Soltesz 2013). More recently, the positional mapping of distinct inhibitory subtypes in the developing somatosensory cortex was algorithmically inferred by combining cellular and molecular constraints from protein tissue stains and genetic expression profiles (Keller et al. 2019). Here we use a variety of relevant data from the literature to derive the resultant neuron type populations. Our use of optimization algorithms allows one to solve the cell census problem by leveraging all suitable information relating reported counts to neuronal properties, including location and

transcriptomics, but also electrophysiology, morphology, and any other available empirical evidence. The DG neuron type counts reported is this study should only be considered as preliminary results presented for the sole purpose of providing an in-depth illustration of this methodology. Further research must be conducted before finalizing the appropriate choice of algorithms for each specific use case. The use of simulated annealing for convex optimization is actively investigated (Kalai and Vempala 2006; Abernethy and Hazan 2015). Restricted Newton step methods such as nonlinear least squares (More 1978) and the Trust-Region-Reflective Algorithm (Coleman and Li 1996; Gill et al. 1981) may also be used to optimize overdetermined systems. Exploring multiple algorithms with different parameter settings is essential to identify the combination of implementation details yielding the best results.

Ultimately, however, the robustness of results will depend on the quality and quantity of the available data. Producing more complete and useful results will thus require feeding further constraints to the optimization algorithms. Possible sources of additional constraints include forthcoming results of ongoing experiments, more thorough data mining of existing literature, and assumptions based on domain expert knowledge. The amount of relevant empirical evidence in neuroscience has been increasing over the years and this growth is widely expected to continue in the foreseeable future. For the time being, the presented count estimates are likely inflated due to the exclusion of as-yet undiscovered neuron types as well as of types too vaguely described in publications for

inclusion in Hippocampome.org. Future analyses could account for these neurons by including additional decision variables representing unknown types in each layer.

It is noteworthy that two distinct solutions were found in the case of dentate gyrus neuron types counts with equal objective value but considerable differences in population sizes. The fact that the midpoint of these two solution vectors also constitutes a solution of equal objective value suggests the existence of an infinite set of optimal solutions. A possible approach to better selecting a solution within this set would be to add a regularization term (Hoerl and Kennard 1970; Tibshirani 1996; Wang et al. 2006) on the cell counts and bias them towards zero. Here we simply interpret the results as possible ranges of plausible values, in line with the multiple animal strains and varied age groups corresponding to the various constraints utilized in the optimization. As a positive side effect, even an underdetermined system can be useful in identifying the most under-constrained target unknowns in need of additional experimental evidence. Researchers can leverage this information to design specific experiments for revealing the missing information.

In this study, we assessed the reliability of available data at the level of experimental categories (e.g. stereology vs. morphological ratios) and employed this information in assigning weights. Certain source articles, however, may contain useful information to quantify the reliability of individual datasets, such as a standard error for their measurements (e.g. Buckmaster and Jongen-Relo 1999). Such an analysis also provides a

method to weigh the constraints and corresponding equations in the optimization. Out of the 32 publications utilized in the presented application to the dentate gyrus neuron type census, 14 reported a standard error or similar estimates of variance. While here we simply utilized averages from all information sources, a future improvement might consist of quantifying the uncertainty in the morphological ratios with a statistical model.

In summary, operations research offers a powerful approach to generating quantitative estimates and a measurable error for the unknown counts of cell types. Although we illustrated this method to estimate neuronal counts in the dentate gyrus, this technique can be extended to quantify the neuron type population size in any region of the brain. In future work, we aim to extend this methodology for estimating populations of neuronal types beyond the dentate gyrus and throughout the entire hippocampal formation, including areas CA1, CA2, CA3, the subiculum, and the entorhinal cortex.

Acknowledgements

This project is supported in parts by Grants R01NS39600 and U01MH114829. The authors are grateful to Drs. Diek Wheeler, Keivan Moradi, and Padmanabhan Seshaiyer for their help and many useful discussions.

CHAPTER THREE: QUANTIFICATION OF NEURON TYPES IN THE RODENT HIPPOCAMPAL FORMATION BY DATA MINING AND NUMERICAL OPTIMIZATION (SAROJINI M. ATTILI … GIORGIO A. ASCOLI)

Quantifying the population sizes of distinct neuron types in different anatomical regions is an essential step towards the goal of building a brain map. Although there are estimates for the total neuronal populations in different species today, the distinct number as well as the definition of each neuron type is still a subject for research in the scientific community. Hippocampome.org is an open-source knowledge base with morphological, physiological, and molecular information for 122 neuron types in the rodent hippocampal formation. Although this framework effectively identifies all known neuron types in this system, in conjunction with their molecular and electrophysiological properties, their population sizes remain largely unknown. The goal of this study is to contribute quantified estimates for the distinct neuron types of Hippocampome.org by leveraging data from existing literature and using numerical optimization. We report the total neuronal estimates for the six subregions – dentate gyrus, CA1-3, subiculum, and entorhinal cortex to be 2,958,500. The highest numbers are from dentate gyrus at 1,197,548, and the lowest are from the subregion CA2 at 29,493. In addition to fulfilling this goal, we conducted sensitivity analysis to obtain ranges and examine the reliability of our estimates, studied the distribution of biomarkers and firing patterns, and analyzed population estimates and numerical densities of four broad groups with respect to volumes across the hippocampal formation. Our analysis indicated that dendritic targeting neuron densities significantly decreased with increasing volume. Population

estimates and related code including mined data have been made available on the current version of Hippocampome.org.

Keywords: Hippocampus, Neuron types, Rodent, Numerical optimization, Operations Research

Introduction

The brain is a highly complex organ encompassing an extraordinary quantity and diversity of cells, for which there is still a substantial amount of knowledge needed to fully understand the brain functionality. For instance, despite having information on the different anatomical parcels, cell types, and molecular and electrophysiological mechanisms of the brain (H Damasio, 1995), compared to the anatomical characterization and research on the different high-level parcels in the brain (Nolte 1993; Taubert et al., 2010; Lange et al., 1997), the knowledge on low-level components such as neuron types is limited. As an example, we still lack a complete census or taxonomy of brain cell types (Mukamel and Ngai, 2019), even though there is notable progress in recent times with developments in a single cell transcriptome, epigenome profiling, and anatomical methods.

Attaining a comprehensive census of distinct neuron types is fundamental to understanding the functionality of the brain and generating working models. Mathematical modeling, statistical analysis, and exploratory data mining are powerful techniques necessary to understand the general principles of brain function (Mott et al., 2018). Building upon this notion, the US Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative aims to develop new tools to accelerate and

deepen our understanding of the human brain function in health and disease (The BRAIN Initiative, 2014). In addition, the National Institutes of Health launched the BRAIN Initiative Cell Census Network (BICCN), a consortium of research projects tasked with generating a comprehensive molecular and anatomical cellular "parts list" within a three-dimensional common reference mouse whole-brain atlas (Ecker et al., 2017).

The knowledge repository Hippocampome.org, which contains the beginnings of a parts list for the hippocampal formation, accumulates information on the anatomical, chemical, and biophysical properties of rodent hippocampal neurons (Wheeler et al., 2015). The resource primarily defines neuron types from the six hippocampal formation subregions, dentate gyrus (DG), CA3, CA2, CA1, subiculum (SUB), and entorhinal cortex (EC), based on their neurotransmitter and the spatial distributions of their axons and dendrites (Hamilton et al., 2016). As of version 1.9 (Sanchez-Aguilera, et al., 2021), 122 neuron types have been identified and new neuron types will be added as knowledge increases in the field. The ultimate goal of Hippocampome.org is to create biologically plausible computational models of the hippocampus. To achieve this goal, one of the key pieces of information needed is the quantified estimates of the classified neuronal types. The present study aims at obtaining these estimates using computational methods, while also relying on values obtained from the existing scientific literature. The addition of numerical estimates for the 122 neuron types would substantially enhance the already rich Hippocampome.org knowledge repository and would help the community advance

towards the first biologically plausible simulation of a whole functional region of the mammalian brain.

In terms of attempts to understand the brain, extensive research has been conducted in understanding principal neurons in some regions but relatively fewer interneurons (Cembrowski et al., 2016; ESL Faber et al., 2001; Canto and Witter, 2012; S Goebbels et al., 2006; B Tahvildari and Alonso, 2005; Suzuki and Bekkers, 2011; Ehrlich et al., 2012; Rasmussen et al., 1996) have been studied. Among the six hippocampal subregions of interest, the DG and area CA1 have been most widely researched (Armstrong et al., 2011, Buckmaster et al., 1992, Ceranik et al., 1997, Han et al., 1994, Lubke et al., 1998, Mott et al., 1997, Grady et al. 2003, Fitting et al. 2009, Mulders et al. 1997, Ramsden et al. 2003, Mcbain et al., 1994, Sik et al., 94, vida et al., 1998, Svoboda et al., 1999, Bezaire et al., 2016, West et al., 1991, Hosseini-Sharifabad and Nyengaard, 2007) followed by CA2, CA3 and the EC (Rapp and Gallagher, 1996, Kaae et al., 2012, Lister et al., 2006, Fitting et al., 2009, Kohus et al, 2016; Szabo et al., 2014; Losonczy et al., 2004; Gulyas et al., 2010; Szabadics et al., 2010; Mercer et al., 2007). The SUB is the least researched area (Harris and Stewart, 2000; Mulders et al., 1997, Lister et al., 2006, Andrade 2000, Fitting et al., 2009, Kim et al., 2017, Bjerke et al., 2021) among the six subregions and has only three neuron types classified on Hippocampome.org. Although the other areas have a higher number of neurons classified, there is no neuronal quantification reported for many of these classified types.

Traditional methods of quantification, such as stereology and modern methods that use computation, are generally used to report layer based total counts or principal cell counts (Grady et al. 2003, West et al., 1991, Murakami et al., 2018, Ero et al., 2018, Attili et al., 2019). Individual interneuron counts have been scarcely reported (Jiao et al., 2007, Ero et al., 2018, Bezaire et al., 2016). However, there are abundant electrophysiological studies that experiment with a wider variety of neuron types and many of these results can be used to derive ratios between the quantities of individual cell types (McBain et al., 1994; Sik et al., 1994, Buhl et al., 1994, Hajos and Mody 1997, Kohus et al., 2016; Szabo et al., 2014). Apart from these studies, there are computational modeling papers that report the number of principal cells and other well studied interneurons such as basket cells, neurogliaforms, etc. (Bezaire et al., 2016). Another important sources of data are papers that report proportions or quantities of neurons from hippocampal layers that are positive to certain biomarkers of interest (Jinno and Kosaka, 2006; Kim et al., 2017, Kosaka et al., 1987, Bezaire et al., 2016). Every such relevant piece of evidence has been transformed into linear equations and numerically optimized to arrive at the population estimates in this study.

Methods

This work is focused on computing numerical estimates of the 122 neuron types classified by Hippocampome.org. We made use of scientific data from the existing literature to come up with the population estimates. Our methodology includes heavy data mining, data transformation, and numerical optimization. On a high level, we can divide the work into a three-step approach that includes data mining or literature search;

equation generation, which entails the transformation of the mined data; and numerical optimization, which finally produces the estimated counts for each neuron type of interest.

To begin with, we relied heavily on the existing scientific literature that was available from credible sources. Hundreds of journal articles were mined to identify any references to one or more neuron types that corresponded to the ones defined on the Hippocampome.org. As there is no standardization of nomenclature of the hippocampal neurons, it is common that two different researchers give different names to the same 'type.' Hence, during the data mining phase, the first step was to identify the neuron type in the article and map it to the appropriate type on Hippocampome.org, where the locations of the soma, axon, and dendrites were the primary factors in mapping the cell type. In rare cases, where the morphological patterns were identical, understanding biomarker reactivity was important in the identification/mapping process.

The data mining process began with a structured search. Every neuron type on Hippocampome.org contains the list of scientific sources or 'evidence' information, figures, other names it has been referred to as, its electrophysiological properties, and connectivity patterns. We reviewed each article listed under the 'evidence' page for each neuron type on Hippocampome.org. A search was conducted through the reference articles for each of these papers for any relevant sources. Another search was conducted to find relevant articles that cited these papers of our interest. Thus, a 'cited search' and a

'citing search' were performed. Apart from this structured method, articles of interest were searched for on Google scholar or other credible sources.

Identification of neuron types

The 122 neuron types have further been divided into 198 types based on their somatic location. In order to translate research data into a computable format, we assigned a variable for each neuron type with exclusive soma/axon/dendrite location and biomarker reactivity in every subregion of the hippocampus. In other words, we divide one neuron type into multiple types if it has somata in more than one location/layer. For example, Hippocampome.org defines Basket cells in the CA1 with somata in stratum pyramidale (SP) or stratum oriens (SO). In this case, we divide the cell type into two types – SP-Basket and SO-Basket. During the data mining process, it was noted that information regarding neuronal groups in some layers that were not classified on Hippocampome.org, was available in the literature (Ero et al., 2018). Nine additional neuronal groups were added to our study – CA2 SLM excitatory, CA2 SLM inhibitory, CA2 SO inhibitory, CA2 SR inhibitory, CA3 SLM excitatory, interneurons in the polymorphic layer of the subiculum, deep layer interneurons of the entorhinal cortex, medial entorhinal cortex layer I interneurons, and lateral entorhinal layer I interneurons. Upon dividing the types by soma as described earlier and adding the 9 additional neuronal groups, we ended up with 207 neuron types and corresponding variables for numerical optimization. The list of variables assigned for neuronal types based on our classification system can be found in the supplementary table. A total of 207 types have been categorized from six

hippocampal subregions as follows: 22 types in the DG, 37 types in area CA3, 9 types in area CA2, 64 types in area CA1, 5 types in the SUB and 70 types in the EC.

Leveraging existing scientific literature

Leveraging existing scientific literature: Data mined from various sources were recorded in a structured format that included low level details, such as the location of the text of interest, species information, interpretation of the text, author, source name, and so on. Rodent scaling rules (Herculano-Houzel et al., 2006) were used to integrate mouse data with most of the available information that is specific to rats – a multiplier of 2.44 was used to convert mouse data to rat. Each of the interpretations was converted into a linear algebraic equation that was later normalized and used as part of an objective function.

We added weights to each equation based on the experimental method described in the source. Stereology-based measurements and image-processing calculations were considered more reliable than morphological ratios and indirect inferences. This is because stereology and image processing are designed to obtain accurate population counts, while morphological ratios typically come from experiments, which use unclear sampling methodologies, and inferences are based on uncertain assumptions. Therefore, stereology/image-processing constraints were weighted 10:1 against morphological ratios, indirect inference constraints, and biomarker data.

Data transformation

Each equation is transformed into a normalized least squared form, to which a weight was added, i.e., $r = \sum_i^n \left(w_i \left(x_i^{optimization} - x_i^{data}\right)/x_i^{data}\right)^2$, where r is residual, i is the equation number, n is the total number of equations, w is the weight, and x is the neuron count. These weighted least squared residuals for each subregion are composed into an objective function, which would then be optimized to obtain estimates for the unknown variables – neuron types. We describe the low-level details of equation construction – linear formulation and transformation into normalized form in the 'Methods' section of our previously published work (Attili et al., 2020). Table 8 lists the number of neuron types for each subregion, number of equations generated from the different sources and the corresponding residual errors.

Table 8 The number of neuron types (including the 9 neuronal groups added), variables, equations by type and the residual errors for the six subregions of the hippocampal formation.

Subregion	**Number of neuron types**	**Number of variables**	**Total Number of equations**			**Residual Error**
			Sum	Ratio	Inequality	
DG	18	22	115			11.44%
			37	28	50	
CA3/CA2	30	45	166			18.65%
			37	31	98	
CA1	42	64	208			21.47%
			41	77	90	

SUB	3	5	30			11.45%
			18	1	11	
EC	31	70	126			22.02%
			47	39	40	

Algorithm selection and optimization

After a careful study and testing of several optimization algorithms, we decided to use PSwarm (Vaz and Vicente, 2007) for this work. PSwarm is a global optimizer that combines pattern search (Lewis and Torczon, 2002) and particle swarm (Kennedy and Eberhart, 1997; Shi, 2001) algorithms for bound and linear constrained problems. We used a lower bound of ten and an upper bound of a million for each neuron type variable. Once the algorithm was finalized and tested, the weighted least squared residuals for each subregion were used to develop an objective function that would undergo optimization. Results from 200 iterations of optimizations were examined and iterations with the lowest error were selected for analysis. In some cases, we saw that multiple neuron types had constant sum; we equally divided this sum and attributed the amount to each of the neuron types. For example, in the multiple iteration report, an interdependency was found between three neuron types from CA3 - Interneuron Specific Oriens, O-LMs in SO, Trilaminars in SO. Their sum was a constant of 6,350 in every iteration. We equi-divided this sum to assign a value of 2117 to each of the three neuron types. The

complete list of variables, equations, interpretations, and sources can be found in the supplementary material.

Ranges

To assess the reliability of results, sensitivity analysis was conducted to determine the range of each neuron type's numerical estimate. Lower and upper bounds were programmatically computed at 5% of the residual error for the subregion for every neuron type. Based on the optimum and ranges, the reliability for each neuron type was determined.

P-values were calculated using Pearson's coefficient and the number of data points to understand significance. During the analysis phase, all p-values in this study have been corrected using the Bonferroni method (Napierala, 2012).

Results

The primary goal of this work is to obtain estimated populations of neuron types from the six subregions of the rodent hippocampus as defined on Hippocampome.org. Our work consists of leveraging results from existing scientific literature and the knowledge contained on Hippocampome.org to compute numerical estimates for the defined neuron types. We made use of applied mathematics and operations research methods for this effort. We described an overview of our methodology in a previous publication (Attili et al., 2020), where we estimated neuronal numbers for the neuron types in the DG. Here we added the use of biomarker data to improve our estimates.

Data sources containing neuron type information needed for this work could be categorized as follows – Stereology, Electrophysiological Ratios, Computational, Marker Inequalities, and Others. Research results yielded quantified estimates of neuron types, where the numbers could be totals per hippocampal subregion, per layer, for a subset of neuron types, or for the neuron type itself. For example, Coulin et al report 1,130 Cajal-Retzius cells in the CA1 region (Coulin et al., 2001), and Murakami and colleagues report 178,123 cells in the CA1 stratum lacunosum-moleculare (SLM) (Murakami et al., 2018). These types of data come from diverse sources including stereology (Grady et al. 2003), computational methods (Murakami et al., 2018), modeling papers (Bezaire et al., 2016), immunohistochemistry-based work (Fuentealba et al., 2008), etc. For example, Bezaire et al estimate the CA1 pyramidal cells to be around 311,500 (Bezaire et al., 2016).

Another type of source are electrophysiological experiments conducted on the rodent hippocampal formation. Often, researchers record from the different subregions or layers of the hippocampus, identify the resultant cell types, and study their properties. From this identification, we derive possible proportions of the neuron types in that location. For example, the following is the text from an experiment that was used to extract an electrophysiological ratio (Gulyas et al., 2010): “From a total number of 61 PV-EGFP cells recorded in different sets of experiments, in 30 cases the biocytin-filled boutons were localized to st. pyramidale and only rarely approached ankyrin G-stained profiles

(Fig. 1c 1–3) suggesting their FSBC [fast-spiking Basket cell] origin. Conversely, in the remaining 31 cases the axonal arbor was densest in st. pyramidale and neighboring st. oriens border, and the boutons formed close appositions with ankyrin G-immunoreactive segments, often in climbing-fiber manner, similar to the axon terminals of AAC [Axo-axonic cells]."

As a first step in determining counts from this quotation, the neuron types were mapped to Hippocampome.org neuron types. As per the description and figures in the paper being mined, it was determined that the neuron types of interest were Basket cells and Axo-axonic cells in the area CA3. The cited paragraph provides further information about the proportions from the recorded results. Here, we interpret that the ratio of Basket to Axo-axonic cells is 30:31. We mined several such electrophysiology based experimental sources for each subregion.

Hippocampome.org provides a list of neuron types that are either positive or negative, or in other cases, 'undetermined' for 20 different biomarkers. We used information from a few articles that published neuronal counts or densities of the neuron types positive to the biomarker/s of interest. These numbers were used to construct linear inequalities in conjunction with biomarker data on Hippocampome.org. For example, Jinno and Kosaka report numerical densities of GABAergic neuron types for eight different biomarkers in the mouse hippocampus (Jinno and Kosaka, 2006).

The biomarker page of Hippocampome.org marks the 122 neuron types as ‘positive,’ ‘negative,’ or ‘undetermined’ for twenty biomarkers (White et al., 2019). These molecular expression profiles are a result of existing evidence from multiple sources of scientific literature and relational inferences. There are several instances with conflicted results regarding biomarker reactivity. In some cases, the neuron type has both positive and negative subtypes to the biomarker of interest – for example, the horizontal Basket cells of region CA1 have mixed reactivity towards parvalbumin (Tricoire et al., 2010; Losonczy et al., 2002; Maccaferri et al., 2000). In other cases, there is no known evidence of reactivity to a biomarker, in which case reactivity to the biomarker is labeled as ‘unknown.’

Since there are a significant proportion of ‘undetermined’ neuron types (positive-negative subtypes, unresolved subtypes, and unknown subtypes) on the Hippocampome.org biomarker page, we could not directly translate data from the literature into an equation, as in the case of sums and ratios. In the case of biomarkers, we translated the data into inequalities, such as the evidence for biomarker reactivity that has been reported by Kim et al. and Jinno and Kosaka (Kim et al., 2017; Jinno and Kosaka, 2006). Kim et al. studied mouse neuron numbers that are positive for parvalbumin, somatostatin, and vasoactive intestinal polypeptide, and Jinno and Kosaka (Jinno and Kosaka, 2006) reported, also in mouse, the numerical densities of neurons in the hippocampus positive for eight different biomarkers by layer. In addition, a few inequalities describing parvalbumin and calbindin reactivity were reported by Bjerke (Bjerke et al., 2021).

These numerical densities for the mouse were converted into neuronal numbers for rat by first multiplying with volumes (Ero et al., 2018) and then applying rodent scaling rules (Herculano Houzel et al., 2006). For example – after conversion into rat neurons, the number of parvalbumin positive neurons in the granule layer of the dentate gyrus is equal to 3,680. As per Hippocampome.org, the relevant neuron types include Axo-axonic, DG SG PV+ Basket cells, and MOLAX, whereas Total Molecular Layer and Outer Molecular Layer neuron types have an undetermined status. Using data from Jinno and Kosaka, we propose that the number of PV positive neurons in the granule layer is no more than 3,680, and the number of PV positive and undetermined neurons combined is no less than 3,680 (Table 9). In total, 290 biomarker inequalities were constructed for the six subregions using existing scientific data. Table 9 provides examples of the three types of equations constructed as part of the data transformation process – sums, ratios, and inequalities.

Table 9 Three types of equations obtained from the data transformation process

Equation Type	Sum	Ratio	Inequality
Region	CA1	CA3	Dentate Gyrus
Source	Coulin, 2001	Gulyas, 2010	Jinno and Kosaka, 2006
Interpretation	CA1 Cajal-Retzius = 1130	Ratio between axoaxonic and basket cells is 30:31	Number of parvalbumin positive neurons in granule layer is less than 3680

Equation	$x_3 = 1130$	$\frac{x_6}{x_8} = \frac{31}{30}$	$x_7 + x_8 \leq 3680$ $x_7 + x_8 + x_{17} + x_{21} + x_{22} \geq 3680$
Normalized form	$\frac{x_3}{1130} - 1 = 0$	$\frac{30x_6}{31x_8} - 1 = 0$	$\frac{x_7 + x_8}{3680} - 1 \leq 0$ $\frac{x_7 + x_8 + x_{17} + x_{21} + x_{22}}{3680} - 1 >= 0$
Least squares form	$\left(\frac{x_3}{1130} - 1\right)^2 = 0$	$\left(\frac{30x_6}{31x_8} - 1\right)^2 = 0$	$max\left(0, \frac{x_7 + x_8}{3680} - 1\right)^2$ $max\left(0, 1 - \frac{x_7 + x_8 + x_{17} + x_{21} + x_{22}}{3680}\right.$
With weights	$r = 10\left(\frac{x_3}{1130} - 1\right)^2$	$r = 1\left(\frac{30x_6}{31x_8} - 1\right)^2$	$r = 1 \times max\left(0, \frac{x_7 + x_8}{3680} - 1\right)^2$ $r = 1 \times max\left(0, 1 - \frac{x_7 + x_8 + x_{17} + x_{21} + x_{22}}{3680}\right.$

Based on this broad classification of literary sources, we mined several hundreds of papers for relevant information on neuron type numbers, ratios, or biomarkers. We used data from 155 out of 500 journal articles that were mined to construct 645 equations. As expected, we saw that there was high amount of evidence and, hence, the largest number of equations for the subregion CA1, which is known to be extensively researched. Many sources reported combined results for subregions CA3 and CA2, which prompted us to

design our equations together for these areas. The least number of papers and subsequently equations happen to be for the SUB.

The compiled equations were used to construct an objective function for each subregion. We then applied the optimization algorithm, PSwarm (Vaz and Vicente, 2007), to obtain estimated populations for 207 neuron types in the hippocampal formation. The obtained results are the optimum (minimized residuals) produced by the algorithm, after considering the data from the 155 research studies. The cell populations of the 207 neuron types, along with ranges and layer specific information, are available in the supplementary data.

The total number of neurons from all subregions was estimated to be 2,958,500. The highest numbers are from the DG at 1,197,548, and the lowest are from the subregion CA2 at 29,493 (Fig. 10A). EC has been divided into the lateral and medial areas. Surprisingly, the lateral EC has a higher number of neurons at 583,002 compared to the medial EC population size at 196,452. Interestingly, in CA1 and CA3, SO had the highest ratio of interneurons, while in CA2, it was SLM.

As a next step after obtaining estimates of the neuron type populations, we conducted sensitivity analysis to determine ranges and examined the proportions across different

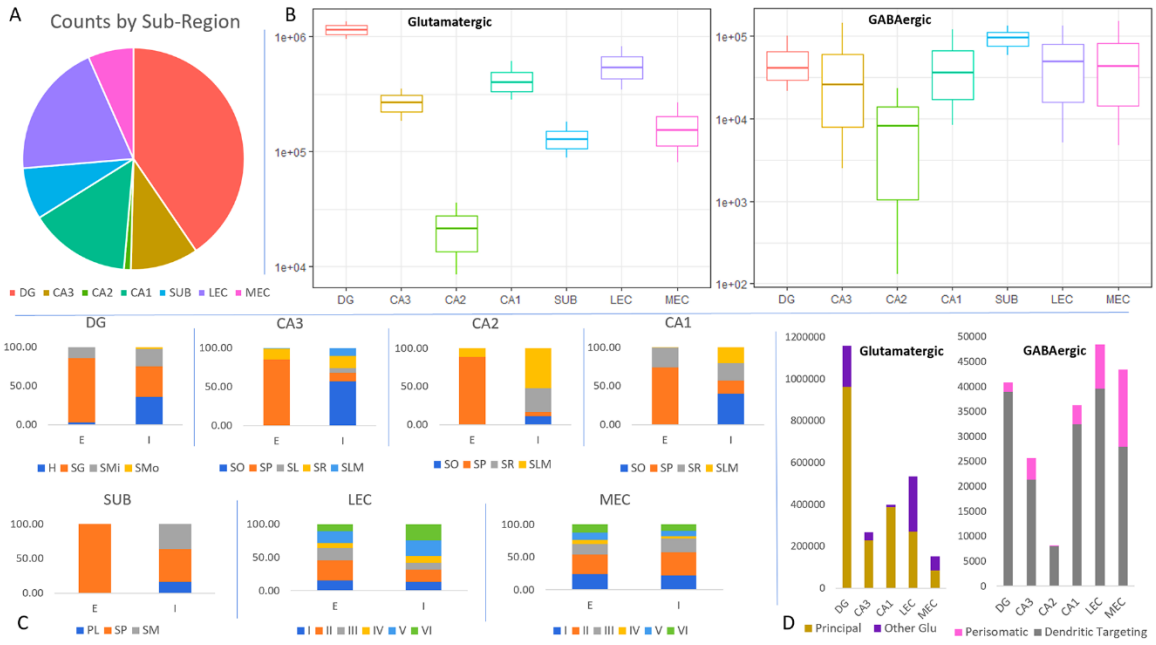

Figure 10 High-level statistics for 122 neuron types and 9 additional neuron groups

A. Proportions of neurons by subregions (EC is divided into lateral (LEC) and medial (MEC) areas in all subpanels of the figure). Total Counts: DG – 1197548, CA3 – 293278, CA2 – 29493, CA1 – 435735, SUB – 222992, LEC – 583002, MEC – 196452; B. Totals and ranges for glutamatergic (left) and GABAergic (right) neurons; Y-axis: Counts in log scale. C. Percentages of neurons in the corresponding layers for excitatory and inhibitory types. Y-axis represents Counts. Layer abbreviations: outer stratum (s.) moleculare (SMo), inner s. moleculare (SMi), s. granulosum (SG), hilus (H); s. lacunosum-moleculare (SLM), s. radiatum (SR), s. lucidum (SL), s. pyramidale (SP), s. oriens (SO); SUB s. moleculare (SM), polymorphic layer (PL); and EC layers I–VI; D. Distribution of high-level groups of neuron types – principal cells vs. other glutamatergic cells and dendritic-targeting vs perisomatic cells; Y-axis represents Counts.

layers and high-level groups. Ranges were estimated for each neuron type as an outcome of our sensitivity analysis (Fig. 10B). Our estimates for interneurons had a wider range indicating a higher error margin. Also, the wider ranges could be due to the fact that the principal cells, which comprise a major portion of the glutamatergic population, are extensively researched compared to the GABAergic populations. Overall, the total GABAergic estimate is 297,803, which is 10% of the total population, although this proportion varies in individual subregions. We examined the proportions of excitatory and inhibitory neuron type populations from individual layers for each subregion (Fig. 10C). As expected, the principal cell layers occupied the highest proportion of cells in the excitatory category. In the inhibitory category, however, an uneven distribution of proportions of the different layers was observed among subregions. To further analyze the population distributions, we divided the neurons into four high level categories – principal cells and all other glutamatergic cells and perisomatic and dendritic-targeting interneurons. Initially, we studied the population distribution of these categories among glutamatergic and GABAergic types (Fig. 10D). There are no non-principal glutamatergic cells classified for the areas - CA2 and the SUB on Hippocampome.org. There is no known evidence regarding the distinction of the GABAergic cell types for the SUB, as of Hippocampome.org v1.8. Thus, the bar plots of Fig. 10D show the

distributions of neuronal numbers for the rest of the subregions, excluding the SUB and glutamatergic CA2.

We saw a wide range between the estimates of interneurons across the subregions. The DG MOCAPs , CA3 QuadD-LMs, CA2 SLM inhibitory types, CA1 Neurogliaforms, SUB SP interneurons, and deep layer interneurons from the EC had the highest estimates. The DG Hilar Ectopic Granule cells CA3 MFA ORDEN , CA2 Bistratified, CA1 Perforant Path-Associated Quad, SUB PL interneurons, and MEC LIII Superficial Trilayer interneurons have the lowest population estimates from their respective subregions.

Figure 11 illustrates neuron type counts and their upper and lower bound ranges for all subregions. Based on the reliability of the numbers, we marked the types in 3 different colors – blue, orange and pink. The blue types hold a higher reliability, these are neuron types that are extensively researched and, in most cases, stereologically or

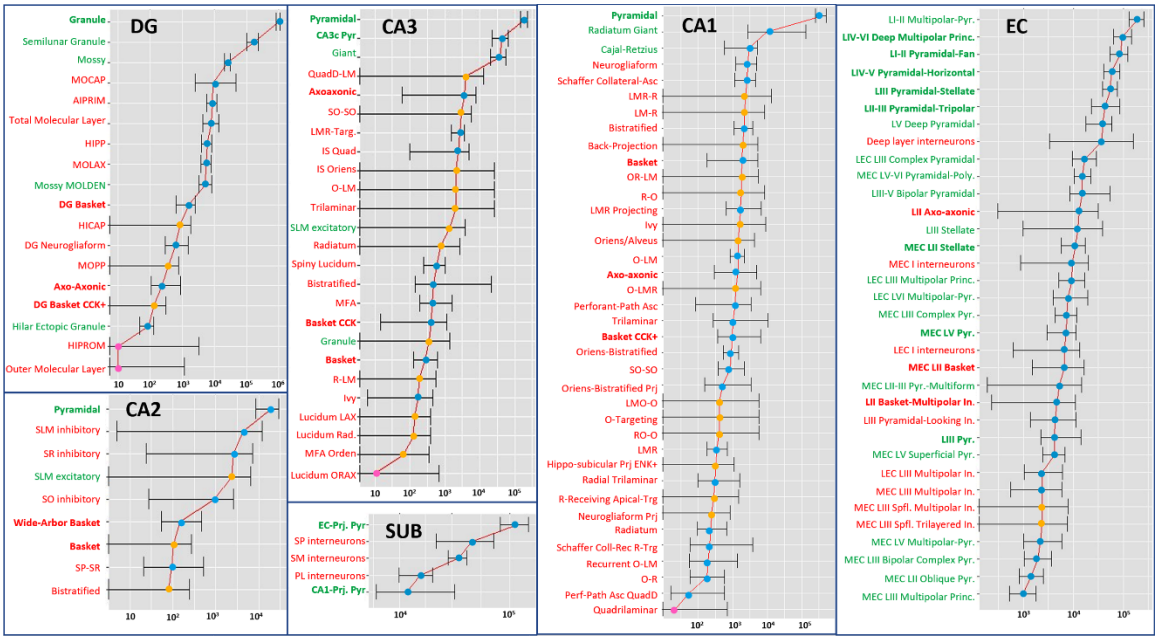

Figure 11 Sensitivity analysis

Neuron type counts and ranges for 122 types. X-axis: Neuronal counts in log scale; Y-axis: Neuron type names. Green text – Excitatory neurons; Red text – Inhibitory neurons; Green bolded text – Principal cells; Red Inhibitory – Perisomatic cells; Green unbolded text – Other glutamatergic cells; Red unbolded text – Dendritic-Targeting cells. Blue dot – High reliability count; Orange dot – Medium reliability count; Pink dot – Low reliability count.

computationally counted. We observed that for some of the neuronal types, the lower bound is reaching 0, which makes the range wider. These are usually the neuron types for which there is no direct evidence, and information from electrophysiological ratios and

layer specific totals were used. Based on these factors, we label these neuron types with a lower level of reliability (these neuron types are labeled orange). Four neuron types are labeled pink as the estimates were the lower bound values provided to the optimization algorithm. These are the DG HIPROMs and DG Outer Molecular Layer interneurons, CA3 Lucidum ORAX, and CA1 Quadrilaminar. We suggest the existence of these neuron types to be further investigated in future studies, considering the sparse evidence for these types in the literature. In all other subregions, the principal cells and the dendritic targeting cells outweigh their excitatory and inhibitory counterparts with respect to quantities.

Hippocampome.org has collated information about 20 biomarkers (White et al., 2019) and 23 firing pattern phenotypes of neuronal types (Komendantov et al., 2019). We plotted this information as a function of our neuron type census data to observe the distributions (Fig. 12). We illustrated interneuron reactivity for nineteen biomarkers (Fig. 12A). We categorized neuron type population distributions layer-wise. The highest number of interneurons in the hippocampal formation are positive to Neuropeptide Y, and the lowest number are positive to enkephalin. Among firing patterns, the adapting spiking (ASP) described the majority of the neuron types. Transient stuttering pattern (TSTUT.SLN, silence preceded by transient stuttering) is exhibited by the least number of neurons. Most of the information for Figs. 12A and 12B comes from the DG and CA1, which are the two most studied areas of the rodent hippocampus. CA2 and the SUB are the least studied areas among the 6 subregions, which is reflected in the figure.

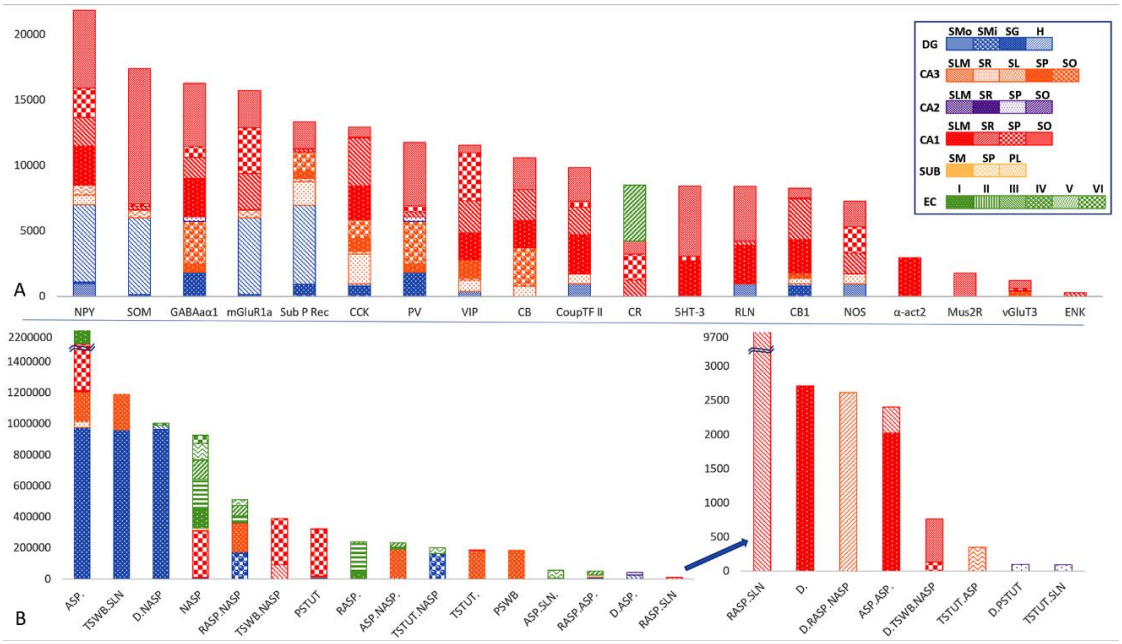

Figure 12 Distribution of neuron counts for biomarker reactivity and firing patterns

A. Distribution of 19 biomarker reactive interneurons in the 26 layers of the six hippocampal subregions; B. Distribution of neurons from the 26 layers of the hippocampal formation pertaining to the 23 firing phenotypes. Firing pattern abbreviations: ASP. – adapting spiking, ASP.ASP. – adapting spiking followed by (slower) adapting spiking, ASP.NASP. – non-adapting spiking preceded by adapting spiking, ASP.SLN – silence preceded by adapting spiking, D. – delayed spiking, D.ASP. – delayed adapting spiking, D.RASP.NASP – non-adapting spiking preceded by delayed fast-adapting spiking, D.NASP – delayed non-adapting spiking, D.PSTUT – delayed persistent stuttering, D.TSWB.NASP – non-adapting spiking preceded by delayed

transient slow-wave bursting, RASP. – fast-adapting spiking, RASP.ASP. – fast-adapting spiking followed by adapting spiking, RASP.NASP – non-adapting spiking preceded by fast-adapting spiking, RASP.SLM – silence preceded by fast adapting spiking, NASP – non-adapting spiking, PSTUT – persistent stuttering, PSWB – persistent slow-wave bursting, TSTUT. – transient stuttering, TSTUT.ASP. – transient stuttering followed by adapting spiking, TSTUT.NASP – non-adapting spiking preceded by transient stuttering, TSTUT.SLN – silence preceded by transient stuttering, TSWB. NASP – non-adapting spiking preceded by transient slow-wave bursting, TSWB.SLN – silence preceded by transient slow-wave bursting.

Next, we investigated the variation of the population counts for the main excitatory and inhibitory families (principal neurons, other glutamatergic cells, perisomatic interneurons, and dendritic-targeting interneurons) as a function of the volumes of each anatomical parcel (subregion and layer) in the hippocampal formation. If each neuronal population has (approximately) constant density across parcels, the population count should be proportional to the anatomical volumes. If, in contrast, the population counts themselves were approximately constant, then the densities should be inversely proportional to the anatomical volumes. We tested these alternative hypotheses by plotting the population counts against parcel volumes and the densities against the inverse of the volumes for each of the four populations (Fig. 4). As general trends, neuron counts

were proportional to anatomical volumes for other (non-principal) glutamatergic neurons and for perisomatic interneurons. In those two cases the densities did not vary with the inverse of volumes. In contrast, in the other populations, principal cells and dendritic-targeting interneurons, the counts did not vary with volumes. Instead, for the dendritic-targeting interneurons, the densities were inversely proportional to the volume with high statical significance ($p<10\text{-}4$ after multiple testing correction). The analysis results were the same when analyzing either rat or mouse data. Thus, we conclude that dendritic-targeting interneurons have constant counts independent of the volume of the anatomical parcel in which their somata reside.

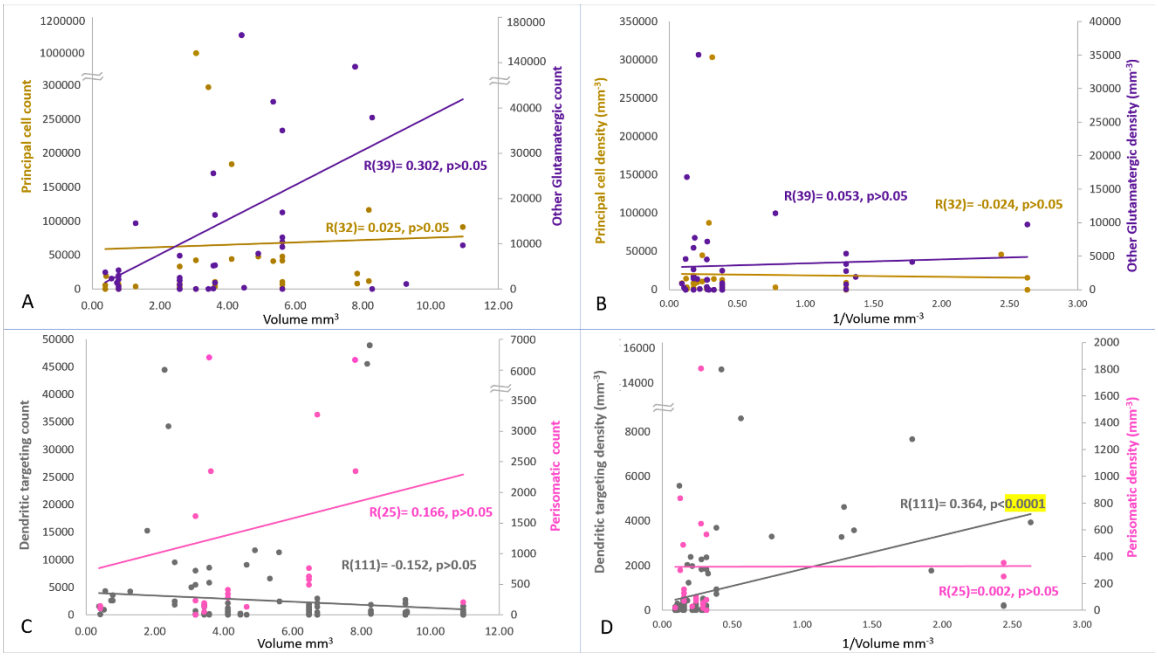

Figure 13 Neuron types counts/densities vs volume

Numerical estimates (A) and numerical densities (B) of principal cells (left Y axis) and other glutamatergic (right Y axis) with respect their corresponding layer-specific volumes (X axis); Numerical estimates (C) and numerical densities (D) of dendritic-targeting (left Y axis) and perisomatic (right Y axis) with respect their corresponding layer-specific volumes (X axis). Number in the parenthesis is n value.

We have made all the code and mined data available through the supplementary material for further research and analysis. The numerical estimates have been made available on the Hippocapome.org site for each neuron type on the neuron specific page (Fig. 14). The 'quantification browse matrix' page displays numerical estimates for both rat and mouse along with the lower and upper bounds of ranges upon hovering over a number. Since these estimates are a result of multiple sources, they will be linked to the evidence page, where the different source and excerpts will be displayed for the corresponding neuron type. Apart from this, the neuron type quantities and ranges are also displayed on the neuron type specific page, which can be accessed from all the Hippocampome.org pages. It should be noted that the estimates are a result of optimization performed using multiple sources and, hence, the displayed counts may not be equal to those on the evidence reported.

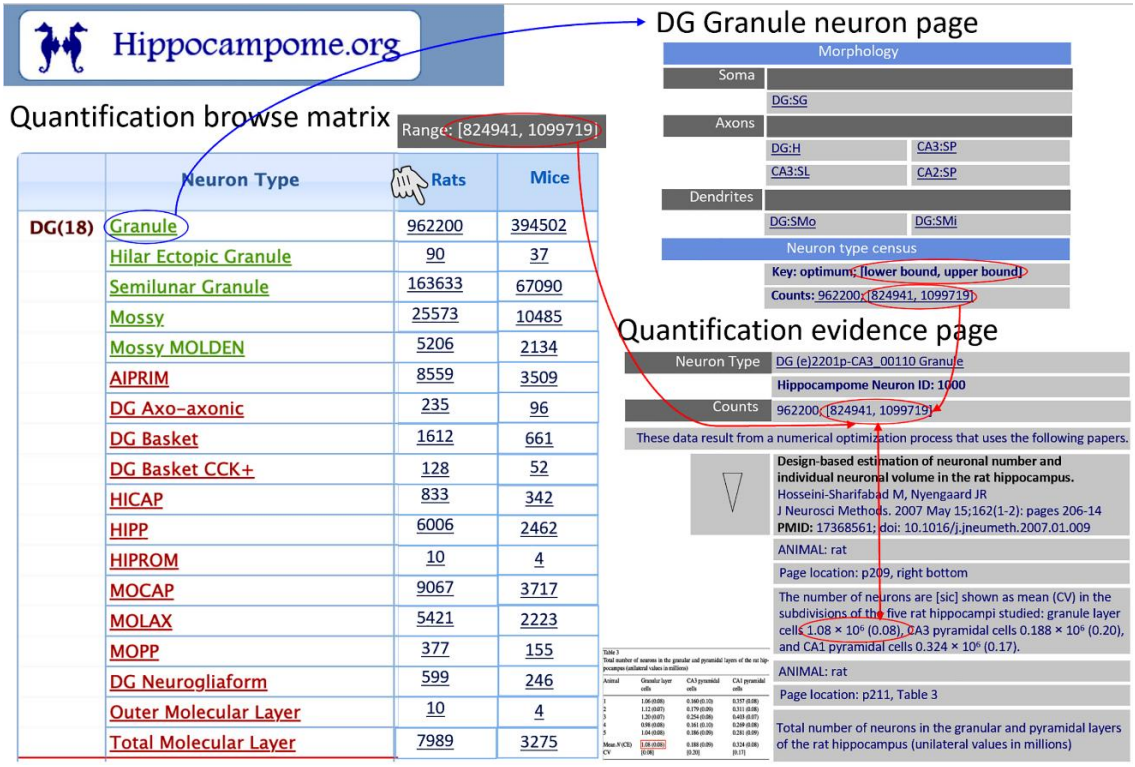

Figure 14 Quantification matrix on Hippocampome.org

Displaying the list of neuron type quantified estimates for rats and mice (left), corresponding neuron type page (top right) and evidence page (bottom right).

Discussion

Estimates for neuron type populations were obtained through the described computational methods. However, it must be noted that the results are largely a reflection of the amount of research conducted on the specific cell types. Since our method heavily relies on the

existing scientific literature, we considered the data mining phase as an important beginning. A careful step by step process was followed for the literature search, with the goal of obtaining every single relevant journal article or scientific report for all the neuron types of interest. It was clear from our data mining process that some areas of the brain, like the CA1 stratum pyramidale and the DG granule layer, for instance, were extensively studied by multiple teams and, hence, there were several results or evidence for neurons from these regions. In contrast, areas such as the SUB and CA2, which is a relatively small area of the hippocampus between CA1 and CA3, have been less studied compared to the rest of the hippocampal areas.

Principal cells being the pathways of communication between the different subregions are well studied (Wang et al., 2006; Barkai and Hasselmo, 1994; Yang et al., 1996; Hestrin et al., 1990). Interneurons that reside in the principal cell layers, such as Basket and Axo-axonal cells, which provide synapses to the principal cells, have also been well researched compared to interneurons that reside in non-principal layers. Through the data mining process, we obtained many results from the principal cell layer neuron types and fewer results from non-principal cell layer types for the six subregions.

Neurons in the principal cell layers were mostly counted using stereology (Hossieni-Sharifabad and Nyengaard, 2007; Fitting et al., 2009, Grady et al., 2003) and more recently using computational methods (Murakami et al., 2018, Ero et al., 2018, Attili et al., 2019). These types of results were used to put together layer sum equations, which

were important for the optimization process, since they helped create an estimated upper bound for that set of neuron types. In some cases though, we ended up with multiple layer sums that had a wide range. For example, West et al. report 380,000 CA1 principal layer neurons (West et al., 1991) , whereas Fitting et al. report 262,181 (Fitting et al., 2009). Yet another study reports 650,000 CA1 principal cell layer neurons for non-impaired aged rats and 570,000 from select young control rats (Rasmussen et al., 1996). It should be noted that the rat species, age, or experimental conditions vary between the studies. Similarly, we found that the ratios between the same set of neurons, from two different studies, were never the same. For example, the ratio between DG SMo MOPP and Neurogliaform cells has been derived to be 6:11 (0.54) as per Armstrong (Armstrong et al., 2011) and 14:17 (0.82) as per Ceranik (Ceranik et al., 1997).

Considering the diversity in the results reported in the literature, we aimed to find an optimum using mathematical optimization methods. In cases where we had evidence for certain neuron types but no corresponding neuron type on Hippocampome.org, we added a variable for the unknown and integrated it into the optimization process. For example, there is no known type for excitatory neurons in the SLM layer of CA3, but Ero et al. give a numerical estimate for these cells (Ero et al., 2018). Dividing Hippocampome.org cell types by layers and adding non-existent neuron types where evidence was found, we came up with a total of 207 neuron types and 645 equations (Table 8).

For each of the six subregions, 200 iterations of optimization were run, and those with the lowest residual were chosen for analysis. Analysis led to the discovery of interdependences between some neuron types. Whenever there were such dependencies between neuron types, their sums always remained constant. We equally divided their sums by the number of neuron types that have dependencies to arrive at their population estimates. The complete set of results including data from multiple iterations is made available in the supplementary material.

Sensitivity analysis let us examine the reliability of our estimates based on how the upper and lower limits ranged with respect to the optimum. Biomarker (White et al., 2019) and firing pattern data (Komendantov et al., 2019) were used to map the neuronal estimates derived from the numerical optimization. This gave is an overview of the proportions of neuron types that exhibited reactivity to certain biomarkers or certain firing pattern/s. Hippocampal cell population distributions (neurons and glia together) were observed to be proportional to the layer specific areas from images in our previous work (Attili et al., 2019). It was interesting to discover that the dendritic targeting numerical densities were significantly decreasing with increasing layer-specific volumes, after undergoing statistical correction. It should be noted that this group is the largest among the four groups (principal cells, other glutamatergic, perisomatic interneurons and dendritic targeting) with 111 neuron types.

The results from this research project - estimates for the 122 neuronal types, as classified by the Hippocampome.org, and any associated source code, are being released to the public for free. We anticipate that our results will be used by future projects that involve the development of computational working models of the rodent brain. Our methodology can be extended to obtain any cell type populations across the different brain regions and species.

Acknowledgements: This project is supported in parts by grants R01NS39600 and U01MH114829. We are grateful to Drs. Padmanabhan Seshaiyer (Professor of Mathematical Sciences, George Mason University) and Siva Venkadesh (from the author's lab) for critical discussions.

Conflict of interest: The authors have no conflict of interest to declare.

Author Contributions: SMA performed initial research, methodology generation, data mining, data transformation, numerical optimization, analysis, manuscript writing, tables and figures (10-14) generation. KM contributed to data mining, numerical optimization, and editing of manuscript, figures and tables. DWW contributed to figure 14 generation, and editing of manuscript, figures and tables. Both KM and DWW provided their subject matter expertise during critical discussions. This research study was conducted under the guidance of GAA who supervised all aspects of the project.

REFERENCES

Abercrombie M (1946) Estimation of nuclear population from microtome sections. The Anatomical Record 94:239-247. https://doi:10.1002/ar.1090940210

Allen Data Production (2011) Allen mouse brain atlas technical white paper: in situ hybridization data production.

http://help.brain - map.org/download/attachments/2818169/ABADataProductionProcesses.pdf. Accessed 15 Feb 2017

Amaral DG, Ishizuka N, Claiborne B (1990) Chapter 1 Chapter Neurons, numbers and the hippocampal network. Progress in Brain Research Understanding the Brain Through the Hippocampus the Hippocampal Region as a Model for Studying Brain Structure and Function. 1-11. doi:10.1016/s0079-6123(08)61237-6

Andrade JP, Madeira M, Paula-Barbosa M (2000) Sexual dimorphism in the subiculum of the rat hippocampal formation. Brain Research 875:125-137. https://doi:10.1016/s0006-8993(00)02605-6

Andrey P, Kiêu K, Kress C, et al (2010) Statistical Analysis of 3D Images Detects Regular Spatial Distributions of Centromeres and Chromocenters in Animal and Plant Nuclei. PLoS Computational Biology 6. https://doi:10.1371/journal.pcbi.1000853

Ascoli GA, Alonso-Nanclares L, Anderson S, et al (2008) Petilla terminology: Nomenclature of features of GABAergic interneurons of the cerebral cortex. Nature Reviews Neuroscience 9:557-568. https://doi:10.1038/nrn2402

Baldwin SA, Gibson T, Callihan CT, Sullivan PG, Palmer E, Scheff SW (1997) Neuronal Cell Loss in the CA3 Subfield of the Hippocampus Following Cortical Contusion Utilizing the Optical Disector Method for Cell Counting. Journal of Neurotrauma 14:385-398. https://doi:10.1089/neu.1997.14.385

Bahney J, Bartheld CS (2017). The Cellular Composition and Glia-Neuron Ratio in the Spinal Cord of a Human and a Nonhuman Primate: Comparison With Other Species and Brain Regions. The Anatomical Record 301:697-710. https://doi:10.1002/ar.23728

Bayer S, Yackel J, Puri P (1982) Neurons in the rat dentate gyrus granular layer substantially increase during juvenile and adult life. Science 216:890-892. https://doi:10.1126/science.7079742

Bezaire MJ, Raikov I, Burk K, Vyas D, Soltesz I (2016) Interneuronal mechanisms of hippocampal theta oscillations in a full-scale model of the rodent CA1 circuit. ELife 5. https://doi:10.7554/elife.18566

Bhanu B, Peng J (2000) Adaptive integrated image segmentation and object recognition. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews) 30:427-441. https://doi:10.1109/5326.897070

Boss BD, Peterson GM, Cowan WM (1985) On the number of neurons in the dentate gyrus of the rat. Brain Research 338:144-150. https://doi:10.1016/0006-8993(85)90257-4

Boyce RW, Gundersen HJ (2018) The Automatic Proportionator Estimator Is Highly Efficient for Estimation of Total Number of Sparse Cell Populations. Frontiers in Neuroanatomy 12. https://doi:10.3389/fnana.2018.00019

Bray M, Vokes MS, Carpenter AE (2015) Using CellProfiler for Automatic Identification and Measurement of Biological Objects in Images. Current Protocols in Molecular Biology. https://doi:10.1002/0471142727.mb1417s109

Brun VH, Solstad T, Kjelstrup KB, Fyhn M, Witter MP, Moser EI, Moser M (2008) Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. Hippocampus 18(12): 1200-1212. doi:10.1002/hipo.20504

Calhoun ME, Kurth D, Phinney AL, et al (1998) Hippocampal neuron and synaptophysin-positive bouton number in aging C57BL/6 mice. Neurobiology of Aging 19:599-606. https://doi:10.1016/s0197-4580(98)00098-0

Canto CB, Wouterlood FG, Witter MP (2008) What Does the Anatomical Organization of the Entorhinal Cortex Tell Us? Neural Plasticity 2008:1-18. https://doi:10.1155/2008/381243

Erö C, Gewaltig M, Keller D, Markram H (2018) A Cell Atlas for the Mouse Brain. Frontiers in Neuroinformatics 12. https://doi:10.3389/fninf.2018.00084

Fitting S, Booze RM, Hasselrot U, Mactutus CF (2009) Dose-dependent long-term effects of Tat in the rat hippocampal formation: A design-based stereological study. Hippocampus. https://doi:10.1002/hipo.20648

García-Cabezas MÁ, John YJ, Barbas H and Zikopoulos B (2016) Distinction of Neurons, Glia and Endothelial Cells in the Cerebral Cortex: An Algorithm Based on Cytological Features. Frontiers in Neuroanatomy 10. https://doi: 10.3389/fnana.2016.00107

Giocomo LM, Hasselmo ME (2008) Time constants of h current in layer II stellate cells differ along the dorsal to ventral axis of medial entorhinal cortex. Journal of Neuroscience 28:9414-25. https://doi: 10.1523/JNEUROSCI.3196-08.2008

Grady MS, Charleston JS, Maris D, Witgen BM, Lifshitz J (2003) Neuronal and Glial Cell Number in the Hippocampus after Experimental Traumatic Brain Injury: Analysis by Stereological Estimation. Journal of Neurotrauma 20:929-941. https://doi:10.1089/089771503770195786

Goldenberg TM, Bakay RA, Ribak CE (1995) Electron microscopy of cell islands in layer II of the primate entorhinal cortex. The Journal of Comparative Neurology 355:51-66. https://doi:10.1002/cne.903550108

Gonzalez-Riano C, Tapia-González S, García A, Muñoz A, DeFelipe J, Barbas C (2017) Metabolomics and neuroanatomical evaluation of post-mortem changes in the hippocampus. Brain Structure and Function 222:2831-2853. https://doi:10.1007/s00429-017-1375-5

Häder D (2001) Image analysis: methods and applications. CRC Press, Florida

Hasselmo ME, Stern CE (2015) Current questions on space and time encoding. Hippocampus 25:744-752. https://doi:10.1002/hipo.22454

Hedreen JC (1998a) Lost caps in histological counting methods. The Anatomical Record 250:366-372. https://doi:10.1002/(sici)1097-0185(199803)250:3<366::aid-ar11>3.3.co;2-v

Hedreen JC (1998b) What was wrong with the Abercrombie and empirical cell counting methods? A review. The Anatomical Record, 250:373-380. https://doi:10.1002/(sici)1097-0185(199803)250:3<373::aid-ar12>3.0.co;2-l

Herculano-Houzel S, Lent R (2005) Isotropic Fractionator: A Simple, Rapid Method for the Quantification of Total Cell and Neuron Numbers in the Brain. Journal of Neuroscience 25:2518-2521. https://doi:10.1523/jneurosci.4526-04.2005

Herculano-Houzel S, Mota B, Lent R (2006) Cellular scaling rules for rodent brains. PNAS 103:12138-12143. https://doi: 10.1073/pnas.0604911103

Herculano-Houzel S, Ribeiro P, Campos L, Silva AV, Torres LB, Catania KC, Kaas JH (2011) Updated Neuronal Scaling Rules for the Brains of Glires (Rodents/Lagomorphs). Brain, Behavior and Evolution 78:302-314. https://doi:10.1159/000330825

Herculano-Houzel S, Watson C, Paxinos G (2013) Distribution of neurons in functional areas of the mouse cerebral cortex reveals quantitatively different cortical zones. Frontiers in Neuroanatomy 7. https://doi:10.3389/fnana.2013.00035

Hosseini-Sharifabad M, Nyengaard JR (2007) Design-based estimation of neuronal number and individual neuronal volume in the rat hippocampus. Journal of Neuroscience Methods 162:206-214. https://doi:10.1016/j.jneumeth.2007.01.009

Hu T, Xu Q, Lv W, Liu Q (2017) Touching Soma Segmentation Based on the Rayburst Sampling Algorithm. Neuroinformatics 15:383-393. https://doi:10.1007/s12021-017-9336-y

Insausti AM, Megías M, Crespo D, et al (1998) Hippocampal volume and neuronal number in Ts65Dn mice: a murine model of down syndrome. Neuroscience Letters 253:175-178. https://doi:10.1016/s0304-3940(98)00641-7

Insausti R, Herrero MT, Witter MP (1998) Entorhinal cortex of the rat: Cytoarchitectonic subdivisions and the origin and distribution of cortical efferents. Hippocampus 7(2):146-183. https://doi:10.1002/(sici)1098-1063(1997)7:23.0.co;2-l

Insel TR, Landis SC, Collins FS (2013) The NIH BRAIN Initiative. Science 340:687-688. https://doi:10.1126/science.1239276

Jinno S, Fleischer F, Eckel S, Schmidt V, Kosaka T (2007) Spatial arrangement of microglia in the mouse hippocampus: A stereological study in comparison with astrocytes. Glia 55(13):1334-1347. https://doi:10.1002/glia.20552

Jones AR, Overly CC, Sunkin SM (2009) The Allen Brain Atlas: 5 years and beyond. Nature Reviews Neuroscience 10:821-828. https://doi:10.1038/nrn2722

Kaae SS, Chen F, Wegener G, Madsen TM, Nyengaard JR (2012) Quantitative hippocampal structural changes following electroconvulsive seizure treatment in a rat model of depression. Synapse 66:667-676. https://doi:10.1002/syn.21553

Kandel ER (2004) The Molecular Biology of Memory Storage: A Dialog Between Genes and Synapses. Bioscience Reports 24:475-522. https://doi:10.1007/s10540-005-2742-7

Kandel ER, Markram H, Matthews PM, Yuste R, Koch C (2013) Neuroscience thinks big (and collaboratively). Nature Reviews Neuroscience 14:659-664. https://doi:10.1038/nrn3578

Kayasandik CB, Labate D (2016) Improved detection of soma location and morphology in fluorescence microscopy images of neurons. Journal of Neuroscience Methods 274:61-70. https://doi:10.1016/j.jneumeth.2016.09.007

Kim Y, Yang GR, Pradhan K, et al (2017) Brain-wide Maps Reveal Stereotyped Cell-Type-Based Cortical Architecture and Subcortical Sexual Dimorphism. Cell 171. https://doi:10.1016/j.cell.2017.09.020

Kitamura T, Pignatelli M, Suh J, Kohara K, Yoshiki A, Abe K, Tonegawa S (2014) Island Cells Control Temporal Association Memory. Science 343(6173): 896-901. doi:10.1126/science.1244634

Lamprecht M, Sabatini D, Carpenter A (2007) CellProfiler™: free, versatile software for automated biological image analysis. BioTechniques 42:71-75. https://doi:10.2144/000112257

Latorre A, Alonso-Nanclares L, Muelas S, Peña J, Defelipe J (2013) Segmentation of neuronal nuclei based on clump splitting and a two-step binarization of images. Expert Systems with Applications 40:6521-6530. https://doi:10.1016/j.eswa.2013.06.010

Lau C, Ng L, Thompson C, et al (2008). Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. BMC Bioinformatics 9:153. https://doi:10.1186/1471-2105-9-153

Lister JP, Tonkiss J, Blatt GJ, Kemper TL, Debassio WA, Galler JR, Rosene DL (2006) Asymmetry of neuron numbers in the hippocampal formation of prenatally malnourished and normally nourished rats: A stereological investigation. Hippocampus 16:946-958. https://doi:10.1002/hipo.20221

Long JM, Kalehua AN, Muth NJ et al (1998) Stereological analysis of astrocyte and microglia in aging mouse hippocampus. Neurobiology of Aging 19:497-503. https://doi:10.1016/s0197-4580(98)00088-8

Luengo-Sanchez S, Bielza C, Benavides-Piccione R, Fernaud-Espinosa I, Defelipe J, Larrañaga P (2015) A univocal definition of the neuronal soma morphology using Gaussian mixture models. Frontiers in Neuroanatomy 9. https://doi:10.3389/fnana.2015.00137

Maechler M (2016) Package 'diptest' (Tech.). Retrieved June 15, 2018, from https://cran.r-project.org/web/packages/diptest/diptest.pdf. Accessed 15 June 2018

Malberg JE, Eisch AJ, Nestler EJ, Duman RS (2000) Chronic Antidepressant Treatment Increases Neurogenesis in Adult Rat Hippocampus. The Journal of Neuroscience 20:9104-9110

Meyer HS, Wimmer VC, Oberlaender M, Kock CP, Sakmann B, Helmstaedter M (2010) Number and Laminar Distribution of Neurons in a Thalamocortical Projection

Column of Rat Vibrissal Cortex. Cerebral Cortex 20:2277-2286. https://doi:10.1093/cercor/bhq067

Miki T, Satriotomo I, Li H et al (2005) Application of the physical disector to the central nervous system: Estimation of the total number of neurons in subdivisions of the rat hippocampus. Anatomical Science International 80:153-162. https://doi:10.1111/j.1447-073x.2005.00121.x

Moser EI, Moser M, Mcnaughton BL (2017) Spatial representation in the hippocampal formation: A history. Nature Neuroscience 20:1448-1464. https://doi:10.1038/nn.4653

Mulders W, West M, Slomianka L (1997) Neuron numbers in the presubiculum, parasubiculum, and entorhinal area of the rat. The Journal of Comparative Neurology 385:83-94. https://doi:10.1002/(sici)1096-9861(19970818)385:1<83::aid-cne5>3.0.co;2-8

Murakami TC, Mano T, Saikawa S et al (2018) A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing. Nature Neuroscience 21:625-637. https://doi:10.1038/s41593-018-0109-1

Nakatomi H, Kuriu T, Okabe S et al (2002) Regeneration of Hippocampal Pyramidal Neurons after Ischemic Brain Injury by Recruitment of Endogenous Neural Progenitors. Cell 110:429-441. https://doi:10.1016/s0092-8674(02)00862-0

Otsu N (1979) A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics 9:62-66. https://doi:10.1109/tsmc.1979.4310076

Peng H, Roysam B, Ascoli GA (2013) Automated image computing reshapes computational neuroscience. BMC Bioinformatics 14:293. https://doi:10.1186/1471-2105-14-293

Quan T, Zheng T, Yang Z et al (2013) NeuroGPS: Automated localization of neurons for brain circuits using L1 minimization model. Scientific Reports 3. https://doi:10.1038/srep01414

Rajkowska G (2000) Postmortem studies in mood disorders indicate altered numbers of neurons and glial cells. Biological Psychiatry 48:766-777. https://doi:10.1016/s0006-3223(00)00950-1

Rajkowska G, Clarke G, Mahajan G et al (2016) Differential effect of lithium on cell number in the hippocampus and prefrontal cortex in adult mice: a stereological study. Bipolar Disorders 18:41-51. https://doi:10.1111/bdi.12364

Ramsden M, Berchtold NC, Kesslak JP, Cotman CW, Pike CJ (2003) Exercise increases the vulnerability of rat hippocampal neurons to kainate lesion. Brain Research 971:239-244. https://doi:10.1016/s0006-8993(03)02365-5

Rapp PR, Gallagher M (1996) Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. Proceedings of the National Academy of Sciences 93:9926-9930. https://doi:10.1073/pnas.93.18.9926

Rasmussen T, Schliemann T, Sørensen JC, Zimmer J, West MJ (1996) Memory impaired aged rats: No loss of principal hippocampal and subicular neurons. Neurobiology of Aging 17:143-147. https://doi:10.1016/0197-4580(95)02032-2

Ray S, Brecht M (2016) Structural development and dorsoventral maturation of the medial entorhinal cortex. ELife 5. doi:10.7554/elife.13343

Ropireddy D, Bachus S, Ascoli G (2012) Non-homogeneous stereological properties of the rat hippocampus from high-resolution 3D serial reconstruction of thin histological sections. Neuroscience 205:91-111. https://doi:10.1016/j.neuroscience.2011.12.055

Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic Imaging 13:146. https://doi:10.1117/1.1631315

Schindelin J, Rueden CT, Hiner MC, Eliceiri KW (2015) The ImageJ ecosystem: An open platform for biomedical image analysis. Molecular Reproduction and Development, 82:518-529. https://doi:10.1002/mrd.22489

Schmitz C, Hof P (2005) Design-based stereology in neuroscience. Neuroscience 130:813-831. https://doi:10.1016/j.neuroscience.2004.08.050

Schmitz C, Eastwood BS, Tappan SJ et al (2014). Current automated 3D cell detection methods are not a suitable replacement for manual stereologic cell counting. Frontiers in Neuroanatomy 8. https://doi:10.3389/fnana.2014.00027

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9:671-675. https://doi:10.1038/nmeth.2089

Sherwood CC, Stimpson CD, Raghanti MA et al (2006) Evolution of increased glia-neuron ratios in the human frontal cortex. Proceedings of the National Academy of Sciences 103:13606-13611. https://doi:10.1073/pnas.0605843103

Sousa N, Madeira MD, Paula-Barbosa MM (1998) Effects of corticosterone treatment and rehabilitation on the hippocampal formation of neonatal and adult rats. An unbiased stereological study. Brain Research 794: 199-210. https://doi:10.1016/s0006-8993(98)00218-2

Sun C, Kitamura T, Yamamoto J, et al (2015) Distinct speed dependence of entorhinal island and ocean cells, including respective grid cells. Proceedings of the National Academy of Sciences 112:9466-9471. https://doi:10.1073/pnas.1511668112

Sunkin SM, Ng L, Lau C et al (2012) Allen Brain Atlas: An integrated spatio-temporal portal for exploring the central nervous system. Nucleic Acids Research 41(D1). https://doi:10.1093/nar/gks1042

Tapias V, Greenamyre JT (2014) A Rapid and Sensitive Automated Image-Based Approach for In Vitro and In Vivo Characterization of Cell Morphology and Quantification of Cell Number and Neurite Architecture. Current Protocols in Cytometry 68. https://doi:10.1002/0471142956.cy1233s68

Vinet J, Lemieux P, Tamburri A, Tiesinga P, Scafidi J, Gallo V, Sík A. (2010) Subclasses of oligodendrocytes populate the mouse hippocampus. European Journal of Neuroscience 31(3):425-438. https://doi:10.1111/j.1460-9568.2010.07082.x

West MJ, Slomianka L, Gundersen HJ (1991) Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the

optical fractionator. The Anatomical Record 231:482-497. https://doi:10.1002/ar.1092310411

Wheeler DW, White CM, Rees CL, Komendantov AO, Hamilton DJ, Ascoli GA (2015) Hippocampome.org: A knowledge base of neuron types in the rodent hippocampus. ELife 4. https://doi:10.7554/elife.09960

Witter M (2011) Entorhinal cortex. Scholarpedia, 6:4380.

Wu H, Rassoulpour A, Goodman JH, Scharfman HE, Bertram EH, Schwarcz R (2005) Kynurenate and 7-Chlorokynurenate Formation in Chronically Epileptic Rats. Epilepsia 46:1010-1016. https://https://doi:10.1111/j.1528-1167.2005.67404.x

Zhang D, Liu S, Song Y, Feng D, Peng H, Cai W (2018) Automated 3D Soma Segmentation with Morphological Surface Evolution for Neuron Reconstruction. Neuroinformatics 16:153-166. https://https://doi:10.1007/s12021-017-9353-x

Abernethy, J., & Hazan, E. (2015). Faster convex optimization: simulated annealing with an efficient universal
barrier. http://proceedings.mlr.press/v48/abernethy16.pdf. Accessed 02 October 2019.

Armañanzas, R., & Ascoli, G. A. (2015). Towards the automatic classification of neurons. *Trends in Neuroscience,*
38(5), 307–318.

Armstrong, C., Szabadics, J., Tamas, G., & Soltesz, I. (2011). Neurogliaform cells in the molecular layer of
the dentate gyrus as feed-forward gamma-aminobutyric acidergic modulators of entorhinal-hippocampal
interplay. *The Journal of Comparative Neurology, 519*(8), 1476–1491.

Ascoli, G. A.,&Wheeler, D.W. (2016). In search of a periodic table of the neurons: Axonal-dendritic circuitry
as the organizing principle—Patterns of axons and dendrites within distinct anatomical parcels provide
the blueprint for circuit-based neuronal classification. *BioEssays, 38*(10), 969–976.

Attili, S. M., Silva, M. F. M., Nguyen, T., & Ascoli, G. A. (2019). Cell numbers, distribution, shape, and
regional variation throughout the murine hippocampal formation from the adult brain Allen Reference
Atlas. *Brain Structure and Function, 224,* 2883–2897.

Audet, C.,&Dennis, J. E., Jr. (2003). Analysis of generalized pattern searches. *SIAM Journal on Optimization,*

13(3), 889–903.

Bartheld, C. S. V. (2001). Comparison of 2-D and 3-D counting: The need for calibration and common sense.

Trends in Neurosciences, 24(9), 504–506. https://doi.org/10.1016/s0166-2236(00)01960-3.

Bayer, S., Yackel, J., & Puri, P. (1982). Neurons in the rat dentate gyrus granular layer substantially increase

during juvenile and adult life. *Science, 216,* 890–892.

Bezaire, M. J., Raikov, I., Burk, K., Vyas, D., & Soltesz, I. (2016). Interneuronal mechanisms of hippocampal

theta oscillations in a full-scale model of the rodent CA1 circuit. *ELife*.

https://doi.org/10.7554/eLife.18

566.001.

Bezaire, M. J., & Soltesz, I. (2013). Quantitative assessment of CA1 local circuits: Knowledge base for

interneuron-pyramidal cell connectivity. *Hippocampus, 23*(9), 751–785.

https://doi.org/10.1002/hipo.22

141.

Bhanu, B., & Peng, J. (2000). Adaptive integrated image segmentation and object recognition. *IEEE Trans*

Syst Man Cybern Part C (Appl Rev), 30, 427–441.

Bota, M.,&Swanson, L.W. (2007). The neuron classification problem. *Brain Research Reviews, 56*(1), 79–88.

Buckmaster, P. S.,&Jongen-Relo,A. L. (1999).Highly specific neuron loss preserves lateral inhibitory circuits
in the dentate gyrus of kainate-induced epileptic rats. *Journal of Neuroscience, 19*(21), 9519–9529.

Byrd, R. H., Gilbert, J. C., & Nocedal, J. (2000). A trust region method based on interior point techniques for
nonlinear programming. *Mathematical Programming, 89*(1), 149–185.

Calhoun, M. E., Kurth, D., & Phinney, A. L. (1998). Hippocampal neuron and synaptophysin-positive bouton
number in aging C57BL/6 mice. *Neurobiology of Aging, 19,* 599–606.

Ceranik, K., Bender, R., Geiger, J. R., Monyer, H., Jonas, P., Frotscher, M., et al. (1997). A novel type of
GABAergic interneuron connecting the input and the output regions of the hippocampus. *Journal of*
Neuroscience, 17(14), 5380–5394.

Coleman, T. F., & Li, Y. A. (1996). Reflective newton method for minimizing a quadratic function subject to
bounds on some of the variables. *SIAM Journal on Optimization, 6*(4), 1040–1058.

Conn, A. R.,Gould, N. I. M.,&Toint, P. (1997).Aglobally convergent augmented lagrangian barrier algorithm
for optimization with general inequality constraints and simple bounds. *Mathematics of Computation,*
66(217), 261–288.
Ecker, J. R., Geschwind, D. H., Kriegstein, A. R., Ngai, J., Osten, P., Polioudakis, D., et al. (2017). The BRAIN
initiative cell census consortium: Lessons learned toward generating a comprehensive brain cell atlas.
Neuron, 96, 542–557.
Erö, C., Gewaltig, C., Keller, M., & Markram, D. (2018). A cell atlas for the mouse brain. *Frontiers in*
Neuroinformatics. https://doi.org/10.3389/fninf.2018.00084.
Fitting, S., Booze, R. M., Hasselrot, U., & Mactutus, C. F. (2009). Dose-dependent longterm effects of Tat in
the rat hippocampal formation: A design-based stereological study. *Hippocampus*. https://doi.org/10.10
02/hipo.20648.
Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization*. Cambridge: Academic Press.
Grady, M. S., Charleston, J. S., Maris, D.,Witgen,B.M.,&Lifshitz, J. (2003).Neuronal and glial cell number in

the hippocampus after experimental traumatic brain injury: Analysis by stereological estimation. *Journal*

of Neurotrauma, 20(10), 929–941.

Hamilton, D. J., Shepherd, G. M., Martone, M. E., & Ascoli, G. A. (2012). An ontological approach to

describing neurons and their relationships. *Frontiers in Neuroinformatics*.

https://doi.org/10.3389/fninf.

2012.00015.

Hamilton, D. J.,White, C. M., Rees, C. L., Wheeler,D.W.,&Ascoli, G. A.

(2017).Molecular fingerprinting of

principal neurons in the rodent hippocampus: A neuroinformatics approach. *Journal of Pharmaceutical*

and Biomedical Analysis, 144(10), 269–278.

Han, Z. S. (1994). Electrophysiological and morphological differentiation of chandelier and basket cells in

the rat hippocampal formation: A study combining intracellular recording and intracellular staining with

biocytin. *Neuroscience Research, 19*(1), 101–110.

Herculano-Houzel, S. (2009). The human brain in numbers: A linearly scaled-up primate brain. *Frontiers in*

Human Neuroscience. https://doi.org/10.3389/neuro.09.031.2009.

Herculano-Houzel, S., Bartheld,C. S.V., Miller,D. J.,&Kaas, J.H. (2015).Howto count cells: The advantages
and disadvantages of the isotropic fractionator compared with stereology. *Cell and Tissue Research,*
360(1), 29–42.

Herculano-Houzel, S., Mota, B., & Lent, R. (2006). Cellular scaling rules for rodent brains. *PNAS, 103,*
12138–12143.

Herculano-Houzel, S., Ribeiro, P., Campos, L., Valotta da Silva, A., Torres, L. B., Catania, K. C., et al.
(2011). Updated neuronal scaling rules for the brains of Glires (rodents/lagomorphs). *Brain, Behavior*
and Evolution, 78, 302–314.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems.
Technometrics, 12(1), 55–67.

Hosseini-Sharifabad, M., & Nyengaard, J. R. (2007). Design-based estimation of neuronal number and individual
neuronal volume in the rat hippocampus. *Journal of Neuroscience Methods, 162,* 206–214.

Husmann, K., Lange, A., &Spiegel, E. (2017). The R package optimization: Flexible global optimization with

simulated-annealing. Researchgate.net. Accessed 01 August 2019.

Insausti, A. M.,Megas,M.,&Crespo, D. (1998). Hippocampal volume and neuronal number in Ts65Dnmice:

A murine model of down syndrome. *Neuroscience Letters, 253,* 175–178.

Januszewski, M., Kornfeld, J., Li, P. H., Pope, A., Blakely, T., Lindsey, L., et al. (2018). High-precision

automated reconstruction of neurons with flood-filling networks. *Nature Methods, 15*(8), 605–610.

Kaae, S. S., Chen, F., Wegener, G., Madsen, T. M., & Nyengaard, J. R. (2012). Quantitative hippocampal

structural changes following electroconvulsive seizure treatment in a rat model of depression. *Synapse*

(New York, N. Y.), 66, 667–676.

Kalai, A. T., & Vempala, S. (2006). Simulated annealing for convex optimization. *Mathematics of Operations*

Research, 31(2), 253–266. https://doi.org/10.1287/moor.1060.0194.

Keller, D., Meystre, J., Veettil, R. V., Burri, O., Guiet, R., Schurmann, F., et al. (2019). A derived positional

mapping of inhibitory subtypes in the somatosensory cortex. *Frontiers in Neuroanatomy*. https://doi.org/

10.3389/fnana.2019.00078.

Laarhoven, P. J. M. V., & Aarts, E. H. L. (1987). Simulated annealing. In *Simulated annealing: Theory and*

applications (pp. 77–98). Dordrecht: Springer.

Lawson, C. L., & Hanson, R. J. (1995). *Solving least squares problems. Classics in applied mathematics*.

Philadelphia: SIAM.

Lister, J. P., Tonkiss, J., Blatt, G. J., Kemper, T. L., Debassio,W. A., Galler, J. R., et al. (2006). Asymmetry of

neuron numbers in the hippocampal formation of prenatally malnourished and normally nourished rats:

A stereological investigation. *Hippocampus, 16,* 946–958.

Lubke, J., Frotscher, M., & Spruston, N. (1998). Specialized electrophysiological properties of anatomically

identified neurons in the hilar region of the rat fascia dentata. *Journal of Neurophysiology, 79*(3),

1518–1534.

Moradi, K., & Ascoli, G. A. (2019). A comprehensive knowledge base of synaptic electrophysiology in the

rodent hippocampal formation. *Hippocampus*. https://doi.org/10.1101/632760.

More, J. J. (1978). The Levenberg–Marquardt algorithm: Implementation and theory. *Lecture Notes in Mathematics*

Numerical Analysis, 45, 105–116. https://doi.org/10.1007/bfb0067700.

Morgan, R. J., Santhakumar, V., & Soltesz, I. (2007).Modeling the dentate gyrus. *Progress in Brain Research,* *163*, 639–658.

Mott, D. D., Turner, D. A., Okazaki, M. M., & Lewis, D. V. (1997). Interneurons of the dentate-hilus border of the rat dentate gyrus: Morphological and electrophysiological heterogeneity. *Journal of Neuroscience,* *17*(11), 3990–4005.

Mulders, W., West, M., & Slomianka, L. (1998). Neuron numbers in the presubiculum, parasubiculum, and entorhinal area of the rat. *Journal of Comparative Neurology, 385,* 83–94.

Mullen, M. (2015). The Stark–Parker algorithm for bounded-variable least squares. https://cran.rproject.org/web/packages/bvls/bvls.pdf. Accessed 01 August 2019.

Mullen, M., & van Stokkum, H. M. (2015). The Lawson–Hanson algorithm for non-negative least squares (NNLS). https://cran.r-project.org/web/packages/nnls/nnls.pdf. Accessed 01 August 2019.

Murakami, T.C., Mano, T., Saikawa, S., Horiguchi, S. A., Shigeta, D., et al. (2018).Athree-dimensional singlecell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing. *Nature*

Neuroscience, 21, 625–637.

Peng, H., Hawrylycz, M., Roskams, J., Hill, S., Spruston, N., Meijering, E., et al. (2015). BigNeuron: Largescale

3D neuron reconstruction from optical microscopy images. *Neuron, 87*(2), 252–256.

Peng, H., Roysam, B., & Ascoli, G. A. (2013). Automated image computing reshapes computational neuroscience.

BMC Bioinformatics, 14, 293.

Ramsden, M., Berchtold, N. C., Kesslak, J. P., Cotman, C. W., & Pike, C. J. (2003). Exercise increases the

vulnerability of rat hippocampal neurons to kainate lesion. *Brain Research, 971,* 239–244.

Rapp, P. R., & Gallagher, M. (1996). Preserved neuron number in the hippocampus of aged rats with spatial

learning deficits. *Proceedings of the National Academy of Sciences of the United States of America,*

93(18), 9926–9930.

Rasmussen, T., Schliemann, T., Sorensen, J. C., Zimmer, J., & West, M. J. (1996). Memory impaired aged

rats: No loss of principal hippocampal and subicular neurons. *Neurobiology of Aging, 17,* 143–147.

Russ, J. C., & Deho, R. T. (2001). *Practical stereology*. New York: Kluwer Academic.

Shepherd,M., G., Marenco, Luis,Hines, L., M., et al. (2019, February 7).Neuron names:Agene- and property based
name format, with special reference to cortical neurons. *Frontiers*, https://www.frontiersin.org/
articles/10.3389/fnana.2019.00025/full. Accessed 24 October 2019.

Sousa, N., Madeira, M. D., & Paula-Barbosa, M. M. (1998). Effects of corticosterone treatment and rehabilitation
on the hippocampal formation of neonatal and adult rats. An unbiased stereological study. *Brain*
Research, 794, 199–210.

Stark, P. B., & Parker, R. L. (1993). Bounded-variable least-squares: An algorithm and applications. http://
digitalassets.lib.berkeley.edu/sdtr/ucb/text/394.pdf. Accessed 01 August 2019.

Tasic, B., Yao, Z., Smith, K. A., Graybuck, L., Nguyen, T., Bertagolli, D., et al. (2018). Shared and distinct
transcriptomic cell types across neocortical areas. *Nature*, *563*(7729), 72–78.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical*
Society: Series B, 58(1), 267–288.

Wang, L., Gordon, M. D., & Zhu, J. (2006). Regularized least absolute deviations regression and an effcient

algorithm for parameter tuning. IEEE.
https://ieeexplore.ieee.org/abstract/document/4053094. Accessed
01 August 2019.
West, M. J., Slomianka, L., & Gundersen, H. J. (1991). Unbiased stereological estimation of the total number
of neurons in the subdivisions of the rat hippocampus using the optical fractionator. *The Anatomical*
Record, 231(4), 482–497.
Wheeler, D.W., et al. (2015). Hippocampome.org: A knowledge base of neuron types in the rodent hippocampus.
Elife, 4, 09960.
White, C. M., Rees, C. L., Wheeler, D.W., Hamilton, D. J., & Ascoli, G. A.
(2019).Molecular expression profiles
of morphologically defined hippocampal neuron types: Empirical evidence and relational inferences.
Hippocampus. https://doi.org/10.1002/hipo.23165.
Williams, P. A., Larimer, P., Gao, Y., & Strowbridge, B. W. (2007). Semilunar granule cells: Glutamatergic
neurons in the rat dentate gyrus with axon collaterals in the innermolecular layer. *Journal of Neuroscience,*
27(50), 13756–13761.

Woodson,W.,Nitecka, L.,&Ben-Ari,Y. (1989).Organization of theGABAergic system in the rat hippocampal

formation: A quantitative immunocytochemical study. *The Journal of Comparative Neurology, 280*(2),

254–271.

Xiang, Y., Gubian, S., Suomela, B., & Hoeng, J. (2013). Generalized simulated annealing for global optimization:

The GenSA package. *The R Journal, 5,* 13.

Andrade JP, Madeira M, Paula-Barbosa M (2000) Sexual dimorphism in the subiculum of the rat hippocampal formation. Brain Research 875:125-137. https://doi:10.1016/s0006-8993(00)02605-6

Armstrong, C., Szabadics, J., Tamas, G., Soltesz, I. (2011). Neurogliaform cells in the molecular layer of the dentate gyrus as feed-forward gamma-aminobutyric acidergic modulators of entorhinal-hippocampal interplay. The Journal of Comparative Neurology, 519(8), 1476-1491.

Attila I. Gulyas, Gergely G. Szabo, Istvan Ulbert, Noemi Holderith, Hannah Monyer, Ferenc Erdelyi, Gabor Szabo, Tamas F. Freund, and Norbert Hajos (2010). Parvalbumin-Containing Fast-Spiking Basket Cells Generate the Field Potential Oscillations Induced by Cholinergic Receptor Activation in the Hippocampus.

Attili†, Sean T. Mackesey†, Giorgio A. Ascoli 2020. Operations Research Methods for Estimating the Population Size of Neuron Types.

Attili, S. M., Silva, M. F. M., Nguyen, T., & Ascoli, G. A. (2019). Cell numbers, distribution, shape, and regional variation throughout the murine hippocampal formation from the adult brain Allen Reference Atlas. Brain Structure and Function. 224:2883–2897.

Barkai, E., & Hasselmo, M. E. (1994). Modulation of the input/output function of rat piriform cortex pyramidal cells. Journal of Neurophysiology, 72(2), 644–658. https://doi.org/10.1152/jn.1994.72.2.644

Bezaire, M. J., Raikov, I., Burk, K., Vyas, D., Soltesz, I. (2016) Interneuronal mechanisms of hippocampal theta oscillations in a full-scale model of the rodent CA1 circuit. ELife, https://doi.org/10.7554/eLife.18566.001

Buckmaster, P. S., Strowbridge, B. W., Kunkel, D. D., Schmiege, D. L., Schwartzkroin, P. A. (1992). Mossy cell axonal projections to the dentate gyrus molecular layer in the rat hippocampal slice. Hippocampus, 2(4), 349-362.

Buhl E. H., Katalin Halasy, Peter somogyi 1994 Diverse sources of hippocampal unitary inhibitory postsynaptic potentials and the number of synaptic release sites

Canto, C. B., & Witter, M. P. (2011). Cellular properties of principal neurons in the rat entorhinal cortex. II. The medial entorhinal cortex. Hippocampus, 22(6), 1277–1299. https://doi.org/10.1002/hipo.20993

Cembrowski, M. S., Wang, L., Sugino, K., Shields, B. C., & Spruston, N. (2016). Hipposeq: a comprehensive RNA-seq database of gene expression in hippocampal principal neurons. ELife, 5. https://doi.org/10.7554/elife.14997

Ceranik, K., Bender, R., Geiger, J. R., Monyer, H., Jonas, P., Frotscher, M., et al. (1997). A novel type of GABAergic interneuron connecting the input and the output regions of the hippocampus. Journal of Neuroscience, 17(14), 5380-5394.

Coulin C (2001). Stereological Estimates of Total Neuron Numbers in the Hippocampus of Adult Reeler Mutant Mice: Evidence for an Increased Survival of Cajal-Retzius Cells

Damasio, H. (2005). Human brain anatomy in computerized images. Oxford University Press.

Eberhart, & Yuhui Shi. (n.d.). Particle swarm optimization: developments, applications and resources. Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546). https://doi.org/10.1109/cec.2001.934374

Ecker, J. R., Geschwind, D. H., Kriegstein, A. R., Ngai, J., Osten, P., Polioudakis, D., et al. (2017). The BRAIN initiative cell census consortium: lessons learned toward generating a comprehensive brain cell atlas. Neuron, 96, 542-557.

Ehrlich, D. E., Ryan, S. J., & Rainnie, D. G. (2012). Postnatal development of electrophysiological properties of principal neurons in the rat basolateral amygdala. The Journal of Physiology, 590(19), 4819–4838. https://doi.org/10.1113/jphysiol.2012.237453

Ero, C.,Gewaltig, M.,Keller, D.,Markram, H. (2018). A Cell Atlas for the Mouse Brain. Frontiers in Neuroinformatics, https://doi.org/10.3389/fninf.2018.00084

Faber, E. S. L., Callister, R. J., & Sah, P. (2001). Morphological and Electrophysiological Properties of Principal Neurons in the Rat Lateral Amygdala In Vitro. Journal of Neurophysiology, 85(2), 714–723. https://doi.org/10.1152/jn.2001.85.2.714

Fitting S, Booze RM, Hasselrot U, Mactutus CF (2009) Dose-dependent long-term effects of Tat in the rat hippocampal formation: A design-based stereological study. Hippocampus. https://doi:10.1002/hipo.20648

Fuentealba P., Ryohei Tomioka, Yannis Dalezios, La´szlo´ F.Ma´rton, Michele Studer, Kathleen Rockland, Thomas Klausberger, and Peter Somogyi (2008). Rhythmically

Active Enkephalin-Expressing GABAergic Cells in the CA1 Area of the Hippocampus Project to the Subiculum and Preferentially Innervate Interneurons

Goebbels, S., Bormuth, I., Bode, U., Hermanson, O., Schwab, M. H., & Nave, K.-A. (2006). Genetic targeting of principal neurons in neocortex and hippocampus of NEX-Cre mice. Genesis, 44(12), 611–621. https://doi.org/10.1002/dvg.20256

Grady, M. S., Charleston, J. S., Maris, D., Witgen, B. M., Lifshitz, J. (2003). Neuronal and Glial Cell Number in the Hippocampus after Experimental Traumatic Brain Injury: Analysis by Stereological Estimation. Journal of Neurotrauma, 20(10), 929-941.

Hajos N. and Mody I. (1997) Synaptic Communication among Hippocampal Interneurons: Properties of Spontaneous IPSCs in Morphologically Identified Cells

Hamilton et al., 2016 Name-calling in the hippocampus (and beyond): coming to terms with neuron types and properties. Brain Informatics 2017 Mar;4(1):1-12; doi:10.1007/s40708-016-0053-3.

Han, Z. S. (1994). Electrophysiological and morphological differentiation of chandelier and basket cells in the rat hippocampal formation: a study combining intracellular recording and intracellular staining with biocytin. Neuroscience Research, 19(1), 101-110.

Harris E, Stewart M December 2000. Propagation of synchronous epileptiform events from subiculum backward into area CA1 of rat brain slices

Herculano-Houzel, S., Mota, B. & Lent, R. (2006). Cellular scaling rules for rodent brains. PNAS. 103,12138–12143.

Herculano-Houzel, S., Ribeiro P., Campos L., Valotta da Silva A., Torres L.B., Catania K.C., et al. (2011). Updated neuronal scaling rules for the brains of Glires (rodents/lagomorphs). Brain, Behavior and Evolution, 78, 302-314.

Hestrin, S., Nicoll, R. A., Perkel, D. J., & Sah, P. (1990). Analysis of excitatory synaptic action in pyramidal cells using whole-cell recording from rat hippocampal slices. The Journal of Physiology, 422(1), 203–225. https://doi.org/10.1113/jphysiol.1990.sp017980

Hosseini-Sharifabad, M., & Nyengaard, J. R. (2007). Design-based estimation of neuronal number and individual neuronal volume in the rat hippocampus. J Neurosci Methods. 162:206–214.

Ingvild E. Bjerke, Sharon C. Yates, Arthur Laja, Menno P. Witter, Maja A. Puchades, Jan G. Bjaalie, Trygve B. Leergaard (2021). Densities and numbers of calbindin and parvalbumin positive neurons across the rat and mouse brain.

Jinno S, Kosaka T (2006). Cellular architecture of the mouse hippocampus: A quantitative aspect of chemically defined GABAergic neurons with stereology

Kaae, S. S., Chen, F., Wegener, G., Madsen, T. M., & Nyengaard, J. R. (2012). Quantitative hippocampal structural changes following electroconvulsive seizure treatment in a rat model of depression. Synapse. 66:667–676.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., Hudspeth, A. J., Mack, S. (2013). Principles of neural science. NY: McGraw-Hill Medical.

Kennedy, J., & Eberhart, R. C. (n.d.). A discrete binary version of the particle swarm algorithm. 1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation. https://doi.org/10.1109/icsmc.1997.637339

Kim, Y., Yang, G. R., Pradhan, K., Venkataraju, K. U., Bota, M., García del Molino, L. C., … Osten, P. (2017). Brain-wide Maps Reveal Stereotyped Cell-Type-Based Cortical Architecture and Subcortical Sexual Dimorphism. Cell, 171(2). https://doi.org/10.1016/j.cell.2017.09.020

Kohus Z., S. Kali, L. Rovira-Esteban, D. Schlingloff, O. Papp, T. F. Freund, N. Hajos and A. I. Gulyas. Properties and dynamics of inhibitory synaptic communication within the CA3 microcircuits of pyramidal cells and interneurons expressing parvalbumin or cholecystokinin

Komendantov, et al., 2019 Quantitative firing pattern phenotyping of hippocampal neuron type. Scientific Reports 2019 Nov 29;9(1):17915; doi: 10.1038/s41598-019-52611-w.

Kosaka T, Katsumaru H, Kiyoshi Hama, Jang-Yen Wu and Claus W. Heizmann (1987). GABAergic neurons containing the Ca2+-binding protein parvalbumin in the rat hippocampus and dentate gyrus

Lange, N., Giedd, J. N., Xavier Castellanos, F., Vaituzis, A. C., & Rapoport, J. L. (1997). Variability of human brain structure size: ages 4–20 years. Psychiatry Research: Neuroimaging, 74(1), 1–12. https://doi.org/10.1016/s0925-4927(96)03054-5

Lewis, R. M., & Torczon, V. (2002). A Globally Convergent Augmented Lagrangian Pattern Search Algorithm for Optimization with General Constraints and Simple Bounds. SIAM Journal on Optimization, 12(4), 1075–1089. https://doi.org/10.1137/s1052623498339727

Lister, J. P., Tonkiss, J., Blatt, G. J., Kemper, T. L., Debassio, W. A., Galler, J. R., & Rosene, D. L. (2006). Asymmetry of neuron numbers in the hippocampal formation of prenatally malnourished and normally nourished rats: a stereological investigation. Hippocampus. 16:946–958.

Losonczy A, Zhang L, Shigemoto R, Somogyi P, Nusser Z J Physiol, (2002). Cell type dependence and variability in the short-term plasticity of EPSCs in identified mouse hippocampal interneurones. 542 (Pt 1), pages: 193 – 210

Losonczy* A., A´ gota A. Biro´ , and Zoltan Nusser* (2004). Persistently active cannabinoid receptors mute a subpopulation of hippocampal interneurons

Lubke, J., Frotscher, M., Spruston, N. (1998). Specialized electrophysiological properties of anatomically identified neurons in the hilar region of the rat fascia dentata. Journal of Neurophysiology, 79(3), 1518-1534.

Maccaferri, G., David, J., Roberts, B., Szucs, P., Cottingham, C. A., & Somogyi, P. (2000). Cell surface domain specific postsynaptic currents evoked by identified GABAergic neurones in rat hippocampus in vitro. The Journal of Physiology, 524(1), 91–116. https://doi.org/10.1111/j.1469-7793.2000.t01-3-00091.x

McBain C J, DiChiara T J, and Kauer J A 1994. Activation of lhetabotropic Glutamate Receptors Differentially Affects Two Classes of Hippocampal Interneurons and Potentiates Excitatory Synaptic Transmission

Mercer A., Hayley L. Trigg, and Alex M. Thomson. Characterization of Neurons in the CA2 Subfield of the Adult Rat Hippocampus

Mott, D. D., Turner, D. A., Okazaki, M. M., Lewis, D. V. (1997). Interneurons of the dentate-hilus border of the rat dentate gyrus: morphological and electrophysiological heterogeneity. Journal of Neuroscience, 17(11) 3990-4005.

Mott M. C., Joshua A. Gordon,Walter J. Koroshetz. The NIH BRAIN Initiative: Advancing neurotechnologies, integrating disciplines

Mukamel, E. A., & Ngai, J. (2019). Perspectives on defining cell types in the brain. Current Opinion in Neurobiology, 56, 61–68. https://doi.org/10.1016/j.conb.2018.11.007

Mulders W, West M, Slomianka L (1997) Neuron numbers in the presubiculum, parasubiculum, and entorhinal area of the rat. The Journal of Comparative Neurology 385:83-94. https://doi:10.1002/(sici)1096-9861(19970818)385:1<83::aid-cne5>3.0.co;2-8

Murakami, T. C., Mano, T., Saikawa, S., Horiguchi, S. A., Shigeta, D., et al. (2018). A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing. Nature Neuroscience, 21, 625-637.

Napierala M. A. (2012). What is the Bonferroni correction?

Nolte, J., & Sundsten, J. (2009). The human brain: an introduction to its functional anatomy. Mosby-Elsevier.

Ramsden, M., Berchtold, N. C., Kesslak, J. P., Cotman, C. W., & Pike, C. J. (2003). Exercise increases the vulnerability of rat hippocampal neurons to kainate lesion. Brain Res. 971:239–244.

Rapp, P. R., Gallagher, M. (1996). Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. Proceedings of the National Academy of Sciences of the United States of America, 93(18), 9926-9930.

Rasmussen, T., Schliemann, T., Sørensen, J. C., Zimmer, J., & West, M. J. (1996). Memory impaired aged rats: No loss of principal hippocampal and subicular neurons. Neurobiology of Aging, 17(1), 143–147. https://doi.org/10.1016/0197-4580(95)02032-2

Sanchez-Aguilera, A., Wheeler, D. W., Jurado-Parras, T., Valero, M., Nokia, M. S., Cid, E., Fernandez-Lamo, I., Sutton, N., García-Rincón, D., de la Prida, L. M., & Ascoli, G. A. (2021). An update to Hippocampome.org by integrating single-cell phenotypes with circuit function in vivo. *PLOS Biology*, *19*(5). https://doi.org/10.1371/journal.pbio.3001213

Sanes, D. H., Reh, T. A., & Harris, W. A. (2007). Development of the nervous system. Amsterdam: Elsevier Acad. Press.

Sik, A., Ylinen, A., Penttonen, M., & Buzsaki, G. (1994). Inhibitory CA1-CA3-hilar region feedback in the hippocampus. Science, 265(5179), 1722–1724. https://doi.org/10.1126/science.8085161

Suzuki, N., & Bekkers, J. M. (2011). Two Layers of Synaptic Processing by Principal Neurons in Piriform Cortex. Journal of Neuroscience, 31(6), 2156–2166. https://doi.org/10.1523/jneurosci.5430-10.2011

Svoboda K. R., Cathy E. Adams, and Carl R. Lupica (1999). Opioid Receptor Subtype Expression Defines Morphologically Distinct Classes of Hippocampal Interneurons

Szabadics J., Csaba Varga, Janos Brunner, Kang Chen, and Ivan Soltesz (2010). Granule Cells in the CA3 Area

Szabo G. G., Orsolya I. Papp, Zoltan Mate,2 Gabor Szabo, and Norbert Hajos (2014). Anatomically Heterogeneous Populations of CB1 Cannabinoid Receptor-Expressing Interneurons in the CA3 Region of the Hippocampus Show Homogeneous Input–Output Characteristics

Tahvildari, B., & Alonso, A. (2005). Morphological and electrophysiological properties of lateral entorhinal cortex layers II and III principal neurons. The Journal of Comparative Neurology, 491(2), 123–140. https://doi.org/10.1002/cne.20706

Taubert, M., Draganski, B., Anwander, A., Muller, K., Horstmann, A., Villringer, A., & Ragert, P. (2010). Dynamic Properties of Human Brain Structure: Learning-Related Changes in Cortical Areas and Associated Fiber Connections. Journal of Neuroscience, 30(35), 11670–11677. https://doi.org/10.1523/jneurosci.2567-10.2010

Tecuatl C., Diek W. Wheeler, Nate Sutton, and Giorgio A. Ascoli (2021). Comprehensive Estimates of Potential Synaptic Connections in Local Circuits of the Rodent Hippocampal Formation by Axonal-Dendritic Overlap

Tricoire L, Kenneth A. Pelkey, Michael I. Daw, Vitor H. Sousa, Goichi Miyoshi, Brian Jeffries, Bruno Cauli, Gord Fishell and Chris J. Mcbain (2010). Common origins of hippocampal ivy and nitric oxide synthase expressing neurogliaform cells.

Vaz A.I.F. and Vicente L.N., A particle swarm pattern search method for bound constrained global optimization, Journal of Global Optimization, 39 (2007) 197-219. (published version - report).

Vida * I.†, Katalin Halasy*‡, Csaba Szinyei *, Peter Somogyi * and Eberhard H. Buhl * (1998). Unitary IPSPs evoked by interneurons at the stratum radiatum—stratum lacunosum-moleculare border in the CA1 area of the rat hippocampus in vitro

Wang, Y., Markram, H., Goodman, P. H., Berger, T. K., Ma, J., & Goldman-Rakic, P. S. (2006). Heterogeneity in the pyramidal network of the medial prefrontal cortex. Nature Neuroscience, 9(4), 534–542. https://doi.org/10.1038/nn1670

West, M. J., Slomianka, L., Gundersen, H. J. (1991). Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator. The Anatomical Record, 231(4), 482-497.

Wheeler et al., 2015. Hippocampome.org: A knowledge base of neuron types in the rodent hippocampus. eLife 2015;4:e09960.

White, et al., 2019 Molecular Expression Profiles of Morphologically Defined Hippocampal Neuron Types: Empirical Evidence and Relational Inferences. Hippocampus 2020

Yang, C. R., Seamans, J. K., & Gorelova, N. (1996). Electrophysiological and morphological properties of layers V-VI principal pyramidal cells in rat prefrontal cortex in vitro. The Journal of Neuroscience, 16(5), 1904–1921. https://doi.org/10.1523/jneurosci.16-05-01904.1996

BIOGRAPHY

Sarojini Manjusha Attili received her Bachelor of Individualized Study in Information Technology Management from George Mason University in 2005. She was employed as a Quality Assurance Analyst, Requirements Lead and Project Manager in different companies including Unisys and Deloitte from 2004 to 2013. She received her Master of Science in Bioinformatics and Computational Biology from George Mason University in 2016.